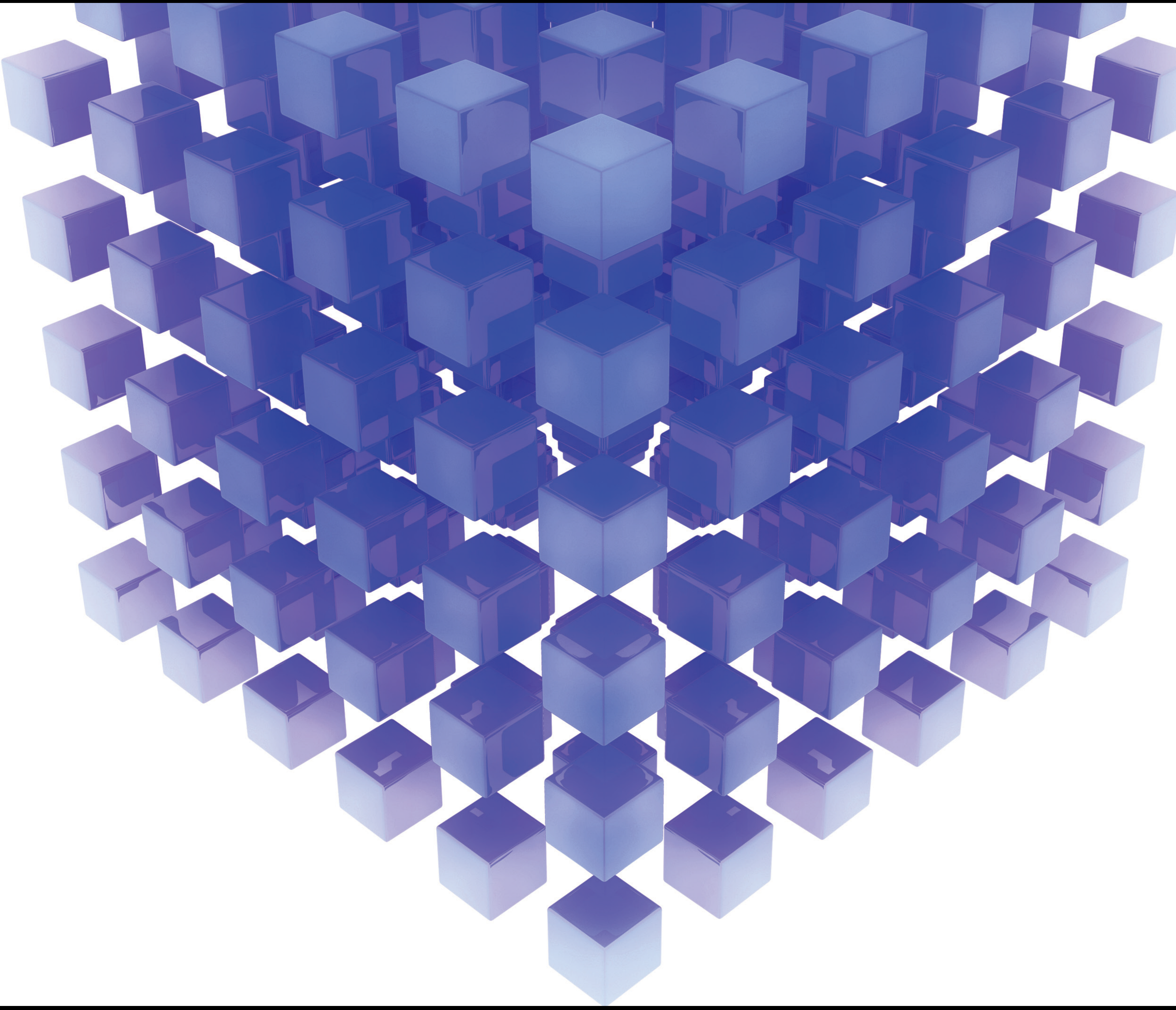
Mathematical Problems in Engineering

Big Data Modelling of Engineering and Management 2021

Lead Guest Editor: Wen-Tsao Pan
Guest Editors: Shiang-Hau Wu, Wei-Lin Xiao, and Yi-Wen Zhang



Big Data Modelling of Engineering and Management 2021

Mathematical Problems in Engineering

Big Data Modelling of Engineering and Management 2021

Lead Guest Editor: Wen-Tsao Pan
Guest Editors: Shiang-Hau Wu, Wei-Lin Xiao, and Yi-Wen Zhang

Paolo Boscariol, Italy
Daniela Boso, Italy
Guillermo Botella-Juan, Spain
Abdesselem Boulkroune, Algeria
Boulaïd Boulkroune, Belgium
Fabio Bovenga, Italy
Francesco Braghin, Italy
Ricardo Branco, Portugal
Julien Bruchon, France
Matteo Bruggi, Italy
Michele Brun, Italy
Maria Elena Bruni, Italy
Maria Angela Butturi, Italy
Bartłomiej Błachowski, Poland
Dhanamjayulu C, India
Raquel Caballero-Águila, Spain
Filippo Cacace, Italy
Salvatore Caddemi, Italy
Zuowei Cai, China
Roberto Caldelli, Italy
Francesco Cannizzaro, Italy
Maosen Cao, China
Ana Carpio, Spain
Rodrigo Carvajal, Chile
Caterina Casavola, Italy
Sara Casciati, Italy
Federica Caselli, Italy
Carmen Castillo, Spain
Inmaculada T. Castro, Spain
Miguel Castro, Portugal
Giuseppe Catalanotti, United Kingdom
Alberto Cavallo, Italy
Gabriele Cazzulani, Italy
Fatih Vehbi Celebi, Turkey
Miguel Cerrolaza, Venezuela
Gregory Chagnon, France
Ching-Ter Chang, Taiwan
Kuei-Lun Chang, Taiwan
Qing Chang, USA
Xiaoheng Chang, China
Prasenjit Chatterjee, Lithuania
Kacem Chehdi, France
Peter N. Cheimets, USA
Chih-Chiang Chen, Taiwan
He Chen, China
Kebing Chen, China
Mengxin Chen, China
Shyi-Ming Chen, Taiwan
Xizhong Chen, Ireland
Xue-Bo Chen, China
Zhiwen Chen, China
Qiang Cheng, USA
Zeyang Cheng, China
Luca Chiapponi, Italy
Francisco Chicano, Spain
Tirivanhu Chinyoka, South Africa
Adrian Chmielewski, Poland
Seongim Choi, USA
Gautam Choubey, India
Hung-Yuan Chung, Taiwan
Yusheng Ci, China
Simone Cinquemani, Italy
Roberto G. Citarella, Italy
Joaquim Ciurana, Spain
John D. Clayton, USA
Piero Colajanni, Italy
Giuseppina Colicchio, Italy
Vassilios Constantoudis, Greece
Enrico Conte, Italy
Alessandro Contento, USA
Mario Cools, Belgium
Gino Cortellessa, Italy
Carlo Cosentino, Italy
Paolo Crippa, Italy
Erik Cuevas, Mexico
Guozeng Cui, China
Mehmet Cunkas, Turkey
Giuseppe D'Aniello, Italy
Peter Dabnichki, Australia
Weizhong Dai, USA
Zhifeng Dai, China
Purushothaman Damodaran, USA
Sergey Dashkovskiy, Germany
Adiel T. De Almeida-Filho, Brazil
Fabio De Angelis, Italy
Samuele De Bartolo, Italy
Stefano De Miranda, Italy
Filippo De Monte, Italy

José António Fonseca De Oliveira Correia, Portugal
Jose Renato De Sousa, Brazil
Michael Defoort, France
Alessandro Della Corte, Italy
Laurent Dewasme, Belgium
Sanku Dey, India
Gianpaolo Di Bona, Italy
Roberta Di Pace, Italy
Francesca Di Puccio, Italy
Ramón I. Diego, Spain
Yannis Dimakopoulos, Greece
Hasan Dinçer, Turkey
José M. Domínguez, Spain
Georgios Dounias, Greece
Bo Du, China
Emil Dumic, Croatia
Madalina Dumitriu, United Kingdom
Premraj Durairaj, India
Saeed Eftekhar Azam, USA
Said El Kafhali, Morocco
Antonio Elipe, Spain
R. Emre Erkmen, Canada
John Escobar, Colombia
Leandro F. F. Miguel, Brazil
FRANCESCO FOTI, Italy
Andrea L. Facci, Italy
Shahla Faisal, Pakistan
Giovanni Falsone, Italy
Hua Fan, China
Jianguang Fang, Australia
Nicholas Fantuzzi, Italy
Muhammad Shahid Farid, Pakistan
Hamed Faroqi, Iran
Yann Favennec, France
Fiorenzo A. Fazzolari, United Kingdom
Giuseppe Fedele, Italy
Roberto Fedele, Italy
Baowei Feng, China
Mohammad Ferdows, Bangladesh
Arturo J. Fernández, Spain
Jesus M. Fernandez Oro, Spain
Francesco Ferrise, Italy
Eric Feulvarch, France
Thierry Floquet, France
Eric Florentin, France
Gerardo Flores, Mexico
Antonio Forcina, Italy
Alessandro Formisano, Italy
Francesco Franco, Italy
Elisa Francomano, Italy
Juan Frausto-Solis, Mexico
Shujun Fu, China
Juan C. G. Prada, Spain
HECTOR GOMEZ, Chile
Matteo Gaeta, Italy
Mauro Gaggero, Italy
Zoran Gajic, USA
Jaime Gallardo-Alvarado, Mexico
Mosè Gallo, Italy
Akemi Gálvez, Spain
Maria L. Gandarias, Spain
Hao Gao, Hong Kong
Xingbao Gao, China
Yan Gao, China
Zhiwei Gao, United Kingdom
Giovanni Garcea, Italy
José García, Chile
Harish Garg, India
Alessandro Gasparetto, Italy
Stylianos Georgantzinos, Greece
Fotios Georgiades, India
Parviz Ghadimi, Iran
Ștefan Cristian Gherghina, Romania
Georgios I. Giannopoulos, Greece
Agathoklis Giaralis, United Kingdom
Anna M. Gil-Lafuente, Spain
Ivan Giorgio, Italy
Gaetano Giunta, Luxembourg
Jefferson L.M.A. Gomes, United Kingdom
Emilio Gómez-Déniz, Spain
Antonio M. Gonçalves de Lima, Brazil
Qunxi Gong, China
Chris Goodrich, USA
Rama S. R. Gorla, USA
Veena Goswami, India
Xunjie Gou, Spain
Jakub Grabski, Poland

Antoine Grall, France
George A. Gravvanis, Greece
Fabrizio Greco, Italy
David Greiner, Spain
Jason Gu, Canada
Federico Guarracino, Italy
Michele Guida, Italy
Muhammet Gul, Turkey
Dong-Sheng Guo, China
Hu Guo, China
Zhaoxia Guo, China
Yusuf Gurefe, Turkey
Salim HEDDAM, Algeria
ABID HUSSANAN, China
Quang Phuc Ha, Australia
Li Haitao, China
Petr Hájek, Czech Republic
Mohamed Hamdy, Egypt
Muhammad Hamid, United Kingdom
Renke Han, United Kingdom
Weimin Han, USA
Xingsi Han, China
Zhen-Lai Han, China
Thomas Hanne, Switzerland
Xinan Hao, China
Mohammad A. Hariri-Ardebili, USA
Khalid Hattaf, Morocco
Defeng He, China
Xiao-Qiao He, China
Yanchao He, China
Yu-Ling He, China
Ramdane Hedjar, Saudi Arabia
Jude Hemanth, India
Reza Hemmati, Iran
Nicolae Herisanu, Romania
Alfredo G. Hernández-Diaz, Spain
M.I. Herreros, Spain
Eckhard Hitzer, Japan
Paul Honeine, France
Jaromir Horacek, Czech Republic
Lei Hou, China
Yingkun Hou, China
Yu-Chen Hu, Taiwan
Yunfeng Hu, China
Can Huang, China
Gordon Huang, Canada
Linsheng Huo, China
Sajid Hussain, Canada
Asier Ibeas, Spain
Orest V. Iftime, The Netherlands
Przemyslaw Ignaciuk, Poland
Giacomo Innocenti, Italy
Emilio Insfran Pelozo, Spain
Azeem Irshad, Pakistan
Alessio Ishizaka, France
Benjamin Ivorra, Spain
Breno Jacob, Brazil
Reema Jain, India
Tushar Jain, India
Amin Jajarmi, Iran
Chiranjibe Jana, India
Łukasz Jankowski, Poland
Samuel N. Jator, USA
Juan Carlos Jáuregui-Correa, Mexico
Kandasamy Jayakrishna, India
Reza Jazar, Australia
Khalide Jbilou, France
Isabel S. Jesus, Portugal
Chao Ji, China
Qing-Chao Jiang, China
Peng-fei Jiao, China
Ricardo Fabricio Escobar Jiménez, Mexico
Emilio Jiménez Macías, Spain
Maolin Jin, Republic of Korea
Zhuo Jin, Australia
Ramash Kumar K, India
BHABEN KALITA, USA
MOHAMMAD REZA KHEDMATI, Iran
Viacheslav Kalashnikov, Mexico
Mathiyalagan Kalidass, India
Tamas Kalmar-Nagy, Hungary
Rajesh Kaluri, India
Jyotheeswara Reddy Kalvakurthi, India
Zhao Kang, China
Ramani Kannan, Malaysia
Tomasz Kapitaniak, Poland
Julius Kaplunov, United Kingdom
Konstantinos Karamanos, Belgium
Michal Kawulok, Poland

Irfan Kaymaz, Turkey
Vahid Kayvanfar, Qatar
Krzysztof Kecik, Poland
Mohamed Khader, Egypt
Chaudry M. Khalique, South Africa
Mukhtaj Khan, Pakistan
Shahid Khan, Pakistan
Nam-Il Kim, Republic of Korea
Philipp V. Kiryukhantsev-Korneev, Russia
P.V.V Kishore, India
Jan Koci, Czech Republic
Ioannis Kostavelis, Greece
Sotiris B. Kotsiantis, Greece
Frederic Kratz, France
Vamsi Krishna, India
Edyta Kucharska, Poland
Krzysztof S. Kulpa, Poland
Kamal Kumar, India
Prof. Ashwani Kumar, India
Michal Kunicki, Poland
Cedrick A. K. Kwuimy, USA
Kyandoghere Kyamakya, Austria
Ivan Kyrchei, Ukraine
Márcio J. Lacerda, Brazil
Eduardo Lalla, The Netherlands
Giovanni Lancioni, Italy
Jaroslaw Latalski, Poland
Hervé Laurent, France
Agostino Lauria, Italy
Aimé Lay-Ekuakille, Italy
Nicolas J. Leconte, France
Kun-Chou Lee, Taiwan
Dimitri Lefebvre, France
Eric Lefevre, France
Marek Lefik, Poland
Yaguo Lei, China
Kauko Leiviskä, Finland
Ervin Lenzi, Brazil
ChenFeng Li, China
Jian Li, USA
Jun Li, China
Yueyang Li, China
Zhao Li, China
Zhen Li, China
En-Qiang Lin, USA
Jian Lin, China
Qibin Lin, China
Yao-Jin Lin, China
Zhiyun Lin, China
Bin Liu, China
Bo Liu, China
Heng Liu, China
Jianxu Liu, Thailand
Lei Liu, China
Sixin Liu, China
Wanquan Liu, China
Yu Liu, China
Yuanchang Liu, United Kingdom
Bonifacio Llamazares, Spain
Alessandro Lo Schiavo, Italy
Jean Jacques Loiseau, France
Francesco Lolli, Italy
Paolo Lonetti, Italy
António M. Lopes, Portugal
Sebastian López, Spain
Luis M. López-Ochoa, Spain
Vassilios C. Loukopoulos, Greece
Gabriele Maria Lozito, Italy
Zhiguo Luo, China
Gabriel Luque, Spain
Valentin Lychagin, Norway
YUE MEI, China
Junwei Ma, China
Xuanlong Ma, China
Antonio Madeo, Italy
Alessandro Magnani, Belgium
Toqeer Mahmood, Pakistan
Fazal M. Mahomed, South Africa
Arunava Majumder, India
Sarfraz Nawaz Malik, Pakistan
Paolo Manfredi, Italy
Adnan Maqsood, Pakistan
Muazzam Maqsood, Pakistan
Giuseppe Carlo Marano, Italy
Damijan Markovic, France
Filipe J. Marques, Portugal
Luca Martinelli, Italy
Denizar Cruz Martins, Brazil

Yong (Aaron) Tan, United Kingdom
Marco Antonio Taneco-Hernández, Mexico
Lu Tang, China
Tianyou Tao, China
Hafez Tari, USA
Alessandro Tasora, Italy
Sergio Teggi, Italy
Adriana del Carmen Téllez-Anguiano, Mexico
Ana C. Teodoro, Portugal
Efstathios E. Theotokoglou, Greece
Jing-Feng Tian, China
Alexander Timokha, Norway
Stefania Tomasiello, Italy
Gisella Tomasini, Italy
Isabella Torcicollo, Italy
Francesco Tornabene, Italy
Mariano Torrisi, Italy
Thang nguyen Trung, Vietnam
George Tsiatas, Greece
Le Anh Tuan, Vietnam
Nerio Tullini, Italy
Emilio Turco, Italy
Ilhan Tuzcu, USA
Efstratios Tzirtzilakis, Greece
FRANCISCO UREÑA, Spain
Filippo Ubertini, Italy
Mohammad Uddin, Australia
Mohammad Safi Ullah, Bangladesh
Serdar Ulubeyli, Turkey
Mati Ur Rahman, Pakistan
Panayiotis Vafeas, Greece
Giuseppe Vairo, Italy
Jesus Valdez-Resendiz, Mexico
Eusebio Valero, Spain
Stefano Valvano, Italy
Carlos-Renato Vázquez, Mexico
Martin Velasco Villa, Mexico
Franck J. Vernerey, USA
Georgios Veronis, USA
Vincenzo Vespri, Italy
Renato Vidoni, Italy
Venkatesh Vijayaraghavan, Australia
Anna Vila, Spain
Francisco R. Villatoro, Spain
Francesca Vipiana, Italy
Stanislav Vítek, Czech Republic
Jan Vorel, Czech Republic
Michael Vynnycky, Sweden
Mohammad W. Alomari, Jordan
Roman Wan-Wendner, Austria
Bingchang Wang, China
C. H. Wang, Taiwan
Dagang Wang, China
Guoqiang Wang, China
Huaiyu Wang, China
Hui Wang, China
J.G. Wang, China
Ji Wang, China
Kang-Jia Wang, China
Lei Wang, China
Qiang Wang, China
Qingling Wang, China
Weiwei Wang, China
Xinyu Wang, China
Yong Wang, China
Yung-Chung Wang, Taiwan
Zhenbo Wang, USA
Zhibo Wang, China
Waldemar T. Wójcik, Poland
Chi Wu, Australia
Qiuhong Wu, China
Yuqiang Wu, China
Zhibin Wu, China
Zhizheng Wu, China
Michalis Xenos, Greece
Hao Xiao, China
Xiao Ping Xie, China
Qingzheng Xu, China
Binghan Xue, China
Yi Xue, China
Joseph J. Yame, France
Chuanliang Yan, China
Xinggang Yan, United Kingdom
Hongtai Yang, China
Jixiang Yang, China
Mijia Yang, USA
Ray-Yeng Yang, Taiwan

Contents

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 2557865, 12 pages
https://doi.org/10.1155/2022/2557865

Research Article

A Novel Methodology for Credit Spread Prediction: Depth-Gated Recurrent Neural Network with Self-Attention Mechanism

Xiao Liu[1] **Rongxi Zhou**[2] **Daifeng Qi**[3] **and Yahui Xiong**[4]

[1]*School of Economics and Management, North China University of Technology, Beijing 100144, China*
[2]*School of Banking and Finance, University of International Business and Economics, Beijing 100029, China*
[3]*Peking University HSBC Business School, Shenzhen 518055, Guangdong, China*
[4]*No.8 Department, 32180 Army, Beijing 100072, China*

Correspondence should be addressed to Yahui Xiong; yy11xyh@163.com

Received 27 August 2021; Revised 3 January 2022; Accepted 3 May 2022; Published 9 August 2022

Academic Editor: Yiwen Zhang

This paper develops a depth-gated recurrent neural network (DGRNN) with self-attention mechanism (SAM) based on long-short-term memory (LSTM)\gated recurrent unit (GRU) \Just Another NETwork (JANET) neural network to improve the accuracy of credit spread prediction. The empirical results of the U.S. bond market indicate that the DGRNN model is more effective than traditional machine learning methods. Besides, we discovered that the Depth-JANET model with one gated unit performs better than Depth-GRU and Depth-LSTM models with more gated units. Furthermore, comparative analyses reveal that SAM significantly improves DGRNN's prediction performance. The results show that Depth-JANET neural network with SAM outperforms most other methods in credit spread prediction.

1. Introduction

Credit spread is the risk premium demanded by investors of credit bonds over the yield of risk-free bonds of the same maturity, which is the basis of credit bond pricing and risk management. By grasping the future trend of credit spread, stakeholders involved in the bond market can make decisions more scientifically. For instance, investors can improve the accuracy of transactions. Financiers can choose time scientifically, and regulators can properly prevent and control financial risks. Besides, credit spread can be utilized to monitor macroeconomics and warn governments. However, because the bond market is usually regarded as a complex system [1], existing techniques cannot perform well in predicting the credit spread accurately. Therefore, it is necessary to theoretically and empirically discuss how to improve the credit spread's prediction accuracy.

As a representative technology of artificial intelligence, deep learning methods have developed rapidly in recent years [2–4]. Deep neural networks have become the most advanced forecasting method in finance due to its outstanding performance in time-series prediction [5, 6]. They have been widely used to predict indicators, such as stock prices, exchange rates, gold prices, and housing prices [5, 7–9].

Many studies show that deep neural networks can effectively fit complex nonlinear relationships between input variables with a higher fitting degree, reducing the overfitting of shallow foundations and local extremum problems. Besides, deep neural networks have no restrictions on the form of input variables. Therefore, all relevant information can be included. Particularly, deep neural networks can perform generalized learning based on data characteristics, weakening irrelevant information while learning heterogeneous information.

Existing literature on the prediction of the credit spread is mainly based on linear models [6, 10]. Although the deep learning methods can help improve the accuracy of credit spread prediction, it is not clear which algorithm has the best prediction performance according to the "No Free Lunch Theorem" proposed by Wolpert. Thus, it is also worthy of in-depth investigation of the performance of deep learning algorithms in credit spread prediction [11].

This paper aims to construct a depth-gated recurrent neural network with self-attention mechanism (SAM-DGRNN) to predict credit spreads in the U.S. corporate bond market. The main contributions are as follows: (1) to apply the XGBoost algorithm to integrate the selected credit spread determinants and extract the feature variables with the highest importance of prediction. (2) To construct depth-gated recurrent neural networks based on LSTM/GRU/JANET and compare them with three traditional nonlinear machine learning models (e.g., support vector regression, multilayer perception, and the random forest approaches) and a linear model (e.g., the vector autoregressive model). The comparative analysis of their prediction effects supports the superiority of deep learning methods in predicting credit spreads. (3) To construct a depth-gated recurrent neural network with self-attention mechanism (SAM-DGRNN) to explore the effectiveness of SAM in credit spread prediction.

The remainder of this paper is organized as follows. Section 2 provides a brief background. Section 3 discusses literature review related to the deep learning field. Section 4 introduces the theoretical methods and methodology for constructing the model of predicting the credit spread. Section 5 presents the experimental results. Section 6 is the conclusion section of this paper.

2. Background

2.1. Long-Short-Term Memory (LSTM) Neural Network. LSTM neural network has three gated units: input gate, forget gate, and output gate. The gated units allow information to affect recurrent neural networks at each moment selectively. Each gate outputs a value between 0 and 1. The value refers to how much information can be passed (0 means "no information can pass and one means "all information is allowed to pass"). The forget gate controls what information is discarded or saved from the cell state, the input gate controls how much new information is added to the cell state, and the output gate controls which part of the cell state will be output. The schematic diagram of LSTM neural network structure is shown in Figure 1. The update rules are shown in equations (1) to (6).

First, the forget gate discards useless historical information:

$$f_t = \sigma\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right). \tag{1}$$

Second, the input gate updates the state with input data and historical information:

$$i_t = \sigma\left(W_{ix}x_t + W_{ih}h_{t-1} + b_i\right), \tag{2}$$

$$\widetilde{c}_t = \tanh\left(W_{\widetilde{c}x}x_t + W_{\widetilde{c}h}h_{t-1} + b_{\widetilde{c}}\right), \tag{3}$$

$$c_t = f_t \times c_{t-1} + i_t \times \widetilde{c}_t, \tag{4}$$

Third, the output gate outputs current information:

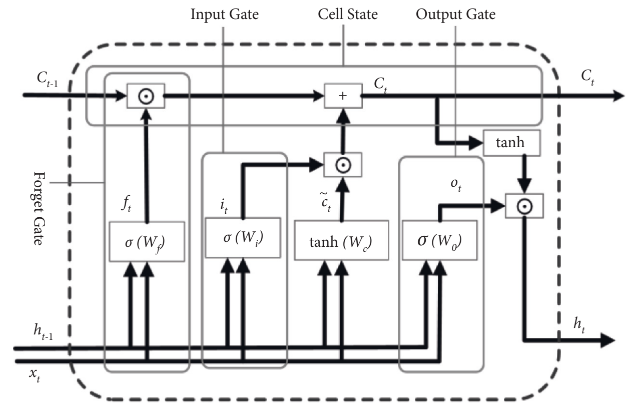


Figure 1: Structure of LSTM neural network.

$$o_t = \sigma\left(W_{ox}x_t + W_{oh}h_{t-1} + b_o\right), \tag{5}$$

$$h_t = o_t \times \tanh\left(c_t\right), \tag{6}$$

where x_t is the input vector at time t, h_{t-1} is the output vector at time $t-1$, $W_{fx}, W_{fh}, W_{ix}, W_{ih}, W_{\widetilde{c}x}, W_{\widetilde{c}h}, W_{ox}, W_{oh}$ are the weight matrixes, $b_f, b_i, b_{\widetilde{c}}, b_o$ are the bias vectors, σ is the logistic sigmoid function with the form of $\sigma(x) = (1+e^{-x})^{-1}$, tanh is the hyperbolic tangent activation function, f_t, i_t, o_t are the states of forget gate, input gate, and output gate at time t, respectively, and c_t is the state of the memory unit at time t.

2.2. Gated Recurrent Unit (GRU) Neural Network. GRU neural network consists of two gated units. The update gate is used to control the degree to which previous state information is brought into the current state. The smaller its value is, the less information it brings and the smaller the impaction the current hidden layer is. The reset gate is used to control the degree of state information that is ignored at the previous moment. The larger its value is, the less information is overlooked. GRU neural network synthesizes the input gate and the forget gate in LSTM neural network into a single update gate and combines the cell state and the hidden state. These features not only maintain the advantages of LSTM in solving long-term dependency problems but also lead to a more straightforward structure, with fewer parameters and higher training efficiency. The schematic diagram of GRU neural network structure is shown in Figure 2. Update rules are shown in equations (7) to (10).

First, the reset gate determines the degree of the alternative state $\widetilde{h}_t$ depending on the previous state h_{t-1}:

$$r_t = \sigma\left(W_{rx}x_t + W_{rh}h_{t-1} + b_r\right), \tag{7}$$

$$\widetilde{h}_t = \tanh\left(W_{\widetilde{h}x}x_t + W_{\widetilde{h}h}r_t h_{t-1} + b_{\widetilde{h}}\right). \tag{8}$$

Second, the update gate determines the weights of historical information inheriting from the previous state h_{t-1} and new information the current alternative state accepts:

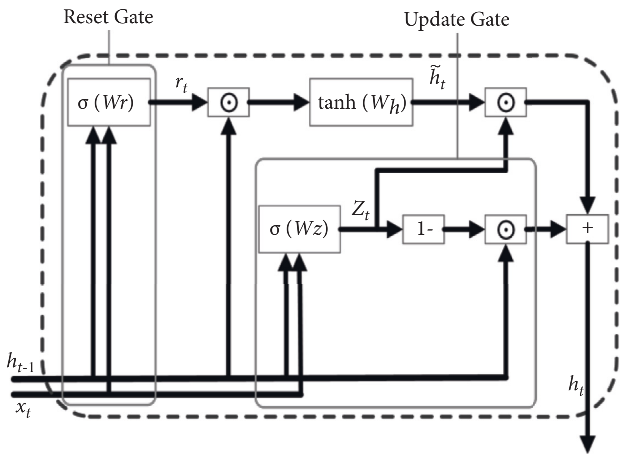


Figure 2: Structure of GRU neural network.

$$z_t = \sigma\left(W_{zx}x_t + W_{zh}h_{t-1} + b_z\right), \tag{9}$$

$$h_t = \left(1 - z_t\right) \times h_{t-1} + z_t \times \tilde{h}_t, \tag{10}$$

where x_t is the input vector at time t, h_{t-1} is the output vector at time $t-1$, $b_r, b_{\tilde{h}}, b_z$ are the bias vectors, $W_{rx}, W_{rh}, W_{\tilde{h}x}, W_{\tilde{h}h}, W_{zx}, W_{zh}$ are the weight matrixes, σ is the logistic sigmoid function, tanh is the hyperbolic tangent activation function, and r_t, z_t are the output state of the reset gate and update gate at time t, respectively.

2.3. Just Another NETwork (JANET) Neural Network. JANET neural network, with dramatically less training time, performs better on multiple benchmark data sets than LSTM neural network. The schematic diagram of JANET neural network structure is shown in Figure 3. The update rules are shown in equations (11) to (14).

$$f_t = \sigma\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right), \tag{11}$$

$$\tilde{c}_t = \tanh\left(W_{\tilde{c}x}x_t + W_{\tilde{c}h}h_{t-1} + b_{\tilde{c}}\right), \tag{12}$$

$$c_t = f_t \times c_{t-1} + \left(1 - f_t\right)\tilde{c}_t, \tag{13}$$

$$h_t = c_t, \tag{14}$$

where x_t is the input vector at time t, h_{t-1} is the output vector at time $t-1$, $W_{fx}, W_{fh}, W_{\tilde{c}x}, W_{\tilde{c}h}$ are the weight matrixes, $b_f, b_{\tilde{c}}$ are the bias vectors, σ is the logistic sigmoid function, tanh is the hyperbolic tangent activation function, f_t is the output state of the forget gate at time t, and c_t is the state of memory unit at time t.

2.4. Self-Attention Mechanism (SAM). SAM is an improvement in the basic attention mechanism. It uses an attention mechanism to dynamically generate the weights of different connections between the input and output of the same layer (not the model's final output) to obtain this layer's output. SAM considers the logical connection between the upper and lower sequences. Assume that the input

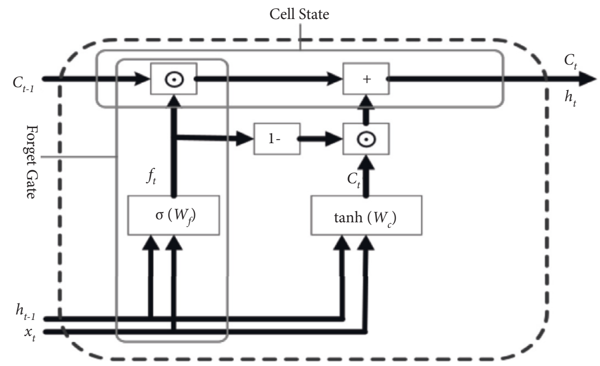


Figure 3: Structure of JANET neural network.

sequence of a layer is $X = [x_1, x_2, \ldots, x_N]$, and the output sequence is $H = [h_1, h_2, \ldots, h_N]$ with the same length. First, deduce three vector sequences through a linear transformation:

$$\begin{aligned} Q &= W_Q X \in R^{d_3 \times N}, \\ K &= W_K X \in R^{d_3 \times N}, \\ V &= W_V X \in R^{d_3 \times N}, \end{aligned} \tag{15}$$

where Q, K, V are the query vector sequence, key vector sequence, and value vector sequence, respectively, and W_Q, W_K, W_V are the learning parameters.

The output vector is

$$\begin{aligned} h_i &= att\left((K, V), q_i\right) = \sum_{j=1}^{N} a_{ij}v_j = \sum_{j=1}^{N} \operatorname{soft\,max}\left(s\left(k_j, q_i\right)\right)v_j \\ &= \sum_{j=1}^{N} \frac{\exp\left(s\left(k_j, q_i\right)\right)}{\sum_i \exp\left(s\left(k_i, q_i\right)\right)}v_j. \end{aligned} \tag{16}$$

where $(K, V) = [(k_1, y_1), (k_2, y_2), \ldots (k_N, y_N)]$ is the key-value pair, representing the input information. $i, j \in [1, N]$ are the positions of input and output vector sequences, and the connection weight α_{ij} is dynamically generated by the attention mechanism. softmax ensures that the sum of all weights is 1.

3. Related Work

Deep learning methods have been widely used in many fields. As one of the classic deep learning models, long-short term memory (LSTM) neural network has great advantage in mining long-term dependencies of sequence data. It was first proposed by Hochreiter and Schmidhuber to solve long-term memory problems in recurrent neural network by considering the "gated units" [12]. Wang et al. applied LSTM to speech enhancement and proposed a LSTM convolution network, which includes transpose convolution and jump connection [13]. Ma et al. introduced convolution operation into traditional LSTM and proposed a CLSTM learning algorithm to extract time-frequency information and obtained features through convolution [14]. Petmezas et al.

combined LSTM with convolutional neural networks (CNN) and proposed the CNN-LSTM model. Through CNN, signal features are transmitted to LSTM to realize dynamic memory [15]. Yu et al. applied LSTM to a nonlinear system model and proposed an improved depth LSTM. Combining the strengths of LSTM and multilayer perception, the stability of the training method is verified by the Lyapunov function. At the same time, the model is preferable to other existing models in a nonlinear system [16].

Due to many parameters involved, the LSTM neural network performs a lower training efficiency. To improve this drawback, Cho et al. proposed a more simplified gated recurrent unit (GRU) neural network based on LSTM neural network and proved that the prediction performance of the GRU Neural network is better than that of standard LSTM neural network [17]. Particularly, GRU can significantly simplify the structure of LSTM, reduce the number of parameters, and greatly shortens the training time. Liu et al. used GRU to replace the LSTM in the neural programmer interpreter for changing the core structure [18]. Based on the classification results of LSTM and full convolution network LSTM-FCN, Nelsayed et al. found that GRU has higher classification accuracy and simpler hardware implementation in time-series classification problems, which are of smaller architecture and less computation [19]. Wu et al. combined GRU with CNN to propose a GRU-GNN hybrid neural network model. In the GRU-GNN model, GRU is responsible for extracting the feature vector of time-series data, and CNN extracts the feature vector of high-dimensional data [20]. Pan et al. applied the GRU-GNN combined model to the water level prediction of the Yangtze River. Through the 30-year water level data of the Yangtze River and comparative analysis, it is confirmed that the model is superior to wavelet neural network, LSTM, and statistically integrated moving average autoregressive model ARIMA [21]. Given the excellent performance of the GRU neural network after eliminating redundant gates, Westhuizen and Lasenby further explored the necessity of three gated units in the LSTM neural network to build more efficient models [22]. They proposed a JANET (Just Another NETwork) with only a forget gate and chronologically initialized bias terms.

Attention mechanisms are widely used in neuroscience and computational neuroscience. This common mechanism comes from the fact that many animals only focus on specific parts of their vision to give enough response. Therefore, many neural computing studies have concluded that people only need the most relevant information, rather than all information, for further neural processing. In recent years, this mechanism has also been widely used in deep learning research, such as image re-rolling and voice recognition. Recent studies have found that considering the self-attention mechanism in deep learning can effectively extract the most critical information for current tasks to enhance predictive power. Attention mechanism has become one of the most important topics in the deep learning literature following the research by Vaswani et al. [23]. Zhao et al. designed a long-short term memory (LSTM) neural network structure model with attention mechanism based on the dynamic sequence in the internet financial market [24]. The empirical results showed that their attention mechanism model outperformed others. Chen and Ge applied an LSTM neural network with attention mechanism to predict the stock price trend in Hong Kong and achieved satisfactory prediction results [25].

In the training of deep neural network models, the gradient vanishing and overfitting often result in unsatisfactory learning effects. Studies have shown that batch-normalization (B. N.) method can alleviate the gradient vanishing by pulling the data back to a standard normal distribution with a mean of 0 and a variance of 1 [26]. Furthermore, Dropout can prevent overfitting to a certain extent by preventing neuronal coadaptation during the training phase [27]. However, improper use of both methods will generate the opposite effect. Li et al. found that placing Dropout in all B. N. layers or modifying Dropout's formula to reduce the sensitivity of variance could improve the coordination between B. N. and Dropout [28]. Luo et al. suggested that by adopting differentiable learning, the switchable-normalization method (S. N.) could determine the appropriate normalization operation for each normalization layer in a deep network [29]. As a result, it is more advantageous than B. N. in avoiding gradient disappearance. Therefore, in our deep neural network design, we add the S. N. layer and Gaussian Dropout layer to optimize its structure. A reasonable combination of the S. N. layer and Dropout layer will improve the performance of the neural networks.

4. Methodology

4.1. Depth-Gated Recurrent Neural Network with Self-Attention Mechanism (SAM-DGRNN). We add the S. N. layer and Gaussian dropout layer to optimize its structure in the deep neural network. A reasonable combination of the S. N. layer and dropout layer will improve the performance of the neural networks. Specifically, the main structure of the depth-gated recurrent neural network constructed in this paper includes an attention mechanism layer, a three-layer LSTM/GRU/JANET neural layer, and two fully connected layers (of which the first neural layer has 128 neurons, the second has 64 neurons, the third has 32 neurons, and the two fully connected layers have 32 neurons and one neuron, respectively). An S.N. layer is added in front of each LSTM/GRU/JANET neural layer. A Gaussian Dropout layer is added at the back of the LSTM/GRU/JANET neural layer, and the drop rate is set to 0.2. The structure of the deep LSTM/GRU/JANET neural network is shown in Figure 4, and the neural network structure is shown in the dotted box.

4.2. Training Method, Loss Function, and Optimizer Selection. We apply the mini-batch gradient descent method to train the deep learning neural network. In order to predict future credit spreads, the mean square error (MSE) is selected in the loss function. We choose Adam optimizer (adaptive moment estimation) to perform optimization training. Compared with other self-adaptive learning rate algorithms, the Adam algorithm is more robust in selecting hyperparameters, with higher training efficiency, and can generate more effective results [30]. The experimental environment of this paper is shown in Table 1.

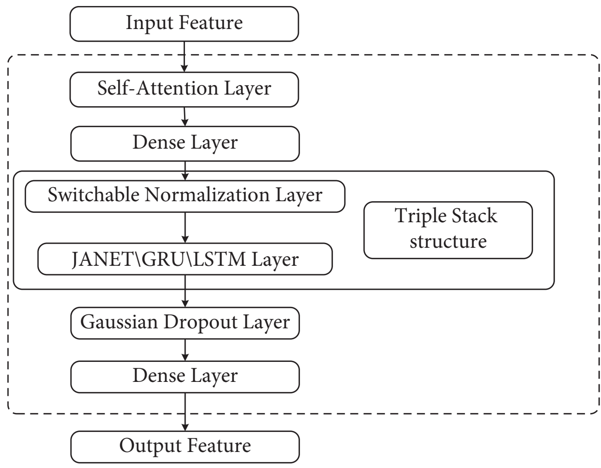


Figure 4: Depth-gated recurrent neural network with self-attentional mechanism.

Table 1: Experiment environment.

Operating system	Windows 7
Processor	Intel(R) core (TM) i7-2760QM
Random-access memory	8.00 GB
Version of python	Python 3.7.0
Version of spyder	Spyder 3.3.1
Version of keras	Keras 2.2.4
Version of tensorflow	Tensorflow 1.13.1
Optimizer	Adam optimizer

4.3. Control Models. In the depth-gated recurrent neural network model, we employ the rolling-window prediction method and use the data of n trading days to predict the credit spread on the next day. This paper first integrates indicators with XGBoost (extreme gradient boosting) algorithm, extracting the predictor variables with higher importance ranking. XGBoost algorithm combines random forest algorithm, which can further reduce calculation complexity. When dealing with a large amount of data, XGBoost can operate parallel and divide the data according to different characteristics to form a tree sequence. This algorithm is simpler and more effective. It can transfer complex data to an orderly and concise arrangement form. Then the feature variables with a higher importance ranking are selected as the model input.

To comprehensively evaluate the prediction effect of depth-gated recurrent neural network, one benchmark deep learning model RNN and three traditional machine learning models (support vector machines (SVR), multilayer perceptron (MLP), and random forest (RF)) in financial prediction are selected as nonlinear control models. Research has put VAR as a linear control model. [6]. Deep RNN selection and depth-gated recurrent neural network have the same structure. The parameter combination in SVR is set as "Radial Basis Function (RBF), penalty parameter $C = 1$, gamma = auto". We also select the classic MLP neural network with three layers. The prediction methods of RNN, SVR, MLP, and R. F. are consistent with depth-gated recurrent neural network. The prediction idea of a VAR model is as follows: first, the stationarity of all sequences is comprehensively judged by the ADF test, KPSS test, and P. P. test and decide whether to carry out the corresponding order difference to obtain the stationary sequence according to the test results. Second, the VAR model is established. The order of the VAR model is determined by integrating AIC and BIC information criteria. Third, the VAR model was estimated, and the model's stability was tested. Finally, the credit spread sequence is predicted based on the stable VAR model. The prediction flowchart is shown in Figure 5.

5. Empirical Prediction Analysis

5.1. Variable Selection. We collect daily closing data from 2009 to 2019. The 2517 trading days during this period are divided into a training set (includes the first 85% of trading days) and a test set (includes the remaining 15% of trading days).

Table 2 shows variables in the literature that have been verified as significant credit spread determinants. The credit spread sequence is a forecast indicator, and additional variables are used as the characteristics to predict credit spreads. The detailed indicators are discussed as follows.

Risk-free interest rate term structure: the risk-free interest rate is an important variable in the structural model. The information contained in the shape of the riskless yield curve can improve the prediction performance of credit spreads [6].

Credit spread term structure: the credit spread curve's level, slope, and curvature are the principal variables for predicting the future credit spread [6].

Fama-French factor returns: credit spreads indicate the extra compensation of holding risky assets as an analogy to stock risk premiums. Therefore, financial markets would transfer the explanatory power of stock returns, represented by Fama-French factor returns, to the bond market [31].

Return on Stock Index: stocks are also yield-producing securities. The equity market is the most plausible alternative to the fixed-income market, and equity market indexes measure capital market investment levels. Therefore, the return on the stock index could be relevant to corporate bond credit spreads [32].

Volatility of Stock Index: VIX Index, often referred to as the market's "fear gauge," can be correlated with credit spreads, which capture the future probability of default as a common forward-looking risk metric. As a result, stock market volatility is a significant variable for explaining credit spread changes [33].

Exchange rate: the prevailing economic theory, such as uncovered interest rate parity, suggests that there should be an empirical relationship between exchange rates and interest rates. Given exchange rate fluctuations, foreign investors will be attracted to invest in U.S. corporate bonds. The foreign exchange rate is a heretofore overlooked variable for explaining credit spread changes [33].

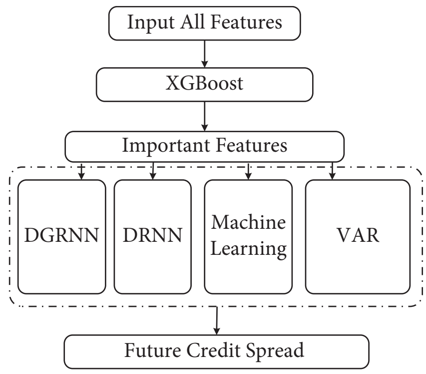


Figure 5: Prediction flowchart.

Table 2: Variables and data.

Variables			Data
Credit spread			ICE BofAML US corporate master option-adjusted spread[$]
Financial market factors	Credit spread term structure	Level factor	The first principle component extracted From a large cross section of credit spreads of various maturities and various ratings[$]
		Slop factor	The second principal component extracted From a large cross section of credit spreads of various maturities[$]
		Curve factor	The third principle component extracted From a large cross section of credit spreads of various maturities[$]
	Risk-free interest rate term structure	Level factor	The long-term factor of Nelson and Siegel term structure decomposition extracted from a large cross section of treasury yields of various maturities[$$]
		Slop factor	The short-term factor of Nelson and Siegel term structure decomposition extracted from a large cross section of treasury yields of various maturities[$$]
		Curve factor	The mid-term factor of Nelson and Siegel term structure decomposition extracted from a large cross section of treasury yields of various maturities[$$]
	Fama-French factor returns		Excess return on the market Small-minus-big return High-minus-low return
	Return on stock index		Return on S&P 500
	Volatility of stock index		Volatility of S&P 500
	Exchange rate		U.S. Dollar index
	Oil prices		Crude oil prices: West Texas intermediate (WTI)
	TED spread		Difference between 3-month LIBOR based on U.S. dollars and 3-month treasury bill
	Swap spread		Difference between10-year swap rate and 10-year treasury yield
	Commodity price index		RJ/CRB index

Note: Data are obtained from https://fred.stlouisfed.org; $ ICE BofAML US Corporate 1–3\3–5\5–7\7–10\10–15\15+ \AAA\AA\A\BBB\BB\B\CCC Option-Adjusted Spread;**$$** 3\6-month and 1\2\3\5\10-year treasury yields.

Oil Prices: energy prices, as the cost of economic activities, are captured by oil prices to study their influence on credit spreads.

TED spread: TED spread captures additional macroeconomic and interest rate information from international fixed-income markets. LIBOR and U.S. treasury yields are often used to price complex financial derivative products, and the difference is an important predictive variable [34].

Swap spread: swap spread is highly correlated with credit spreads because it is a proxy for credit rate. Since the swap market is more well developed and liquid than corporate bonds, swap rates may provide a forward indication of credit spreads. Credit spreads will increase with swap spreads [35].

Commodity Price Index: it is widely used to analyze price fluctuations in commodity markets and macroeconomy. CPI index is a better indicator of inflation [6].

The importance score based on the XGBoost algorithm is shown in Figure 6. The y-axis represents the feature, x-axis represents the importance score, and the score is between 0 and 1.

Figure 6 shows the mutual information of selected features. To avoid disturbance from insignificant features, with 0.01 as the cut-off point of importance score, we select ten features with the highest mutual information from the

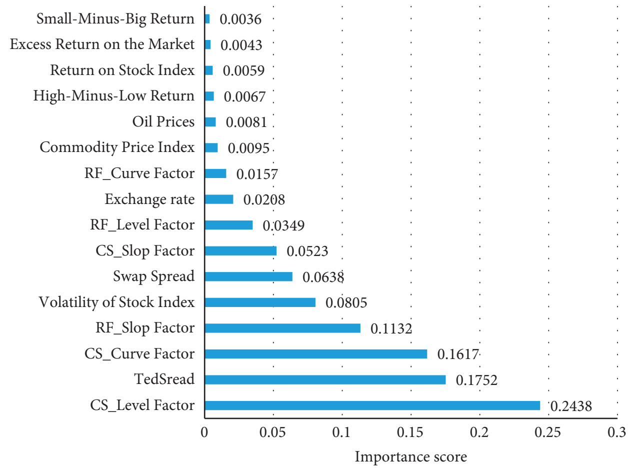


Figure 6: Mutual information for selected features. (There is no doubt that the term structure of credit spread is the most important factor to predict credit spread. In order to visualize the importance of the other factors in the diagram, we only examine the importance of the remaining factors here.)

raw feature set. Features with higher mutual information are more helpful to determine future spreads.

5.2. Model Evaluation and Result Comparison

5.2.1. Evaluation Indexes for Prediction Results. In this paper, we apply three indicators, including MAE (mean absolute error), MAPE (mean absolute percentage error) and RSR (The classification of RSR values by Moriasi et al. (2007): when $\text{RSR} \leq 0.5$, the prediction performance is excellent; when $0.5 \leq \text{RSR} \leq 0.6$, the prediction performance is good; when $0.6 \leq \text{RSR} \leq 0.7$, the prediction performance is at an average level; when $\text{RSR} > 0.70$, the prediction performance is poor) (root mean square error (RMSE) divided by standard deviation), to evaluate prediction accuracy. The smaller the value is, the higher accuracy the prediction will have. SDAPE (standard deviation of mean absolute percentage error) is used to evaluate the prediction stability. The smaller the SDAPE value is, the better the prediction stability will be. The calculation formulas of evaluation indicators are shown in formulas (17)–(21):

$$\text{MAE} = \frac{1}{N} \times \sum_{i=1}^{N} (y_i - \hat{y}_i), \tag{17}$$

$$\text{MAPE} = \frac{1}{N} \times \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \tag{18}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}, \tag{19}$$

$$\text{RSR} = \frac{\text{RMSE}}{\text{STD}}, \tag{20}$$

$$\text{SDAPE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \left(\left| \frac{y_i - \hat{y}_i}{y_i} \right| - \text{MAPE} \right)^2 \times 100\%}, \tag{21}$$

where y_i and $\hat{y}_i$ represent the actual value and the predicted value of credit spreads, respectively, STD is the standard derivation of the actual value, and N is the sample size.

Although the indicators are widely used to compare prediction accuracy, their values alone are insufficient to determine models' prediction ability. We also conduct D. M. statistical tests with these indicators as the basic loss function to analyze statistical significance [36]. The idea of the D. M. statistical test is as follows. For a set of actual time series$\{y_t\}_{t=1}^{T}$, the estimated values for the two models are $\{\hat{y}_{it}\}_{t=1}^{T}$ and $\{\hat{y}_{jt}\}_{t=1}^{T}$, whose error sequences are $\{e_{it}\}_{t=1}^{T}$ and $\{e_{jt}\}_{t=1}^{T}$, and whose loss functions are $g(y_t, \hat{y}_{it}) \equiv g(e_{it})$ and $g(y_t, \hat{y}_{jt}) \equiv g(e_{jt})$, respectively. As a result, their relative loss functions can be expressed as $d_t = g(e_{it}) - g(e_{jt})$. The null hypothesis is that the two models' prediction abilities are not different, expressed as $E(d_t) = 0$. If the loss-differential series $\{d_t\}_{t=1}^{T}$ is covariance stationary and short memory, then standard results may be used to deduce the asymptotic distribution of the sample mean loss differential.

We have $\sqrt{T}(d_{\text{mean}} - \mu) \xrightarrow{d} N(0, 2\pi f_d(0))$, where $d_{\text{mean}} = 1/T \sum_{t=1}^{T} d_t$ is the sample mean loss differential. $f_d(0) = 1/2\pi \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau)$ is the spectral density of the loss differential at frequency 0, and $\gamma_d(\tau) = E[(d_t - \mu)(d_{t-\tau} - \mu)]$ is the autocovariance of the loss differential at displacement γ, and τ is the population mean loss differential. The formula of $f_d(0)$ shows that the correction for serial correlation can be substantial, even if the loss differential is only weakly serially correlated, due to the accumulation of the autocovariance terms.

Because in large samples the sample means loss differential d_{mean} is approximately normally distributed with mean μ and variance $2\pi f_d(0)/T$, the obvious large-sample $N(0,1)$ statistic for testing the null hypothesis of equal forecast accuracy is $\text{DM} = d_{\text{mean}} / (2\pi \hat{f}_d(0)/T)$, where $\hat{f}_d(0)$ is a consistent estimate of $f_d(0)$. If the absolute value of D. M. statistic is significantly greater than the critical value, the null hypothesis is rejected, indicating that the two models' predictive abilities are significantly different.

5.2.2. Prediction Performance of Depth-Gated Recurrent Neural Network (DGRNN). We have the following discoveries about repeated experiments. (a) The deep learning model is sensitive to the number of traversals (the value of hyperparameter epochs); the prediction effectiveness of the same model is in a U-shaped relationship with epochs value; when $\text{epochs} = 100 \pm 10$, the deep learning models perform best in the experiments; the machine learning models are not sensitive to hyperparameter epochs. In this paper, the number of traversals is 100 ($\text{epochs} = 100$). (b) Several representative values (1, 5, 20, 60, 120, 180, and 250) were selected to test the sensitivity of the hyperparameter look_back, which determines the number of trading days

Table 3: Statistical performance indicators of the prediction models.

	DGRNN				Machine learning			VAR
	DJANET	DGRU	DLSTM	DRNN	SVR	MLP	RF	
MAE	**5.6770**	7.7149	8.4230	11.9192	12.0704	10.1098	9.8887	12.7386
DM	—	1.8761*	1.9799**	2.5902***	2.7406***	2.4763**	2.1659**	3.1721***
MAPE	**0.0479**	0.0635	0.0728	0.1046	0.2271	0.0863	0.0763	0.3826
D.M.	—	1.6973*	1.9477*	3.7901***	5.6809***	2.7217***	2.0705**	8.3050***
RSR	**0.4198**	0.5549	0.5995	0.8277	1.6920	0.7263	0.7076	1.8340
SDAPE	**0.0316**	0.0437	0.0442	0.0570	0.1154	0.0526	0.0471	0.2008

Note:*, **, and *** indicate that D.M. statistics are significant at the 10%, 5%, and 1% levels, respectively; DJANET is the benchmark in the D.M. test; boldface represents the optimal value under different evaluation criteria; the following tables are the same.

Table 4: Results of the prediction models (loss function = "MAE")

	True value	DJANET	DGRU	DLSTM	DRNN	SVR	MLP	RF	VAR
9/2017	112.89	114.51#	109.59	118.56	115.36	108.56	116.97	117.694	121.44
10/2017	107.45	106.21	107.89#	113.95	112.15	106.05	113.28	112.76	115.04
11/2017	104.42	105.73	105.31#	110.77	108.05	101.11	106.83	109.22	112.30
12/2017	100.40	100.97#	102.42	107.45	105.83	97.19	105.64	105.71	111.58
1/2018	94.71	94.03	98.69	101.68	100.32	94.72#	101.68	100.51	106.28
2/2018	96.52	95.70#	103.21	106.88	103.22	99.78	98.77	111.14	103.21
3/2018	109.66	104.35	111.59	116.20	111.28	110.32#	110.56	112.88	114.01
4/2018	112.47	112.63#	118.29	123.56	116.87	117.08	115.34	124.02	116.36
5/2018	115.95	116.33#	110.80	113.83	110.56	113.03	117.52	111.91	122.35
6/2018	123.95	122.61	117.84	121.34	118.13	117.28	124.91#	116.67	131.73
7/2018	122.95	124.95#	117.51	119.97	117.96	117.32	127.17	117.46	131.63
8/2018	117.78	116.48#	113.22	115.88	115.23	112.03	122.08	115.09	127.66
9/2018	115.94	117.47	112.53	113.81	113.92	109.91	121.54	116.97#	125.37
10/2018	117.09	115.53	113.99	113.60	113.21	110.01	116.28#	123.65	122.99
11/2018	132.20	130.91#	125.84	130.23	125.76	124.81	130.04	130.63	135.14
12/2018	151.47	150.50#	149.44	158.19	146.64	150.15	149.32	152.88	152.51
1/2019	150.14	151.70	150.50#	159.20	148.31	148.73	147.88	149.66	154.39

Table 5: Results of the different lengths of time steps (loss function = "SDAPE")

	DGRNN			
	DJANET	DGRU	DLSTM	DRNN
5	0.0467	0.0593	0.0949	0.1176
20	0.0738	0.1313	0.1321	0.1400
60	0.0783	0.1005	0.1019	0.1105
120	0.0856	0.0836	0.0977	0.1155
180	0.0719	0.0894	0.0925	0.0976

used to predict the credit spreads of the next day. We find that the prediction effect of the same model and the value of look_back show W-type characteristics. All models perform best when look_back = 5, indicating that the historical data of the previous five trading days already contain enough information. Too few trading days result in insufficient information, while too many bring extra noise. Therefore, we set the look_back parameter to 5 in the subsequent analysis. The prediction results with the parameter combination [epochs, look_back] as [100, 5] are shown in Tables 3 and 4.

As can be seen from Table 3, all D. M. test results reject the null hypothesis, indicating that the predictive power of these models is significantly different. Depth-gated recurrent neural networks (DJANET/DGRU/DLSTM) are superior to nongated deep recurrent neural networks (DENN), such as the traditional machine learning models (SVR/MLP/RF) and linear prediction model (VAR), in credit spread prediction of U.S. bond market in terms of accuracy and stability. Furthermore, the RSR value of the DGRNN model is less than 0.6, suggesting that the DGRNN model also performs better in absolute dimensions according to the classification standard of RSR metric value by Moriasi et al. [37]. Besides, D.M. test results also show that the null hypothesis is rejected at least at the significance level of 10%, indicating that the predictive power of the DJANET model is significantly different from that of the other models. In the deep learning model with the gated unit mechanism, the DJANET model with one gated unit performs better than the DGRU and DLSTM models, which have more gated units. Furthermore, the DGRU model with two gated units is better than the DLSTM model with three gated units.

Table 6: Impact of SAM on prediction results

	SAM-DJANET	DJANET	SAM-DGRU	DGRU	SAM-DLSTM	DLSTM
MAE	**2.8122**	5.6770	3.2107	7.7149	5.8179	8.4230
DM	2.7254***		3.2561***		2.2883**	
MAPE	**0.0215**	0.479	0.0275	0.0635	0.0502	0.0728
DM	3.1083***		4.1697***		2.1168**	
RSR	**0.3009**	0.4198	0.3733	0.5549	0.4311	0.5995
SDAPE	**0.0219**	0.0316	0.0260	0.0437	0.0322	0.0442
Time (s)	389.65	306.24	494.68	391.04	574.88	471.60

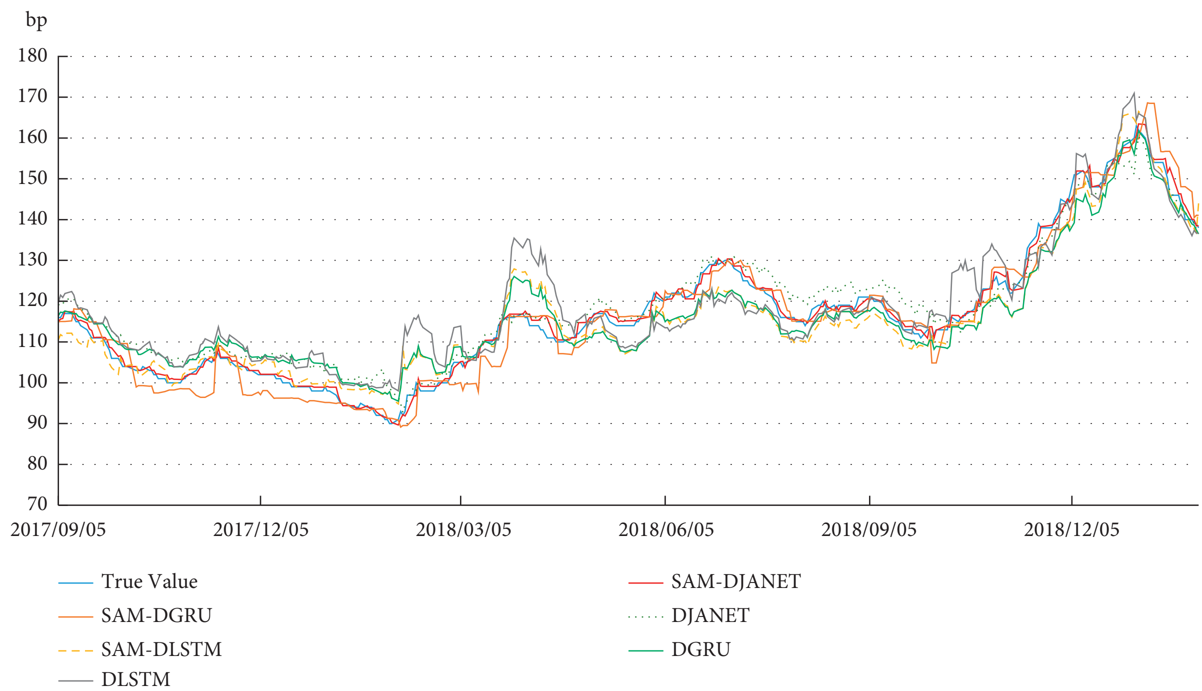


Figure 7: Comparison of DGRNN prediction curves on the prediction set.

As shown in Table 4, in 366 trading days (17 months), there are nine best-predicted values (the value with "#" in the table, the deviation of them from their actual values is the smallest) in the DJANET model and 12 best-predicted values in DGRNN model, accounting for 70.59% of the total values, showing that DGRNN model has a better prediction effect.

We compared the performance of models DJANET, DGRU, DLSTM, and DRNN at different lengths of time steps in Table 5, where SDAPE (standard deviation of mean absolute percentage error) is used to evaluate the prediction stability. The smaller the SDAPE value is, the better the prediction stability will be. As we can see, the values of DJANET are larger when the lengths of time steps are from 5 to 120 in Table 5. However, the value of DJANET is lower when the length of time steps is 180. DJANET has only a forget gate and chronologically initializes bias terms. The short-term prediction is effective and relatively loses the long-term information content. Credit spreads are cyclical, so the results will be similar to those in the short term when the ultralong term is 180 days. At the same time, there are different performances on DGRU and DLSTM. For example, the values of DGRU are lower when the lengths of time steps are from 20 to 120, and the values of DLSTM are lower when the lengths of time steps are from 20 to 180. DGRU is a variant of DLSTM, and their performance is equal in many tasks. DLSTM can learn the characteristics of long-term trend series, so its prediction efficiency is improved with the increase in time length. We discover that no matter how much the length of time steps is, the Depth-JANET model with one gated unit performs better than Depth-GRU and Depth-LSTM models, which have more gated units. However, as the length of time steps increases, the prediction accuracy will decrease.

5.3. *Effectiveness of SAM.*

Table 6 reports the predicted performance of the depth-gated recurrent neural network with SAM. It can be seen from Table 6 that the four evaluation indicators of the SAM-DGRNN model are all smaller than the DGRNN model, and the D. M. test results also show that the null hypothesis is rejected at least at the significance level of 5%, indicating that depth-gated recurrent neural network with SAM performs better than the models without the mechanism. It also suggests that SAM can improve the

Table 7: Prediction performance of DGRNN models (train_ratio = 0.75).

	SAM-DJANET	DJANET	SAM-DGRU	DGRU	SAM-DLSTM	DLSTM
MAE	**3.031**	6.202	3.924	7.747	7.056	10.560
DM	2.977***		2.635***		2.472**	
MAPE	**0.026**	0.058	0.032	0.067	0.063	0.086
DM	3.063***		2.894***		2.117**	
RSR	**0.311**	0.452	0.380	0.602	0.477	0.790
SDAPE	**0.022**	0.032	0.025	0.046	0.034	0.053

Table 8: Prediction performance of DGRNN models (sample = [2015, 2018]).

	SAM-DJANET	DJANET	SAM-DGRU	DGRU	SAM-DLSTM	DLSTM
MAE	**2.949**	5.942	3.645	7.814	6.879	9.329
DM	2.819***		3.108***		2.066**	
MAPE	**0.024**	0.057	0.030	0.070	0.061	0.079
DM	3.326***		2.934***		1.947**	
RSR	**0.307**	0.451	0.378	0.632	0.468	0.706
SDAPE	**0.022**	0.031	0.026	0.046	0.033	0.051

performance of depth-gated recurrent neural network in predicting credit spreads.

Figure 7 further shows the fitting curve of the DGRNN model on the prediction set in the U.S. credit spread. It can be seen that the SAM-DJANET curve fits best.

Li believed that simply assigning statistical data to random without testing its certainty and randomness will lead to large deviations between the predicted results and the actual values [38]. In other words, the prediction outcome and accuracy are largely dependent on a reasonable prediction model and the randomness of the original data of the predicted variable. If the original data exhibits a certain logical change and is less random, adopting an appropriate prediction model will inevitably improve prediction performance and have higher accuracy. Therefore, following the method in Tang et al. [39], this paper applies the Ljung-Box statistic to conduct a random independence test on the original credit spread data. The test results show that its Ljung-Box statistic value is 3519.1, and the p-value is 0.00, and the original hypothesis is rejected at a 1% significant level. The results indicate that the original data are not randomly independent, laying a statistical foundation for exploring the best fitting model. The proper models can be extrapolated extensively, and prediction results in the test set should be very close to the actual real value. What is more, these models can extract the original data information. Furthermore, their residual sequences are white noise sequences, which meet random independence.

To further test the prediction extrapolation ability of the model, we perform a Ljung-Box test on the residual sequence of the prediction set in the SAM-DJANET model. The Ljung-Box statistic value is 3.07, and the p-value is 0.19, indicating that the null hypothesis cannot be rejected even at the 10% significance level. The results suggest that the residual sequence is a white noise sequence, confirming that depth-gated recurrent neural network with self-attention mechanism (SAM-DGRNN) has good predictive extrapolation ability and rationality.

5.4. Robustness Test. This paper conducts a robustness test from the following two aspects: (1) setting the cutting point of the training set and prediction set to 3 : 1, and the results are shown in Table 7; (2) shortening the sample interval to 2015–2018 and the results are shown in Table 8. According to the evidence in Tables 7 and 8, the empirical results of this paper are robust.

6. Conclusions

Traditional prediction data and technologies are insufficient to forecast the credit spread accurately, particularly when using big data. Considering the nonlinear changes in credit spreads, this paper introduces a deep learning algorithm to build depth-gated recurrent neural network with a self-attention mechanism. Additionally, it compares various prediction methods. We choose multiple evaluation indicators and randomness tests for predicted variables' original data to conduct a comparative analysis. The conclusions are as follows.

First, traditional intelligent algorithms such as machine learning and deep learning can capture nonlinear relationships better than linear algorithms. The prediction results of credit spread prediction indicate that deep learning models (LSTM, GRU, and JANET) and traditional machine learning models (SVR, MLP, and RFR) are better than the VAR model. Second, deep learning models with gated unit mechanisms are extremely advantageous in mining the long-term dependence of sequence data. The results show that the deep learning models with gated unit mechanisms have better prediction accuracy and higher stability than those without gated. Third, when predicting credit spreads, JANET, the latest deep learning model that has only one gated unit, excels in prediction efficiency, accuracy, and stability compared with the earlier models, which have more gated units, such as LSTM and GRU. GRU model with two gated units is superior to the LSTM model with three gating units. Fourth, deep learning models with SAM can efficiently filter out critical information to the current task. The prediction of the credit spreads in the U.S. bond market shows that the deep learning model based on the attention mechanism has better prediction performance than that without the mechanism.

In summary, by comparing each model's prediction results and robustness tests through statistical performance indicators, we confirm that depth-gated recurrent neural network (DGRNN) is an effective prediction method for the U.S. bond credit spread. SAM-DGRNN model can further improve prediction performance. Among the three gated recurrent neural network models, the SAM-DJANET has the highest prediction accuracy, stability, and efficiency. The prediction results can provide a reference for the decision-making of market participants and regulatory authorities in the U.S. bond market.

However, we consider some factors that may affect the credit spread, and there may be other influencing factors. So, we can explore adding more relevant variables to improve the forecasting effect. We can further find a more accurate model in a certain type of credit spread according to the maturity, rating, and industry. In addition, these principles and forecasting methods can extend to the relevant problems of financial time series.

Data Availability

The data used to support the findings of the study were obtained from https://fred.stlouisfed.org.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research was supported by the National Natural Science Foundation of China under Grant no. 71871062.

References

[1] R. X. Zhou, S. N. Du, M. Yu, and F. Yang, "Pricing credit spread option with Longstaff-Schwartz and GARCH models in Chinese bond market," *Journal of Systems Science and Complexity*, vol. 28, no. 6, pp. 1363–1373, 2015.

[2] A. Shresthaand and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, Article ID 53040, 2019.

[3] G. Shan, H. Wang, W. Liang, and K. Chen, "Robust encoder-decoder learning framework for offline handwritten mathematical expression recognition based on a multi-scale deep neural network," *Science China Information Sciences*, vol. 64, no. 3, Article ID 139101, 2021.

[4] J. Wu and H. Li, "Deep ensemble reinforcement learning with multiple deep deterministic policy gradient algorithm," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4275623, 12 pages, 2020.

[5] A. S. Temür, M. Akgün, and G. Temür, "Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models," *Journal of Business Economics and Management*, vol. 20, no. 5, pp. 920–938, 2019.

[6] C. N. V. Krishnan, P. H. Ritchken, and J. B. Thomson, "Predicting credit spreads," *Journal of Financial Intermediation*, vol. 19, no. 4, pp. 529–563, 2010.

[7] J. Zhai, C. Yang, and G. Tao, "Deep learning for price movement prediction using convolutional neural network and long short-term memory," *Mathematical Problems in Engineering*, vol. 2020, Article ID 2746845, 13 pages, 2020.

[8] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.

[9] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: a hybrid model integrating lstm with multiple garch-type models," *Expert Systems with Applications*, vol. 103, no. 8, pp. 25–37, 2018.

[10] G. Tauchen and H. Zhou, "Realized jumps on financial markets and predicting credit spreads," *Journal of Econometrics*, vol. 160, no. 1, pp. 102–118, 2011.

[11] D. H. Wolpert, "The lack of a priori distinctions between learning algorithms," *Neural Computation*, vol. 8, no. 7, pp. 1341–1390, 1996.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] Z. Y. Wang, T. Zhang, Y. Y. Shao, and B. Ding, "LSTM-convolutional-BLSTM encoder- decoder network for minimum mean-square error approach to speech enhancement," *Applied Acoustics*, vol. 172, Article ID 107647, 2021.

[14] M. Ma and Z. Mao, "Deep-convolution-based LSTM network for remaining useful life prediction," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1658–1667, 2021.

[15] G. Petmezas, K. Haris, L. Stefanopoulos et al., "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets," *Biomedical Signal Processing and Control*, vol. 63, Article ID 102194, 2021.

[16] W. Yu, J. Gonzalez, and X. O. Li, "Fast training of deep LSTM networks with guaranteed stability for nonlinearsystem modeling," *Neurocomputing*, vol. 422, pp. 85–94, 2021.

[17] K. Cho, B. V. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, and H. Schwenk, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, https://arxiv.org/abs/1406.1078.

[18] W. Liu, Q. Wang, Y. L. Zhu, and H. Chen, "GRU: optimization of NPI performance," *The Journal of Supercomputing*, vol. 76, no. 5, pp. 3542–3554, 2020.

[19] A. Nelsayed, S. Maida, and M. Bayoumi, "Gated recurrent neural networks empirical utilization for time series Classification," in *Proceedings of the IEEE Int Congress on Cybermat*, pp. 1207–1210, IEEE, Atlanta. New York, NY, USA, July 2019.

[20] L. Z. Wu, C. Kong, X. H. Hao, and W. Chen, "A short-term load forecasting method based on GRU-CNN hybrid neural network model," *Mathematical Problems in Engineering*, vol. 2020, Article ID 1428104, 10 pages, 2020.

[21] M. Y. Pan, H. N. Zhou, J. Y. Cao et al., "Water level prediction model based on GRU and CNN," *IEEE Access*, vol. 8, Article ID 60090, 2020.

[22] J. Westhuizen and J. Lasenby, "The unreasonable effectiveness of the forget gate," 2018, https://arxiv.org/abs/1804.04849.

[23] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 6000–6010, 2017.

[24] H. K. Zhao, L. K. Wu, and Z. Li, "Dynamic forecast of Internet financial market based on deep neural network structure," *Journal of Computer Research and Development*, vol. 56, no. 8, pp. 1621–1631, 2019.

[25] S. Chen and L. Ge, "Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction," *Quantitative Finance*, vol. 19, no. 9, pp. 1507–1515, 2019.
[26] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, https://arxiv.org/abs/1502.03167.
[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
[28] X. Li, S. Chen, X. L. Hu, and J. Yang, "Understanding the disharmony between dropout and batch normalization by variance shift," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 2682–2690, Long Beach, CA, USA, June 2019.
[29] P. Luo, J. M. Ren, Z. L. Peng, R. Zhang, and J. Li, "Differentiable learning-to-normalize via switchable normalization," 2018, https://arxiv.org/abs/1806.10779.
[30] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2017, https://arxiv.org/abs/1412.6980.
[31] E. J. Elton, M. J. Gruber, D. Agrawal, and C. Mann, "Explaining the rate spread on corporate bonds," *The Journal of Finance*, vol. 56, no. 1, pp. 247–277, 2001.
[32] I. Loncarski and P. G. Szilagyi, "Empirical analysis of credit spread changes of us corporate bonds," *International Review of Financial Analysis*, vol. 24, no. 3, pp. 12–19, 2012.
[33] E. Clark and S. Baccar, "Modelling credit spreads with time volatility, skewness, and kurtosis," *Annals of Operations Research*, vol. 262, no. 2, pp. 431–461, 2018.
[34] L. Wang and C. H. Nie, "Forecasting ability of Chinese bond spread to macroeconomic indicators," *Management Review*, vol. 31, no. 5, pp. 66–76, 2019.
[35] D. Duffie and M. Huang, "Swap rates and credit quality," *The Journal of Finance*, vol. 51, no. 3, pp. 921–949, 1996.
[36] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 134–144, 2002.
[37] D. N. Moriasi, J. G. Arnold, and M. W. VLiew, "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations," *Transactions of the ASABE*, vol. 50, no. 3, pp. 885–900, 2007.
[38] J. C. Li, "On the randomness of statistical data," *China Statistics*, vol. 66, no. 7, pp. 24-25, 2018.
[39] X. B. Tang, R. Zhang, and L. X. Liu, "Forecast of second-hand housing price in Beijing based on bat algorithm SVR model," *Statistical Research*, vol. 35, no. 11, pp. 71–81, 2018.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 3588822, 14 pages
https://doi.org/10.1155/2022/3588822

Research Article

Data Modelling in Human Resource Management: Influencing Factors of Employees' Job Satisfaction

Mei-Er Zhuang [1] and Wen-Tsao Pan [1,2]

[1] *Guangdong University of Foreign Studies, Guangzhou 510006, China*
[2] *Hwa Hsia University of Technology, New Taipei 220, China*

Correspondence should be addressed to Mei-Er Zhuang; 990883130@qq.com

Received 21 May 2021; Revised 23 December 2021; Accepted 6 January 2022; Published 8 February 2022

Academic Editor: Ali Ahmadian

In the digital era, data mining and statistical analysis have been widely used to solve problems, especially in the field of management and engineering. Therefore, we aim to make a new insight of human resource management based on multiple regression modelling and quantile regression modelling. Specifically, the systematic framework of job satisfaction in this research is constructed by three dimensions from the perspective of psychology, namely, the perception of interpersonal relationship, financial compensation, and work conditions. Each dimension consists of two measures which reflect the employees' view towards them. The empirical estimation results show the following. (1) Perceived relationship with managers, perceived rationality of compensation, perceived match degree of job, and perceived autonomy degree of work are all significantly positively correlated with job satisfaction. (2) The effect of perceived rationality of compensation is significantly different between the high quantile and the low quantile. For those with lower perceived rationality of compensation, their job satisfaction is more likely to be affected due to the perceived compensation than those with higher perception. This research enriches the existing theory by constructing a comprehensive framework of the influencing factors of job satisfaction, which provides useful implications of human resource management optimization for enterprises.

1. Introduction

In the past decade, with the decline of birth rate and the intensification of population aging, China's demographic dividend has gradually faded. While the downward trend of global economy has resulted in unemployment for many workers, the shortage of labour supply for companies in different industries has been an increasingly obvious problem. Consequently, it has become a challenge for human resource management to attract adequate and capable employees and avoid losing labour assets. Braham [1] argues that employee turnover can lead to the cost of management and the damage of productivity.

It is of great importance to keep the appropriate employee retention as well as understand the reasons why employees disengage and leave their jobs. One possibility attributed to employee turnover is that employees are not only pulled by better offers but also pushed out of the company they are working in due to relatively low satisfaction [2]. Job satisfaction has been proved to have a significant relationship to organizational employee turnover and has generated widespread interest among researchers [3, 4]. Thus, it is imperative to pay more attention to retaining employees by improving employee job satisfaction.

Job satisfaction has been defined in various ways. In a narrow sense, it mainly refers to the positive feelings and emotions from the appraisal of one's job or job experiences [5]. In a broad sense, job satisfaction can be regarded as the combination of psychological and environmental circumstances that arouses one's subjective reaction to all the issues pertaining to their jobs [6, 7]. In this study, job satisfaction can be described as attitudes towards the perception between what one wants from one's job and what they actually gain from the job as offering or entailing [8].

Existing relevant research is mainly concentrated on the study of determinants and outcomes of job satisfaction from different perspectives, such as the influencing factors of job satisfaction and the effects of job satisfaction [9, 10]. As for the latter, numerous literatures have reached a basically consistent conclusion, that is, job satisfaction has a positive impact on organizational performance through multiple channels. McNeese-Smith [11] indicated that employees with higher level of job satisfaction are likely to be more productive. The higher the job satisfaction is, the better the performance will be. Besides, the impacts of job satisfaction on performance are stable across time [12]. Since job satisfaction is conducive to higher performance, it is vital to have a deeper insight of the influencing factors of employees' job satisfaction, so that employees and employers can achieve better coordination by enhancing the employees' happiness and organizational performance.

Although much research has made contributions to the studies of job satisfaction, most of these previous research is restricted in developed economy or certain industries, lacking in establishing a framework to measure employee job satisfaction levels in a wider range [13]. It should be noted that studies have not clearly and systematically investigated the influencing factors of job satisfaction. Which specific aspects are important for job satisfaction may differ per countries or job type. China is one of the most influential emerging economies in the world. It is also a relation-oriented country with high power distance. According to Hofstede's cross-cultural theory, there are five cultural dimensions. Power distance is one of the dimensions and refers to the recognition of members of the society on the class difference caused by power and wealth [14]. China has unique cultural context which is greatly different from previous studies related to job satisfaction. Therefore, the purpose of this study is to identify the potential factors of job satisfaction in the context of China by applying a newly modified theoretical and structural framework.

2. Literature Review

A plethora of research has been published pertaining to employees' satisfaction including but not solely focused on personal characteristics, professional accomplishments, income, or their relationship with other staff [15]. According to the existing literature, a vast amount of research on the issue of job satisfaction has been documented based on different perspectives. Particularly, the influencing factors of job satisfaction can be divided into four categories: demographic characteristics, individual endowments, contextual situations, and psychological conditions [16, 17]. Demographic characteristics refer to basic attributes that cannot be changed manually such as gender [18] and age [19]. Individual endowments include educational background such as academic level or career-training experience [20] Contextual situations include work environment, communication atmosphere, compensation and benefits, insurance, and security [21]. Psychological conditions include adaptability and job stress [22]. Many studies have confirmed that these factors mentioned above have significant impacts on job satisfaction to some extent [23].

In addition to the above classification of factors affecting job satisfaction, there are also other classification methods or evaluations methods [24]. Some scholars proposed that the determinants of job satisfaction can be grouped into three categories according to the perspectives of personal characteristics, job characteristics, and organizational characteristics, while others argued that the determinants should be divided into two kinds, namely, endogenous intrinsic factors and exogenous extrinsic factors [25, 26]. On the one hand, the combination of endogenous psychological factors is regarded as a closed system, such as psychological contract, sense of support, trust, emotional intelligence, and mental health [27]. It mainly integrates individual characteristics, focusing on the role of psychological determinants from a deeper internal perspective [28]. On the other hand, the combination of exogenous factors is regarded as an open system, focusing on the work conditions and social environment perspective [29]. Studies related to the exogenous factors mainly demonstrate the effect of work atmosphere, leadership style, professional positions, or working hours on job satisfaction. For example, common environmental stressors in the work environment can be stressful to staff and influence job satisfaction [30, 31].

Research on job satisfaction has been developing greatly in the past few decades, but it remains unclear which types of determinants exert most effects on the employees' attitude partly because the traditional framework lacked multidimensional analysis. Due to the complication of the mechanism between job satisfaction and influencing factors, related determinants have not been fully explored and validated up to now, which provides the research gap and the reason why more investigation is needed to make a deeper insight on this field.

Therefore, making the most of the paradigm of endogenous and exogenous research, we construct a more comprehensive framework of job satisfaction from the perspective of both endogenous and exogenous factors. Based on the individual perception of employees, we set up two-level variables for analysis on the secondary indicators in each dimension. Specifically, we focused on the following dimensions: interpersonal relationship, financial compensation, and work conditions. In each dimension, we construct two variables involving the perception of employees as the determinants of job satisfaction.

Drawing on the data of Chinese General Social Survey (CGSS), we extracted variables that meet these three dimensions for empirical analysis and explored the estimation effects of these six factors on job satisfaction to clarify the mechanism, thereby creating a more comprehensive explanation of the relationship between the three-dimensional determinants and job satisfaction.

3. Hypothesis Development

3.1. Interpersonal Relationship. Interpersonal relationship of employees in the workplace can be subdivided into two categories, namely, the relationship with managers and the relationship with colleagues [32]. Managers in this context mainly include supervisors, team leaders, and any superiors that have a higher position level than the employee themselves in the organization while the colleagues refer to those co-workers that are in the same or similar position level in the organization.

Relationship may have a "double-edged sword" effect on employees' performance. Individual-level relationship may increase job satisfaction and performance, while group-level relationship may weaken employees' sense of procedural justice and reduce their work performance [33]. Relationship with supervisors and colleagues has been estimated to have significant influence on job satisfaction in different fields such as healthcare industry [34]. Employees who are from dominant groups in the organization tend to be more satisfied with their jobs because their managers and co-workers tend to provide them with more feedback and support [35]. Scholars have also revealed that employees at the establishment career stage are more likely to expect to become accepted as equal and regular members of the organization by building an effective relationship with co-workers and supervisors, thus learning organizational norms and values better [36]. Although research on social decision making has already shown that trust on a third party affects the individual's cooperative behavior [37], studies have suggested that employees' perception of both co-workers and supervisory support may bring different outcomes, which means that employees' trust towards their co-workers and supervisors varies in general working place [38]. Hence, interpersonal relationship with managers and colleagues may influence employees' perception differently.

On the one hand, perceived relationship with managers can be defined as a kind of supervisor-subordinate relation based on employees' perception of interaction with leaders and supervisors [39]. According to Hofstede's theory, China's power distance gap is relatively huge and the concentration degree and dictatorship degree of power in Chinese society are high. Furthermore, leadership behaviors can vary over two domains according to the leadership theory proposed by Hersey and Blanchard [40]: task-oriented and relation-oriented leadership behaviors. China is a typical relation-oriented society where informal systems often play a potentially huge role. Hence, specific leadership behaviors such as coordinating and structuring and also how supervisors communicate may have more significant impacts on employee job satisfaction [41]. For example, from the frontline staff perspective, if managers fail to provide constant support and understanding, they will perceive a poor leadership that undermines their trust and job satisfaction [42]. As the director of power and the distributor of resources, leaders have the right to determine the allocation of resources in the organization and may be closely related to the vital interests of each employee, which means that getting the appreciation of leadership is particularly important [43]. If employees have a better relationship with managers who have dominant power and resources, they may have potential competitive advantages in terms of resource acquisition and benefit distribution, which is conducive to gaining opportunities for career development [44]. This can account for the reason why employees that perceived their relationships with supervisors more positive show greater levels of job satisfaction than those who do not hold such beliefs [45]. In that case, employees getting along well with managers may have more self-belief and even a sense of superiority, so their job satisfaction will be correspondingly higher [46].

Therefore, the first hypothesis is proposed as follows.

H1a: perceived interpersonal relationship with managers has a positive effect on job satisfaction.

On the other hand, perceived relationship with colleagues refers to the peer relation of co-workers that are in the same or similar position level in the organization, represented by friendliness and kindness from colleagues. Studies have found that co-worker support is considered to be predictors of the employees' behavior [47]. However, such influences may differ with different genders, and studies have proved that perceived co-worker support was more strongly related to organization commitment which is typical outcome of job satisfaction for women than for men [48]. Relationships with colleagues are mutually supportive in a situation where co-workers are cooperative and work collaboratively. Co-worker support was significantly related to personal accomplishment [49]. Those having good interpersonal relationship with colleagues are also easier to overcome difficulties due to knowledge sharing when they encounter challenges at work. Employees experienced close and friendly relationship in the occupational team as a source of strength that enabled them to overcome different hurdles [50]. In other words, when faced with trouble, those having co-worker support are more likely to obtain enough valuable information and timely help, so that they are more likely to find a solution to the problem in a shorter time, which helps to increase the probability of success, thereby generating more positive emotions and improving job satisfaction [51]. Otherwise, when relationships with colleagues were strained, employees' ideas and creativity would be hindered, which may harm their job satisfaction [52]. Research has also been shown that teamwork was positively associated with job satisfaction by creating positive interaction. Good interpersonal relationship with colleagues may help to form a closer emotional connection [53]. When employees are in a bad mood and need care, they are more likely to be comforted by colleagues if they have good interpersonal relationship, which can reduce depression and conflicts in workplace [54]. Such a good state of interpersonal relationship may have an impact on the effectiveness and efficiency of communication, social exchanges, and emotional venting among employees and contributes to carry out more harmonious cooperation and form a good working atmosphere [55]. Good atmosphere created a supportive environment and a sense of safety and security, which can reduce the pressure of life and work to a certain extent [56]. When one felt supported and backed by colleagues, their job satisfaction may be well improved.

Therefore, the second hypothesis is proposed as follows.

H1b: perceived interpersonal relationship with colleagues has a positive effect on job satisfaction.

3.2. Financial Compensation. Compensation refers to the salaries, income, rewards, insurance, and any monetary payment and financial benefits employees received from their jobs. Earlier studies believed that compensation did not have a significant impact on job satisfaction [57, 58]. In some studies related to hotel internship program, researchers found that compensation composed of overtime pay, fringe and benefits, and salary did not significantly affect internship satisfaction [59, 60].

However, more and more studies have argued that compensation level is positively correlated with job satisfaction, which can be explained by the resource preservation theory [61, 62]. In their research, improvement in compensation represents an increase in the material resources occupied by employees. Psychological motivation of employees would be enhanced, thus accumulating higher job satisfaction [63]. Some other scholars also show convincing evidence of the assumption that stable income is important for the attractiveness of job [64].

Nevertheless, some scholars held different views and presented opposite conclusions, suggesting that salary level is negatively correlated with job satisfaction [65]. Besides, some complicated mechanism may be existing in the relationship between compensation and job satisfaction. For example, it could be a type of nonlinear relationship, which can be described as "inverted U-shaped" curve [66]. These diverse and seemingly contradictory research results reflect the fact that the relationship between compensation and job satisfaction is complex and still needs further investigation.

On the one hand, perceived rationality of compensation can be defined as the perception and judgment of compensation based on their ability and expectations. Numerous studies have suggested that perceptions of fairness play a vital role in the service encounter [67]. Studies highlight how compensation can influence their job satisfaction [68]. When exploring the effects of compensation, most of the existing studies are value-oriented and merely focus on the amount of compensation rather than the perception of employees. In fact, the perception of income and the specific value of income are not equal. In other words, perception of compensation is not the same as the amount of compensation itself, which should not be ignored. On the basis of Adam's [69] equity theory, job satisfaction can be determined by the employee's input-income ratio and those of the referents. The level of an employee's job satisfaction may be affected by the assessment based on their own contribution and compensation [70].

Thus, the employee's subjective assessment of his or her input-related reward can determine his or her job satisfaction [71]. If people with average incomes feel that their income is inadequate and does not reach a proper level they should be, their perceived rationality of compensation becomes weak, and then their job satisfaction may be relatively low. On the contrary, for people whose income level is not very high, if they believe that their income is in line with their own contribution and meets their expectations, then their sense of reasonableness of income will be strong, which may be easier to arouse a stronger feeling of job satisfaction.

In short, the perception of reasonableness of compensation is not equal to the absolute value of compensation and the higher perceived rationality of compensation may have a positive impact on job satisfaction. Therefore, it is necessary to distinguish the absolute value of compensation from the perceived rationality of compensation.

Therefore, the third hypothesis is proposed as follows.

H2a: perceived rationality of compensation has a positive effect on job satisfaction.

On the other hand, the level of an employee's job satisfaction can also be affected by a comparison with the contribution and financial compensation of others. Previous theory claims that people's job satisfaction is not only related to personal absolute compensation but also more closely related to people's sense of fairness and equity in distribution [72].

Perceived equity of distribution can be defined as the perception of fairness of the job compensation, which is the subjective judgment of the employees on the fairness of the organization's resource distribution [73]. Studies have shown that both the external fairness and internal fairness of compensation will have a significant impact on employee job satisfaction [74]. In other words, an employee may feel a psychological conflict associated with his or her compensation when an equivalent colleague receives a higher income, and this discrepancy can decrease the employee's job satisfaction. McLoughlin and Carr [75] suggested that employees tend to be less satisfied with their job because of the inequalities they experience in rewards.

Perceived equity of distribution is important, and it is reported that equal and fair compensation system positively influences job satisfaction of a multicultural workforce [76]. Research suggested that fairness of salary in centralized public procurement systems is the key factor of job satisfaction [77].

However, the relationship between perceived equity of distribution and job satisfaction in the context of Chinese workplace has not been fully estimated.

Therefore, the fourth hypothesis is proposed as follows.

H2b: perceived equity of distribution has a positive effect on job satisfaction.

3.3. Work Conditions. Work conditions mainly refer to the combination of objective and subjective issues related to workload, working period, workplace environment, promotion opportunity, and work stress [78]. In our study, we focus on two important components of work conditions, namely, perceived match between ability and job [79] and perceived autonomy [80].

Person-job fit is defined as a match between individual knowledge, skills, abilities, and the job requirements [81]. It is usually presented as the compatibility between the employee and the tasks that are expected to be accomplished in exchange for employment [82, 83].

Researchers suggested that employees' work attitude was affected by the perceived person-organization fit [84]. In an analysis for female managers, scholars found that the higher the individual-organization matching is, the more the job satisfaction would be [85]. The lower person-job match represented by the perception of overqualification of employees would cause employees to feel frustrated because the work they are engaged in cannot give full play to their own skills, which makes them more likely to lose interest in work [86].

To be more specific, we use the term "match degree of ability" to depict the degree to which employees' job fits his or her ability. A high match degree of ability indicates that the knowledge and skills required for a given job are highly related to those provided by their education, experience, and abilities [87]. In such situations, job-related knowledge is strongly associated with potential job performance and may affect job satisfaction. Otherwise, when employees' previous knowledge or skills cannot be applied in the existing work fully or when the employees feel that they have difficulties in adapting to the current job or exerting their original talents and experience on the work task, there will be a sense of mismatch, resulting in lower job satisfaction. Lawler [88] suggested that if employees perceive that reward allocation is unfair or unrelated to the level of employee contribution, it would not be possible to sustain—in the long run—managerial practices based on empowerment. Under those conditions, employee motivation would decrease and interest towards empowerment would eventually be eroded [89]. Even if some employees have good qualifications, the mismatch between employees' occupation and their abilities may still lead to relatively worse performance to some extent, thus causing the lack of self-confidence and job satifaction.

Therefore, the fifth hypothesis is proposed as follows.

H3a: perceived match degree of ability has a positive effect on job satisfaction.

On the other hand, autonomy degree of work provides another indicator of the source of employee satisfaction in terms of work conditions [90]. Hackman and Oldham [91] defined job autonomy as the degree to which the job provides substantial freedom, independence, and discretion to the employee in scheduling the work and in determining the procedures to be used in carrying it out. In other words, autonomy degree of work can be defined as the level of freedom employees experience in terms of decision making at the workplace [92]. In this study, we use the term "autonomy degree of work" to depict the degree to which the job provides freedom and independence for employees' developments and work-life balance.

Self-determination theory believes that people have basic psychological needs for autonomy, sense of belonging and relatedness, and competence. When good working conditions are provided and the three innate needs are met, employees are more likely to work under the drive of internal motivation, thus exerting their potential and creating positive work performance. Employees who believe they have greater autonomy to make decisions at work are also shown to be more satisfied with their jobs [93]. The domain of autonomy is reflected in conflicts which arise from leadership style, management practices, and decision-making processes. Job autonomy is aligned with job resource which seeks to prevent the negative impact job demands will bring. While job demands concern themselves with the cost in physiology, social, psychological, or organizational sides of the job like emotional demands, job resources lessen the impact of job demands and their costs to stimulate some level of learning, growth, and development. Therefore, absence of job autonomy raises the negativities of absenteeism, stress, repetitive strain, and ill health, whereas presence of job autonomy leads to higher employee job satisfaction [94].

Research on the influence mechanism of job autonomy on job satisfaction has been quite consistent. Findings have mostly suggested that job autonomy leads to job satisfaction on the same assumption. A lack of autonomy will result in higher levels of stress which in turn can lead to dissatisfaction in one's work [95].

It has been confirmed that autonomy in work process [96], flexible working hours [97], and autonomy in workload [98] all have positive impacts on satisfaction. Studies have shown that working hours affect job satisfaction by changing employee perceptions of work context [99]. Some scholars have pointed out that flexible work plans can significantly reduce employees' work-family conflict and turnover tendency and improve perceived job autonomy, job satisfaction, organizational commitment, and job performance [100]. Compatibility between family life and working hours, namely, the reduction of conflicts between work and leisure time, is clearly conducive to satisfaction [101]. It can be inferred that work-life balance contributes to job satisfaction while unsatisfactory working conditions lead to work-family conflicts, time pressure, emotional exhaustion, and time stress, which ultimately result in high turnover rates [102].

Based on the analysis, it can be inferred that if the employers enhance the autonomy and freedom of employees and support them in arranging their time and work tasks reasonably, it can largely alleviate the time conflicts of employees at work, thereby enhancing work flexibility degree and adaptability, which improves the job satisfaction of employees.

Therefore, the sixth hypothesis is proposed as follows.

H3b: perceived autonomy degree of work has a positive effect on job satisfaction.

The research framework is shown as Figure 1.

4. Research Method

4.1. Model Construction. Based on the three dimensions of determinants of job satisfaction, namely, interpersonal relationship, financial compensation, and work conditions, a regression model is established as follows:

$$\begin{aligned} \mathrm{JS} = \beta_0 + \beta_1\mathrm{MANAG} + \beta_2\mathrm{COLLEA} + \beta_3\mathrm{RATION} \\ + \beta_4\mathrm{EAUITY} + \beta_5\mathrm{MATCH} + \beta_6\mathrm{AUTON} + \varepsilon. \end{aligned} \quad (1)$$

On the left side, JS represents the dependent variable, namely, job satisfaction. On the right side, six independent variables are represented by acronyms. To be more specific, MANAG stands for relationship with

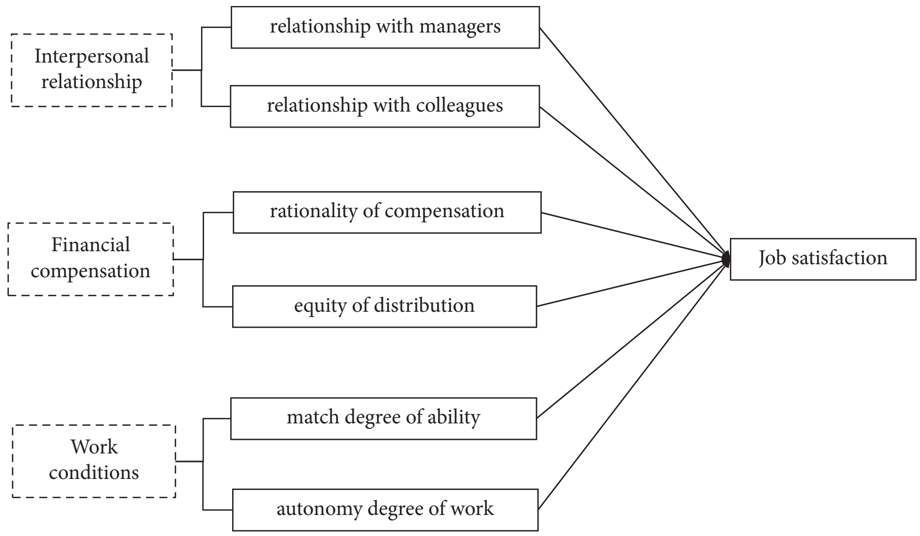


Figure 1: Research framework.

managers, and COLLEA stands for relationship with colleagues. RATION is represented by perceived rationality of compensation. EQUITY stands for perceived equity of distribution, and MATCH stands for perceived match degree of job. AUTON stands for perceived autonomy degree of work. $\beta_1, \beta_2 \ldots \beta_6$ are the corresponding regression coefficients. β_0 is a constant term, and ε represents the random error.

4.2. Data Source. Drawing on the data form Chinese General Social Survey (CGSS), we select and construct the corresponding scale to measure the variables and conduct empirical analysis to test the hypotheses. CGSS is China's first national and continuous large-scale social survey project. CGSS aims to collect data from Chinese people and all aspects of Chinese society regularly and systematically, summarizing the long-term trend of social change. The main purpose is to explore social issues of great theoretical and practical significance and promote the openness and sharing of domestic social science research.

Up to now, the sample data in CGSS in 2015 are considered the latest version of the project. CGSS 2015 adopts random sampling method covering 28 provinces (excluding Tibet, Xinjiang, Hainan, Hong Kong, Macao, and Taiwan). It is worth noting that 10968 residents from 478 natural villages and 83 prefecture-level cities were interviewed across130 counties and 369 townships, which makes the sample size very large and convincing. Content of the questionnaire of CGSS 2015 not only includes the description of objective characteristics but also includes subjective questions, which is suitable for the requirements of this research.

4.3. Measures. The method of preprocessing of the raw data and the detailed descriptive information of each item in the scale are introduced in this section (see Table 1).

There are three steps in the preprocessing. First, remove missing values and invalid data. As some interviewees did not fill in all the questionnaire, some of the data were incomplete. Therefore, invalid samples with missing values in the key variables must be eliminated to ensure the scientific and rigorous data analysis. Second, since the CGSS2015 questionnaire is divided into several modules and some of the questions in some modules are only for specific groups of participants, the applicable objects of each question are not completely the same. The items we need for this research originated from module A, module B, and module D. However, module A and module B are full samples of 10968 because those participants can answer all the questions, while module D has only part of the interviews for the reason that some of them are not suitable for the questionnaire, which means that the data sample size extracted from the three modules of CGSS 2015 is quite different. In order to ensure the accuracy of the empirical analysis, the data need to be matched and paired. The specific method is based on whether the D module of the questionnaire is answered or not. Those corresponding samples that did not answer question related to job satisfaction are eliminated, and only those valid samples that have answered the main questions at the same time are retained, so that the sample size of each question item is basically balanced. Third, we processed the control variables including gender and education level in the social demographic attributes. We construct 0-1 dummy variable for gender and education level.

5. Result Analysis

Stata is used to conduct empirical estimation on sample data. First, the descriptive statistics are demonstrated, followed by the correlation analysis. Then, hypothesis testing is conducted through multiple regression, and

Table 1: Scales and items of variables.

Variables	Items	Scales
JS	Are you satisfied with your current job?	This question is rated using a 5-point Likert-type scale format. 1 = "very satisfied," 2 = "satisfied," 3 = "neutral," 4 = "dissatisfied," and 5 = "very dissatisfied"
MANAG	What do you think of your relationship with your managers?	This question is rated using a 5-point Likert-type scale format. 1 = "very good," 2 = "good," 3 = "neutral," 4 = "bad," and 5 = "very bad"
COLLEA	What do you think of your relationship with your colleagues?	This question is rated using a 5-point Likert-type scale format. 1 = "very good," 2 = "good," 3 = "neutral," 4 = "bad," and 5 = "very bad"
RATION	Considering your abilities and work status, is your current income reasonable?	This question is rated using a 4-point Likert-type scale format. 1 = "very reasonable," 2 = "reasonable," 3 = "unreasonable," and 4 = "very unreasonable"
EQUITY	Do you think the social distribution method is fair?	This question is rated using a 5-point Likert-type scale format. 1 = "very fair," 2 = "fair," 3 = "neutral," 4 = "unfair," and 5 = "very unfair"
MATCH	How much of your past work experience/or skills can be used in your current work?	This question is rated using a 4-point Likert-type scale format. 1 = "very little," 2 = "a little bit," 3 = "some," and 4 = "very much"
AOTON	How is your daily work schedule?	This question is rated using a 3-point Likert-type scale format. 1 = "free," 2 = "neutral," and 3 = "fixed"
Controls	Gender (GEN) Education level (EDU)	Female = 0; male = 1 Senior high school and above = 0; junior high school and below = 1

finally the robustness check through quantile regression modelling is illustrated.

5.1. Descriptive Statistics. According to statistics, the gender ratio of the sample is basically balanced, with 49.23% of women and 50.77% of men. In terms of educational level, the proportion of people with higher education background is slightly smaller. Nearly 47% of participants graduated from high school or have a higher degree.

As shown in Table 2, the variance inflation factor (VIF) of each explanatory variable is very small, and the mean value of variance inflation factor is 1.39, indicating that there is no multicollinearity problem in the variables. The mean value of MANAG is about 3.4, indicating that the sample's perception of the manager's relationship is roughly at the upper-middle level. The mean of the variables COLLEA, RATION, and EQUITY is in the range of 2.0 to 2.3, indicating that the sample's perception of these issues is relatively worse. The mean of the variable MATCH is about 3.1, reflecting that more interviewees think that some of their work experience and skills can be applied to the current jobs. The mean of that measure is about 2.5, which is close to the highest value of this question. Due to the negative rating of the question related to AOTON, most of the sample respondents showed that they cannot freely decide their daily work arrangements.

In addition, variables in this study have high standard deviations (SDs), indicating that the measurement method used in the sample is effective, and the heterogeneity in the selected variables is also obvious.

5.2. Correlation Analysis. Before regression analysis, we conducted a preliminary analysis of the correlation between each independent variable and the dependent variable. It can be seen from the second column of Table 3 that in the case of univariate analysis, except for the control variable of gender, the other variables are all significantly correlated with the dependent variable at the 5% statistical level. Since the rating rules of the two variables EQUITY and MATCH are opposite to the rules of other dependent variable, the negative sign before the coefficient of EQUITY and MATCH represents a positive correlation actually. In other words, from the results of the correlation analysis, the dependent variable is significantly positively correlated with all independent variables, and they are all positively correlated. However, correlation analysis is not enough to confirm whether there is a robust causal relationship between variables. Therefore, multiple regression analysis is needed to estimate the parameters.

5.3. Multiple Regression Modelling. As shown in Table 4, we use stepwise multiple regression to test all hypotheses separately. From columns (1), (4), (5), and (7), the regression coefficient of MANAG is positive at the significance level of 1%, which indicates that the better the employee's perception of the relationship with managers, the higher the satisfaction level, so hypothesis H1a is verified. On the other hand, regression coefficient of COLLEA is not significant, indicating that the perceived relationship with colleagues may not have such a strong impact on job satisfaction. Hypothesis H1b cannot be supported.

From columns (2), (5), (6), and (7), regression coefficient of RATION is significant at the 1% statistical level, indicating that the more reasonable the employee's perceived compensation is, the higher the satisfaction level is. Hypothesis H2a is verified.

However, the regression coefficient of EQUITY is not significant, indicating that the effect of distribution fairness on job satisfaction is not so obvious as expected. Hypothesis H2b has not been supported.

Table 2: Descriptive statistics and multicollinearity diagnosis.

Variables	N	Mean	SD	VIF	1/VIF
MANAG	782	3.337596	1.059603	2.13	0.469180
COLLEA	506	2.23913	0.6837074	2.11	0.473927
RATION	511	2.064579	0.6460572	1.05	0.952748
EQUITY	758	2.340369	0.5417627	1.04	0.958860
MATCH	778	3.120823	1.004906	1.02	0.977118
AOTON	765	2.508497	0.9089396	1.01	0.988677

Table 3: Correlation analysis results.

Variables	JS	MANAG	COLLEA	RATION	EQUITY	MATCH	AOTON	GEN	EDU
JS	1								
MANAG	0.279*	1							
COLLEA	0.246*	0.722*	1						
RATION	0.336*	0.042	0.029	1					
EQUITY	−0.104*	0.013	−0.028	−0.201*	1				
MATCH	−0.086*	−0.084	−0.051	−0.019	0.022	1			
AOTON	0.160*	0.132*	0.101*	0.095*	0.002	−0.035	1		
GEN	−0.073	0.004	0.010	−0.049	0.054	0.040	−0.099*	1	
EDU	0.156*	0.035	0.013	0.022	0.010	−0.058	−0.102*	−0.009	1

Note. *denotes the significance level at 5% (*means $P \leq 0.05$).

Table 4: Multiple regression analysis results.

Variables	JS						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MANAG	0.294***			0.248***	0.279***		0.246***
	(0.089)			(0.088)	(0.083)		(0.082)
COLLEA	0.144			0.167	0.130		0.151
	(0.095)			(0.093)	(0.089)		(0.087)
RATION		0.625***			0.661***	0.679***	0.622***
		(0.068)			(0.074)	(0.073)	(0.073)
EQUITY		−0.038			−0.060	−0.048	−0.073
		(0.037)			(0.040)	(0.040)	(0.040)
MATCH			−0.170***	−0.175***		−0.146***	−0.147***
			(0.046)	(0.048)		(0.044)	(0.045)
AOTON			0.197***	0.171***		0.154***	0.127**
			(0.056)	(0.060)		(0.053)	(0.056)
GEN	−0.110	−0.119*	−0.108	−0.069	−0.073	−0.076	−0.044
	(0.085)	(0.072)	(0.083)	(0.083)	(0.079)	(0.078)	(0.078)
EDU	0.049	0.308***	0.120	0.014	0.058	0.096	0.032
	(0.088)	(0.072)	(0.085)	(0.088)	(0.082)	(0.080)	(0.082)
Constant	2.280***	1.884***	3.231***	2.387***	0.966***	1.811***	1.205***
	(0.161)	(0.222)	(0.187)	(0.241)	(0.272)	(0.283)	(0.312)
N	502	752	576	491	491	559	481
R^2	0.082	0.138	0.054	0.127	0.227	0.195	0.259

Note. The numbers in brackets are robust standard errors. *, **, and ***represent the significance level at 10%, 5%, and 1%, respectively (*means $P \leq 0.1$, **means $P \leq 0.05$, and ***means $P \leq 0.01$).

From columns (3), (4), (6), and (7), regression coefficients of MATCH and AOTON are both significant at the 1% statistical level, indicating that the perception of job match and autonomy both have significant impacts on job satisfaction. Hypotheses H3a and H3b have been verified.

6. Discussion of the Basic Regression Results

Based on the results of basic regression, it can be found that whether interpersonal relationships will affect employees' job satisfaction depends on the specific types of interpersonal relationships. Perceived relationship with managers contributes to job satisfaction while the effects of perceived relationship with colleagues is not significant. The inherent logic may be as follows.

On the one hand, if employees have better relationship with managers who have dominant power and resources in the organization, employees may think that they are more likely to be recognized and appreciated by their supervisors. In that case, they will have a more positive attitude on themselves, believing that they not only have the opportunity to get promotion for development but also have more competitive advantages in terms of resource acquisition and benefit rewards. With the help of such a promising expectation, employees' confidence will be strengthened and their psychological pressure may be reduced. Besides, employees may think that they are more likely to be favored and supported by their managers, thus getting more attention and respect from their managers, which is consistent with the findings of previous studies [103]. These findings mirrored those obtained in an earlier published study, which suggested that job dissatisfaction was caused by lack of respect from the supervisor. Previous studies have shown that relationship with supervisors and team leaders plays a more significant role by the way of communication. Different relationship with their managers of leaders may give them the impression that the employees are respected or not [104].

On the other hand, as for the relationship with colleagues, although it may also affect the mood and state of employees at work, its scope of influence may be relatively limited because employees may tend to directly associate the quality of relationship with a specific person rather than the job. In other words, the focus of relationship with colleagues will be put on the individual itself instead of the community. Therefore, the perception of relationships with colleague does not significantly affect job satisfaction as expected.

According to the results of basic regression, we can also find that perceived rationality of compensation has positive effects on job satisfaction while perceived equity does not have such significant effects. Compensation does play a vital role in the job satisfaction. The positive and significant relationship between pay and job satisfaction supports the findings in similar studies [105]. However, the regression results showed that perceived equity of distribution is not so important as the perceived rationality of compensation. One possibility is that the items of perceived equity of distribution in our scale are more inclined to ask their perception of fairness in social level rather than organizational level. Consequently, this measure was not totally accurate and consistent with our research purpose, which is an aspect that needs improvement in future study.

Furthermore, perceived match degree of job and perceived autonomy of work both have positive effects on job satisfaction. When employees feel that their knowledge, skills, and work experience can be effectively used in their existing work, their sense of competency will be effectively met, and they will be driven by their internal motivation to work to achieve their potential and create better performance and higher job satisfaction. In that case, they may think that they are indeed suitable for the current job and can finish related tasks well. This sense of job match can greatly reduce the anxiety and stress that may exist in employees, thereby improving job satisfaction. Findings are consistent with previous research which found that satisfaction in job training is impacted by the perceived compatibility between training knowledge and job requirements [106].

Meanwhile, in terms of job autonomy, the estimation also shows that it is a strong predictor which showed a positive association with job satisfaction. The results indicated that the more autonomy the employees had, the more satisfied they were with their jobs. When employees' perceived autonomy of work is high, it means that employees will have greater flexibility at work, which may reduce the constraints of working hours. Alleviating employees' time conflicts and work schedules can help to better balance work and life, thereby increasing the job satisfaction of employees and making the job satisfaction higher. Employees do not achieve post-training satisfaction because they expect job training to improve job performance, result in salary increase, and ensure promotion [107].

6.1. Quantile Regression Modelling. In order to estimate the influencing factors of job satisfaction in depth, we use quantile regression to conduct the heterogeneity test of different quantile areas.

Quantile regression splits the data into multiple quantile points according to the dependent variable and further estimates the relationship under different quantile points. The main purpose of quantile regression is to analyze the trend of the influence of the independent variable on the dependent variable and to check the robustness of the regression model.

We use the lowest 20% as the low quantile and the highest 20% as the high quantile. The analysis results are shown in Table 5 and Figure 2. It can be seen that the P value of RATION is equal to (or less than) 0.05 both in the low quantile and the high quantile ($P \leq 0.05$), which rejects the null hypothesis and proves that there is a significant difference between the low quantile and the high quantile of perceived rationality of compensation. Besides, the P value of MATCH and AOTON is less than 0.05 in the low quantile ($P \leq 0.05$), suggesting that compared with the average interviewees, those with lower perception of match degree of job and autonomy degree of work are significantly different. The P value of MANAG is less than 0.05 in the high quantile ($P \leq 0.05$) and reveals that the perception of relationship with managers is indeed playing a more significant role in job satisfaction when they think they are really having a good connection with managers.

Here we further discuss the very variable that passed the heterogeneity test ($P \leq 0.1$), namely, perceived rationality of compensation (see Table 6). Since the value of the perceived rationality of compensation is rated in a reverse way, the high quantile interval represents a lower perception of rationality, and vice versa. It can be seen from Table 5 that in the high quantile interval, the results of quantile regression are significantly higher than the results of ordinary linear regression, indicating that ordinary linear regression has a lower estimation of the effects of perceived rationality of compensation. In other words, ordinary linear regression underestimates the impact of perceived rationality of compensation on job satisfaction. Specifically, when the perceived rationality of compensation is low, employees will be more concerned about compensation, so the effects of compensation on job satisfaction become more obvious. In the low quantile range, there exists an opposite situation, especially in the 0.2–0.4 interval. The results of quantile

Table 5: Quantile regression modelling results.

	JS	Coef.	Bootstrap Std. Err.	t	$P>t$	(95% conf.)	(Interval)
q20	MANAG	0.2	0.1418238	1.41	0.159	−0.0786781	0.4786781
	COLLEA	0.2	0.1703923	1.17	0.241	−0.1348141	0.5348141
	RATION	0.4	0.2037519	1.96	0.050*	−0.0003644	0.8003644
	EQUITY	$-1.75e-16$	0.0695895	−0.00	1.000	−0.1367405	0.1367405
	MATCH	−0.2	0.0866263	−2.31	0.021*	−0.3702173	−0.0297827
	AOTON	0.2	0.0874006	2.29	0.023*	0.0282612	0.3717388
	_cons	0.8	0.5179869	1.54	0.123	−0.2178236	1.817824
q80	MANAG	0.4285714	0.2065284	2.08	0.039*	0.0227514	0.8343914
	COLLEA	$-7.18e-15$	0.1531282	−0.00	1.000	−0.3008909	0.3008909
	RATION	0.8571429	0.1718596	4.99	0.000**	0.5194456	1.19484
	EQUITY	−0.1428571	0.1041957	−1.37	0.171	−0.3475976	0.0618833
	MATCH	−0.1428571	0.0996767	−1.43	0.152	−0.3387178	0.0530035
	AOTON	0.1428571	0.124361	1.15	0.251	−0.1015072	0.3872215
	_cons	1.285714	0.8899919	1.44	0.149	−0.4630844	3.034513

Note. *and **represent the significance level at 5% and 1%, respectively (*means $P\le 0.05$; **means $P\le 0.01$).

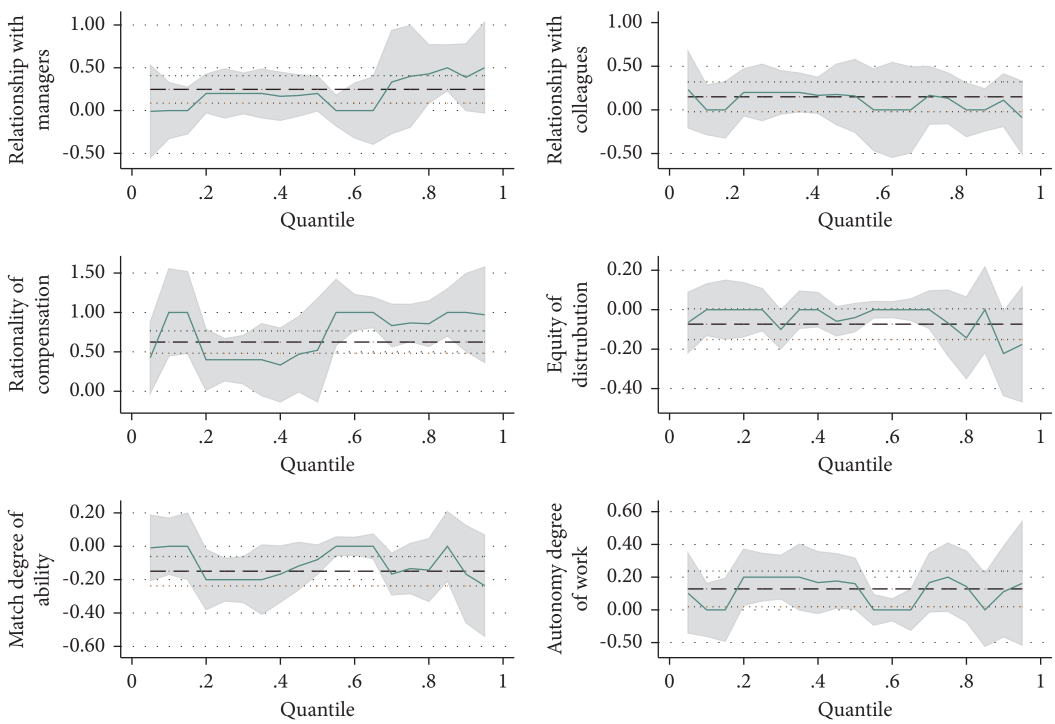


Figure 2: Quantile regression interval difference.

Table 6: Heterogeneity test: the difference between high quantile and low quantile.

	MANAG	COLLEA	RATION	EQUITY	MATCH	AOTON
F	1.01	0.97	3.48	1.61	0.22	0.17
Prob > F	0.32	0.325	0.06*	0.21	0.64	0.68

Note. *represents the significance level at 10% *means $P\le 0.01$).

regression are significantly lower than those of ordinary linear regression, indicating that ordinary linear regression has a high estimation of the parameters of samples with higher perceived rationality of compensation, that is, ordinary linear regression overestimates the impact of the higher sample of income reasonableness perception. When employees think that their income is reasonable, their attention to income will be reduced, and they may turn to consider career prospects and interpersonal relationships. In other aspects, they may also consider whether their work is in line with their own interests, skills, majors, and experiences. In short, employees with high perception of rationality of compensation may be affected by other incentives rather than compensation. It has been consistent with previous studies that regard salary as the most important factor in retaining such staff, who feel that they are not paid sufficiently to reward their particular qualification [108, 109].

To sum up, for those with worse perception of compensation rationality, their job satisfaction is more likely to be affected by their compensation while job satisfaction of employees with a better perception of income rationality is relatively less affected by compensation.

7. Conclusions and Implications

Based on the sample data of CGSS 2015, we construct a more comprehensive research framework with three dimensions, namely, interpersonal relationship, financial compensation, and work conditions. We analyze the influencing factors of job satisfaction systematically from the perspective of employee's psychological perception. The results demonstrate that perceived relationship with managers, perceived rationality of compensation, perceived match degree of ability, and perceived autonomy degree of work are all significantly positively correlated with job satisfaction. Besides, the effect of perceived rationality of compensation is significantly different between the high quantile and the low quantile. For those with lower perceived rationality of compensation, their job satisfaction is more likely to be affected due to the perceived rationality of compensation than those with higher perception of compensation.

This study sheds light on human resource management in the business practice. First, supervisors should establish proper management modes and use their affinity to interact well with employees. In this way, they can make employees feel themselves in the process of communicating and hold a better perception of interpersonal relationship with managers, thus helping employees to improve their job satisfaction. We recommend that supervisors become more aware of whether and how their behaviors influence employees' job satisfaction. Especially providing specific instructions and using two-way communication seem important to help employees deal with their insecurities and to offer them support. Second, it is important to identify employees' perception of compensation by various ways such as informal communication or questionnaire. As for those employees that regard their compensation unreasonable, it is necessary to adopt corresponding measures to help them realize that the income they receive is in line with their labour commitments or to adjust their compensation of appropriate subsidies, thereby reducing their psychological imbalance and improving employee job satisfaction. Third, managers should pay more attention to employees' characteristics and release the potential of employees by providing employees with a suitable platform and opportunities to match their abilities and fully develop their talents. Fourth, it is essential to give employees appropriate autonomy when arranging work tasks, allowing them to have more flexibility in the work process, thereby increasing employees' satisfaction.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Innovative Research Project of Guangdong University of Foreign Studies.

References

[1] L. Braham, *The 7 Hidden Reasons Employees Leave: How to Recognize the Subtle Signs and Act before It's Too Late*, Amacom, New York, NY, USA, 1st edition, 2005.

[2] G. Cai, "Liu Jianxiong. Attribution, autonomy and job satisfaction," *Management World*, no. 1, pp. 133–142+167, 2013.

[3] C. W. Barrow, "Employee turnover: implications for hotel managers," *FIU Hospitality Review*, vol. 8, no. 1, pp. 24–31, 1990.

[4] M. R. Testa, "Organizational commitment, job satisfaction, and effort in the service environment," *Journal of Psychology*, vol. 135, no. 2, pp. 226–36, 2001.

[5] P. C. Smith, L. M. Kendall, and C. L. Hulin, *Measurement of Satisfaction in Work and Retirement*, Rand McNally, Chicago, IL, USA, 1975.

[6] D. A. Harrison, D. A. Newman, and P. L. Roth, "How important are job attitudes? meta-analytic comparisons of integrative behavioral outcomes and time sequences," *Academy of Management Journal*, vol. 49, no. 2, pp. 305–325, 2006.

[7] R. Hoppock, *Job Satisfaction*, Harper & Row, New York, NY, USA, 1935.

[8] E. A. Locke, "What is job satisfaction?" *Organizational Behavior & Human Performance*, vol. 4, no. 4, pp. 309–336, 1969.

[9] A. C. Keith, N. Warshawsky, D. Neff, V. Loerzel, and J. Parchment, "Factors that influence nurse manager job satisfaction: an integrated literature review," *Journal of Nursing Management*, vol. 29, no. 3, pp. 373–384, 2021.

[10] H. Lu, Y. Zhao, and A. While, "Job satisfaction among hospital nurses: a literature review," *International Journal of Nursing Studies*, vol. 94, pp. 21–31, 2019.

[11] D. K. McNeese-Smith, "The influence of manager behavior on nurses' job satisfaction, productivity, and commitment," *The Journal of Nursing Administration*, vol. 27, no. 9, pp. 47–55, 1997.

[12] F. Fu, X. Lu, and Z. Wang, "High satisfaction leads to high performance, or high performance leads to high satisfaction? a 4-year leadership-subordinate pairing study," *Psychological Science*, vol. 41, no. 4, pp. 897–903, 2018.

[13] I. Huđek, P. Tominc, and K. Širec, "The impact of social and cultural norms, government programs and digitalization as entrepreneurial environment factors on job and career satisfaction of freelancers," *Sustainability*, vol. 13, no. 2, p. 779, 2021.

[14] G. Hofstede, *Culture's Consequences: International Differences in Work-Related Values*, Sage, Newcastle upon Tyne, UK, 1984.

[15] J. Li, R. T. Roessler, P. D. Rumrill, and E. R. Ahmed, "Factors influencing job satisfaction for employed adults with multiple sclerosis," *Rehabilitation Counseling Bulletin*, vol. 61, no. 1, pp. 28–40, 2017.

[16] J. K. Eskildsen, K. Kristensen, and A. H. Westlund, "Work motivation and job satisfaction in the Nordic countries," *Employee Relations*, vol. 26, no. 2, pp. 122–136, 2003.

[17] S. You, A. Y. Kim, and S. A. Lim, "Job satisfaction among secondary teachers in Korea," *Educational Management Administration & Leadership*, vol. 45, no. 2, pp. 284–297, 2017.

[18] A. E. Clark, "Job satisfaction and gender: why are women so happy at work?" *Labour Economics*, vol. 4, no. 4, pp. 341–372, 1997.

[19] Y. Luo, "Gender and job satisfaction in urban China: the role of individual, family, and job characteristics," *Social Indicators Research*, vol. 125, no. 1, pp. 289–309, 2016.

[20] A. E. Clark, "Job satisfaction in britain," *Journal of Industrial Relations*, vol. 32, no. 4, pp. 189–217, 1996.

[21] L. Chang and W. Kelly, "Individual employment characteristics of hotel employees that play a role in employee satisfaction and work retention," *International Journal of Hospitality Management*, vol. 29, no. 3, pp. 344–353, 2010.

[22] D. Scott, "Easton, leila dal santo, najwa S safadi, kim hokanson. job stressors and solutions: perspectives of social workers in the occupied Palestinian territories," *Human Service Organizations*, vol. 43, no. 1, pp. 1–15, 2021.

[23] Y. Xiao and S. Chen, "Increase income or increase the sense of fairness? a survey of job satisfaction of Chinese urban and rural labor," *Population and Economy*, no. 2, pp. 78–91, 2019.

[24] A. Ijadi Maghsoodi, I. Azizi-ari, Z. Barzegar-Kasani, M. Azad, E. K. Zavadskas, and J. Antucheviciene, "Evaluation of the influencing factors on job satisfaction based on combination of PLS-SEM and F-multimoora approach," *Symmetry*, vol. 11, no. 1, p. 24, 2019.

[25] R. M. Han, P. Carter, J. D. Champion, Lee, and D. Joseph, "Relationships among factors affecting advanced practice registered nurses' job satisfaction and intent to leave: a systematic review," *Journal of the American Association of Nurse Practitioners*, vol. 30, no. 2, pp. 101–113, 2018.

[26] T. Penconek, K. Tate, A. Bernardes et al., "Determinants of nurse manager job satisfaction: a systematic review," *International Journal of Nursing Studies*, vol. 118, Article ID 103906, 2021.

[27] P. Zhou and X. Jiang, "Work pressure and job satisfaction of state-owned enterprise managers: the moderating effect of psychological capital," *East China Economic Management*, vol. 27, no. 9, pp. 134–137, 2013.

[28] D. V. Jan, "The psychologization of humanitarian aid: skimming the battlefield and the disaster zone Hist," *Human science*, vol. 24, no. 3, pp. 103–122, 2011.

[29] D. Applebaum, S. Fowler, N. Fiedler et al., "The impact of environmental factors on nursing stress, job satisfaction, and turnover intention," *The Journal of Nursing Administration: The Journal of Nursing Administration*, vol. 40, no. 7/8, pp. 323–328, 2010.

[30] P. Liu, X. Wang, R. Liu, and W. Xia, "Research on the mechanism of the influence of leadership style on job satisfaction: taking employee relationship as the mediating variable," *Chinese Management Science*, vol. 21, no. S1, pp. 75–80, 2013.

[31] W. Wu, "The impact of working hours on professional well-being: an empirical analysis based on three typical occupations," *China Industrial Economics*, no. 3, pp. 130–145, 2016.

[32] F. M. Hussien and M. La Lopa, "The determinants of student satisfaction with internship programs in the hospitality industry: a case study in the USA," *Journal of Human Resources in Hospitality & Tourism*, vol. 17, no. 4, pp. 502–527, 2018.

[33] Bo Fu, G. Yu, and X. Liang, "The "double-edged sword" effect of the practice of subordinate relationship on employee performance: a cross-level analysis," *Scientific Research Management*, vol. 40, no. 8, pp. 273–283, 2019.

[34] D. Morton, C. Bowers, L. Wessels, A. Koen, and J. Tobias, "Job satisfaction of registered nurses in a private critical care unit in the Eastern Cape: a pilot study," *Health SA Gesondheid*, vol. 25, pp. 1–9, 2020.

[35] E. E. Kossek, S. A. Lobel, and A. J. Brown, "Human resource strategies to manage workforce diversity," in *Handbook of Workplace Diversity*, A. M. Konard, P. Prasad, and J. M. Pringle, Eds., pp. 54–74, Sage, Thousand Oaks, CA, USA, 2005.

[36] S. Alrawahi, S. F. Sellgren, N. Alwahaibi, S. Altouby, and M. Brommels, "Factors affecting job satisfaction among medical laboratory technologists in University Hospital, Oman: an exploratory study," *The International Journal of Health Planning and Management*, vol. 34, pp. e763–e775, 2019.

[37] M. T. Cardador, G. B. Northcraft, K. W. Rockmann, and B. C. Grant, "Characteristics of affected third parties and cooperative behavior in social dilemmas," *The Journal of Social Psychology*, vol. 156, no. 6, pp. 565–580, 2016.

[38] M. K. Nalla, S. Y. Paek, and S. S. Lim, "The influence of organizational and environmental factors on job satisfaction among security guards in Singapore," *Australian and New Zealand Journal of Criminology*, vol. 50, no. 4, pp. 548–565, 2017.

[39] N. Tomaževič and A. Aristovnik, "Factors of trust in immediate leaders: an empirical study in police service environment," *International Journal of Environmental Research and Public Health*, vol. 16, no. 14, p. 2525, 2019.

[40] P. Hersey and K. H. Blanchard, *Management of Organizational Behavior: Utilizing Human Resources*, Prentice-Hall, Hoboken, NJ, USA, 3rd edition, 1979.

[41] A. K. Havig, A. Skogstad, M. Veenstra, and T. I. Romøren, "The effects of leadership and ward factors on job satisfaction in nursing homes: a multilevel approach," *Journal of Clinical Nursing*, vol. 20, no. 23-24, pp. 3532–3542, 2011.

[42] S. P.-p. Choi, S. M.-c. Pang, K. Cheung, and T. K.-s. Wong, "Stabilizing and destabilizing forces in the nursing work environment: a qualitative study on turnover intention," *International Journal of Nursing Studies*, vol. 48, no. 10, pp. 1290–1301, 2011.

[43] M. A. Van der Wal, J. Schönrock-Adema, F. Scheele, N. R. Schripsema, A. D. C. Jaarsma, and J. Cohen-Schotanus, "Supervisor leadership in relation to resident job satisfaction," *BMC Medical Education*, vol. 16, no. 1, p. 194, 2016.

[44] L. Stringer, "The link between the quality of the supervisor-employee relationship and the level of the employee's job satisfaction," *Public Organization Review*, vol. 6, no. 2, pp. 125–142, 2006.

[45] F. Herzberg, "One more time: how do you motivate employees," *Harvard Business Review*, vol. 46, no. 1, pp. 53–62, 1968.

[46] B. Gilbreath ∗ and P. G. Benson, "The contribution of supervisor behaviour to employee psychological well-being," *Work & Stress*, vol. 18, no. 3, pp. 255–266, 2004.

[47] G. Abekah-Nkrumah and J. Nkrumah, “Perceived work environment and patient-centered behavior: a study of selected district hospitals in the central region of Ghana,” *PLoS One*, vol. 16, no. 1, Article ID e0244726, 2021.
[48] M. M. Sloan, “Gender differences in commitment to state employment: the role of coworker relationships,” *Public Personnel Management*, vol. 46, no. 2, pp. 170–187, 2017.
[49] H. Kroll, T. Macaulay, and M. Jesse, “A preliminary survey examining predictors of burnout in pain medicine physicians in the United States,” *Pain Physician*, vol. 5, pp. E689–E696, 2016.
[50] S. S. Kim, “The effect of social contexts and formation of individualism-collectivism orientation on knowledge sharing intention: the case of workers in Korea,” *Journal of Knowledge Management*, vol. 24, no. 2, pp. 196–215, 2020.
[51] G. Ulrich, A. Homberg, S. Karstens, K. Goetz, and C. Mahler, “Job satisfaction of young professionals in health care,” *Das Gesundheitswesen*, vol. 81, no. 2, pp. 99–105, 2019.
[52] A. Oven and B. Domajnko, “Job satisfaction and creativity at work among occupational therapy practitioners: a mixed-methods study,” *Work*, vol. 69, no. 4, pp. 1351–1362, 2021.
[53] J. F. Volkwein and K. Parmley, “Comparing administrative satisfaction in public and private universities,” *Research in Higher Education*, vol. 41, no. 1, pp. 95–116, 2000.
[54] S. Xu, L. Tao, H. Huang, J. Little, and L. Huang, “Pediatric nurses’ turnover intention and its association with calling in China’s tertiary hospitals,” *Journal of Pediatric Nursing*, vol. 52, p. e51, 2020.
[55] P. Brough and R. Frame, “Predicting police job satisfaction and turnover intentions: the role of social support and police organisational variables,” *New Zealand Journal of Psychology*, vol. 33, no. 1, pp. 8–18, 2004.
[56] C. B. V. Balinbin, K. T. R. Balatbat, A. N. B. Balayan et al., “Occupational determinants of compassion satisfaction and compassion fatigue among Filipino registered nurses,” *Journal of Clinical Nursing*, vol. 29, no. 5-6, pp. 955–963, 2020.
[57] G. A. Adams and T. A. Beehr, “Turnover and retirement: a comparison of their similarities and differences,” *Personnel Psychology*, vol. 51, no. 3, pp. 643–665, 1998.
[58] M. Cho, “Student perspectives on the quality of hotel management internships,” *Journal of Teaching in Travel & Tourism*, vol. 6, no. 1, pp. 61–76, 2006.
[59] E. Marinakou and C. Giousmpasoglou, “An investigation of student satisfaction from hospitality internship programs in Greece,” *Journal of Tourism and Hospitality Management*, vol. 1, no. 3, pp. 103–112, 2013.
[60] H. Qu, X. Y. Leung, S. Huang, and J. He, “Factors affecting hotel interns’ satisfaction with internship experience and career intention in China,” *Journal of Hospitality, Leisure, Sports and Tourism Education*, vol. 28, Article ID 100311, 2021.
[61] A. Malka and J. A. Chatman, “Intrinsic and extrinsic work orientations as moderators of the effect of annual income on subjective well-being: a longitudinal study,” *Personality and Social Psychology Bulletin*, vol. 29, no. 6, pp. 737–746, 2003.
[62] Z. Wang and H. Jiang, “An empirical study on the influence of individual characteristics of Chinese employees on company satisfaction,” *Nankai Management Review*, no. 1, pp. 101–106, 2004.
[63] M. C. Ellickson and K. Logsdon, “Determinants of job satisfaction of municipal government employees,” *Public Personnel Management*, vol. 31, no. 3, pp. 343–358, 2002.
[64] Y. Ohara, Y. Nomura, Y. Yamamoto et al., “Job attractiveness and job satisfaction of dental hygienists: from Japanese dental hygienists’ survey 2019,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 2, p. 755, 2021.
[65] Y. Yang, M. Li, T. Xiong, and H. Song, “The influence of the total remuneration of public institutions in Beijing on job satisfaction-the mediating role of salary fairness,” *Journal of Beijing Administration Institute*, no. 1, pp. 76–83, 2017.
[66] H. Wang, J. Yang, and L. Ye, “Research on the curve mechanism of salary level and job satisfaction,” *Economics and Management*, vol. 41, no. 7, pp. 105–120, 2019.
[67] P. G. Patterson, E. Cowley, and K. Prasongsukarn, “Service failure recovery: the moderating impact of individual-level cultural value orientation on perceptions of justice,” *International Journal of Research in Marketing*, vol. 23, no. 3, pp. 263–277, 2006.
[68] M. Alameddine, M. Baroud, S. Kharroubi et al., “Investigating the job satisfaction of healthcare providers at primary healthcare centres in Lebanon: a national cross-sectional study,” *Health and Social Care in the Community*, vol. 25, no. 6, pp. 1805–1816, 2017.
[69] J. S. Adams, “Towards an understanding of inequity,” *Journal of Abnormal and Social Psychology*, vol. 67, no. 5, pp. 422–436, 1963.
[70] U. Hytti, T. Kautonen, and E. Akola, “Determinants of job satisfaction for salaried and self-employed professionals in Finland,” *International Journal of Human Resource Management*, vol. 24, no. 10, pp. 2034–2053, 2013.
[71] O. Dousin, N. Collins, T. Bartram, and P. Stanton, “The relationship between work-life balance, the need for achievement, and intention to leave: m,” *Journal of Advanced Nursing*, vol. 77, no. 3, pp. 1478–1489, 2021.
[72] J. S. Adams, “Inequity in social exchange,” *Advances in Experimental Social Psychology*, vol. 2, no. 4, pp. 267–299, 1966.
[73] M. P. Miceli and P. W. Mulvey, “Consequences of satisfaction with pay systems: two field studies,” *Industrial Relations*, vol. 39, no. 1, pp. 62–87, 2000.
[74] B. Wang, “Research on the relationship between salary fairness, personality traits and job satisfaction,” *Scientific Research Management*, no. 3, pp. 91–100, 2011.
[75] D. McLoughlin and S. C. Carr, “Equity sensitivity and double demotivation,” *The Journal of Social Psychology*, vol. 137, no. 5, pp. 668–670, 1997.
[76] D. Al, A. Mohammed, M. Awais Bhatti, and A. Syah Juhari, *Do Psychological Diversity Climate, HRM Practices, and Personality Traits (Big Five) Influence Multicultural Workforce Job Satisfaction and Performance? Current Scenario, Literature Gap, and Future Research Directions*, SAGE Open, New York, NY, USA, 2019.
[77] Y. Wang, J. Liu, J. Zuo, and R. Rameezdeen, “Ways to improve the project management efficiency in a centralized public procurement system,” *Engineering Construction and Architectural Management*, vol. 27, no. 1, pp. 168–185, 2019.
[78] W. Lyu, Y. Zheng, C. Fonseca, and J. Z. Zhao, “Public-Private partnership transformation and worker satisfaction: a case study of sanitation workers in H-city, China,” *Sustainability*, vol. 12, no. 13, p. 5479, 2020.
[79] A. M. Saks and B. E. Ashforth, “A longitudinal investigation of the relationships between job information sources, applicant perceptions of fit, and work outcomes,” *Personnel Psychology*, vol. 50, no. 2, pp. 395–426, 1997.

[80] A. Bieńkowska and K. Tworek, "Job performance model based on employees' dynamic capabilities (EDC)," *Sustainability*, vol. 12, no. 6, p. 2250, 2020.

[81] J. R. Edwards, "Person-job fit: a conceptual integration, literature review, and methodological critique," in *International Review of Industrial and OrganIzational Psychology*, C. L. Cooper and I. T. Robertson, Eds., vol. Volume 6, pp. 283–357, John Wiley & Sons, Oxford, UK, 1991.

[82] A. L. Kristof, "Person-organization fit: an integrative review of its conceptualizations, measurement, and implications," *Personnel Psychology*, vol. 49, no. 1, pp. 1–49, 1996.

[83] M. A. Chilton, B. C. Hardgrave, and D. J. Armstrong, "Person-job cognitive style fit for software developers: the effect on strain and performance," *Journal of Management Information Systems*, vol. 22, no. 2, pp. 193–226, 2005.

[84] S. Firfiray and M. Mayo, "The lure of work-life benefits: perceived person-organization fit as A mechanism explaining job seeker attraction to organizations," *Human Resource Management*, vol. 56, no. 4, pp. 629–649, 2017.

[85] J. Luo, Y. Yang, and C. Sun, "Research on the individual-organization matching law of the career growth of female high-level talents in enterprises," *Management Review*, vol. 30, no. 12, pp. 187–199, 2018.

[86] B. Erdogan, T. N. Bauer, J. M. Peiró, and D. M. Truxillo, "Overqualified employees: making the best of a potentially bad situation for individuals and organizations," *Industrial and Organizational Psychology*, vol. 4, no. 2, pp. 215–232, 2011.

[87] J. Robst, "Education and job match: the relatedness of college major and work," *Economics of Education Review*, vol. 26, no. 4, pp. 397–407, 2007.

[88] E. E. Lawler, *Strategic Pay*, Jossey-Bass, San Francisco, CA, USA, 1990.

[89] A. Zuraik and L. Kelly, "The role of CEO transformational leadership and innovation climate in exploration and exploitation," *European Journal of Innovation Management*, vol. 22, no. 1, pp. 84–104, 2019.

[90] B. Dai and B. Akey-Torku, "The influence of managerial psychology on job satisfaction among healthcare employees in Ghana," *Healthcare*, vol. 8, no. 3, p. 262, 2020.

[91] J. R. Hackman and G. R. Oldham, "Development of the job diagnostic survey," *Journal of Applied Psychology*, vol. 60, no. 2, pp. 159–170, 1975.

[92] M. I. Loft and C. S. Jensen, "What makes experienced nurses stay in their position? A qualitative interview study," *Journal of Nursing Management*, vol. 28, pp. 1305–1316, 2020, http://gfzgf7f12bbd3c09549a1s99060wvkwnwk6n0f.fxbh.librra.gdufs.edu.cn/10.1111/jonm.13082.

[93] Y. Jo and L. T. Hoover, "Source of job satisfaction among South Korean police officers," *International Journal of Police Science and Management*, vol. 14, no. 2, pp. 136–153, 2012.

[94] A. B. Bakker and E. Demerouti, "The job demands-resources model: state of the art," *Journal of Managerial Psychology*, vol. 22, no. 3, pp. 309–328, 2007.

[95] F. P. Morgeson and S. E. Humphrey, "The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work," *Journal of Applied Psychology*, vol. 91, no. 6, pp. 1321–1339, 2006.

[96] S. Mohammed and L. C. Angell, "Surface- and deep-level diversity in workgroups: examining the moderating effects of team orientation and team process on relationship conflict," *Journal of Organizational Behavior*, vol. 25, no. 8, pp. 1015–1039, 2004.

[97] J. Liu, C. Liu, and R. Liu, "Research on the performance of flexible working hours of employees in high-tech enterprises: a mediation model based on work autonomy," *Science and Technology Management Research*, no. 19, pp. 119–124, 2015.

[98] L. Luo and X. Zhou, "Overtime, workload autonomy and utility on the existence of working hours constraints," *Economic Trends*, no. 3, pp. 59–73, 2019.

[99] L. Tian, "The five-dimensional exogenous cause of job satisfaction: which is more important?" *Foreign Economics and Management*, vol. 41, no. 9, pp. 75–90, 2019.

[100] M. Wu, Y. Yang, and X. Wang, "The impact of flexible work plans on employee psychology and behavior: a meta-analysis test," *Economics and Management*, vol. 42, no. 7, pp. 126–140, 2020.

[101] K. Namasivayam and X. Zhao, "An investigation of the moderating effects of organizational commitment on the relationships between work-family conflict and job satisfaction among hospitality employees in India," *Tourism Management*, vol. 28, no. 5, pp. 1212–1223, 2007.

[102] M. Deery and L. Jago, "Revisiting talent management, work-life balance and retention strategies," *International Journal of Contemporary Hospitality Management*, vol. 27, no. 3, pp. 453–472, 2015.

[103] L. L. Andersen, D. Fishwick, E. Robinson, N. M. Wiezer, Z. Mockałło, and V. Grosjean, "Job satisfaction is more than a fruit basket, health checks and free exercise: cross-sectional study among 10,000 wage earners," *Scandinavian Journal of Public Health*, vol. 45, no. 5, pp. 476–484, 2017.

[104] D. M. Harmening, B. M. Castleberry, and M. E. Lunz, "Technologists report overall job satisfaction: 10-year retrospective study examines career patterns," *Laboratory Medicine*, vol. 25, no. 12, pp. 773–775, 1994.

[105] L. Lum, J. Kervin, K. Clark, F. Reid, and W. Sirola, "Explaining nursing turnover intent: job satisfaction, pay satisfaction, or organizational commitment?" *Journal of Organizational Behavior*, vol. 19, no. 3, pp. 305–320, 1998.

[106] C. A. O'Reilly and J. Chatman, "Organizational commitment and psychological attachment: the effects of compliance, identification, and internalization on prosocial behavior," *Journal of Applied Psychology*, vol. 71, p. 492, 1986.

[107] M. Nedeljković Knežević, M. D. Petrović, S. Kovačić, M. Mijatov, D. B. Vuković, and J. Kennell, "Acting the part: emotional intelligence and job satisfaction as predictors of emotional labor in travel agencies," *Tourism and Hospitality Research*, vol. 21, no. 2, pp. 183–201, 2021.

[108] K. Doig and S. Beck, "Factors contributing to the retention of clinical laboratory personnel," *Clinical Laboratory Science: Journal of the American Society for Medical Technology*, vol. 18, no. 1, pp. 16–27, 2005.

[109] A. Yami, L. Hamza, A. Hassan, C. Jira, and M. Sudhakar, "Job satisfaction and its determinants among health workers in Jimma University Hospital," *Ethiopian Journal of Health Science*, vol. 21, no. 1, pp. 19–27, 2011.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 1819323, 9 pages
https://doi.org/10.1155/2021/1819323

Research Article

An Analysis of Community Group Buying Behavior of Urban Residents Based on Big Data

Nanxin Huang, Kexin Yu, and Cheng Chen

Department of Media, Broadcasting and Television Engineering, University of China, No. 1, Dingfuzhuang East Street, Communication University of China, Beijing 100024, China

Correspondence should be addressed to Nanxin Huang; hnancy@cuc.edu.cn

Received 25 August 2021; Accepted 25 October 2021; Published 27 November 2021

Academic Editor: Wen-Tsao Pan

By using keywords crawled by big data as a survey reference, this research applied latent category clustering method and binary logistic regression model analysis method to analyze the differences in community group buying behaviors of residents from different city scale and summarize the shopping behavior and features of different types of residents, for the purpose of offering advice on different marketing methods for different types of urban residents, so as to realize the precise marketing of community e-commerce and promote the further development of the industry.

1. Introduction

1.1. Research Background. In early 2020, the sudden outbreak of COVID-19 greatly changed people's way of life. Due to the pandemic, residents were forced to isolate at home. Under these circumstances, work and social life can only be conducted through the Internet, making it difficult for people to purchase the daily necessities. Therefore, community group buying industry in China witnessed an explosive growth during the pandemic, as it offers convenience for people to purchase daily necessities. According to iiMedia Research (a consulting agency) [1], the growth rate of community group purchase market increased by more than 100% and the market scale reached 72 billion yuan. Community group buying finally ushered in a new development boom in 2020, since it fell into trough in 2019.

In addition, we also found that the main battlefield of community group buying in the past two years began to develop in the sinking market, and it did not enter the first-tier and new first-tier cities such as Beijing, Shanghai, Guangzhou, and Shenzhen until the first half of 2020 [2]. Therefore, we are thinking of whether residents' purchase intentions and consumption behavior in community group purchase will be significantly different due to the different city scale where they live, and whether the influence factors will also be very different.

Now we are in the postepidemic era; the community group buying industry is accelerating its integration, and more entrepreneurs want to get a share of it, making the competition in this industry increasingly fierce. At the same time, people's reliance on the community group buying has decreased with the improvement of the pandemic [3]. If residents' consumption behavior in community group buying differs significantly due to the scale of the city, then different marketing should be adopted for residents from cities of different scales. This method is more conducive to community e-commerce to achieve precision marketing and promote further development of this industry.

1.2. Current Status of Research Conducted in China and Abroad. Through consulting domestic and foreign literature, we find that people 's consumption behavior is mainly affected by their internal factors such as gender, age, education, income, price awareness, quality awareness, and personal preferences, and it is also affected by external stimuli such as platform and social environment. In the analysis of the causes of the online shopping willingness of fresh agricultural products authored by Chen and Lu [4], it is shown that the basic personal characteristics such as gender, age, and income are significantly related to the results of consumers' buying fresh products online. In the research on the factors affecting consumers' willingness to buy imported fresh fruits online, He [5] found that the main factors affecting consumers' willingness to buy imported fresh fruits online are based on personal cognitive characteristics, such as time saving and labor saving, and the rich varieties of imported fruits. Guo and Xu [6] explored the influence factors of customer online shopping behavior under the background of "Internet +" and found that external environmental factors such as merchant, logistics, website, and commodities are the most direct factors affecting customer satisfaction.

In their research on the influence of online shopping festivals on consumers' online shopping intentions during the "Double Eleven Shopping Carnival," Bai and Liu [7] found that external environmental stimuli such as festival atmosphere and panic buying have a positive effect on consumers' shopping mood and, therefore, enhanced consumers' shopping intention.

In summary, existing research mainly analyzes the factors that affect consumers' online shopping willingness and rarely involves consumer community group buying willingness and behavior based on the perspective of city scale. Therefore, by making community group buying behavior as the carrier and city scale as the perspective, this research applied latent category cluster analysis, chi-square test, and binary logistic regression model to study the impact of different factors on urban residents' consumption behavior.

1.3. Collect the Data with the Octopus Software. Today, in the 21st century, information is growing in an explosive way. The era of big data has long come, and people use a variety of methods to deal with big data, such as cloud storage, cloud computing, and Python. As a product of the high-tech era, big data work as the original driving force for the world economic development. Contemporary industries, social networking, companies, and various industries are inseparable from big data. According to the "Big Data Special" report of Chinese Entrepreneurs, Nongfu Spring uses big data to sell mineral water. In 2011, SAP launched the innovative database platform, SAP Hana, with which real-time reporting could be achieved as compared with the previous data without big data. It can be seen that big data can bring forward-looking decisions and help optimize the allocation of resources. Therefore, the study will analyze the online shopping behavior of urban residents based on big data. Starting from the current background, the authors put forward the research topics, consulted the research situation in China and abroad, further determined the entry point of the research, used Octopus crawler software to collect data, and calculated the data to form cloud map and then analyze which online shopping modes are preferred by residents and how this influences online shopping behavior, products, and other information and thus obtained the characteristics of consumer shopping behavior.

2. Questionnaire Design and Sample Composition of a Community Group Buying Behavior

2.1. Investigation Method. This study adopts a questionnaire survey method to issue online questionnaires. On the one hand, the questionnaires were distributed to people in their area through relatives and friends. On the other hand, the questionnaires were distributed by random search community social groups on the Internet and also distributed randomly on social media platforms such as Weibo and Douban, so as to obtain more valid sample data. A total of 800 questionnaires were issued, and 750 questionnaires were collected, of which 672 were valid questionnaires, with an efficacy rate of 89.60%. All the collected questionnaires were recorded and sorted out. In the end, a total of 672 samples from primary, secondary, and below cities of Beijing, Guangdong province, and Hunan and Henan provinces were obtained for us to understand the basic situation of group buying behavior in urban cities of different scales.

2.2. Questionnaire Design. On the basis of literature review, the questionnaire was designed in combination with the opinions of relevant experts. We set up options such as gender, age, city, and monthly income to understand the basic information of the interviewees. At the same time, we defined the monthly purchase frequency, consumption level, purchase channels, shopping type, and other options to understand consumption behavior characteristics in community group.

This survey uses a presurvey method to evaluate the reliability and validity of the questionnaire and adjusts the content and structure of the questionnaire based on the survey situation and evaluation opinions. After the survey was completed, we checked the completeness of the questionnaire and logic of all questionnaires and removed the unqualified questionnaires.

2.3. Statistical Description of Individual Information. Among the 672 respondents, the proportions of the scale of the cities where residents live were equal. Among them, there were 311 respondents from large-scale cities and 361 from small cities, accounting for 53.7% of the total. More women were interviewed than men, accounting for 56.1% of the total respondents, and the majority were unmarried, accounting for 58.9% of the total number. Respondents were mainly aged among young groups, with 53.9% aged 18–25 years. More respondents were living with others compared to living alone. Among them, those with more than four people living together accounted for 38.4%. In addition, their monthly income was concentrated in the low to medium level. The respondents with an income of 5,000 yuan or less accounted for 66.7% of the total, and those with a monthly income of 10,000 yuan and above accounted for 10.1% of the total. The education level of the interviewees was mostly concentrated in the undergraduate level, accounting for 62.4% of the total. The majority were students, followed by employee of enterprise, public institutions, and individual businesses, accounting for 46.6%, 22.9%, and 9.8% respectively. Among the interviewees in this survey, 287 have participated in community group buying, accounting for 42.7% of the total.

3. Data Processing and Analysis of Group Buying Behavior of Urban Residents

3.1. Data Statistical Tools and Methods. This study used SPSS 25.0 and Mplus 8.0 to process the data, used Mplus to perform latent category cluster analysis, used SPSS to perform chi-square test on the data to check the correlation between the latent categories and the basic information of residents, and then used the binary logistic regression model to test the degree of influence of various basic information on group buying behaviors in different potential types of residents' communities.

3.2. Model Construction and Variable Selection

3.2.1. NVivo Software Analysis. NVivo is powerful qualitative analysis software that can effectively analyze a variety of different data. This research will use big data text analysis software NVivo to make statistics of the collected text data to form a word cloud map and then perform cluster analysis for further research. With these data, the authors find that the larger the proportion of the data area in the word cloud map is, the higher the willingness to do online shopping that consumers have, and the more they are inclined to choose this type of purchasing.

3.2.2. Latent Category Cluster Analysis Model Construction and Variable Selection. Potential category analysis is a mathematical model that describes the interrelationship between a set of categorical variables, and the method of integrated clustering is suitable for exploratory research.

This study first uses Mplus 8.0, based on the perspective of different city scales, and takes the consumption behavior characteristics of community group buying and the scale of permanent cities as observable external variables to conduct exploratory potential category analysis and find out the fit through specific indicators.

In the end, this study set a total of 11 categorical variables, such as urban scale, shopping channels, consumption level, purchase frequency, shopping preference, and important characteristics.

3.2.3. Chi-Square Test Analysis. Chi-square test is a widely used hypothesis testing method. It is used to calculate the degree of difference between the actual observed value and theoretical inferred value of a sample. It is usually expressed by Pearson's chi-square, and asymptotic significance is used as two random variables' statistical indicator of the closeness of the correlation. This study believes that when the asymptotic significance is less than 0.01, it can indicate that there is a strong correlation between two random variables.

In this study, SPSS 25.0 was used to perform a chi-square test to analyze the basic information and potential categories of the interviewees and to initially screen out variables with strong correlations with the potential categories.

3.2.4. Binary Logistic Regression Model Construction and Variable Selection. Binary logistic regression is a linear regression analysis model for binary categorical variables to be explained. It is often used in the fields of data mining and economic forecasting and other fields. This study establishes a binary logistic regression model for each potential category and discusses whether the personal basic information variable that has a strong correlation with the city scale variable has a significant impact on the community group buying behavior of residents in each potential category. Among them, the explanatory variables are the respondent's age, gender, occupation, marital status, monthly income, and the number of people living together, and the interpreted variable is whether to belong to this potential category.

4. Results and Analysis of Group Buying Behavior in Three Urban Communities

4.1. NVivo Analysis. After the text data was collected by Octopus crawler software, NVivo software was used to perform statistical analysis on text big data, count the frequency of word occurrences, and analyze the concerns of the consumer community during group purchases. Figure 1 shows the information that urban residents cared about during online shopping.

All the words in Figure 1 are closely related to community group buying. The results indicate that, because of the pandemic, commodity operation in the market has further developed towards community group buying. "Sink" and "city" reflect that the city scale is changing. The shopping goods are mainly raw and fresh fruits. Shopping channels mainly include Meituan selection optimization, Xingsheng Optimal, and related stores and convenience stores, and then these products are delivered to home to improve online shopping efficiency; thus a supply chain is formed in this way. Customers, cost, capital, commission, and o forth reflect the level of residents' purchase level. Price, demand, quality, and after-sales reflect the factors that consumers cared about when shopping. Community group buying platforms are mainly provided on small program, WeChat, online community, and so on. Analysis is made based on factors like the city scale, monthly purchase frequency, consumption level, purchase channels, shopping types, and so forth, to reflect consumers' purchase features in the community group buying under certain circumstances.

4.2. Cluster Analysis. The study starts with the single-category initial model and selects latent category models from single category to 7 categories to explore the minimum number of potential categories that can fully explain the relationship between the explicit variables of residents' consumption behavior.

Figure 1: Text word cloud diagram of consumer online shopping.

The indicators used in this study are Log (Log likelihood): log likelihood function value, AIC (Akaike information criterion), BIC (Bayesian information criterion) and aBIC (Sample-Size-Adjusted BIC): BIC after sample-size correction, Entropy: Entropy and LMR (Lo-Mendell-Rubin), and BLRT (parametric bootstrap likelihood ratio test): bootstrap-based likelihood ratio test.

Studies have shown that the smaller the values of Log, AIC, BIC, and aBIC, the better the fitting effect of the model; the higher the Entropy value, the higher the accuracy of its latent category classification; the significant LMR and BLRT values indicate that n categories of the model are better than the $n-1$ category model. Table 1 reports the data fit from the single-category model to the 7-category model.

The results in Table 1 show that Log (L) decreases with the increase in the number of categories. The information evaluation indicators AIC and aBIC have minimum values when the model category is 4, and the BLRT value reaches a very significant level ($p < 0.001$), indicating 4 potential categories better than 3 latent category models. Generally speaking, when the number of samples is not more than 1000, it is recommended to judge the fitting effect of the model with the AIC index. In total, 287 samples are analyzed in this study, so AIC can be used as a decision-making indicator for model suitability. According to the analysis results of the 7 models, the AIC value is the lowest when the number of model categories is 4 (3229.823), so this study considers to choose the 4 best-fitting latent models (Class1, Class2, Class3, and Class4).

The cluster icicle diagram is shown in Figure 2. By observing the height of the white strips, we can divide the number of 287 samples into 4 categories. In conclusion, through Mplus, the frequency of Class1 is 42, the frequency of Class2 is 106, the frequency of Class3 is 53, and the frequency of Class4 is 86. The x-axis represents the observation object, and the y-axis represents the frequency that can be divided into each category (Class). Each sample name corresponds to a blue strip, and 287 sample strips have the same length. There is also a white strip between every two sample strips. The length of the strip indicates the degree of similarity between the two samples. The higher the similarity, the longer the length of the white strip.

The average attribution probability matrices of 4 potential categories are shown in Table 2. The average probability distribution of each potential category is between 72% and 90%, indicating that the models with 4 potential categories are reliable.

A comprehensive analysis of Table 3 and Figure 3 shows that 287 urban residents with community group buying behaviors are classified into Class1, Class2, Class3, and Class4. They have the following characteristics:

> 42 urban residents come from Class1 cities, accounting for 14.6%. For Class1 cities, city scale is more evenly distributed and people have high monthly purchase frequency. Residents usually use community group buying APP to conduct community group buying and will not purchase due to the rich variety of goods. Among them, more than 75% of urban residents will buy food and nonfood goods through community group buying.
>
> 106 residents come from Class2 cities, accounting for 36.9%. For Class2 cities, city scale is smaller and has low monthly purchase frequency. Most of the purchase channels are community group buying APP, and residents choose community group buying due to rich variety of goods, time saving, affordable price, quality assurance, and good service quality. All residents will buy nonfood goods through community group buying.
>
> 53 residents belong to Class3 cities, accounting for 15.8%, Class3 cities are with a small city scale, low monthly purchase frequency, and high consumption level. Most of the purchase channels are self-organized purchases through WeChat group chats; because of the rich variety of goods, time saving, better quality assurance, and distribution service, the community group purchase is selected. The characteristics of price concessions are less important and everyone will buy food goods.
>
> 86 residents belong to Class4 cities, accounting for 30.0%. For Class4 cities, the city scale is more evenly distributed. Most of the purchase channels are through community group purchase apps. The consumption level is low. At the same time, residents choose community group buying due to rich varieties of commodities, time saving, affordable price, quality assurance, and good distribution service, and everyone will buy food goods.

It can also be seen from Figure 2 that the consumption behavior characteristics of the two categories of Class2 and Class4 are similar. The main difference lies in the types of goods purchased when conducting community group purchases. Most residents of Class2 live in small-scale cities, and everyone will buy nonfood products through community group buying. Most people in Class4, where the city scale is evenly distributed, will not buy nonfood products through community group purchases. Most residents of Class3 live in small-scale cities, and the response probability of monthly purchase frequency, purchase channel, and consumption level is significantly higher than those of the

Table 1: Potential category model adaptation index of the consumption behavior of community group buying.

Model	K	Log (L)	AIC	BIC	aBIC	Entropy	LMR	BLRT
1	11	−1673.063	3368.126	3408.38	3373.5			
2	23	−1626.824	3299.648	3383.817	3310.88	0.692	0.125	***
3	35	−1587.412	3244.823	3372.905	3261.92	0.721	*	***
4	47	−1567.912	3229.823	3401.819	3252.78	0.679	*	***
5	59	−1556.083	3230.166	3446.076	3258.98	0.722	0.224	0.3333
6	71	−1544.846	3231.693	3491.516	3266.37	0.768	0.2786	0.6667
7	83	−1533.372	3232.744	3536.481	3273.28	0.798	*	0.3077

Note. K is a free estimated parameter value (number of free parameters); $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$, the same below.

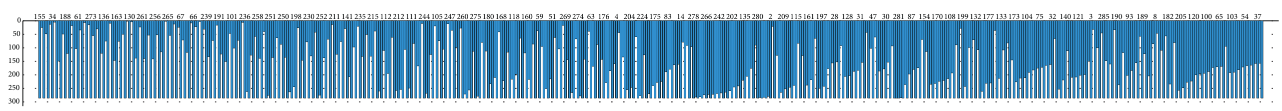


Figure 2: Cluster ice chart.

Table 2: Average ownership rate (column) of subjects (rows) in the 4 potential categories.

Class	C1 (%)	C2 (%)	C3 (%)	C4 (%)
C1	0.90	0.03	0.03	0.04
C2	0.02	0.72	0.10	0.16
C3	0.04	0.00	0.87	0.09
C4	0.06	0.09	0.03	0.82

Table 3: Response probability of 4 potential categories of group purchase consumption behavior in each variable.

Variable name	Variable value	Class1	Class2	Class3	Class4
(A) City scale	1. Large	0.469	0.296	0.153	0.555
	0. Small	0.531	0.704	0.847	0.445
(B) Monthly purchase frequency	1. Low	0.334	0.607	0.795	0.484
	0. High	0.666	0.393	0.205	0.516
(C) Purchase channels	1. Group chat spontaneous purchase	0.278	0.301	0.629	0.183
	0. Use the community group buying APP	0.722	0.699	0.371	0.817
(D) Consumption level	1. High level	0.405	0.522	0.941	0.240
	0. Low	0.595	0.478	0.059	0.760
(E) Do consumers choose to participate in the community group buying due to the rich variety of goods	1. Yes	0.241	0.944	0.979	0.918
	0. No	0.759	0.056	0.021	0.082
(F) Do consumers choose to participate in the community group buying because of saving time	1. Yes	0.470	1.000	0.650	0.952
	0. No	0.530	0.000	0.350	0.048
(G) Do consumers choose to participate in the community group buying due to the affordable price	1. Yes	0.646	1.000	0.486	1.000
	0. No	0.354	0.000	0.514	0.000
(H) Do consumers choose to participate in the community group buying due to quality assurance	1. Yes	0.575	0.924	0.872	0.909
	0. No	0.425	0.076	0.128	0.091
(I) Do consumers choose to participate in the community group buying due to the good distribution service	1. Yes	0.649	1.000	0.961	0.885
	0. No	0.351	0.000	0.039	0.115
(J) Do consumers buy food goods through community group buying	1. Yes	0.847	0.672	1.000	1.000
	0. No	0.153	0.328	0.000	0.000
(K) Do consumers buy nonfood goods through community group buying	1. Yes	0.750	1.000	0.507	0.393
	0. No	0.250	0.000	0.493	0.607

other three categories. The response probability of whether to choose to participate in the community group buying due to the affordable price is significantly lower than those of the other three categories; Class1 urban scale is evenly distributed. The response probability of whether to participate in community group buying due to the rich variety of goods,

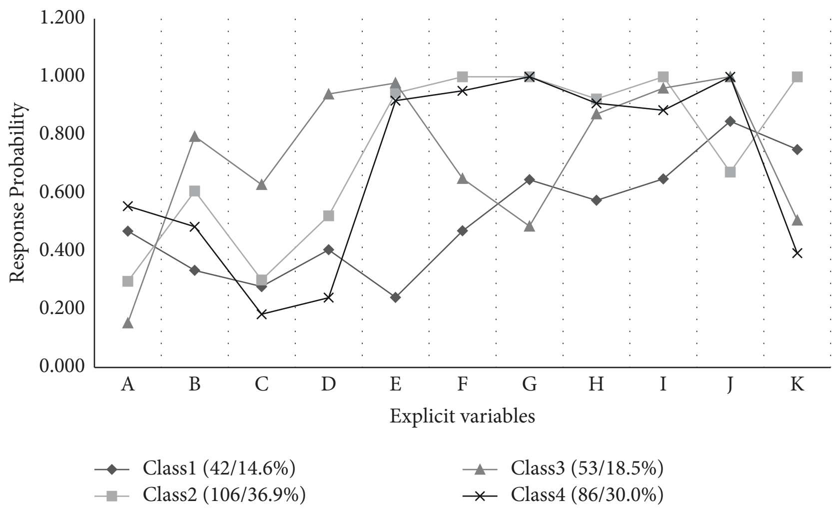


Figure 3: Response probability of the 4 potential categories of group purchase consumption behavior in each variable.

whether to choose to participate in community group buying because of time saving, whether to participate in community group buying because of better quality assurance, and whether to participate in community group buying due to good distribution service are significantly lower than those in the other three categories.

4.3. Inspection and Analysis of the Chi-Square Test. Firstly, the correlation between the two variables is analyzed: the residents' understanding of community group buying and whether they participate in community group buying and city scale.

It can be seen from Table 4 that the level of understanding of community group buying among small-scale urban residents is higher than that of large-scale cities. The number of residents participating in community group buying in small-scale cities (50.1%) is significantly more than that of those in large-scale cities (34.1%).

According to the chi-square test in Table 5, the progressive significance <0.01 reaches a significant level, indicating that the two variables, the respondents' understanding of community group buying and whether they have participated in community group buying, have a strong correlation with the variable of city scale.

Second, it studies the correlation between the basic information of the interviewees and the potential categories and makes a preliminary screening of the basic information variables that affect the group buying consumption behavior of residents' communities. The basic information is the age, gender, occupation, marital status, number of people living together, and education level.

From Table 6, it can be seen that, except for the variable of education level (progressive significance >0.01), the remaining basic information variables have a strong correlation with the potential category.

In conclusion, since the variable of education level has no significant correlation with the potential category, the variable of education level is excluded from the explanatory variable when analyzing the date with the binary logistic regression model.

4.3.1. Analysis of the Binary Logistic Regression Model. In order to further explore the difference of community group consumption behavior of residents from cities of different scale, this study uses the binary logistic regression model on four potential categories with different characteristics by using basic information that is highly correlated with potential categories as independent variables.

According to Tables 7 and 8, the impact of the birth year on all 4 groups was insignificant. This may be caused by the rapid development of Internet big data, as well as the increasing advancement of science and technology in China. Under this background, consumers of different birth years are all affected by the same external environment.

When studying the influence factors of the community group buying behavior of different groups, the authors find the following features:

For Class1 group, monthly income has a significant impact on the community group buying behavior. For Class1 consumers, their monthly income is at the middle and low levels of four groups. Such consumers generally do not have high requirements for the richness of product categories. The focus is on meeting their daily needs, and they pay more attention to the convenience brought by affordable prices.

For Class2 Group, the basic information variables are not significant. This is caused by smaller city scale in which household connection is closer and information exchange is more frequent and information received is similar. Therefore the basic information has less influence on the consumption behavior of this type of group.

For Class3 group, community group consumption behavior, gender, the number of people living together, marital status, monthly income, and occupation all have significant impact on the group. The reason is that

Table 4: Whether residents participated in the community group buying *city scale cross table in the current area. Community group purchase understanding degree *city scale cross table.

Count of the number		City scale: Small city scale	City scale: Large city scale	Total
Understanding degree of community group buying	Know a lot	157 (43.5%)	109 (35.0%)	266
	Do not know very well	144 (39.9%)	120 (38.6%)	264
	Know very well	38 (10.5%)	32 (10.3%)	70
	Do not know at all	22 (6.1%)	50 (16.1%)	72
Total		361	311	672
Whether residents participate in the community group buying in the current area	No	180 (49.9%)	205 (65.9%)	385
	Yes	181 (50.1%)	106 (34.1%)	287
Total		361	311	672

Table 5: Community group purchase understanding degree *city scale with chi-square inspection. Whether to participate in the community group purchase *city scale chi-square inspection in the current area.

Variables		Value	Degree of freedom	Progressive significance (bilateral)
Understanding degree	Pearson's square	18.630	3	***
Participation or not		17.600	1	***

Table 6: Basic information *potential category with chi-square inspection.

Variables		Value	Degree of freedom	Progressive significance (bilateral)
Gender	Pearson's square	14.203	3	**
Year of birth		199.379	138	***
Education level		24.377	12	0.018
Number of people living together		30.899	9	***
Marriage status		45.102	3	***
Monthly income		53.312	18	***
Career		43.650	3	***

Table 7: Model regression results.

The project	Class1	Class2	Class3	Class4
Gender	0.559	0.18	0.001***	0.188
Year of birth	0.904	0.929	0.938	0.88
Number of people living together	0.115	0.162	0.095*	0.429
Marriage status	0.714	0.763	0.002***	0.06*
Monthly income	0.016**	0.108	0.004***	0.102
Career	0.806	0.226	0.046**	0.000***

Note. *** **and *represent the significance levels of 1%, 5%, and 10%, respectively.

Table 8: 4 distribution differences in the potential categories on the basic information variables.

Basic information variable		Class1 ($n = 42$)	Class2 ($n = 106$)	Class3 ($n = 53$)	Class4 ($n = 86$)	Total
Gender	Female	26 (13.76%)	75 (39.68%)	24 (12.70%)	64 (33.86%)	189
	Male	16 (16.32%)	31 (31.63%)	29 (42.65%)	22 (22.45%)	98
Marital status	Unmarried	21 (15.44%)	49 (36.03%)	6 (4.42%)	60 (44.11%)	136
	Married	21 (13.91%)	57 (37.75%)	47 (31.12%)	26 (17.22%)	151
People living together	One person	12 (21.82%)	18 (32.73%)	1 (1.82%)	24 (43.63%)	55
	Two people	3 (10.00%)	12 (40.00%)	5 (16.67%)	10 (33.33%)	30
	Three people	17 (16.50%)	34 (33.01%)	32 (31.07%)	20 (19.42%)	103
	Four people and more	10 (10.10%)	42 (42.42%)	15 (15.16%)	32 (32.32%)	99

Table 8: Continued.

Basic information variable		Class1 ($n = 42$)	Class2 ($n = 106$)	Class3 ($n = 53$)	Class4 ($n = 86$)	Total
Monthly income	RMB 3,000 and less	19 (17.27%)	37 (33.64%)	6 (5.45%)	48 (43.64%)	110
	\$3000–5000	11 (16.18%)	28 (41.17%)	12 (17.65%)	17 (25.00%)	68
	\$5000–8000	7 (10.45%)	22 (32.83%)	27 (40.30%)	11 (16.42%)	67
	RMB 8000–10000	1 (5.88%)	7 (41.18%)	4 (23.53%)	5 (29.41%)	17
	RMB 10,000-RMB 20,000	2 (14.29%)	8 (57.14%)	3 (21.43%)	1 (7.14%)	14
	RMB 20,000-RMB 50,000	1 (10.00%)	4 (40.00%)	1 (10.00%)	4 (40.00%)	10
	RMB 50,000 and above	1 (100.00%)	0	0	0	1
Career	Students	14 (15.38%)	29 (31.87%)	1 (1.10%)	47 (51.65%)	91
	Nonstudent	28 (14.28%)	77 (39.29%)	52 (26.53%)	39 (19.90%)	196

married high-income men are the main shoppers in Class3; and as they have more family members, they need to buy a lot of daily necessities to meet the family needs at each time. Also high-income men are busy with work; the frequency of participating in community group purchases is relatively low, so they cared about time saving and family life quality. Factors like the variety and richness of commodity, product quality, and the speed of delivery service on the community group buying platform are particularly important.

For Class4 group, marital status and occupation have a significant influence on this group, where the individuals are mostly unmarried, and more than half of them are students. The individuals in this group usually only need to ensure their daily life needs, and most of them are living in a collective life and do not have good food storage conditions. So they tend to buy only short-term needs. Therefore their community group purchase frequency is higher, while purchase amount is lower than those of the rest of the groups.

5. Conclusions and Suggestions of Differences in Four Urban Residents

5.1. Main Conclusions. Based on the survey data of community group buying behaviors of 672 respondents in cities of different sizes, latent category clustering analysis and binary logistic regression analysis are carried out, and the following research conclusions are drawn.

The differences in community group purchasing behaviors of urban residents of different sizes are as follows.

Small-scale urban residents account for a larger proportion of the two consumer behavior categories, Class2 and Class3, which is because, in small-scale cities, community group buying is more popular and has a wider audience. Through latent category cluster analysis, the 287 urban residents who have participated in community group purchases are roughly divided into 4 types of consumption behavior, among which the consumption behaviors of urban residents in the Class3 and Class4 categories have obvious grouping characteristics. On the whole, only the year of birth has no significant impact on the four groups. Monthly income has a significant impact on the consumption behavior of Class1 urban residents, and the number of people living together, occupation, marital status, and gender have a significant impact on the consumption behavior of Class3 urban residents. Marital status and occupation have a significant impact on the consumption behavior of Class4 urban residents.

5.2. Countermeasures and Suggestions. On the whole, the vast majority of consumers value the safety of commodity quality and quality of community group purchase platform and the superiority of distribution service. Therefore, strengthening the quality of goods and improving the quality of after-sales service platform will undoubtedly encourage more users to choose community group purchase. The thoughtfulness of distribution service is also a plus. In addition, more than half of the people in the city residents involved in the questionnaire have not used the community group buying, and chi-square test inspection analysis results show that the understanding of community group buying is strongly correlated to the city scale. Residents living in large-scale cities are less involved in community group buying and have lower understanding of it. It is recommended to use big data advertising to increase the promotion of community group buying as a new shopping mode.

Starting from the consumption behavior of urban residents, it is found that consumers of different groups also show different urban scale distribution and shopping behavior characteristics. Therefore, for the community e-commerce company to accurately identify the target customers and achieve accurate marketing, the specific suggestions for four types of urban residents are as follows.

For Class1 group, because they have low demand for the richness of product types and higher purchase frequency and they are more concerned about price benefits, community e-commerce companies can carry out a large promotion of several commodities on the e-commerce platform according to the psychology of such consumers, so as to attract such people to come to consume.

In view of Class2 group having no obvious influencing factors, community e-commerce platforms should be more targeted in advertisement. They can conduct in-depth investigation into all walks of life, explore the more detailed shopping needs of such groups, and provide a clear target direction for future advertisement, as well as discount activities.

For Class3 group, community platforms should increase their dependence and irreplaceability on community group buying, such as more efficient delivery service and better quality assurance, because most of the individuals in this group are keen to use WeChat group chat to organize group purchases on their own. The community e-commerce platform can promote the further growth of the commodity shopping group by attracting more enthusiastic mothers or community store owners to join the group leaders.

Aiming at the Class4 group, based on the characteristics that most of its individuals are students, the e-commerce platform can adopt means such as product promotions and student discounts to develop more potential student users or also set up more community group buying sites on campus to improve the convenience of community group buying so as to attract more students to join in the community group buying.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Ai Media Consulting, *Special Research Report on Chinese Community Group Purchase Industry in the First Half of 2020*, Sydney, Australia, 2021.

[2] C. Jia, "Analysis of the development status of community group buying in the "post-epidemic era"," *Mall Modernization*, vol. 18, pp. 6–8, 2020.

[3] Severe, "Thoughts on the development of the new retail model of community group buying," *Mall Modernization*, vol. 8, pp. 17-18, 2019.

[4] P. Chen and M. Lu, "Analysis of the online shopping intention of fresh agricultural products—based on the survey data of Hengyang city," *Journal of Hengshui Normal College*, vol. 4, pp. 107–112, 2019.

[5] Y. He, "Investigation on the influencing factors of consumers' willingness to import fresh fruit online shopping," *Business Economy*, vol. 1, pp. 99–178, 2020.

[6] X. Guo and J. Xu, "Customer online shopping behavior influence factors under the background of "internet +" research—takes college students as an example," *Logistics Technology*, vol. 10, pp. 42–46, 2020.

[7] B. Xin and K. Liu, "Study on the impact of the "double 11" online shopping atmosphere on consumers' willingness to do online shopping," *Value Engineering*, vol. 2, pp. 57–59, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4013468, 12 pages
https://doi.org/10.1155/2021/4013468

Research Article

Research on Industry Difference and Convergence of Green Innovation Efficiency of Manufacturing Industry in China Based on Super-SBM and Convergence Models

Yongcan Yan, Jian Li, and Yi Xu

School of Management, Tianjin University of Technology, Tianjin 300384, China

Correspondence should be addressed to Jian Li; lj_tjlgdx001@126.com

Received 15 August 2021; Revised 28 October 2021; Accepted 6 November 2021; Published 26 November 2021

Academic Editor: Yiwen Zhang

To accurately grasp the current situation of green innovation efficiency in the manufacturing industry in China, this paper analyzes the differences and convergence characteristics of green innovation efficiency in various industries. Based on the panel data of 29 manufacturing industries in China from 2010 to 2019, the super-slack-based measure (Super-SBM) model measures the green innovation efficiency of manufacturing industries whose evolution characteristics are classified and analyzed from the perspective of technical demand. The Dagum Gini coefficient decomposition method indicates the source of industry differences in green innovation efficiency of the manufacturing industry in China with its convergence characteristics analyzed from the time dimension by constructing σ and β convergence models. The results reveal the improvement of green innovation efficiency of the Chinese manufacturing industry with obvious distinctions among different sectors and the industries with high green innovation efficiency, mostly high-end technology ones. The narrowing overall difference of green innovation efficiency in the manufacturing industry is accompanied by the lowest contribution rate of super-variable density, with the disparities between groups being the main source. It also shows the fluctuation of the intermittent σ convergence characteristics of the national manufacturing industry as a whole and low-end and high-end technology industry groups. However, the entire manufacturing industry and the three groups witness the absolute β convergence trend, with an ununiform convergence rate. The research will provide a reference for further upgrading the efficiency of green innovation in the industry and help to achieve the goals of carbon emission reduction and neutrality with the policy implications for promoting high-quality development of the manufacturing industry.

1. Introduction

With the transformation from high-speed growth to high-quality growth economic in China, the manufacturing industry, as the main body of the national economy, has developed extensively with high pollution and energy consumption. In responding to the global environmental governance and green economic development, a high-quality development goal will be obtained with refined independent innovation capability and cutting-edge core technologies. As a major manufacturing country globally, China sees its manufacturing industry as the big energy consumer and carbon emitter, with the industry being the key for its national energy-saving goal [1]. Under the 2030 carbon peak target and the 2060 carbon-neutral vision, pollution, carbon reduction, and green innovation will abound in the high-quality development of the Chinese manufacturing industry. Recently, major manufacturing countries have intensified their scientific and technological innovation, actively promoting green creation in manufacturing. Despite some world-leading innovative industries and technologies, China still undergoes an unbalanced innovation development with resource and environmental dilemmas in most cases [2]. To achieve a high-quality manufacturing industry, it is necessary to take green transformation as the development goal, innovation and development as the core driving force, green development and innovation drive as the combination point. Also, it is necessary to promote the two-way balance between economic growth and resources and environment to further enhance its green innovation efficiency [3].

Green innovation sustains environmental, ecological, and sustainable innovations. The green manufacturing industry should anchor its development goal, comprehensively grasp the products, processes, technologies, services, and the management of the whole life cycle to realize multidimensional and whole processes to reduce the environmental pressure caused by the manufacturing industry [4]. Therefore, it is important to promote various industries to achieve carbon neutrality by studying the green innovation efficiency of the manufacturing industries and giving some suggestions to different sectors for efficiency improvement [5].

Since the concept of sustainable development made its debut, green innovation has caught much attention of the researchers worldwide. They have done a lot of research on it from different perspectives and yielded fruitful results. Scholars analyzing the green innovation efficiency from regional heterogeneity believe that the efficiency of green innovation in China has increased [6], with a low-efficiency level overall [7]. The absolute and conditional β-space convergences characterize the great spatial differences in the green innovation efficiency among the provinces. Under different production technology conditions, the efficiency of regional green innovation in China decreased from east to west to center [8]. However, some scholars believe that the national green innovation efficiency shows a step-like decline in the eastern, the central, and the western regions, with the gap in green innovation efficiency among the regions narrowed by the reasonable flow of factors, optimized allocation of innovation resources, and stimulated innovation vitality [9]. In addition, scholars who studied the green innovation convergence have indicated the significant trend of absolute and conditional β convergences in the regional green innovation efficiency across the board [9, 10]. However, after studying the rural green efficiency development, the scholars unveiled the gradually enhanced efficiency of rural green development without an absolute β convergence and "catch-up effect" between the regions [11].

By studying the industrial green innovation efficiency, the scholars discovered innovative and unsustainable phenomena in the heavily polluting industries in China, with green efficiency being the key to lowering the overall green innovation efficiency of the industry. Heavy-polluting industries need to strictly implement the environmental regulation policies, increase green technology development and application, and promote green transformation [12, 13]. There are significant distinctions in green innovation efficiency among different pollution-intensive industries. According to the actual development needs, all industries should follow green development and innovation and promote industrial agglomeration innovation and green transformation [14]. The Chinese manufacturing industry shifts toward green innovation with room for green innovation efficiency improvement, but there are significant differences in the east and west, and the regional differences are gradually expanding [15]. The lower green innovation efficiency of all industries than that of the whole manufacturing industry is followed by the higher efficiency of the patent-intensive manufacturing industry than that of nonpatent-intensive counterparts because of industrial heterogeneity [14, 16]. With the increasing implementation effect of favorable policies, China's manufacturing industry has realized the dual-path transformation of emission reduction and efficiency increase [17]. The green innovation efficiency of manufacturing industry in the Yangtze River Economic Belt has been steadily improved. However, there is still a large space for improvement [18]. Some scholars divide the efficiency of green innovation into two stages: research and development (R&D) and achievement transformation, as well as the measurement of the regional efficiency of the high tech manufacturing industry [19]. Luo et al. maintained the disparate green technology innovation efficiency in industries despite the annual swelling in the national green innovation efficiency of the strategic new ones [20]. Li et al. believed that the green innovation efficiency of the Chinese high tech industries evolved from the large differences in low efficiency to high efficiency, with the climbing proportion of high-efficiency provinces [21]. Claudio et al. measured the technological innovation efficiency of the Spanish manufacturing industry from 1992 to 2005 and found big industry differences in the technological innovation efficiency [22].

The summary of the worldwide factors affecting the manufacturing efficiency of green innovation unveiled the different influences of R&D investment, government support, environmental regulation, and enterprise scale. Yi et al. believed that the government R&D subsidies and ecological regulations improved the manufacturing green innovation efficiency in the manufacturing industry [23], with the firm size and industrial structure impeding the green innovation efficiency that is irrelevant to economic openness. Nuryakin et al. exploited the Batik industry in Indonesia to test the factors of green product and green process innovations of Batik enterprises [24]. Based on the relevant data of the development of the German manufacturing industry, Nuryakin et al. adopted the bivariate Probit model to study the factors affecting its green innovation and believed that increasing R&D investment and environmental regulation could promote green innovation [25].

The existing research mainly focuses on three aspects: regional green innovation efficiency [26], industrial green innovation efficiency [27], and the factors of green innovation efficiency [28]. Insufficient research on the difference and convergence of manufacturing green innovation efficiency from the heterogeneity of industry technology demand is accompanied by the universally proposed policy suggestions and promotion paths void of industry pertinence. Therefore, it is of theoretical and practical significance to accurately grasp the green innovation efficiency of the manufacturing industry with differentiation measures to promote the manufacturing industry, coordinated development of innovation and resources, and the environment.

This paper ingeniously divided the 29 domestic manufacturing industries into high-end, middle-end, and low-end technology industries concerning different technical requirements. The measurement of green innovation efficiency of the manufacturing industry from 2010 to 2019 was followed by the obtained sources of differences in that of various sectors with the convergence trend study of industry differences based on s and β convergences. To narrow the industry differences and improve the efficiency of green innovation in the Chinese manufacturing industry, this paper finally puts forward some countermeasures for green innovations and the environment.

The remainder of the paper is organized as follows: section 2: the model detailing, section 3: data presenting, section 4: the empirical result discussing, and section 5: conclusion with policy implications.

2. Methodology and Models Specification

2.1. Super-SBM Model. The super-slack-based measure (Super-SBM) model is a nonradial and nonangle efficiency evaluation model proposed by Tone [29]. Compared with the traditional CCR and BCC models, this one overcomes the relaxation effect of elements, considers the relaxation variables in the objective function, and solves the scheduling problem when the SBM model cannot simultaneously distinguish multiple effective decision-making units. Therefore, the super-SBM model with an undesired output for the green innovation efficiency measurement of the manufacturing industries chimes more with the actual research need [30, 31]. The specific model is as follows:

$$\min \rho^* = \min \frac{1 - 1/W \sum_{w=1}^{W} s_n^x / x_n^k}{1 + \left[1/M + I\left(\sum_{m=1}^{M} s_m^y / y_m^k + \sum_{i=1}^{I} s_i^b / b_m^k \right) \right]}$$

$$s.t. \begin{cases} \sum_{k=1}^{K} z_y^k y_m^k - s_m^y = y_m^k, & m = 1, \ldots, M, \\ \sum_{k=1}^{K} z_y^b b_i^k - s_i^b = b_i^k, & i = 1, \ldots, I, \\ \sum_{k=1}^{K} z_k^x x_w^k - s_w^x = x_w^k, & w = 1, \ldots, W, \\ z_k^i \geq 0, s_m^y \geq 0, s_i^b \geq 0, s_w^x \geq 0, & k = 1, \ldots, K, \end{cases} \tag{1}$$

where ρ^* is the green innovation efficiency value of various industries in the manufacturing industry, s_m^y, s_i^b, s_w^x represent the slack variables, x_w^k, y_m^k, b_i^k, respectively, mean the input element, expected output, and unexpected output of the kth production unit; W, M, and I, respectively, equate the quantity of input factors, expected output, and unexpected output; z_k^x, z_y^k, z_y^b, respectively, denote the weight of the above three indicators.

2.2. Dagum Gini Coefficient Decomposition. Dagum Gini coefficient and its decomposition can measure the sources and contributions of green innovation efficiency development in various industries [32]. It can obtain the changing trend of the overall industry differences of the manufacturing industry in the sample period and reveal the intragroup and intergroup differences of grouped industries [33]. According to the subgroup decomposition method, this method can be divided into intragroup gap, intergroup gap, and supervariable density. The overall Gini coefficient is defined as formula (2) and the Gini coefficients within and between the groups as formulas (3) and (4). Among them, equation (3) represents the Gini coefficient G_{jj} of the industry group j, with (4) representing the Gini coefficient G_{jh} between the industry groups j and h.

$$G = \frac{\sum_{j=1}^{q} \sum_{h=1}^{q} \sum_{l=1}^{n_j} \sum_{r=1}^{n_h} \left| Y_{jl} - Y_{hr} \right|}{2n^2 \overline{Y}}, \tag{2}$$

$$G_{jj} = \frac{1/2\overline{Y}_j \sum_{l=1}^{n_j} \sum_{r=1}^{n_j} \left| Y_{jl} - Y_{jr} \right|}{n_j^2}, \tag{3}$$

$$G_{jh} = \frac{\sum_{l=1}^{n_j} \sum_{r=1}^{n_h} \left| Y_{jl} - Y_{hr} \right|}{n_j n_h \left(\overline{Y}_j + \overline{Y}_h \right)}, \tag{4}$$

where q represents the number of industry groups, n, the number of all industries, Y_{jl} and Y_{hr}, respectively, the green innovation efficiency values of l and r industries in j and h industry groups. n_j and n_h, the number of industries in the corresponding j and h groups, and $\overline{Y}$, the average value of green innovation efficiency of all manufacturing industries. $\overline{Y}_j$ and $\overline{Y}_h$ denote the average value of green innovation efficiency of j and h industry groups.

The results of intraindustry gap G_w, interindustry gap G_{nb}, and hypervariable density G_t can be expressed as follows:

$$G_w = \sum_{j=1}^{q} G_{jj} P_j S_j, \tag{5}$$

$$G_{nb} = \sum_{j=2}^{q} \sum_{h=2}^{j-1} G_{jh} D_{jh} \left(P_j S_h + P_h S_j \right), \tag{6}$$

$$G_t = \sum_{j=2}^{q} \sum_{h=2}^{j-1} G_{jh} \left(P_j S_h + P_h S_j \right) \left(1 - D_{jh} \right). \tag{7}$$

In equation (5), $p_j = n_j/n$, $s_j = n_j\overline{Y}_j / n\overline{Y}$, $j = 1,2, \ldots, q$; in equations (6) and (7), D_{jh} is the relative influence of the green innovation efficiency between the industry groups j and h as shown in equation (8). d_{jh} represents the difference of green innovation efficiency among the industry groups, and the mathematical expectation of the sum of all the sample values of $Y_{jl}-Y_{hr}>0$ in the industry groups j and h as shown in equation (9). p_{jh} represents the supervariable first order matrix and is the mathematical expectation of the sum of all sample values of $Y_{hr}-Y_{jl}>0$ as shown in equation (10).

$$D_{jh}=\frac{\left(d_{jh}-p_{jh}\right)}{\left(d_{jh}+p_{jh}\right)}, \tag{8}$$

$$d_{jh}=\int_0^{\infty} dF_j(Y)\int_0^{Y}(Y-x)dF_h(x), \tag{9}$$

$$p_{jh}=\int_0^{\infty} dF_h(Y)\int_0^{Y}(Y-x)dF_j(x). \tag{10}$$

In equations (9) and (10), the functions F_j and F_h represent the cumulative density distribution functions of the industry groups j and h.

2.3. Convergence Analysis Method. To analyze how green innovation efficiency differences in the manufacturing industries evolve, this paper applied σ and β convergences to investigate that of the green innovation efficiency in the Chinese manufacturing industry [34, 35].

σ convergence test model: σ convergence can be understood as a process with a continuous decline of the dispersion degree of green innovation efficiency in different industries over time. In this paper, the coefficient of variation method was used, and its calculation equation is as follows:

$$\sigma=\frac{\sqrt{\sum_l^{N_j}\left(PS_{lj}-\overline{PS_{lj}}\right)/N_j}}{\overline{PS_{lj}}}, \tag{11}$$

where j represents the industry group, l, the industry included in the industry group, N_j, the number of industries included in the industry group j, and $\overline{PS_{lj}}$, the mean value of the green innovation efficiency of the industry group j.

β convergence test model: β convergence means that as time goes by, the industries that are low in green innovation efficiency but high in growth rate will overtake the efficient industries with a bridged gap and a consistent level. The different application preconditions enable the division of β convergence: absolute and conditional β convergences. In this paper, the industry convergence trend of the green innovation efficiency in the manufacturing industry is mainly studied based on the absolute β convergence without the influence of other factors on the industry green innovation efficiency. The absolute β convergence model is as follows:

$$\ln\left(\frac{PS_{l,t+1}}{PS_{l,t}}\right)=\alpha+\beta\ln\left(PS_{l,t}\right)+\mu_l+\eta_t+\varepsilon_{lt}, \tag{12}$$

where l represents the industry ($l=1, 2, \ldots, N$), t represents the time ($t=1, 2, \ldots, T$). $PS_{l,t+1}$, $PS_{l,t}$, respectively, equate the green innovation efficiency of industry l in t+1 and t periods; $PS_{l,t+1}/PS_{l,t}$ denotes the annual growth rate of the green innovation efficiency of l industry from t to t+1. β is the convergence coefficient. If $\beta<0$ and passes the significance test, then it indicates that β convergence exists in the green innovation efficiency of the Chinese manufacturing industry with its convergence rate expressed as $\nu=-\ln(1+\beta)/T$. If $\beta>0$ and passes the significance test, then divergence exists. μ_l represents the individual effect of the industry, η_t, the time effect, and ε_{lt}, the interference terms obeying independently and identically distributed.

σ convergence emphasizes that the difference of green innovation efficiency in the manufacturing industries will become smaller with time, while β convergence is more focused on describing the convergence process from the angle of catching up than σ convergence, which can not only get the convergence situation of green innovation in the manufacturing industries, but also get the convergence speed.

3. Indicators and Data

The green innovation efficiency index should be selected in a scientific, objective, and truthful way with the real manufacturing efficiency in green innovation and the index data available. Therefore, this paper took the panel data of 29 industries in the Chinese manufacturing industry from 2010 to 2019 as samples and selected the workforce, capital, and energy input indicators to measure the green innovation efficiency of the manufacturing industry. Among them, the proxy variable of human input was the full-time equivalent of R&D personnel with the capital input characterized by three variables: internal expenditure of R&D funds, expenses of new product development funds, and technology introduction and transformation funds. Technology introduction and transformation funds equated the total cost of various industries with the energy input characterized by the total energy consumption.

The expected and unexpected outputs mainly measured the output indicators of green innovation. In this paper, two were selected as the expected output indicators: (1) the sales revenue of new products reflecting the market value transformation results of the industries; (2) the number of patent applications mirroring the independent innovation results of the industries. The pollutant index reproducing the natural environment of various industries in the innovative R&D activities was the unexpected output with industrial sulfur dioxide emissions, wastewater emissions, and general industrial solid waste production used to measure the environmental impact of green innovation activities in the manufacturing industry.

The index data in this paper are obtained from the relevant statistical yearbooks with their authenticity and reliability underpinning the research of this paper. The related indicators of human input, capital input, and expected output are from the 2011–2020 China statistical yearbook of science and technology. The energy input index comes from the 2011–2020 China energy statistical yearbook with the undesired output indicators from the 2011–2020 China environmental statistics yearbook. The descriptive statistics of the relevant indicators in 2019 are shown in Table 1.

Technology drives further innovation and quality development of the manufacturing industry, whose balanced

Table 1: The descriptive statistics of the evaluation index of the green innovation efficiency of manufacturing industry in 2019.

	Variables	Max	Min	Mean	Standard deviation
Input variable	Full-time equivalent of R&D personnel (man-year)	543781	4256	104812	115387
	Intramural expenditure on R&D (10 000 yuan)	24480937	303865	4652828	5261353
	Expenditure on new products development (10 000 yuan)	36778181	293783	5757977	7441695
	Total expenditure on technology introduction and technological transformation (10 000 yuan)	9609934	30016	1347702	2142693
	Total energy consumption (104 tce)	65387	192	9232	16243
Expected output variable	Sales revenue of new products (10 000 yuan)	441509516	3540521	72037052	94928552
	Patent applications (piece)	204836	3277	35018	47170
Unexpected output variable	Industrial sulfur dioxide Emission (ton)	1037198	8	101176	234231
	Industrial waste water discharged (10 000 tons)	76977	240	14832	20702
	Common industrial solid wastes generated (10 000 tons)	56269	11	5463	12153

economic and environmental benefits hinge on the former with the latter's diversified call for ununiform technology needs. To better compare the green innovation efficiency of the manufacturing industries under different technology needs, this paper analyzed the overall differences of the green innovation efficiency among the disparate groups in the manufacturing industries. Based on the manufacturing industry classification in the organization for economic co-operation and development and relevant research, 29 manufacturing industries were divided into three groups, concerning R&D investment intensity: low-end, mid-end, and high-end technology industries. The low-end technology industry includes the processing of food from agricultural products (C13), manufacture of foods (C14), manufacture of liquor, beverages, and refined Tea (C15), manufacture of tobacco (C16), manufacture of textile (C17), manufacture of textile, wearing apparel, and accessories (C18), manufacture of leather, fur, feather, and related products and footwear (C19), processing of timbers and manufacture of wood, bamboo, rattan, palm, and straw products (C20), manufacture of furniture (C21), manufacture of paper and paper products (C22), printing and reproduction of recording media (C23), manufacture of articles for culture, education, arts and crafts, sport, and entertainment activities (C24), and other manufacture (C41). Mid-end technology industry includes the processing of petroleum, coal, and other fuels (C25), manufacture of rubber and plastic products (C29), manufacture of non-metallic mineral products (C30), smelting and pressing of ferrous metals (C31), smelting and pressing of nonferrous metals (C32), and manufacture of metal products (C33). High-end technology industry includes the manufacture of raw chemical materials and chemical products (C26), manufacture of medicines (C27), manufacture of chemical fibers (C28), manufacture of general purpose machinery (C34), manufacture of special purpose machinery (C35), manufacture of automobiles (C36), manufacture of railway, ship, aerospace, and other transport equipment (C37), manufacture of electrical machinery and apparatus (C38), manufacture of computer, communication, and other electronic equipment (C39), and manufacture of measuring instrument and machinery (C40). Because of the various statistical calibers in the statistical yearbooks of different years, the utilization of waste resources (C42) and metal products, machinery and equipment repair (C43) are incomprehensive during the sample study period. Hence, these two industries are excluded.

4. Results and Discussions

4.1. Efficiency Measurement and Analysis of Green Innovation in the Manufacturing Industry. By considering the Super-SBM model with the unexpected output and max data envelopment analysis software, the green innovation efficiency values of 29 industries, three major industry groups, and national manufacturing industries in China from 2010 to 2019 are calculated as shown in Table 2. For a clearer analysis of the development difference evolution in the green innovation efficiency about the three types of industry groups under different technical requirements, the time-series change diagram of the efficiency is drawn from the standpoint of the country and three major industry groups as shown in Figure 1. In addition, according to the efficiency classification basis of references [19], the efficiency value of green innovation is divided into high efficiency (≥0.9), medium efficiency (0.5–0.9), and low efficiency (≤0.5).

From an industry perspective, the average value of green innovation efficiency shows that over 75% of the Chinese manufacturing industries have a green innovation efficiency value of less than 1 in a DEA invalid state. In addition, the green innovation efficiency values of 29 industries are significantly different from each other. Only 7 industries, including C36, C38, C39, C40, C16, C21, and C24, owned over 1 green innovation efficiency value with a forefront efficiency throughout the study. The green innovation of 10 industries, C13, C17, C19, C22, C25, C32, C28, C34, C35, and C37, has reached an effective level in individual years without high efficiency all the time because of the influence of unstable factors in other years. Other invalid industries during the sample study period witnessed the five industries at the bottom: C31, C15, C30, C14, and C26; the green innovation efficiency values fluctuate between 0.2 and 0.35, with nearly 65%–80% room for improvement. It identified the redundancy in the innovation investment of these five industries, insufficient innovation output capacity, highly unexpected output, serious environmental pollution, and low green development level. The state should strengthen the innovation

Table 2: The calculation results of green innovation efficiency of 29 manufacturing industries from 2010 to 2019.

Industry codes	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean	Rank
C13	0.269	0.253	0.350	0.306	0.324	1.312	0.322	0.347	0.377	0.373	0.423	18
C14	0.281	0.189	0.270	0.286	0.263	0.266	0.279	0.280	0.284	0.296	0.269	26
C15	0.215	0.158	0.231	0.225	0.193	0.198	0.189	0.241	0.314	0.334	0.230	28
C16	1.124	1.185	1.198	1.256	1.229	1.218	1.198	1.348	1.287	1.327	1.237	2
C17	0.450	0.403	0.525	1.028	1.010	1.063	1.032	1.039	0.413	0.381	0.734	11
C18	1.003	0.560	1.023	1.014	1.042	1.040	1.035	1.033	1.035	1.048	0.983	8
C19	0.527	1.018	1.016	1.003	0.530	0.637	0.688	0.567	1.097	1.141	0.822	9
C20	0.393	0.274	0.350	0.313	0.349	0.338	0.286	0.290	0.402	0.457	0.345	23
C21	1.172	1.195	1.141	1.128	1.008	1.103	1.153	1.151	1.145	1.110	1.131	5
C22	0.235	0.188	0.239	0.329	0.242	0.440	0.369	0.427	1.023	0.524	0.402	19
C23	0.274	0.249	0.380	0.413	0.418	0.415	0.448	0.546	0.545	0.588	0.428	17
C24	1.152	1.117	1.140	1.118	1.128	1.121	1.067	1.048	1.044	1.029	1.097	6
C41	0.400	0.312	0.239	0.277	0.300	0.352	0.399	0.414	0.500	0.470	0.366	21
C25	0.135	0.112	0.165	0.235	0.192	1.096	0.216	0.214	1.025	1.029	0.442	16
C29	0.336	0.273	0.385	0.370	0.334	0.352	0.412	0.409	0.446	0.459	0.378	20
C30	0.216	0.166	0.239	0.272	0.266	0.292	0.245	0.299	0.339	0.353	0.269	27
C31	0.168	0.125	0.167	0.201	0.170	0.193	0.231	0.285	0.309	0.312	0.216	29
C32	0.237	0.230	0.237	0.353	0.335	0.374	0.414	1.003	1.023	1.006	0.521	15
C33	0.303	0.344	0.368	0.328	0.333	0.359	0.399	0.379	0.411	0.435	0.366	22
C26	0.180	0.178	0.229	0.291	0.225	0.314	0.289	0.356	0.432	0.360	0.285	25
C27	0.321	0.279	0.302	0.284	0.290	0.293	0.347	0.332	0.360	0.371	0.318	24
C28	0.264	0.312	0.415	1.006	1.000	0.443	0.427	1.000	1.039	1.076	0.698	12
C34	0.384	0.529	0.473	0.531	0.463	0.499	0.709	0.623	0.754	1.023	0.599	14
C35	0.423	0.499	0.561	0.608	0.532	0.600	1.094	1.034	1.069	1.029	0.745	10
C36	1.105	1.115	1.076	1.046	1.051	1.052	1.083	1.079	1.070	1.069	1.075	7
C37	0.458	0.502	0.466	0.462	0.472	0.564	0.797	1.006	0.661	1.023	0.641	13
C38	1.212	1.060	1.180	1.095	1.139	1.250	1.159	1.082	1.108	1.090	1.137	4
C39	1.299	1.236	1.234	1.139	1.140	1.139	1.188	1.091	1.091	1.085	1.164	3
C40	1.009	1.309	1.168	1.259	1.225	1.135	1.187	1.549	1.529	1.634	1.300	1
Low-end technology industry	0.577	0.546	0.623	0.669	0.618	0.731	0.651	0.672	0.728	0.698	0.651	(2)
Mid-end technology industry	0.233	0.208	0.260	0.293	0.272	0.444	0.319	0.432	0.592	0.599	0.365	(3)
High-end technology industry	0.665	0.702	0.711	0.772	0.754	0.729	0.828	0.915	0.911	0.976	0.796	(1)
All industries	0.536	0.530	0.578	0.627	0.593	0.671	0.644	0.706	0.763	0.774	0.642	—

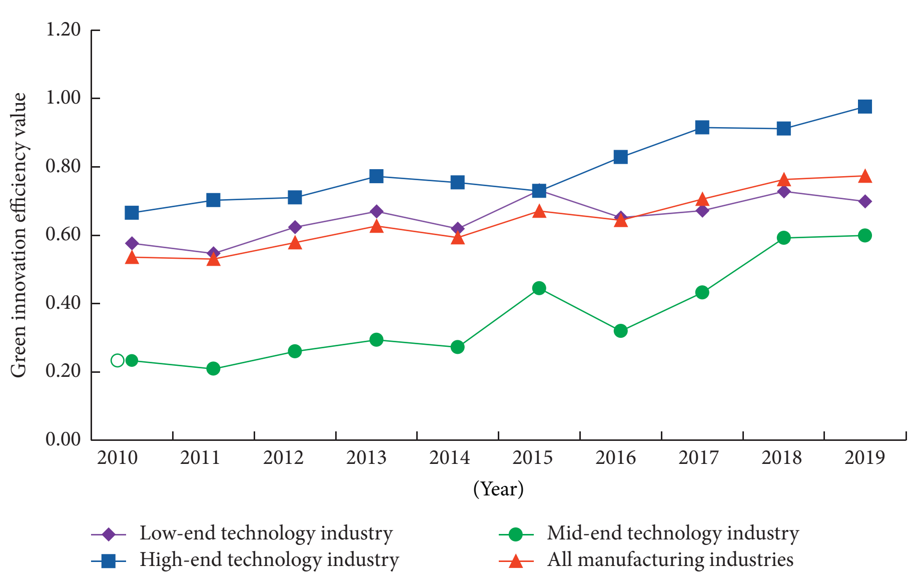


Figure 1: The time-series diagram of green innovation efficiency of the whole country and three major industry groups.

output in these industries and promote the development of green industries through technological innovation and energy conservation and emission reduction.

In view of different technology industry groups, during the sample study period, the mean value of the manufacturing green innovation efficiency in the three industry groups witnessed a high-to-low order: high-end technology industry group (0.796) > low-end technology industry group (0.651) > middle technology industry group (0.356). All of them with a DEA inefficiency possess different respective improvement spaces. The middle-end technology industry group with the lowest green innovation efficiency mainly covers the industries with high pollution and energy consumption, such as petroleum, coal, rubber, and metal processing. To improve the green innovation efficiency of the middle-end technology industry group, the most important thing is to control the pollution emissions of various industries from the perspective of reducing the unexpected output. In addition, Figure 1 unveils the consistency between the green innovation efficiency value of low-end technology industry groups and the national manufacturing industry changes, with the former witnessing a slow fluctuation upward trend and a stable medium efficiency level. The value of green innovation efficiency of the mid-end technology industry group increased from 0.233 in 2010 to 0.599 in 2019. Although it is far from the production frontier, the green innovation efficiency of the mid-end industry with a breaking-neck progress is overtaking the other two. Although the green innovation efficiency of the high-end technology industry group dropped slightly in 2014 and 2015, its value with the overall transformation jumped from 0.665 in 2010 to 0.976 in 2019, shifting from medium to high efficiency. The green innovation represents the core of the high-end technology industry that demands remarkable technological innovations and advanced technologies, better industry advancement environment, and the green development of the whole life cycle concerning products, production, sales, and transportation.

From the national level, the average green innovation efficiency of the manufacturing industry from 2010 to 2019, 0.642, is in the middle green innovation efficiency and does not reach the DEA effective level, indicating 35.8% improvement space. The change trend chart in Table 1 and Figure 1 unveils the swelling of the national green innovation efficiency of the manufacturing industry from 0.536 to 0.774 during the sample period with an obvious fluctuation upward trend and a certain gap with the optimal DEA efficiency. It shows that in the past decade, the policy on green innovation and manufacturing industry development has yielded desirable results with much attention paid to this area. Putting innovation at the core of the overall development of the manufacturing industry provides a favorable environment for its innovation and development and constantly pushes it from high-speed to high-quality development. As China steadily advances green innovation in the manufacturing industry according to its annual average growth rate during the sample study period, the manufacturing industry is expected to achieve a DEA effectiveness in 2026.

4.2. Industry Difference and Decomposition of Green Innovation Efficiency in the Manufacturing Industry. To further explain the development difference of green innovation efficiency in the manufacturing industry and reveal the overall difference and its source in the industry, the Dagum Gini coefficient method is used to measure the industry gap and subgroup decomposition division of green innovation efficiency in the Chinese manufacturing industry and three technology industry groups from 2010 to 2019 as shown in Figure 2.

The overall difference and evolution trend of green innovation efficiency in the Chinese manufacturing industry from 2010 to 2019 are displayed in Figure 2(a). Its overall gap fluctuated and declined with the general Gini coefficient G falling from 0.373 in 2010 to 0.264 in 2019. The prevalent contracted disparities between the manufacturing industries indicate that the manufacturing enterprises emphasized green innovation, the main driving force for manufacturing transformation and development nationwide. China has enjoyed certain fruits in promoting the green transformation and high-quality development of the manufacturing sector.

Figure 2(b) reveals the intraindustry differences and evolution trend of the green innovation efficiency in the manufacturing industry with the higher intragroup difference of a low-end technology industry group than that of the other two counterparts. The average Gini coefficient order of various industry groups: low-end technology industry group (0.319) > high-end technology industry group (0.256) > mid-end technology industry group (0.208). According to the evolution trend of each industry group, the low-end technology industry group showed a trend of slight fluctuation and decline with the biggest difference in 2011. The largest discrepancy in the low-end technology industry can be understood by the balanced development of green innovation in the high-efficiency industry and the massive environmental pollution in the lower one with more daunting polarization left in the low-end technology industry. The large change range in the intraindustry gap of the mid-end technology industry group reveals an inconspicuous consistent change trend. Before 2014, the smallest difference of green innovation efficiency in the mid-end technology industry group fluctuated greatly between 2014 and 2017, and later, it crept up, with the biggest difference in 2015. However, given the low efficiency and unbalanced differences in the mid-end technology industry group, it is necessary to move toward high-efficiency and balanced development through coordinated technological innovation and green advancement. The biggest difference in the high-end technology industry group in 2010 was followed by a continuous decline trend with a large drop unveiling slight discrepancies within the high-end group. As all industries emphasize green innovation progress, the favorable environment of mutual assistance and promotion within the industry group was established, pushing the high-end technology industry group to strengthen innovation and control pollution simultaneously.

Figure 2(c) shows the interindustry differences and evolution trend of green innovation efficiency in the manufacturing industry. The evolution trend unveiled a consistent decreasing fluctuation in the differences among the industry groups with various fluctuation amplitudes. The largest difference recorded between the mid-end technology industry group and the high-end counterpart saw a drop

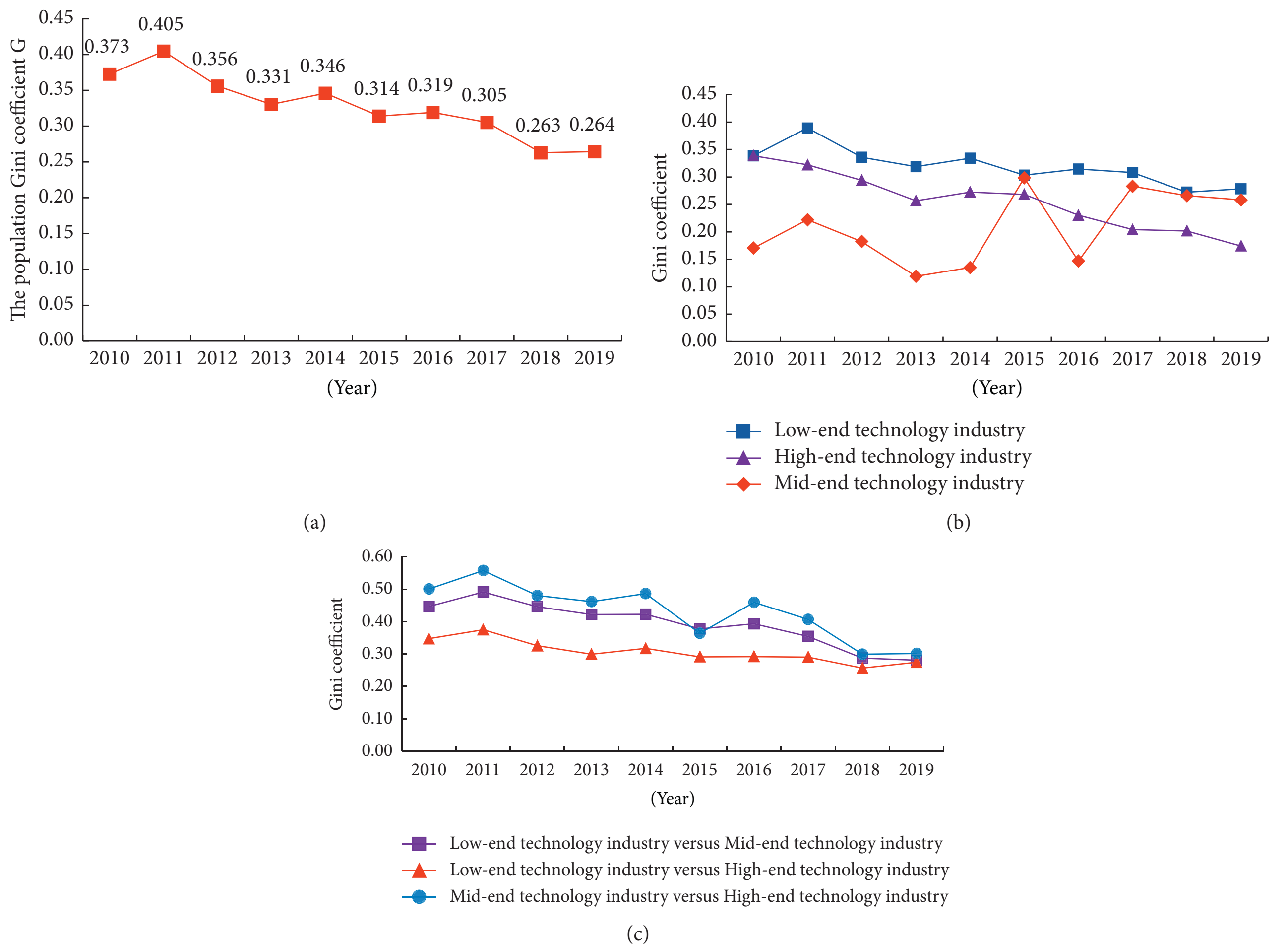


Figure 2: The time-series diagram of green innovation efficiency of the whole country and three major industry groups. (a) Overall differences in green innovation efficiency in manufacturing industries. (b) Intraindustry differences in green innovation efficiency of manufacturing industry. (c) Interindustry differences in green innovation efficiency of manufacturing.

from 0.501 in 2010 to 0.302 in 2019, the fastest decline among the three groups. The smallest difference between the low-end and high-end groups witnessed a small change range and a slow decline rate. The same between the low-end and mid-range technology industry groups gradually narrowed over time. In 2019, the intergroup difference of the industry groups was approximately 0.3, with a progressively same level of difference, specifically. The differences between the industry groups knitted with the development environment of each one. The high-end technology industry groups excel at innovation and development, with great leading technological advantages and industrial integration. In comparison, the middle and low-end groups are uncompetitive in resources such as research and development and environmental pollution treatment, with subsequent large differences among groups incurred.

The source decomposition and contribution results of industrial differences in the green innovation efficiency of the Chinese manufacturing industry are shown in Table 3. The contribution rate of different sources indicates the discrepancies between the groups mainly caused by the overall green innovation efficiency inconsistency in the Chinese manufacturing industry with an average contribution rate of 38.528%. The intragroup differences are the second source of overall inconsistencies with an average contribution rate of 33.159%. The lowest contribution rate of the supervariable density registered a moderate rate of only 28.313%. From the evolution trend, the overall contribution rate of the intragroup differences showed a steady change trend with the contribution rate ranging from 32.01% to 34.857%. The contribution rate of intergroup difference, the largest before 2014 with a small fluctuation range has experienced a W-shaped change, with a great fluctuation since 2015. The contribution rate of the supervariable density variance, the lowest before 2014, with a small fluctuation spectrum, showed an M-shaped change trend with a fluctuation range increased from 2015 to 2019. It should be further explained that the overall difference of the green innovation efficiency in the manufacturing industry changed from the supervariable density difference in 2015 and 2018, reflecting the contribution rate of the cross-overlap of different industry groups to the overall difference.

4.3. Convergence Analysis of Green Innovation Efficiency of Chinese Manufacturing Industry. For a more accurate

Table 3: Source decomposition and contribution results of the industrial differences in the green innovation efficiency of manufacturing industry.

Years	Total G	Contribution			Contribution rate (%)		
		G_w	G_{nb}	G_t	G_w	G_{nb}	G_t
2010	0.373	0.126	0.143	0.104	33.867	38.291	27.842
2011	0.405	0.135	0.171	0.099	33.379	42.252	24.369
2012	0.356	0.119	0.137	0.099	33.490	38.561	27.949
2013	0.331	0.108	0.136	0.087	32.766	41.010	26.224
2014	0.346	0.114	0.147	0.085	32.886	42.620	24.495
2015	0.314	0.109	0.070	0.134	34.857	22.412	42.731
2016	0.319	0.102	0.147	0.070	32.028	45.935	22.037
2017	0.305	0.098	0.134	0.074	32.010	43.818	24.172
2018	0.263	0.090	0.083	0.090	34.091	31.736	34.173
2019	0.264	0.085	0.102	0.077	32.218	38.649	29.133

Note. G_w is the intragroup difference. G_{nb} is the intergroup difference. G_t is the supervariable density difference, satisfying $G = G_w + G_{nb} + G_t$.

Table 4: σ value of green innovation efficiency of manufacturing industry in China and three major industry groups.

Years	National manufacturing industry	Low-end technology industry group	Mid-end technology industry group	High-end technology industry group
2010	0.727	0.666	0.331	0.655
2011	0.776	0.767	0.435	0.614
2012	0.677	0.650	0.368	0.567
2013	0.621	0.618	0.231	0.484
2014	0.650	0.639	0.277	0.519
2015	0.579	0.567	0.734	0.511
2016	0.589	0.594	0.306	0.441
2017	0.562	0.583	0.669	0.408
2018	0.483	0.515	0.571	0.391
2019	0.494	0.530	0.549	0.379

examination of the evolution trend of green innovation efficiency in various industries, the paper focuses on the convergence mechanism analysis of the green innovation efficiency across multiple manufacturing industries, resting on the study of green innovation efficiency level and difference decomposition in the Chinese manufacturing industry.

4.3.1. σ Convergence Analysis of Green Innovation Efficiency in Manufacturing Industry. In this paper, the coefficient of variation method is used to analyze the σ convergence of the green innovation efficiency of the Chinese manufacturing industry during the observation period. The results shown in Table 4 unveil that except the middle-end technology industry group, the national manufacturing industry, the low-end group, and the high-end group underwent a downward trend of fluctuation. The results show the inconsistent convergence of σ in the sample period with the features of intermittent convergence with changes. The expanded differences in some years existed with the unchanged overall downward trend. As far as the mid-end technology industry group is concerned, the green innovation efficiency of various industries exhibited an inconsistent convergence trend with irregular and divergent variation trends in 2011, 2014, 2015, and 2017. The gradually decreasing σ value from 2017 to 2019 witnessed the signs of σ convergence with the higher coefficient of 0.549 in 2019 than that of 2010–2014.

4.3.2. Absolute β Convergence Analysis of Green Innovation Efficiency in Manufacturing Industry. The STATA software is used in this paper for data analysis with the regression results shown in Table 5. The green innovation efficiency of the whole country and three major industry groups in the sample period witnessed inconsistency between its absolute β convergence model and the original hypothesis during model estimation. Therefore, the fixed effects model tested the absolute β convergence of the green innovation efficiency in the manufacturing industry.

The regression results in Table 4 revealed the β coefficient of green innovation efficiency of the national manufacturing industry and the three major industry groups below 0, which have passed the 1% significance level test. It shows that under the similar external environment and influencing factors, the green innovation efficiency of the whole manufacturing industry and the three major industry groups in China have an absolute β convergence phenomenon with a narrowing industry gap. The result chimes with the difference decomposition of the Dagum Gini coefficient mentioned above. The convergence speed of the national manufacturing industry is 0.0668 with the convergence speeds of low, medium, and

Table 5: Absolute β convergence regression results of green innovation efficiency of manufacturing industry and industry groups in China.

Coefficients	National manufacturing industry	Low-end technology industry group	Mid-end technology industry group	High-end technology industry group
β	-0.4872^{***} (-8.25)	-0.6288^{***} (-6.73)	-0.4290^{***} (-3.21)	-0.3727^{***} (-4.43)
α	-0.2766^{***} (-6.34)	-0.3715^{***} (-5.77)	$-.4238^{**}$ (-2.47)	-0.0936^{**} (-2.35)
Time effect	YES	YES	YES	YES
Individual effect	YES	YES	YES	YES
Sample numbers	261	117	54	90
R-squared	0.2275	0.3057	0.1795	0.1987
F-value	68.04^{***}	45.35^{***}	10.28^{***}	19.59^{***}

Note. β is the coefficient of observation. ***, **, and * mean significant at the level of 1%, 5% and 10% respectively. Figures in parentheses are T values. α is the constant term.

high-end technology industry groups to be 0.0991, 0.0560, and 0.0466, respectively. It means there is a slower convergence speed in the high-end group with a higher green innovation efficiency followed by the middle-end one and with the low-end counterpart enjoying the fastest convergence speed. It indicates that the faster growth of the industries with a low green innovation efficiency in the manufacturing industry than that of the industries with a high efficiency has gained a certain catch up momentum, with the green innovation efficiency of different manufacturing industries converging to the same steady-state level over time.

5. Conclusions and Policy Recommendations

Based on the panel data of 29 industries in the national manufacturing industry from 2010 to 2019, this paper analyzes the green innovation efficiency, industry development differences, and the convergence mechanism of manufacturing industries with the super-SBM model, Dagum Gini coefficient decomposition method, and convergence model. The main research conclusions are as follows:

Firstly, the manufacturing green innovation efficiency nationwide increases incessantly with an obvious efficiency difference between 29 industries, three-quarters of the industries in low efficiency, and a large green innovation efficiency improvement space.

Secondly, the green innovation efficiency of different technology industry groups with an annual growth and significant edges in the high-end technology industry group basically shifted from medium to high efficiency. The mid-end and low-end technology industry groups still focus on the green innovation development of the Chinese manufacturing industry.

Thirdly, the σ and β convergences from the time dimension reveal that the green innovation efficiency of the national manufacturing industry and the low-end and high-end technology industry groups all show an intermittent σ convergence trend with the minimum and divergent convergence characteristics of the middle-end ones. For an absolute β convergence, there is a significant trend of an absolute β convergence in the green innovation efficiency of the national manufacturing industry and the three industry groups with diversified convergence rates.

Given the above conclusions with the actual green innovation efficiency of the manufacturing industries in China, to transform the manufacturing industry, enhance its green innovation capability, and achieve its high-quality development, this paper puts forward the following policy recommendations:

The government should formulate appropriate green innovation policies to enhance the green innovation capability of its manufacturing industry that is set apart from that of other sectors with differentiated development strategies for related industries. Relevant departments need to transform the government functions to give full play to the guiding role of the government in promoting green technology innovation of enterprises. For the low-end technology industry group, the strengthened informatization investment should be accompanied by improved product technology, energy efficiency, and environmental protection through policy support. For the middle one, appropriate policies should be adopted to adjust traditional production and operation modes, by means of advanced technology to improve the production process of enterprises and their green management implementation and form a green manufacturing system. For the high-end counterpart, government financial subsidies will accelerate the research and development of innovative technologies and comprehensively promote its high-quality development. The high-end will lead the national manufacturing industry to make breakthroughs in innovation and green development.

All industries should work together to improve the green innovation environment that boosts industrial transformation. To improve its overall efficiency, the national manufacturing industry should focus on optimizing the environment for green innovation, and promoting all industries to uphold the green innovation development, with increasing awareness of in all sectors. The state should increase funding and support for green innovation with the wide application of green technologies. All industries should build an exchange platform for sharing green technologies and exchanging innovative talents in the manufacturing industry, promote open innovation, scientific and technological cooperation to ensure the effective transformation of green innovation achievements in various sectors. Moreover, the manufacturing industry transformation should be encouraged from policies, talents, technology, and environment with foreign advanced green innovation development models.

However, this research has some limitations. In the analysis of β convergence characteristics, this paper mainly studies the industry convergence trend of green innovation efficiency of manufacturing industry with absolute β convergence, without considering the external environment of the development of various industries. In future studies, the influence of various external factors on the industry green innovation efficiency will be considered, and further analysis through conditional β convergence model to grasp the industry convergence characteristics of the green innovation efficiency of the manufacturing industry in China will be made more comprehensively.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that this manuscript has not been published elsewhere or under consideration in any journal. The authors declare no conflicts of interest.

Acknowledgments

This research was funded by the Tianjin Science and Technology Program (18ZLZDZF00190); Tianjin Science and Technology Program (18ZLZXZF00390); and Humanities and Social Sciences Planning Fund Project of the Ministry of Education (20YJA630085).

References

[1] B. L. Jin and Y. Han, "Influencing factors and decoupling analysis of carbon emissions in China's manufacturing industry," *Environmental Science and Pollution Research*, vol. 1, pp. 1–20, 2021.

[2] Y. Fu, A. Supriyadi, T. Wang, L. Wang, and G. T. Cirella, "Effects of regional innovation capability on the green technology efficiency of China's manufacturing industry: evidence from listed companies," *Energies*, vol. 13, no. 20, Article ID 5467, 2020.

[3] J.-D. Wurlod and J. Noailly, "The impact of green innovation on energy intensity: an empirical analysis for 14 industrial sectors in OECD countries," *Energy Economics*, vol. 71, pp. 47–61, 2018.

[4] M. Saunila, J. Ukko, and T. Rantala, "Sustainability as a driver of green innovation investment and exploitation," *Journal of Cleaner Production*, vol. 179, pp. 631–641, 2018.

[5] Y. Wang and Y. Yang, "Analyzing the green innovation practices based on sustainability performance indicators: a Chinese manufacturing industry case," *Environmental Science and Pollution Research*, vol. 28, no. 1, pp. 1181–1203, 2021.

[6] C. C. Lv, C. H. Shao, and Y. Cui, "Spatio-temporal evolution and influencing factors of Green innovation efficiency in China," *Research on Financial and Economic Issues*, vol. 12, pp. 50–57, 2020.

[7] Y. Luo and P. Chen, "On the threshold effect of environmental regulation upon the improvement of China's industrial green innovation efficiency," *Journal of Northeastern University*, vol. 2, pp. 147–154, 2018.

[8] N. Shen and J. J. Zhou, "A study on China's green innovation efficiency evaluation and functional mechanism based on hybrid DEA and SEM model," *Journal of Industrial Engineering and Engineering Management*, vol. 4, pp. 46–53, 2018.

[9] Y. W. Lv, Y. X. Xie, and X. J. Lou, "Study on the space-time transition and convergence trend of China's regional green innovation efficiency," *The Journal of Quantitative & Technical Economics*, vol. 5, pp. 78–97, 2020.

[10] M. G. Liu, "Research on spatial distribution and convergence of green innovation efficiency in regional innovation system," *Journal of Industrial Technological Economics*, vol. 4, pp. 10–18, 2017.

[11] C. Yu, L. Wenxin, S. U. Khan et al., "Regional differential decomposition and convergence of rural green development efficiency: evidence from China," *Environmental Science and Pollution Research*, vol. 27, no. 18, pp. 22364–22379, 2020.

[12] C. Wu, S. W. Yang, P. C. Tang, T. Wu, and S. K. Fu, "Construction of the efficiency promotion model of green innovation in China's heavy polluted industries," *China population, resources and environment*, vol. 28, pp. 40–48, 2018.

[13] Z. Fang, H. Bai, and Y. Bilan, "Evaluation research of green innovation efficiency in China's heavy polluting industries," *Sustainability*, vol. 12, no. 1, p. 146, 2019.

[14] D. Li and T. Zeng, "Are China's intensive pollution industries greening? An analysis based on green innovation efficiency," *Journal of Cleaner Production*, vol. 259, Article ID 120901, 2020.

[15] W. Wang, B. Yu, X. Yan, X. Yao, and Y. Liu, "Estimation of innovation's green performance: a range-adjusted measure approach to assess the unified efficiency of China's manufacturing industry," *Journal of Cleaner Production*, vol. 149, pp. 919–924, 2017.

[16] Z. Feng, B. Zeng, and Q. Ming, "Environmental regulation, two-way foreign direct investment, and green innovation efficiency in China's manufacturing industry," *International Journal of Environmental Research and Public Health*, vol. 15, no. 10, pp. 2–22, 2018.

[17] P. B. Wan, M. Yang, and L. Chen, "How do environmental technology standards affect the green transition of China's manufacturing industry—a perspective from technological transformation," *China Industrial Economics*, vol. 9, pp. 118–136, 2021.

[18] W. H. Huang and M. D. Wang, "Measurement of green technology innovation efficiency of manufacturing industry in Yangtze River Economic Belt," *Statistics and Decisions*, vol. 19, pp. 61–63, 2021.

[19] Q. Deng, S. Zhou, and F. Peng, "Measuring green innovation efficiency for China's high-tech manufacturing industry: a network dea approach," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–13, 2020.

[20] Q. Luo, C. Miao, L. Sun, X. Meng, and M. Duan, "Efficiency evaluation of green technology innovation of China's strategic emerging industries: an empirical analysis based on malmquist-data envelopment analysis index," *Journal of Cleaner Production*, vol. 238, Article ID 117782, 2019.

[21] J. Li, N. N. Li, and Q. M. Yuan, "Study on the spatial-temporal differences and influencing factors of green innovation efficiency in high-tech industry," *Forum on Science and Technology in China*, vol. 4, pp. 92–101, 2020.

[22] C. Claudio, B. Cristina, and G. Teresa, "You can't manage right what you can't measure well: technological innovation efficiency," *Research Policy*, vol. 42, pp. 1239–1250, 2013.

[23] M. Yi, Y. Wang, M. Yan, L. Fu, and Y. Zhang, "Government R&D subsidies, environmental regulations, and their effect on green innovation efficiency of manufacturing industry: evidence from the Yangtze River economic Belt of China," *International Journal of Environmental Research and Public Health*, vol. 17, no. 4, Article ID 1330, 2020.
[24] N. Nuryakin and T. Maryati, "Green product competitiveness and green product success. why and how does mediating affect green innovation performance?" *Entrepreneurship and Sustainability Issues*, vol. 7, no. 4, pp. 3061–3077, 2020.
[25] M. Frondel, J. Horbach, and K. Rennings, "What triggers environmental management and innovation? Empirical evidence for Germany," *Ecological Economics*, vol. 66, no. 1, pp. 153–160, 2008.
[26] C.-L. Miao, M.-M. Duan, Y. Zuo, and X.-Y. Wu, "Spatial heterogeneity and evolution trend of regional green innovation efficiency--an empirical study based on panel data of industrial enterprises in China's provinces," *Energy Policy*, vol. 156, pp. 112370–112378, 2021.
[27] L. Li, L. Lei, and D. Han, "Regional green innovation efficiency in high-end manufacturing," *Journal of Coastal Research*, vol. 82, pp. 280–287, 2018.
[28] S. Yin, N. Zhang, and B. Li, "Enhancing the competitiveness of multi-agent cooperation for green manufacturing in China: an empirical study of the measure of green technology innovation capabilities and their influencing factors," *Sustainable Production and Consumption*, vol. 23, pp. 63–76, 2020.
[29] K. Tone, "A slacks-based measure of super-efficiency in data envelopment analysis," *European Journal of Operational Research*, vol. 143, no. 1, pp. 32–41, 2002.
[30] F. Gökgöz and E. Erkul, "Investigating the energy efficiencies of European countries with super efficiency model and super SBM approaches," *Energy Efficiency*, vol. 12, pp. 601–618, 2018.
[31] L. Dai, X. Mu, C.-C. Lee, and W. Liu, "The impact of outward foreign direct investment on green innovation: the threshold effect of environmental regulation," *Environmental Science and Pollution Research*, vol. 28, no. 26, pp. 34868–34884, 2021.
[32] C. Dagum, "A new approach to the decomposition of the gini income inequality ratio," *Empirical Economics*, vol. 22, pp. 515–531, 1998.
[33] Z. Chen, X. Li, and X. Xia, "Measurement and spatial convergence analysis of China's agricultural green development index," *Environmental Science And Pollution Research*, vol. 28, no. 16, pp. 19694–19709, 2021.
[34] H. Zhang, Z. Geng, R. Yin, and W. Zhang, "Regional differences and convergence tendency of green development competitiveness in China," *Journal of Cleaner Production*, vol. 254, Article ID 119922, 2020.
[35] S. Xu, T. Wu, and Y. Zhang, "The spatial-temporal variation and convergence of green innovation efficiency in the Yangtze River Economic Belt in China," *Environmental Science and Pollution Research*, vol. 27, no. 21, pp. 26868–26881, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 3176504, 12 pages
https://doi.org/10.1155/2021/3176504

Research Article

Research on the Prewarning Model of Relationship Risk Levels in Industry Collaborative Innovation Alliances across Provinces in China

Liufang Yu and Caiyun Chen

School of Management, Wuhan Polytechnic University, Wuhan 430031, China

Correspondence should be addressed to Liufang Yu; 12474@whpu.edu.cn

Received 20 August 2021; Accepted 30 September 2021; Published 21 October 2021

Academic Editor: Shianghau Wu

The governments need beforehand to perceive the innovative relationship risk because they are one of the innovation subjects in those industry collaborative innovation alliances. However, it is difficult for innovation subjects to quantify the risks for industry collaborative innovation alliances due to the complexity, nonlinear, and dynamic condition. This paper firstly constructs an ordered logistic model, uses the following as independent variables: the collaborative degree, the ratio of science technology expenditure to GDP, the ratio of education expenditure to GDP, the ratio of finances to GDP, and uses the levels of risk as the dependent variable. Then, this paper uses the panel data of 30 provinces in China (Hainan is not included) from 2010 to 2018 to fit the model. Based on the fitting results, the research has gained the relationship risk prewarning model in industry collaborative innovation alliances by using the collaborative degree as an independent variable. The governments at all levels can use this relationship risk prewarning model to percept risk levels and reckon the corresponding probability which exists in industry collaborative innovation alliances. Furthermore, there are regional influences existing in the prewarning relationship risk levels in industry collaborative alliances. The east and middle areas have significant regional influence, but it does not exist among west areas and others. The governments at all levels may consider the regional differences.

1. Introduction

The risk evaluation at the microlevel needs the professional knowledge and practical experience accumulation of core experts in the industry, and the basic data acquired is a real subjective cognition of innovation subject at the microlevel on the level of relationship risk to adapt to the decision-making at the microlevel. However, the internal participants of the industrial collaborative innovation alliance have been stratified, and the mesosubjects hope to have a direct prewarning of the level of relationship risk within the industry. In particular, under the major background of "the development of an innovation-driven economy", innovation has been a crucial driving force of the development of the local economy. Governments at all levels positively participate in collaborative innovation alliances and turn to be the important subjects of the alliances to lead other subjects to take part in the collaborative innovation alliance and connect one collaborative innovation alliance after another within the region. As the important subject of the collaborative innovation alliance and the node subject in the collaborative innovation network, the government should focus on the relationships between subjects in the alliance to have a prewarning of relationship risk to lead the new direction of collaborative innovation in the region in the future. The industries within regions and the provinces and cities across the country could predict the self-related relationship risk level of collaborative innovation alliances [1]. Therefore, it is necessary to have a prewarning of risk level and the probability of occurrence of each risk level in their respective regions with mesodata to help the practice community clearly realize and judge the relationship risk of the complicated organization and collaborative innovation alliance and make a scientific judgment and decision.

By far, there is less knowledge on the relationship risk of collaborative innovation alliance, and there has been no existing method for reference for the understanding of the field and the effective prevention of risk. The common risk prewarning models are as follows: univariate model, ZETA model, logistic model, probit model, neural network model, and entropy model (including system entropy, relative entropy, management entropy, and even risk entropy). These models have their own strengths and weakness. The entropy model is widely used in the study of alliance risk prewarning, but it still needs to acquire data with a subjective judgment which reduces its scientificity of conclusions. The logistics model does not need to satisfy the statistical assumptions such as normal distribution and homogeneity of variance as a linear model. There should be approximate treatment in the process of calculation, but the risk level prewarning does not need to be too precise, so it has been applied to much more fields [2]. Based on the comparison of the risk prewarning models, the study tries to simulate the panel data in recent 5 years with logistic regression to explore the influencing factors of the relationship risk level of industrial collaborative innovation alliance to form a risk prewarning model of industrial collaborative innovation alliance based on mesodata, so as to assist governments at all level to have a prewarning of the risk level of collaborative innovation alliance and the probability of occurrence of each level in the region.

2. Construction of the Relationship Risk Prewarning Model for Industrial Collaborative Innovation Alliances

2.1. Building Up the Ordered Logistic Model. The binary logistic regression could be built as a risk prewarming model for risk prewarming because the dependent variable values are taken as 0 and 1 to show the two situations, having risk and having no risk, so that it could be applied to the field of risk management. The relationship risk of industrial collaborative innovation alliance is at five levels: 1 (very low), 2 (low), 3 (general), 4 (high), and 5 (very high). The binary logistic regression is not suitable for the risk prewarning of industrial collaborative innovation alliance because the dependent variable values are only taken as 0 and 1; however, in the ordered logistic model, there are multiple observed values of dependent variables with sorting results, so the ordered logistic regression model could be built as a model for relationship prewarning of the industrial collaborative innovation alliance. The general expression of ordered logistics is as follows:

$$y^* = X\beta + \varepsilon, \varepsilon \mid X \sim \mathrm{Logit}(0, 1). \tag{1}$$

y^* is the latent variable of dependent variable Y, and X means the vector of an independent variable x_i. β is an estimated parameter vector and ε is the random error term. When $w_i, i = 1, 2, 3, 4, 5$, is set as critical value (threshold) and the value y depends on the comparison result between y^* and critical value, the expression of value y is

$$y = \begin{cases} 1. y^* \leq \omega_1 \\ 2. \omega_1 < y^* \leq \omega_2 \\ 3. \omega_2 < y^* \leq \omega_3 . \\ 4. \omega_3 < y^* \leq \omega_4 \\ 5. \omega_4 < y^* \leq \omega_5 \end{cases} \tag{2}$$

In expression (2), $y = 1, 2, 3, 4, 5$ means the very low relationship risk of collaborative innovation alliance (harmonious partnership among members), the low relationship risk of collaborative innovation alliance (having difficulty in the cooperation and communication among members but it is apt to be overcome), general relationship risk of collaborative innovation alliance (general cooperation and communication among members), high relationship risk of collaborative innovation alliance (needing a long-term communication and negotiation), and very high relationship risk of collaborative innovation alliance (bad partnership among members and being on the verge of disintegration). According to the conditional probability knowledge in Mathematics, the corresponding equations of y to X are as follows:

$$\begin{cases} P(y = 1 \mid X) = P(y^* \leq \omega_1 \mid X) = \varphi(\omega_1 - X\beta) \\ P(y = i \mid X) = P(\omega_{i-1} < y^* \leq \omega_i \mid X) = \varphi(\omega i - X\beta) - \varphi(\omega_{i-1} - X\beta) \quad i = 2, 3, 4. \\ P(y = 5 \mid X) = P(y^* > \omega_5 \mid X) = 1 - \varphi(\omega_5 - X\beta) \end{cases} \tag{3}$$

The distribution function is the logistic one.

2.2. The Independent Variables in the Ordered Logistic Model

2.2.1. Collaborative Degree. Collaborative innovation alliance is a strategic behavior to achieve certain objectives with a plan, so it is a "social collaboration." It needs the alliance subject to achieve the organization from being disorder to order, from lowly order degree to the high order degree through the application of their advantages. The process of alliance subject to use their advantages is a collaborative one, so the degree of collaborative highly affects the efficiency of collaborative and the entire relationship between the subjects. Generally speaking, the subjects would have a high cognition and higher trust with each other when the degree of collaborative is higher, so there would be less opportunistic behavior of the subjects. The higher collaborative degree means the stronger resource integration ability of the entire innovation alliance. There are much more resources to be combined and a higher collaborative effect, so the total benefits of collaborative innovation that can be shared by all innovation subjects should be large to relatively reduce the economic interest conflict between innovation subjects. In system theory, a collaborative degree means the degree of collaborative and consistency of subsystems or system elements in the development process of the system, which describes the collaborative degree among subsystems or system elements in the system. In the collaboration, the ordered parameter would be transformed from one phase-transformation state to another state by describing the subsystems or system elements in the evolution process of the system, and it could be used to represent the order structure and type of system. The collaborative degree of the ordered variable "set" could display the overall collaborative degree of the evolving new structure. Therefore, the collaborative degree could be used to evaluate the degree of collaborative from both perspectives of system theory or the synergetic. The evaluation of collaborative degree could be deemed as a measuring instrument for synergetic innovation of industry-university-research (IUR). The measurement of the collaborative ability in the synergetic innovation system of IUR in a certain period could reflect the degree of the collaborative of synergetic innovation of IUR [3]. The collaborative degree is used for measuring the degree of collaborative among the alliance subjects while the degree of collaborative would affect the relationship between subjects. Therefore, the collaborative degree could be deemed as one of the variables of risk level prewarming. In this research, the collaborative degree is an independent variable as shown in Table 1.

2.2.2. Other Independent Variables. In the collaborative innovation alliance, the government is a leading subject since it affects the entire innovation environment and provides new factors for innovation alliance through innovation policy. The government could effectively gather resources to coordinate all innovation subjects through leading to effectively control the relationship between the subjects of collaborative innovation alliance. Generally speaking, governments could achieve their role in innovation alliance through the buildup of innovation environment, such as financial capital investment, innovative talent policy, and financial policy for local innovation. When the innovation factors of a collaborative innovation alliance are sufficient and there are much more resources with high quality to be collaborated by innovation subject, the effect of mutual collaborative would be better and the innovation subjects would get along well with each other because the effect of mutual collaborative would be higher than that of independent innovation. Hence, the influencing factors of the relationship risk level of collaborative innovation alliance are as follows: science and technology expenditure ratio, education expenditure ratio, and financial loan balance ratio. The science and technology expenditure ratio is calculated by the ratio of science and technology expenditure to the GDP of each region in the current year, and the factor affects the supply of innovation capital. The education expenditure ratio is calculated by the ratio of the absolute number of local face-to-face education expenditures to the GDP of each region in that year, and the factor affects the supply of innovation talents. The financial loan balance ratio is calculated by the ratio of the balance of bank loans in various regions to the GDP of various regions in the current year, and the factor would affect the supply of innovation funds and social support for innovation activities. In this research, the ratio of science and technology to GDP, the ratio of education to GDP, and the ratio of finance to GDP are other independent variables as shown in Table 1.

Table 1: Variables in the relationship risk prewarning model.

Variables	Sign	Unit	Type
Collaborative degree	X1	Decimal	—
Ratio of science and technology to GDP	X2	Decimal	—
Ratio of education to GDP	X3	Decimal	—
Ratio of finance to GDP	X4	Decimal	—

2.2.3. Standardized for Independent Variables. As seen from the entire industrial collaborative innovation alliance, the interaction among colleges and universities, scientific research institutions, and agencies and the direct interaction between government and these subjects would be shown in the collaborative degree of independent variables, which would be seen from the index selection in subsequent collaborative measurement while the interaction between the innovation subject, government, and other subjects would be shown through other independent variables. There are different value dimensions of all influencing factors, so the variables above should be standardized. The standardization method in the research is adopted with the range method. The first part in equation (4) is the positive index and the second part of equation (4) is the negative one.

$$S_j\left(x_{ji}\right) = \begin{cases} \dfrac{x_{ji} - \beta_{ji}}{\alpha_{ji} - \beta_{ji}}, & i \in \left[1, l_1\right] \\ \\ \dfrac{\alpha_{ji} - x_{ji}}{\alpha_{ji} - \beta_{ji}}, & i \in \left[l_1, n\right] \end{cases}. \tag{4}$$

The source of data is the *Statistical Yearbook of Chinese Science and Technology, the Statistical Yearbook of Chinese High-Tech Industry, the Compilation of Scientific and Technological Statistical Data of Colleges and Universities, the Statistical Yearbook of Chinese Torch*, and the statistical yearbooks of relevant provinces from 2010 to 2018. However, in the process of model verification, the data of collaborative degree of some independent variables and the data of relationship risk level need to be acquired through a certain approach. Here is an introduction to the acquisition process of the fitting data of the two variables.

2.3. Acquisition of Collaborative Degree Data

2.3.1. Evaluation Method of the Collaborative Degree. As for the evaluation tools of the interaction degree for the subjects of collaborative innovation, many scholars have mentioned the collaborative degree many times. The collaborative degree is one of the effective tools to measure the cross-organizational collaborative innovation effect and it could be used to represent the degree of collaborative and consistency of various innovation elements in the compound system [3]. In the study of the collaborative innovation mechanism, the collaborative degree of collaborative innovation system means the degree of consistency of interaction between collaborative subjects in the process of cooperation and the degree of a behavioral collaborative of subjects in the system; the evaluation of collaborative degree could be deemed as a measuring instrument for synergetic innovation [4–6]. With reference to the outcomes of subsequent research on the expansion of the compound system in different studies, this paper builds up a collaborative degree model suitable for the compound system of the collaborative innovation alliance. It is supposed that the compound system of the collaborative innovation alliance is S, and the subsystem of the collaborative innovation alliance is $S_j\,(j = 1, 2, 3, 4)$. S_1 is a subsystem of technology intermediary service; S_2 is a subsystem of colleges and universities; S_3 is a subsystem of the scientific research institution; S_4 is a subsystem of industry. The ordered variable is needed in the entire collaborative process of innovation alliance to describe that x_{ji} basis could be divided into two types for the impact of dependent variables: the positive influencing factor and the negative one. When x_{ji} is the positive influencing factor, the larger its value is taken, the higher the order degree of the system would be. When x_{ji} is the negative influencing factor, the larger its value is taken, the lower the order degree of the system would be. The subvariable of the ordered variable is $S_j(x_{ji}) = \begin{cases} (x_{ji}-\beta_{ji})/(\alpha_{ji}-\beta_{ji}), & i \in [1, l_1] \\ (\alpha_{ji}-x_{ji})/(\alpha_{ji}-\beta_{ji}), & i \in [l_1, n] \end{cases}$. The influencing factors built in the compound system of collaborative innovation alliance are positive and negative. With the consideration that the subvariable of the ordered variable would be positive after the processing to not affect the following processing, it should be $S_j(x_{ji}) = (x_{ji}-\beta_{ji})/(\alpha_{ji}-\beta_{ji}) \times 0.9 + 0.1$, $S_j(x_{ji}) \in [0.1, 1]$. The measurement of the order degree of the ordered parameter of a general subsystem could be used with the geometric average method and the linear weighted average method. The subjectivity could not be overcome when the weight is confirmed with a linear weighted average method, so the geometric average method is used to measure the order degree of subsystem integration fitting subsystem: $s_j(x_j) = \sqrt[4]{\prod_{i=1}^{4} s_j(x_{ji})}$. According to the evolution of the compound system from disorder to order, it is set with an initial moment as t_0. If the time setting t_0 of data acquisition was set to be 2009, the order degree of t_0 in all subsystems would be $d_j^0(x_j)$, $j = 1, 2, 3, 4$, the next time of evolution process of the compound system is t_1, and the order degree of time t_1 is $d_j^1(x_j)$, $j = 1, 2, 3, 4$, so t_1 is set to be 2010. Similarly, in the next round calculation, t_1 would be 2011 when t_0 is taken as 2010, so as to conclude the order degree of the compound system from 2010 to 2018. The collaborative degree of collaborative innovation alliance is $S(X) = \theta \sqrt[4]{|\prod_{j=1}^{4}[d_j^1(x_j) - d_j^0(x_j)|}$. The parameter θ is to tune the negative and the positive.

$\theta = \min[d_j^1(x_j) - d_j^0(x_j) \neq 0/|\min[d_j^1(x_j) - d_j^0(x_j) \neq 0_j|$, $j = 1, 2, 3, 4$. Ifjthe value of $S(X)$ was bigger, it would mean the optimal collaborative degree of the compound system of collaborative innovation alliance. As the order degree of each subsystem fluctuates differently and exchanges information, materials, and energy with each other, the overall collaborative degree can be positive or negative.

2.3.2. Selection of the Subsystems. One collaborative innovation alliance is a complicated system with diversified subjects and protruding heterogeneity, but the majority of the industry agree that the collaborative innovation alliance is a network innovation organization with collaborative and interaction between diversified subjects, including the core subjects of colleges and universities, incorporations, and research institutions and the auxiliary subjects of governments, financial organizations, intermediary organizations, and innovation platforms. However, the study thinks that collaborative innovation should be a self-organizing system that all innovation subjects keep cooperating with each other and all innovation factors are recycling ceaselessly, and it would attract the exit or entry of all innovation subjects for the open characteristics of the system. The subject of collaborative innovation alliance would not be constant forever, so it would be impossible to focus on the entire microsubject when describing the collaborative process of the entire system. In the practice, it should be described with the subsystem according to the major classification. For example, some researchers divide the IUR technology allocation into three subsystems: subsystem of industry, a subsystem of colleges and universities, and subsystem of research and development [7, 8]. However, along with the profound carryout of collaborative innovation, all innovation subjects have refined and professional distribution in collaborative innovation alliances, so the collaborative innovation system is attracting the participation of various innovation subjects with open characteristics. The subsystem of technology service shows its talent as a crucial bridge to connect all innovation subjects and makes innovation alliance focus on it gradually. Scholars start to have a study on the subsystem of technology serving as a newly born subsystem, and they find that the subsystem has a finer lower-level subsystem composition, such as subsystem of talents and a subsystem of venues [9, 10].

Based on it, the study chooses the following subsystems as the ones for the compound system of collaborative innovation alliance: a subsystem of industry, a subsystem of technology intermediary service, a subsystem of colleges and universities, and a subsystem of the scientific research institution. The study does not deem government as a subsystem since it is the dominant leader of collaborative innovation. In China, the government would interfere agency, such as financial institutions, by affecting the subjects of collaborative innovation alliance with the factor of innovation capital; it would also affect the subsystem of the industry with tax policy and intervene the factor of innovation talents resources with education policy. The subsystem of government is based on a mixed system, so it would be hard to clarify the boundary with others or analyze the interaction among subsystems if it was deemed as a subsystem. However, the subsystem of government does impose impact on the relationship risk level of collaborative innovation alliance, so the indirect interaction between government and other innovation subjects would be considered in other independent variables when building up the model for relationship risk prewarning of collaborative innovation alliance while the direct relationship between government and others subsystems would be considered in the index to measure the collaborative degree of subsystems [11, 12]. For the selection of indexes of all subsystems on the measurement of collaborative degree, it is shown as follows:

(a) The subsystem of technology service: collaborative innovation alliance has been developed to be a complicated subsystem with the core of knowledge increment and value creation. In the subsystem, it is included with technology trading market, productivity promotion center, and business incubator in the core layer; the technical consultation, scientific and technological novelty search, scientific and technological development, Information Research Institute, and property right exchange in the middle layer; and the technology novelty search, talent market, leasing company, and audit and accounting service organization in the peripheral layer. It has been a necessary bridge for knowledge increment and value creation.

(b) The subsystem of colleges and universities: as an important source to create and spread new knowledge and new technology, colleges, universities, and scientific research institutions could greatly push incorporations to carry out innovation activity. In general, applied colleges and universities could send all kinds of innovation talents for all innovation subjects in the industrial innovation alliance; on the other hand, they could have the cooperation of technical application with the subjects in the alliance with the button of human resources. Knowledge-based colleges and universities with higher innovation levels would work with scientific research institutions to have a breakthrough knowledge innovation for serving the national strategy and social development. All innovation subjects in the industrial alliance would coordinate with colleges and universities to gain the resources they lack to solve their insufficient innovation capability. It is the innovation factor of a collaborative innovation alliance: the supplier of talents, technology, and knowledge.

(c) The subsystem of scientific research institutions: also as an important source to create and spread new knowledge and new technology, scientific research institutions could also great push incorporations to carry out innovation activities. However, the role of scientific research institutions in the collaborative innovation alliance is different from that of the subsystem of colleges and universities. The role of colleges and universities is to provide vast innovation personnel for the subsystem of industry. The innovative knowledge and innovative technology that these personnel are equipped with would provide important innovation factors for the subsystem of industrial innovation; meanwhile, it would achieve the exchange of the materials, ability, and information among the subsystems to provide a crucial motivation for the evolution of the entire compound system. However, besides the scientific research and the creation of innovative knowledge, scientific research institutions would also participate in the formulation of relevant innovation policies or laws, provision of strategic planning, and so on with an identity of expert [6]. It would impose an impact on the evolution of the compound subsystem of the collaborative innovation alliance by affecting the peripheral environment of the alliance. It is the innovation factor of collaborative innovation alliance: the supplier of knowledge.

(d) The subsystem of the industry: economic profit-driving is an important motivator for a collaborative innovation alliance. The alliance is dominated by industry and oriented by market demand, but there are still so many restricting factors between the ability of industry and the demand of the market, so subsystem of the industry could not satisfy the market demand as an independent system and it needs the collaborative with other subsystems to acquire the resources and ability to satisfy the market demand to achieve the economic profit and undertake social responsibility [13, 14].

2.3.3. Selection of Index of Four Subsystems. The index system built in the previous evaluation process is based on the input-output index. Du Biyun et al. measured the collaborative degree of measures the collaborative degree of the IUR technology alliance innovation system in the six provinces of the middle region with a compound system of collaborative degree model, and the ordered parameter selected by them is still the index of the input-output index when confirming the ordered parameter of a subsystem of scientific technology alliance [8]. The compound system of collaborative innovation alliance is social collaborative, while social collaborative has a purpose. The behavior of the collaborative subject is directly controlled by the objectives of the subject, so the evaluation on the collaborative degree should try to begin from the subject behavior. Therefore, it is necessary to choose an ordered parameter of the subsystem in the index of collaborative behavior among the subjects. Some researches begin from a complex system theory and dissipative structure theory to suggest a measurement method of regional collaborative innovation based on collaborative degree-management entropy when studying the measurement of collaborative innovation ability in the region [9]. There are two index systems selected by them: order degree of innovation subject and knowledge transfer degree. The index of order degree of innovation subject is selected with a large number of indicators of interaction between innovation subjects. Through the empirical comparison of collaborative degree and management entropy, it is found that the result of the two models is basically consistent so as to prove the scientificity and effectiveness of measurement. It also provides an effective reference for the index selection to the quantitative measurement of collaborative innovation alliance. Based on these, the study selects the following indexes for all subsystems of collaborative innovation alliance when confirming the empirical data of collaborative degree. Above all, the whole index of collaborative innovation alliances is shown in Table 2, where (a) stands for the subsystem of technology service; (b) stands for the subsystem of colleges and universities; (c) stands for the subsystem of scientific research institutions; and (d) stands for the subsystem of industry.

The source of data is the *Statistical Yearbook of Chinese Science and Technology, the Statistical Yearbook of Chinese High-Tech Industry,* the *Compilation of Scientific and Technological Statistical Data of Colleges and Universities*, the *Statistical Yearbook of Chinese Torch*, and the statistical yearbooks of relevant provinces from 2010 to 2018. Given the inconsistent dimensions of the original data, there would be errors when directly participating in the calculation and processing, so the paper uses the level difference method to standardize the original data.

2.3.4. Measurement of the Collaborative Degree. There are four subsystems in the compound system of collaborative innovation alliance: the subsystem of technology service, the subsystem of colleges and universities, the subsystem of scientific research institutions, and the subsystem of industry. By describing the mutual roles of the four subsystems, the order degree of the four subsystems in 30 provinces (cities) around the country and then the collaborative degree of the compound system are "integrated" through the order degree of the four subsystems, as shown in Table 3, so as to get the data of independent variable collaborative degree for the model of relationship risk prewarning of the collaborative innovation alliances.

3. Acquisition of Relationship Risk Level

3.1. Selecting Original Data Indexes to Confirm Relationship Risk Level. The theoretical study of collaborative innovation alliance could be traced to the IUR cooperation. It is suggested by Etzkowit and Leydesdroff. They emphasize that knowledge could be an increasing factor of the economy, and they focus on the cognition of innovation subjects. Colleges and universities, industries, and governments are mutually independent and interactive to form a dynamic triple helix to push the sustainable growth of the economy. Later, Leydesdroff thought that the uncertainty, complication, and completion of the system could be presented by the mutual information among three subsystems based on the cognition of information entropy, so as to suggest the index to measure the relationship of the triple helix [10]. The "triple helix" means innovation subjects. Then, everybody studies the cooperation and interaction relationship between subjects with the mutual information of "triple helix". To explore the relationship between the IUR collaborative and innovation subject, domestic scholars start to integrate the "triple helix" and data mining technology to measure the relationship between innovation subjects. Cai Xiang and Liu Xiaozheng studied the cooperation relationship of government-university-research with the SCI scientific papers, national science and technology standards, and national scientific research fund as the data of the output structure of "triple helix" [15, 16]. Zhuang Tao made the patent data as the original data of "triple helix" output to study the international cooperation of IUR, and he extended the subject of "triple helix" to be four subjects of international cooperation to measure the partnership among subjects to study the interaction between the government-university-research and the international cooperation organization [16]. Hence, based on the previous literature, the paper refers to the study and extension outcomes of above on "triple helix" with the consideration on the availability of data to choose invention patent as basic data. The paper carries out the study among the collaborative innovation subjects through the algorithm of "triple helix" and information theory knowledge to acquire the original data of relationship risk level during the empirical process. There are three types of patents: invention patent, utility model patent, and appearance design. The reason why to choose invention patent as the original data of algorithm of "triple helix" is that invention patents mean the originality with the highest technical content, so it is more suitable for the study on innovation than that of design patent and utility model patent. Hence, in the collection of basic data, the study only adopts the data of invention patents. The invention patent could be divided into job invention patent and nonjob invention patent. The owners of job invention patents are incorporations, scientific research institutions, colleges and universities, and governments. These subjects are closer to that of the collaborative innovation alliance in the study. In the past few years, in the effective invention patents in China, the proportion of nonjob invention patents is decreasing while that of the job one has been increasing. It has increased from 70.1% in 2006 to be 90.0% in 2014, which is increased by nearly 20% in eight years while the foreign countries always keep the high position of 98% for the past five years with an increasing tendency. It ensures sufficient data. With the consideration of the factors mentioned above, the paper chooses the number of service invention patent applications granted as basic data for the calculation of "triple helix".

3.2. Acquiring the Original Data to Confirm Relationship Risk Level. It confirms to calculate the interaction of innovation alliance with job invention to finalize the basic data of relationship risk level in empirical analysis. The data of job invention is acquired by the website of China National

Table 2: List of the index of four subsystems of the compound system of collaborative innovation alliances.

Subsystem	Index	Symbol	Unit	Type
(a) Subsystem of technology service	Total number of incorporations served	X11	Tens of thousands of incorporations	+
	Total income of service	X12	Million yuan	+
	Increased sales for the incorporation	X13	Million yuan	+
	Input of government	X14	Million yuan	+
(b) Subsystem of colleges and universities	Full-time equivalent of R and D personnel in colleges and universities	X21	Tens of thousands of persons per year	+
	Internal expenditure of R and D funds in colleges and universities	X22	Million yuan	+
	Government funds for the internal expenditure of R and D funds of colleges and universities	X23	Million yuan	+
	Corporate funds internal expenditure of R and D funds of colleges and universities	X24	Million yuan	+
	External expenditure of R and D funds of colleges and universities	X25	Million yuan	+
	Expenditure of domestic colleges and universities for the external expenditure of R and D funds of colleges and universities	X26	Million yuan	+
	Expenditure of domestic incorporations for the external expenditure of R and D funds of colleges and universities	X27	Million yuan	+
	Expenditure of domestic scientific research institutions for the external expenditure of R and D funds of colleges and universities	X28	Million yuan	+
	Number of applications for R and D projects in colleges and universities	X29	Item	+
(c) Subsystem of scientific research institutions	Full-time equivalent of R and D personnel in scientific research institutions	X31	Tens of thousands of persons per year	+
	Internal expenditure of R and D funds in scientific research institutions	X32	Million yuan	+
	Government funds for the internal expenditure of R and D funds of scientific research institutions	X33	Million yuan	+
	Corporate funds internal expenditure of R and D funds of scientific research institutions	X34	Million yuan	+
	External expenditure of R and D funds of scientific research institutions	X35	Million yuan	+
	Expenditure of domestic colleges and universities for the external expenditure of R and D funds of scientific research institutions	X36	Million yuan	+
	Expenditure of domestic incorporations for the external expenditure of R and D funds of scientific research institutions	X37	Million yuan	+
	Expenditure of domestic scientific research institutions for external expenditure of R and D funds of scientific research institutions	X38	Million yuan	+
	Number of applications for R and D projects in scientific research institutions	X39	Item	+
(d) Subsystem of industry	Full-time equivalent of R and D personnel in large and medium-sized industrial incorporations	X41	Tens of thousands of persons per year	+
	Internal expenditure of R and D funds in large and medium-sized industrial incorporations	X42	Million yuan	+
	Government funds for the internal expenditure of R and D funds of large and medium-sized industrial incorporations	X43	Million yuan	+
	Corporate funds internal expenditure of R and D funds of large and medium-sized industrial incorporations	X44	Million yuan	+
	External expenditure of R and D funds of large and medium-sized industrial incorporations	X45	Million yuan	+
	Expenditure of research institutions for the external expenditure of R and D funds of large and medium-sized industrial incorporations	X46	Million yuan	+
	Expenditure of domestic colleges and universities for the external expenditure of R and D funds of large and medium-sized industrial incorporations	X47	Million yuan	+
	Number of projects for R and D projects in large and medium-sized industrial incorporations	X48	Item	+
	Expenditure for new product development	X49	Million yuan	+
	Proportion of incorporations with R and D activities	X50	%	+

Table 3: Relational risk level in industry collaborative innovation alliances from 2010 to 2018.

Area/year	2010	2011	2012	2013	2014	2015	2016	2017	2018
Beijing	1	2	2	2	2	1	1	1	1
Tianjin	4	3	4	3	4	2	2	2	2
Hebei	4	3	4	3	4	2	2	2	2
Shanxi	4	3	4	3	4	1	3	3	3
Inner Mongolia	2	5	5	5	5	1	1	1	1
Liaoning	1	2	2	2	2	1	1	3	3
Jilin	1	2	2	2	1	3	4	3	3
Heilongjiang	1	2	2	2	1	4	3	3	3
Shanghai	3	4	3	4	5	1	2	2	2
Jiangsu	5	1	1	1	3	2	2	2	2
Zhejiang	3	4	3	4	4	5	5	4	4
Anhui	3	1	1	1	3	2	2	2	4
Fujian	5	1	1	1	3	2	5	4	4
Jiangxi	4	3	5	5	5	4	3	2	2
Shandong	3	4	3	4	5	2	2	2	1
Henan	3	4	3	4	5	2	1	1	1
Hubei	4	3	4	5	4	1	1	1	1
Hunan	3	3	4	5	4	1	1	1	1
Guangdong	5	1	1	1	3	5	5	5	4
Guangxi	3	3	3	5	5	4	4	3	5
Chongqing	3	4	3	4	5	2	2	2	1
Sichuan	3	4	3	4	4	2	2	1	1
Guizhou	4	3	4	3	4	1	1	1	3
Yunnan	1	2	2	2	2	1	1	1	3
Tibet	2	4	3	4	5	5	1	2	1
Shanxi	1	2	2	2	1	4	3	3	3
Gansu	1	2	2	2	2	3	4	3	5
Qinghai	2	4	5	4	5	2	2	2	1
Ningxia	5	1	1	1	3	2	5	4	2
Xinjiang	4	3	4	3	4	1	3	3	3

Intellectual Property Administration and the websites of Intellectual Property Administration of all provinces. Through the tools of patent searching and analysis in the websites, I set the searching keyword as follows: "patent applicant + application date + code of province" according to the existing searching method. The patent applicants are represented with "Government (G)," "Incorporation (I)," "University (U)," and "Research Institution (R)" (repeated measurement is allowed). According to the appellation of state-owned incorporations and institutions, it would be categorized into the range of "government" if the "ministry," "bureau," "department," or the name of the organization directly under the government were included in the name of the applicant; it would be classified into the range of "incorporation" if the "company," "plant," "incorporation," and "group" were contained in the name; the name of institutions included with "university" and "college" would be classified into the range of "university" and those with the "academy," "lab," and "institute" would be classified into the range of "research institution." The discrimination of "university" and "research institution" is for the methods of the existing IUR studies because there is an obvious distinction in the functions of middle schools and universities and research institutions in the collaborative innovation alliance. The applied university is mainly for the innovation of talent cultivation while the knowledge-based university is to engage in basic innovation activity for the breakthrough innovation to serve for a national strategy. The applied scientific institution is mainly engaged in technical application, but it would be combined with the demand of governments in all places to be the "think tank" of government to advise the government on policies as an expert while the research institution engaging in basic major innovation activity mainly serves for the national strategy. Hence, there would be a distinction between "university" and "research institution" in the text. Besides, there would be respective reports on technical innovation in the statistical yearbook and annual report. As for the cooperation of innovation subjects, if there were two or three names, the name of the patent would be classified to be eight categories as follows: "University–Incorporation (UI)," "University–Government (UG)," "Incorporation–Government (IG)," and "Research Institution–Incorporation (RI)," "Research Institution–University (RU)," "Research Institution–Government (RG)," "University–Incorporation–Government (UIG)," "University–Research Institution–Incorporation–Government (URIG)." Meanwhile, if it was searched by year according to the application date, the province will be given by number. For example, the number of Guangdong is 44, and the frequency of occurrence is counted in each category so that the data after the statistics would be composed of the original database of cooperative patent application research [17]. The original data acquired by the forms is applied for the triple helix model to gain the marginal probability of each innovation subject. Then, the data of the correspondent triple helix system could be gained by calculating mutual information. The calculation of mutual information of diversified subjects is in need of knowledge of information theory. The collaborative innovation alliance is involved with diversified subjects, and there is a cross-relationship between multiple subjects. The innovation subjects have a collaborative innovation and apply for invention patent for the innovation outcome together so that the mutual information could reflect the cooperation among innovation subjects specifically. According to the information theory, Leydesdorff thinks the data provided by the network would produce a relevant frequency distribution. The relevant frequency distribution could generate one probability distribution: $P_i = f_i/\sum_i f_i$. The average information volume of the probability distribution is defined by scholars to be an information entropy: $E_i = -\sum_i P_i \log_2 (p_i)$.

Under the one-dimensional situation, the information entropy means the product of distribution probability and its negative logarithm. Under the multidimensional situation, we have a measurement by adding substitution. In the study, there are 4 subsystems of the collaborative innovation system and 4 job invention subjects, so the calculation of the information entropy of invention application of subjects in collaborative innovation alliance is adopted with four dimensions. According to the information theory, the calculation formula of E under multi-dimensions is as follows:

$$E(X_1, X_2, \ldots, X_n) = \sum_{\substack{x_1 \in X_1 \\ x_2 \in X_2 \\ \cdots \\ x_n \in X_n}} p(x_1, x_2, \ldots, x_n) \log_2 \frac{p(x_1, x_2 \cdots x_n)}{\prod_{i=1}^{n} p(x_i)}. \tag{5}$$

Referring to the calculation formula in the mutual information of information theory, the calculation formula of the mutual information of innovation subject in the collaborative innovation alliance is as follows:

$$\begin{aligned} T_{urig} = {} & E_u + E_r + E_i + E_g - E_{ur} - E_{ui} - E_{ug} - E_{ri} - E_{rg} \\ & - E_{ig} + E_{urg} + E_{uig} + E_{rig} + E_{uri} - E_{ruig}. \end{aligned} \tag{6}$$

Among these, the evidence of the two-dimensional mutual information is

$$T_{ij} = E_i + E_j - E_{ij}, \quad i, j \in \{u, r, i, g\}. \tag{7}$$

The evidence of the three-dimensional mutual information is

$$\begin{aligned} T_{ijk} = {} & E_i + E_j + E_k - E_{ij} - E_{ik} - E_{jk} + E_{ijk}, \\ & i, j, k \in \{u, r, i, g\}. \end{aligned} \tag{8}$$

Besides, in the calculation of information entropy, the value is directly taken as zero when $p = 0$is encountered, so the value is directly taken as zero when there is 0 in the study. The entire value calculation process is carried out according to different provinces so as to gain the measurement sheets of cooperation relationship of triple helix subject listed according to province (city) (there are a total of 30 provinces and cities, excluding Hainan Province).

According to the cluster analysis method, the *K*-mean cluster analysis in the SPSS software is used to classify the relationship of collaborative subjects into five levels, and the five levels correspond to the five levels of relationship risk of collaborative innovation alliance. The correspondent value would be taken as 1, 2, 3, 4, and 5. The *K*-mean cluster analysis of SPSS19.0 is applied to gain the panel data of the relationship risk level shown in Table 3.

4. Verification of the Model for Relationship Risk Prewarning

With the use of the STATA metrological analysis software, the variable Y means the relationship risk level of collaborative innovation alliance, and the variables X1, X2, X3, and X4 are the collaborative degree, science and technology expenditure ratio, education expenditure ratio, and financial loan balance ratio. The control variables are region and degree. Furthermore, the models in Table 4 are added control variables to the mode through "region", "year", X2, X3, and X4 step by step. The results are shown in Table 4.

Seen from the p value of independent variables, only the collaborative degree of the independent variable passes the significance test. The collaborative degree could pass the significance test among multiple influencing factors directly related to the relationship risk level of collaborative innovation alliance, so it means that the relationship risk level of collaborative innovation alliance could be predicted with the collaborative degree from the perspective of management and statistical metrology. But other variables which are science and technology expenditure ratio, education expenditure ratio, and financial loan balance ratio are not reasonable to prewarning relationship risk levels in those collaborative innovation alliances. Even the variables "region" and "year" are as a controlled variable put in model (5), model (6), and model (7), and the degree of X1 significance reduces on the contrary. Whether the variables "region" and "year" are controlled or not, the p values of X2, X3, and X4 are still not significant.

And then, seen from the model calculation and tests, the X1, independent variable of collaborative degree, is more reasonable when building up the model for the relationship risk level of collaborative innovation alliance, so STATA is used to fit the equation containing only the degree of the collaborative as an independent variable. The fitting result shows that the whole p value is 0.0001 and the p value of the independent variable, collaborative degree, is 0.068, so both of them pass the significance test. From model (2) in Table 4, the test adds a control variable: region; the results are still significant. It shows that the relationship risk prewarning contains region influence. According to the *Statistical Yearbook of Chinese Science and Technology*, the east area is Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan; the middle area is Shanxi, Anhui, Jiangxi, Henan, Hubei, and Hunan; the west area is Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shanxi, Gansu, Qinghai, Ningxia,

Table 4: Regression results.

Variables	(1) y	(2) y	(3) y	(4) y	(5) y	(6) y	(7) y
$x1$	0.9122** (2.40)	0.7733* (1.67)	−1.6618** (−2.08)	−1.9144* (−1.92)	−1.8741* (−1.87)	−1.8642* (−1.86)	−1.8838* (−1.87)
$x2$					−0.2615 (−0.78)	0.5764 (0.54)	0.5028 (0.43)
$x3$						−0.4799 (−0.83)	−0.4583 (−0.77)
$x4$							0.1894 (0.16)
Region	No	Yes	No	Yes	Yes	Yes	Yes
Year	No	No	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.0068	0.1230	0.0383	0.1528	0.1535	0.1543	0.1544
Observations	270	270	270	270	270	270	270

z-statistics in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Table 5: Region influence.

Variables	(1) y	(2) y	(3) y	(4) y	(5) y	(6) y	(7) y
$x1$	−2.2626 (−0.83)	−4.1602** (−2.42)	−2.8947** (−2.15)	3.9414 (0.65)	−4.1187 (−1.54)	−3.7985** (−2.34)	−2.8947** (−2.15)
Region	West	East	Middle	Northeast	West	East	Middle
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.0547	0.0661	0.2271	0.3745	0.0350	0.0536	0.2271
Observations	117	76	54	23	131	85	54

z-statistics in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

and Xinjiang; the east-north area is Liaoning, Jilin, and Heilongjiang. According to Economic Research Journal, the east area is Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, Hainan, and Liaoning; the middle area is Shanxi, Anhui, Jiangxi, Henan, Hubei, and Hunan; the west area is Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shanxi, Gansu, Qinghai, Ningxia and Xinjiang, Jilin, and Heilongjiang [18, 19]. Comparatively, It shows that the region influence takes place in the east area and middle area, but it does not take place in the west area and northeast area in model (1), model (2), and model (3) from Table 5 according to the classification of the *Statistical Yearbook of Chinese Science and Technology*. It shows that the region's influence takes place in the east area and middle area, but it does not take place in the west area in model (5), model (6), and model (7) from Table 5 according to the classification of *Economic Research Journal*. That means the coordinative degree is more relative to the innovation risk level in the east and middle areas. This research is more suitable for east and middle governments to prewarning innovation levels.

If the time variable is controlled, the correspondent cumulative ratio of ordered results in the east area is as follows:

$$\begin{cases} y^* = -4.1602x \\ Y = \begin{cases} 1, y^* \leq -5.8412 \\ 2, -5.8412 < y^* \leq -4.3095 \\ 3, -4.3095 < y^* \leq -3.6397 \\ 4, -3.6397 < y^* \leq -2.4402 \\ 5, y^* > -2.4402 \end{cases} \end{cases} \quad (9)$$

Meanwhile, the odds ratio value in the ordered logistic model means that every increase in the collaborative degree by one unit will lead to an increase of corresponding times of the probability that the relationship risk level will decrease by one level. So, it could be seen from the odds ratio value of the model for relationship risk prewarning that the probability of reducing the risk level by one or more levels will increase by times when the collaborative degree changes by one unit. The probability of the occurrence of each risk level would be predicted from the cumulative ratio of ordered results. The odds ratio is 0.0156 if the time variable is controlled in the east area. So, the probability of reducing the risk level by one or more levels will increase by 0.9844 times. It should be noted that the significant meaning of coefficient in the ordered logistic model is bigger than the meaning of coefficient itself. The significance of the coefficient is relevant to the value of the independent variable and the value of β. Hence, the governments at all levels could predict the relationship risk level of industry collaborative innovation alliance in the region.

5. Conclusions

The paper studies the governments at all levels and how to prewarn the innovation risk levels as they are one of the innovation subjects in industry collaborative innovation alliances. This paper makes some contributions about this point which are as follows: (i) the science and technology expenditure ratio, the education expenditure ratio, and the financial loan balance ratio are not reasonable to prewarning relationship risk grades in those collaborative innovation alliances; (ii) the relationship risk prewarning contains region influence. Furthermore, region influence is different among different areas. It is fit for using the collaborative degree to prewarding the risk degree in the east area and middle area, but it is not fit for the west area or northeast area. Using the collaborative degree to predict the risk levels can be suitable for governments which are indicative innovation subjects in industry innovation alliances, but they need to consider the differences among provinces; (iii) the odds ratio value in ordered logistic regression means that every increase in the collaborative degree by one unit will lead to an increase of corresponding times of the probability that the relationship risk level will decrease by one level. The probability of the occurrence of each risk level would be predicted from the cumulative ratio of ordered results.

Certainly, if any industry in the cross-region could refine all subsystems within the industry, it could also refer to the model of the prewarning to predict the level of relationship risk of internal collaborative innovation in the industry. In the future, along with the dynamic change of collaborative innovation alliances, the number of innovative subjects would be increased or decreased. These changes add the complication and risk of a collaborative innovation alliance. However, only the new subsystem is confirmed through the complicated system of collaborative innovation alliance to predict the relationship risk level of prewarning and the probability of occurrence of each level through the collaborative degree.

Data Availability

The data used to support the findings of this research are included within the article. The source of data is from the Statistical Yearbook of Chinese Science and Technology (2010–2018), the Statistical Yearbook of Chinese High-Tech Industry (2010–2018), the Statistical Yearbook of Chinese Torch from 2010 to 2018, and the Chinese National Intellectual Property Reports from 2010 to 2018. The data of job invention are acquired by the website of China National Intellectual Property Administration and the websites of Intellectual Property Administration of all provinces.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Hubei Province Education Science and Technology Foundation (Grant no. D20201605) and Wuhan Polytechnic University Outstanding Young Teachers Scientific Research Project (Grant no. 2020J07).

References

[1] H. Yan, S. P. Qin, and J. Shi, "Dynamic evaluation regional industry-university-research collaboration," *Systems Engineering*, vol. 38, no. 5, pp. 66–74, 2020.
[2] H. Y. He, Z. W. Wang, and J. J. Cai, "A research on influencing factors of collaborative innovation of enterprises, universities and research institutes in China—an empirical analysis base on ordered logistic model," *East China Economic Management*, vol. 28, no. 9, pp. 106–110, 2014.
[3] X. Wu, Q. F. Wei, and X. Gu, "Research on the collaborative degree measurement in collaborative innovation," *Soft Science*, vol. 29, no. 7, pp. 45–50, 2015.
[4] S. Wei and S. Zhao, "The research of regional collaborative innovation mode in China: based on principal component analysis," *Canadian Social Science*, vol. 8, no. 3, pp. 145–150, 2012.
[5] H. Dong, J. Q. Zeng, and M. R. Shen, "Research on construction and collaborative degree of composite system of industrial innovation—base on ICT industry," *Studies in Science of Science*, vol. 34, no. 8, pp. 1166–1160, 2016.
[6] Y. J. Liu and J. D. Yin, "Comparison research on the level of collaboration innovation across provinces under the strategy of innovation-driven," *Science and Technology Progress and Policy*, vol. 34, no. 16, pp. 33–41, 2015.
[7] L. F. Yu, J. D. Yin, and S. P. Xu, "Heterogeneity characteristics of innovation subject and relationship risk in innovation alliances under supply-side reform," *Science and Technology Progress and Policy*, vol. 34, no. 5, pp. 6–13, 2017.
[8] B. Y. Du, C. H. Niu, and T. Niu, "Comparison research on system coordination degree among IUR alliances in Chinese six provinces," *Modern Management*, vol. 36, no. 2, pp. 175-176, 2015.
[9] F. M. Deng, J. G. Zhang, and X. D. Liang, "Measurement on the capacity of regional collaborative innovation based on coordination degree-management entropy," *Science and Technology Progress and Policy*, vol. 31, no. 5, pp. 113–117, 2014.
[10] L. Leydesdorff and G. Zawdie, "The triple helix perspective of innovation systems," *Technology Analysis and Strategic Management*, vol. 22, no. 7, pp. 789–804, 2010.
[11] K. Nakwa and G. Zawdie, "Structure holes, knowledge intermediaries and evolution of the triple helix system with references to the hard disk drive industry in Thailand," *International Journal of Technology Management and Sustainable Development*, vol. 14, no. 1, pp. 20–47, 2015.
[12] X. Cai and J. Zhao, "Research on the efficiency and influencing factors of university-industry-government synergistic innovation," *Soft Science*, vol. 33, no. 2, pp. 56–60, 2019.
[13] Y. D. Wang and L. Lin, "Measuring the relationship of industry-university-research collaborative innovation in

Central China based on mutual information," *Journal of Intelligence*, vol. 37, no. 12, pp. 195–201, 2018.

[14] N. Hewitt-Dundas, "The role of proximity in university-business cooperation for innovation," *The Journal of Technology Transfer*, vol. 38, no. 2, pp. 195–201, 2013.

[15] X. Cai and X. Z. Liu, "Relationship among university, industry and government based on the perspective of SCI," *China Science and Technology Form*, vol. 22, no. 7, pp. 789–804, 2012.

[16] T. Zhuang and H. Wu, "Based on the number of Patents, Research on how to measure the triple helix among Governments-Industry-University-Research in China—the function of Governments in IUR," *Management World*, vol. 8, no. 8, pp. 175-176, 2013.

[17] H. Y. Chen and Z. L. Feng, "The statistical analysis of the application and the authorization of the service invention in China," *Science and Technology, Industry*, vol. 16, no. 3, pp. 81–86, 2016.

[18] L. Y. Han and C. Y. Du, "Regional difference of the impact of urban households' consumer finance," *Journal of economics research*, vol. 1, pp. 30–42, 2011.

[19] X. Cai and H. J. Zhang, "Research on the spatial effect of industry-university synergetic innovation in China and its influencing mechanism," *Journal of Technical Economics and Management*, vol. 10, pp. 26–32, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 7314687, 24 pages
https://doi.org/10.1155/2021/7314687

Research Article

Vertical Channel Conflict Coordination Strategy of e-Commerce Supply Chain under Platform Brand Empowerment

Di Xiao,[1,2] Qianqian Yang,[1] Qi Sun ,[1,2] and Huimin Fang[1]

[1]*School of Business Administration, Zhejiang Gongshang University, Hangzhou 310018, China*
[2]*Research Center of Modern Business and Trade, Zhejiang Gongshang University, Hangzhou 310015, China*

Correspondence should be addressed to Qi Sun; 344497713@qq.com

Received 23 July 2021; Accepted 16 September 2021; Published 5 October 2021

Academic Editor: Weilin Xiao

We develop a game model for a supply chain consisting of one e-commerce platform, one supplier from other channels, and one retailer. The platform has a well-known brand that can influence consumers' purchase decisions, and it provides good-quality products with high prices, while supplier from other channels provides cheaper products but possibly with low quality, and there may even be some serious quality problems, sometimes leading to serious problems such as "free-riding" behavior by the retailer and reducing the profits of the supply chain members. First, we study the decisions of platform and retailer under centralized decision (CD) scenario, decentralized decision (DD) scenario, cost sharing contract (CS) scenario, and minimum order quantity contract (QC) scenario. Second, we found that channel conflicts have a negative impact on supply chain members under DD scenario; however, CS and QC scenarios can make the optimal empowerment level of platform the same as CD scenario and encourage retailer to order more products from platform. Finally, the improvement effect in QC and CS scenarios is affected by the substitutability of the two products, the coefficient of empowerment cost, and the reaction coefficient of product price on goodwill. Furthermore, we found that under QC scenario, only within an appropriate range can the platform and the retailer achieve a win-win situation.

1. Introduction

In recent years, with the continuous development of information technology, e-commerce platforms have developed rapidly. In the United States, Amazon has become the largest e-commerce company after decades of development and has transformed from an online bookstore to a comprehensive e-commerce platform. Especially, its sales revenue is 280.52 billion dollars in 2019, which is more than 8 times as much as that of 2010 (https://www.touzhibang.com/1132.html). Similarly, Alibaba is in a leading position in China's e-commerce market with 55.9% sales share in 2019 (https://www.iimedia.cn/c1061/71838.html). However, as the competition intensifies, the profit growth of online retailers begins to slow down. For example, though e-commerce sales in the United States reached $599 billion in 2019, the growth rate has slowed in recent years, and even declined in 2017 and 2018 (https://www.199it.com/archives/1148207.html). In order to compensate for the decline in profits from online channels, e-commerce platforms have tried to expand offline channels. The impact of e-commerce platforms and the competitive pressure of peers make offline retail stores face severe challenges, especially in some small cities, towns, and even villages, although there are abundant customer demands in the offline market, retail stores that are run by individuals lack brand influence and management, and they have no advantage in market competition. Therefore, they also hope to get the help of e-commerce platforms to obtain high-quality products and enhance the brand image. Then, e-commerce platforms begin to export brands and information technology to retail stores to expand offline channels, and the retail stores often accept the help of e-commerce platforms in order to increase their profits. We define this type of behavior made by e-commerce platform as "empowerment," which means the platform use big data analysis, brand effects, integrated services, and other means to improve other participants' capabilities in the enhancement of brand influence, data utilization and user perception, so as to improve the overall operational performance of e-commerce supply chains. For example, Alibaba cooperates with offline retailers so that they can use Ali's digital tools and solutions to understand consumers' purchase preferences and optimize the delivery of goods to attract new consumers, instead of just selling products on Ali's platforms.

In practice, there are many forms of empowerment, among which "brand empowerment" is an emerging way. The platform brand empowerment (PBE) mainly refers to the platform empowers the retail store with more well-known e-commerce platform brand, while providing some popular products available only on e-commerce platform's channels, which help retail stores enhance brand influence in order to attract more customers and win customers' trust. At the same time, the e-commerce platform can achieve online and offline integration, such as JD.com formulated the "JD Convenience Store" plan in 2017, followed by Ali and Suning also launched "Tmall Convenience Store" and "Suning Convenience Store" respectively, using their own advantages to empower offline retail stores with brand. Taking JD New Channel as an example, it has abundant commodities and powerful supply chain resources of JD.com and provides professional terminal services and data support for small- and medium-sized retail stores, so it launched JD Convenience Store. It can rely on its own resources to promote cooperation between JD Convenience Store and well-known companies, such as Unilever, to optimize the depth of distribution, brand promotion, and preferential promotion. In addition, it can provide JD Convenience Stores with operational planning based on consumption scenarios, which include product types, store layouts, and marketing plans with characteristics of JD.com, and regular improvement suggestions by analyzing relevant data of surrounding consumers. Through these measures, the retail stores improve their abilities to accurately analyze what kind of products are marketable and provide more personalized service to consumers so that consumers' experience and the efficiency of store are both improved. Furthermore, JD.com also expands its brand's influence and directs more customers' demands from offline to online.

However, PBE is still in the primary stage of development and e-commerce platform does not have enough power to restrict the behaviors of retailer. Due to the existence of opportunism, retailer may order some counterfeit products with cheaper wholesale price from other channels, which are similar to genuine products but of poor quality and sell them to consumers in the name of platform, and the "free-riding" behavior will hurt the interests of consumers; it may even endanger their health and safety, which will also greatly damage the platform's brand goodwill. As a result, there is a channel conflict that means although channel members depend on each other, they often dispute over what kind of benefits they want to get. Obviously, different opinions on goals, roles, and returns often lead to channel conflicts [1]. In practice, JD.com allows retailers to order less than 50% of products from other channels and requires them to pay a warranty of 5,000 to 20,000 yuan to ensure that the products are 100% free of counterfeits, while there are no other specific measures to prevent the inflow of counterfeits. Although the products provided by JD.com are of good quality, the wholesale price is higher, so some retailers order counterfeit products from other channels with lower wholesale prices for their own interests. Once consumers buy these counterfeits from convenience stores, they will think that there is a problem with the quality of JD.com's products, and its brand image will be seriously damaged. Therefore, our study attempts to address the following questions:

(1) How much negative impact will vertical channel conflicts bring to supply chain members?
(2) What are the optimal ordering decisions of retailer and optimal empowerment level of platform in some common scenarios?
(3) Can the supply chain incentive mechanism encourage the retailer to order more products from the platform and reduce vertical channel conflicts?

We construct a single-cycle and two-echelon supply chain consisting of one e-commerce platform, one retailer and one supplier from other channels, and try to solve the channel conflicts under PBE and improve the decisions of platform and retailer. First, we found that channel conflicts affect the decisions of supply chain members and decrease the profits, especially in decentralized decision (DD) scenario. Second, we studied the effect of the contracts on the supply chain with empowerment cost sharing contract (CS) and minimum order quantity contract (QC), where CS means e-commerce platform determines cost sharing ratio and level of PBE, and QC scenario refers to that e-commerce platform determines the level of PBE and designs the reward and punishment mechanism for the retailer's ordering decisions. We conclude that both contracts can induce retailer to accept cooperation, increase the quantity of products that retailer order from the platform, and have an equal role in increasing the level of PBE. Finally, through comparison, we find that CD scenario is always the best choice for the supply chain members. At the same time, QC scenario plays a greater role in reducing channel conflicts; however, it is necessary for the platform to design appropriate reward and punishment mechanism.

The remainder of the paper is organized as follows: Section 2 reviews the relevant research studies. Section 3 mainly introduces the model assumptions and the meanings of various symbols in the article. Section 4 discusses the impact of the empowerment cost sharing contract and minimum order quantity contract on supply chain decision-making and performance. Section 5 compares and analyzes the optimal decisions and profits. Section 6 uses numerical simulation to illustrate the conclusions of the proposed model. Section 7 concludes the article and points out the shortage and suggested the future research directions.

2. Literature Review

This article is closely related to platform empowerment, channel conflict, supply chain contract, and other fields. To emphasize our contribute in the research, we review some representative literature in this part.

2.1. Platform Empowerment. The concept of "empowerment" was first proposed by the field of positive psychology. Later, the researches extended to management, which initially focused on the individual and were considered to delegate decision-making power to subordinates [2]. Then, empowerment on the organizational level was gradually rising, which means the organization provides resources to support employees or teams. With the rapid development of the platform economy, empowerment is applied to platform. Most researchers have studied data empowerment (PDE), which refers to the concept that platform empowers the suppliers, customers, and other participants through intelligence capability, connect capability, and analytic capability [3], and improves their data analysis capabilities and information utilization capabilities.

However, there are few literature studies directly studying PBE. The closest concept to PBE is the retailer's store brand. According to PLMA (Private Label Manufacturers Association), store brand or private brand are the products that carry the retailer's name or private brands. It can help retailers improve bargaining power and increase customers' loyalty [4, 5], while most of them have product quality problems, for example, Wal-Mart and IKEA have experienced quality problems with their own brand products. Three different extended warranty contracts [6] and money-back guarantees provided by retailers [7] can effectively improve the quality of these products and increase supply chain profits.

The existing literatures mainly study the empowerment behaviors from the aspects of psychology and resources, while rarely discuss the impact of empowerment behaviors, especially PBE, on the operational decision-making and performance of e-commerce supply chain members. Our study conducts quantitative analysis and research under the context of PBE by constructing mathematical models to enrich existing research and solve more practical problems.

2.2. Counterfeit Product and Goodwill. Our study is related to the previous works on counterfeit product, which refers to these products that are very similar to the genuine products and sneak into the supply chain for sale at an equivalent price and consumers cannot easily distinguish [8, 9]. Counterfeiters always benefit from their products because they can be sold at high prices with relatively low cost. In particular, when product infringement is deceptive or the counterfeit portion of the market is high, counterfeiters can obtain free rides with improved quality [10]. However, goodwill of enterprise, which is one of the strategic resources and intangible assets for it to form a sustainable competitive advantage, is vulnerable to damage [11, 12]. Most existing studies believe that the level of enterprise's goodwill is affected by the quality of the products and the level of advertising [13–15], when enterprise's goodwill level is higher, the competition based on product quality will be more intense [16] and the higher the degree of product defects will have a greater negative impact [17]. Thus, it is very important to prevent counterfeit products from entering the market and identify parties that may be involved in counterfeiting activities [18]. Qian [19] points that product differences have an impact on counterfeits; consumer purchase intentions [20] and government regulation strategies [21] are the same. In alleviating the problems of counterfeit products, the development of relevant laws [22], website recommendation systems [23], and RFID tags [24] are helpful.

Different from the most studies focusing on the measures taken by downstream enterprises to prevent counterfeiters from entering, this article mainly studies the quality control of the upstream platform on the products ordered by the downstream retailer and the incentive mechanisms established by platform to reduce the motivation of retailer to order counterfeit products from other channels.

2.3. Channel Conflicts. Channel conflict is defined as a situation that one channel member believes that the behaviors of other channel members threaten its profit, their interests are inconsistent, and the decisions are made to maximize their own interests [25, 26]. The causes of channel conflict mainly include different role positioning [27], scarce resources [28], different objectives [29], and so on. Obviously, channel conflicts have a negative impact on channel performance [30, 31], such as reducing the satisfaction [32] and cooperation level [33] among channel members. In the context of e-commerce, the researches mainly focus on the horizontal channel conflict between online channels and offline channels, price competition [34–36], asymmetric information [37], and differences in service quality [38] between the two channels is fierce. Chen et al. [39] think that appropriate local advertising level can reduce conflicts between online and offline channels; adding value to products by retailers [40], adjusting the price difference between channels [41], and implementing segmentation and integration strategies all have the same effect [42].

When channel conflicts occur, traditional contracts cannot coordinate the supply chain. Tsay and Agrawal [43] studied the coordination of supply chain with benefit compensation contract. Also, through the improved revenue sharing contract [44, 45] and option contract [46], channel conflicts can be effectively reduced.

According to the relevant literatures, there are few related to vertical channel conflict; however, we study the vertical channel conflict between e-commerce platform and retailer in offline channel. Our article introduces the supply chain coordination contract into the new scenario of PBE, pays attention to the design of vertical conflict management mechanism, and considers the influence of goodwill on the result of contract coordination.

2.4. Supply Chain Contract

2.4.1. Cost Sharing Contract. As an effective supply chain coordination mechanism, cost sharing contract has been widely used in the field of supply chain management.

Generally, it can achieve supply chain coordination when cost sharing ratio is within a reasonable range [47–49]. Moreover, cost sharing contract can not only improve quality of product in the supply chain [50] but also reduce price competition in dual channels [51].

2.4.2. Minimum Order Quantity Contract. The contract requires the retailers to have a minimum order quantity every time or within a certain period, and they can choose not to order [52, 53]. The researches on the minimum order quantity contract can be divided into two categories. One considers the contract as a necessary constraint, which thinks that the retailer's order quantity must be greater than or equal to the critical value of the minimum order quantity [54–56]. The other is the current mainstream of researches that regards the contract as an incentive constraint, which effectively motivate retailers and suppliers and ultimately realize the coordination of the supply chain [57–59].

Our model first considers whether empowerment cost sharing contract and minimum order quantity contract that is seen as an incentive constraint can coordinate the supply chain compared with DD scenario under the context of PBE, and then studies which of the two contracts can improve the profit of supply chain members more.

This article has the following contributions to existing researches. First, existing researches on platform empowerment mainly focuses on platform data empowerment, while this study studies platform brand empowerment, exploring the retailer's ordering decisions and e-commerce platform's empowerment-level decision through quantitative analysis. Second, though Zhang and Zhang [60] construct a two-echelon supply chain to study channel conflicts with counterfeit products, they mainly discuss the impact of counterfeit products on consumers' perception of quality and price of products. Our research studies the vertical channel conflict between the platform and retailer and considers the impact of counterfeit products on the platform's loss of goodwill. Finally, we extend cost sharing contract and minimum order quantity contract to the vertical channel conflict management mechanism of the e-commerce supply chain in the context of platform brand empowerment and consider the impact of goodwill on the coordination results on the basis of traditional researches.

3. Assumptions and Model Description

We construct a single-cycle two-echelon supply chain model that includes one e-commerce platform (such as JD New Channel), one supplier from other channels, and one retailer (such as JD Convenience Store). We believe that the platform is strong and rich in resources, the products provided are of high quality and qualified, and the price is relatively expensive. However, the products from the supplier of other channels whose management level is low have a certain unqualified rate λ, and a single unqualified product will cause the loss of goodwill on the platform to be c_f. Thus, retailer order q_p units of products from platform at wholesale price w_p and q_s units from other channels at wholesale price w_s, where the ratio to the total order quantity is r. We assume $w_p = w_s + u$ and $w_p > \lambda c_f$, where $u > 0$ denotes that the wholesale price of platform's products is higher than that of other channels; at the same time, it is an exogenous variable and cannot be changed in the short term [55]. The latter is to ensure that the platform is profitable. The structure of supply chain considered is shown in Figure 1.

The market sale prices of the two products are p_p and p_s, respectively. Because counterfeit products and genuine products have similar functions, the substitution rate between them is β, where $\beta \in (0, 1]$ denotes that there must be competition between the two channel products, the smaller the β, the greater the product difference, the less price competition.

In the cooperation of the e-commerce platform and retail store, the platform provides retail store with the level of PBE (t_p) to help it enhance the influence and competitiveness of brand, where t_p is considered as an endogenous variable because the platform needs to make decisions on the level of PBE. With the development of new retail, many e-commerce platforms have raised the exploitation of offline retail stores to a strategic position. PBE gradually presents the characteristics of continuity, diversification, and personalization. For example, in recent years, JD.com has continuously empowered offline retail stores through public relations and marketing activities in cooperation with famous brands, while Ali general provides customized marketing solutions for offline retail stores through "city partners" to improve the uniqueness of Tmall Convenience Store. In addition, $t_p \geq 0$ indicates that at least the brand of the platform does not harm the retailer. The empowerment cost function of the platform is $kt_p^2/2$ [61, 62], where k is the cost coefficient of PBE, and $k > 0$ denotes PBE definitely imposes costs on the platform.

In this article, the market demand is certain, we use the inverse demand function, assuming that the price functions of products provided by the platform and other channels are $p_p = a - q_p - \beta q_s + \gamma g$ and $p_s = a - q_s - \beta q_p + \gamma g$, respectively, which show the competition in the order quantity of products from the two channels and the positive effect of PBE on product price, which are widely adopted by literatures [63–66], especially, $g = \theta t_p$ indicates that PBE helps retailers enhance their brand power, where θ denotes the maximum of retailer's brand power and γ denotes the reaction coefficient of product price to goodwill, which is always nonnegative. Furthermore, we assume that $k(\beta + 1) > \theta^2\gamma^2$ and $k \geq [\theta\gamma(2a - w_s - \lambda c_f) + 2\theta^2\gamma^2]/(2(\beta + 1))$ to ensure that the PBE level is nonnegative, and the value range is [0, 1].

The profit functions of the retailer and e-commerce platform are as follows. Especially, we do not consider the purchase cost of the platform from its suppliers or the production cost of supplier from other channels [60].

$$\Pi_R = (p_p - w_p)q_p + (p_s - w_s)q_s; \tag{1}$$

$$\Pi_p = w_s q_p - \frac{kt_p^2}{2} - \lambda q_s c_f. \tag{2}$$

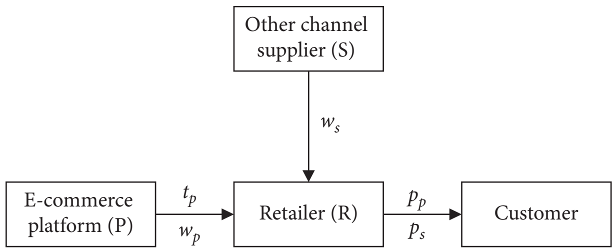


Figure 1: Two-echelon supply chain with vertical channel conflicts.

4. Supply Chain Decisions

In the following part, we discuss the centralized decision scenario as benchmark in Section 4.1 and decentralized decision scenario in Section 4.2, which explores whether PBE is adversely affected by independent decisions of supply chain members. Furthermore, Sections 4.3 and 4.4, respectively, study the role of empowerment cost sharing contract and minimum order quantity contract in reducing vertical channel conflicts.

4.1. CD Scenario. In the CD scenario, the e-commerce platform and retailer operate jointly as a unified whole. This is an ideal decision-making method, in which the decision is globally optimal and provides a benchmark for the improvement of the supply chain members under following scenarios. Decisions on product ordering and the level of PBE are made to maximize the whole supply chain profit.

Note that the total profits of platform and retailer in the CD scenario is as follows, which is the same as that under the CS and QC scenarios.

$$\Pi_{\mathrm{CD}} = p_p q_p + (p_s - w_s) q_s - \frac{k}{2} t_p^2 - \lambda q_s c_f. \tag{3}$$

We conclude the equilibrium solutions in the CD scenario with Proposition 1, and all proofs in this study are provided in the Appendix.

Proposition 1. *For the CD scenario, the equilibrium solutions are as follows:*

(1) The PBE level is $t_p^{CD*} = \theta\gamma(2a - w_s - \lambda c_f)/2K_1$.

(2) The order quantities from the platform and other channels are $q_p^{CD*} = [2ak(1-\beta) + (w_s + \lambda c_f)(2k\beta - \theta^2\gamma^2)]/(4(1-\beta)K_1)$ and $q_s^{CD*} = [2ak\ (1-\beta) - (w_s + \lambda c_f)(2k - \theta^2\gamma^2)]/(4(1-\beta)K_1)$, respectively, where $2ak(1-\beta) > (w_s + \lambda c_f)(2k - \theta^2\gamma^2)$ and $K_1 = k(1+\beta) - \theta^2\gamma^2$, $K_1 > 0$.

Corollary 1. *The monotonicity of equilibrium solutions is as follows:*

(1) q_p^{CD*} increases with c_f, q_s^{CD*} and t_p^{CD*} decrease with c_f

(2) q_p^{CD*}, q_s^{CD*}, and t_p^{CD*} increase with γ, decrease with β and k

Corollary 1 (1) shows that the optimal quantity of products from platform increases with c_f, the reason is that the losses caused by counterfeit products will be greater when c_f increases, so retailer will order more products from the platform in the CD scenario. However, the PBE level decreases with c_f because with the increase of goodwill loss, the platform will reduce the cost by reducing the PBE level, so as to ensure its own profit.

Corollary 1 (2) indicates that the quantity of products from both channels and empowerment level increase with γ but decrease with β and k. The greater the γ, the higher the price of the product, so the profits of retailer and platform will increase when market demand is determined; thus, the quantity of products ordered by retailer from platform will increase, and the platform will also be willing to invest in the PBE. Regarding the substitutability of the two products, the higher it is, the more intense the product competition, and the profit of retailer decreases because of low price. Then, the retailer will buy more products from other supplier with lower wholesale prices. This means that with high product substitution, PBE encourages retailer to order more products from other channels, which will decrease the empowerment level and hurt platform's profit. Also, there is an intuitive conclusion that the empowerment level is inversely related to the k, which is the higher the empowerment cost coefficient, the lower the platform's empowerment level.

4.2. DD Scenario. In the DD scenario, the e-commerce platform and the retailer with decentralized decision are pursuing their own interests to maximize. The corresponding game steps are as follows: (1) The e-commerce platform determines the PBE level (t_p). (2) The retailer decides the order quantity from e-commerce platform (q_p) and from other channels (q_s).

Because the platform is the leader in the game, the second stage of the game is first calculated. From the formula (1), we can get the optimal decisions in Proposition 2.

Proposition 2. *For the DD scenario, the equilibrium solutions are as follows:*

(1) The PBE level is $t_p^{DD*} = \theta\gamma(u + w_s - \lambda c_f)/(2k(\beta+1))$

(2) The equilibrium order quantities from the platform and supplier are

$$q_p^{\mathrm{DD}*} = \frac{\left[(1-\beta)\left(a - w_s + \theta\gamma K_2\left(u + w_s - \lambda c_f\right)\right) - u\right]}{2\left(1-\beta^2\right)},$$

$$q_s^{\mathrm{DD}*} = \frac{\left[(1-\beta)\left(a - w_s + \theta\gamma K_2\left(u + w_s - \lambda c_f\right)\right) + \beta u\right]}{2\left(1-\beta^2\right)}, \tag{4}$$

respectively, where $K_2 = \theta\gamma/(2k(\beta+1))$.

Corollary 2. *The monotonicity of the equilibrium solutions is q_p^{DD*}, q_s^{DD*}, and t_p^{DD*} all increase with γ and decrease with c_f, β, and k.*

Different from the CD scenario, the greater the c_f, the less the product quantity ordered by retailer from e-commerce platform and the lower the empowerment level in the DD scenario; the reason is that e-commerce platform and retailer are chasing their own profits, and retailer does not consider the loss of e-commerce platform and orders more products from other channels with lower wholesale price and quality. Similarly, e-commerce platform reduces the invest in PBE to reduce the costs. As the t_p decreases, the product price of e-commerce platform decreases; thus, the retailer's order quantity from supplier of other channels also reduces.

4.3. CS Scenario. In the scenario, e-commerce platform adopts empowerment cost sharing contract on the basis of DD scenario. Similar to the literatures of Ghosh et al. [67] and Zha et al. [68], we use μ $(0 \leq \mu \leq 1)$ to represent the cost sharing ratio of retailer.

The corresponding game steps are as follows: (1) The brand empowerment cost sharing contract with retailer's sharing ratio (μ) is provided by e-commerce platform; (2) the e-commerce platform decides the PBE level (t_p) and the unit wholesale price (w_s); and (3) the retailer, respectively, decides the product quantity from e-commerce platform (q_p) and supplier of other channel (q_s).

Under the CS scenario, the platform's profit function is

$$\Pi_p^{\mathrm{CS}} = w_p q_p - \frac{k}{2}(1-\mu)t_p^2 - \lambda q_s c_f. \tag{5}$$

The retailer's profit function is

$$\Pi_R^{\mathrm{CS}} = \left(p_p - w_p\right)q_p - \frac{k}{2}\mu t_p^2 + \left(p_s - w_s\right)q_s. \tag{6}$$

According to the decision-making order, we still use the reverse induction method to solve the model, and the following conclusions can be drawn.

Proposition 3. *For the CS scenario, the equilibrium solutions are as follows:*

(1) The PBE level is $t_p^{\mathrm{CS}*} = \theta\gamma(2a - w_s - \lambda c_f)/2K_1$;

(2) The equilibrium product quantities from the e-commerce platform and other channels are

$$\begin{aligned} q_p^{\mathrm{CS}*} &= \frac{2k(1+\beta)\left[(1-\beta)(a-w_s)-u\right] + \theta^2\gamma^2\left[2u + (1-\beta)\left(w_s - \lambda c_f\right)\right]}{4K_1\left(1-\beta^2\right)}, \\ q_s^{\mathrm{CS}*} &= \frac{2k(1+\beta)\left[(1-\beta)(a-w_s)+u\beta\right] - \theta^2\gamma^2\left[(\beta-1)\left(w_s - \lambda c_f\right) + 2u\beta\right]}{4K_1\left(1-\beta^2\right)}. \end{aligned} \tag{7}$$

(3) The empowerment cost sharing ratio of retailer is

$$\mu^{\mathrm{CS}*} = \frac{\theta^2\gamma^2\left(u + w_s - \lambda c_f\right) + k(1+\beta)(2a - u - 2w_s)}{k(1+\beta)\left(2a - w_s - \lambda c_f\right)}. \tag{8}$$

Corollary 3. *The monotonicity of the equilibrium solutions is as follows:*

(1) $q_p^{\mathrm{CS}*}$, $q_s^{\mathrm{CS}*}$, *and* $t_p^{\mathrm{CS}*}$ *decrease with* c_f; $\mu^{\mathrm{CS}*}$ *increases with* c_f

(2) $q_p^{\mathrm{CS}*}$, $q_s^{\mathrm{CS}*}$, $t_p^{\mathrm{CS}*}$, *and* $\mu^{\mathrm{CS}*}$ *increase with* γ *and decreases with* β *and* k

Corollary 3 illustrates that with the increase of c_f, the empowerment cost sharing ratio will increase. In the case of other conditions unchanged, the e-commerce platform will increase the cost sharing ratio with retailer to reduce its own costs.

In addition, $\mu^{\mathrm{CS}*}$ increase with γ, the increase of $\mu^{\mathrm{CS}*}$ inspires e-commerce platform to invest more in PBE; then, the price will be higher and profits of e-commerce platform and retailer will both increase. Moreover, $\mu^{\mathrm{CS}*}$ decrease with k. Under the other conditions unchanged, the e-commerce platform will reduce the PBE level with the k increase and the price will be lower, while the retailer's profit decreases with the increase in cost sharing ratio and retailer will order more products from supplier of other channels in order to increase profit, which leads to a greater loss of goodwill suffered by e-commerce platform, so platform has to lessen the $\mu^{\mathrm{CS}*}$ to reduce losses. Corollary 3 also shows that $\mu^{\mathrm{CS}*}$ decrease with β, the reason is that the increase in β leads to the decrease in product price, and retailer will order more products with lower wholesale price from other channels to ensure its own profits, and e-commerce platform has to decrease cost sharing ratio in order to encourage retailer to order more products from it.

4.4. QC Scenario. In the scenario, e-commerce platform offers minimum order quantity contract to retailer on the basis of DD scenario. We assume that e-commerce platform takes incentive and punishment measures for retailer's order behaviors [57]. When $q_p > T$, the platform rewards $\tau(q_p - T)$ to retailer's overfulfilled order. When $q_p < T$, the platform punishes $\tau(T - q_p)$ to the retailer for incomplete order, where τ is the reward and punishment coefficient, and $0 < \tau < 1$, T is the reward and punishment standard that is set in advance as sales target in the contract between e-commerce platform and retailer.

Under the QC scenario, the profit of platform is written as follows:

$$\Pi_P^{QC} = w_p q_p - \frac{k}{2} t_p^2 - \lambda q_s c_f - \tau\left(q_p - T\right). \quad (9)$$

The retailer's profit function is given by

$$\Pi_R^{QC} = \left(p_p - w_p\right) q_p + \left(p_s - w_s\right) q_s + \tau\left(q_p - T\right). \quad (10)$$

Proposition 4. *For the QC scenario, the equilibrium solutions are as follows:*

(1) The PBE level is $t_p^{QC*} = \theta\gamma\left(2a - w_s - \lambda c_f\right)/2K_1$.

(2) The equilibrium order quantities from the platform and supplier are

$$\begin{aligned} q_p^{QC*} &= \frac{\left[2ak(1-\beta) + \left(w_s + \lambda c_f\right)\left(2k\beta - \theta^2\gamma^2\right)\right]}{4K_1(1-\beta)}, \\ q_s^{QC*} &= \frac{\left[2ak(1-\beta) - \lambda c_f\left(2k\beta^2 - \theta^2\gamma^2(2\beta - 1)\right) - w_s\left(2k - \theta^2\gamma^2\right)\right]}{4K_1(1-\beta)}. \end{aligned} \quad (11)$$

(3) The coefficient of reward and punishment is

$$\tau^{QC*} = u + w_s + \beta\lambda c_f. \quad (12)$$

Corollary 4. *The monotonicity of the equilibrium solutions is as follows:*

(1) q_s^{QC*} *and* t_p^{QC*} *decrease with* c_f; q_p^{QC*} *increases with* c_f

(2) q_p^{QC*}, q_s^{QC*}, *and* t_p^{QC*} *increase with* γ *and decrease with* β *and* k

Corollary 4 shows that q_p^{QC*} increases with c_f and γ but decreases with β and k, which is the same as CD scenario; the greater the c_f, the more the products are ordered from platform. With the c_f increase, the reward and punishment coefficient τ also increases, which encourages retailer to order products from platform, and the risk of goodwill loss to the platform also reduces.

5. Comparative Analysis of Different Supply Chain Scenarios

In different scenarios, the retailer's order decisions and the e-commerce platform's PBE level decision are key decision variables for the supply chain. In this section, we focus on analyzing the relationships between these decision variables and the impact of these variables on the profits of supply chain members.

5.1. Decision Comparisons. To get more insights on decision-making of supply chain members under the different scenarios, we compare the important variables in the following.

Corollary 5. *The optimal PBE levels under different scenarios satisfy the relationship as* $t_p^{QC*} = t_p^{CS*} = t_p^{CD*} > t_p^{DD*}$.

Corollary 5 shows that the optimal PBE level of e-commerce platform can reach the level of CD scenario, and it is greater than that under the DD scenario through the coordination of CS and QC. It means that by implementing cost sharing contract and minimum order quantity contract, e-commerce platform can better exert its PBE level and retailer can also get much support from the platform.

Corollary 6. *The optimal decisions about products quantities ordered by retailer from e-commerce platform under different scenarios satisfy the relationship as* $q_P^{CD*} = q_P^{QC*} > q_P^{CS*} > q_P^{DD*}$.

Corollary 6 reveals that optimal product quantities ordered by retailer from platform under CS and QC scenarios are greater than that under DD scenario. Moreover, under the QC scenario, the optimal product quantity is the same as that under the CD scenario and is greater than that under the CS scenario. Therefore, the e-commerce platform can effectively encourage retailer to order more products from itself through CS and QC, which guarantees product quality sold by retailer to a certain extent, especially QC is more helpful to achieve the goal.

Corollary 7. *The optimal product quantity ordered by retailer from other channels under different scenarios satisfy the following:*

(1) When $0 < \gamma^2 < (\overline{\varphi}/\theta^2)$, *we have* $q_s^{CS*} > q_s^{DD*} > q_s^{QC*} > q_s^{CD*}$

(2) When $(\overline{\varphi}/\theta^2) < \gamma^2 < (k(\beta+1)/\theta^2)$, *we have* $q_s^{CS*} > q_s^{QC*} > q_s^{DD*} > q_s^{CD*}$

From Corollary 7, we find that CS always encourages retailer to order more products from other channels whether the impact of price on goodwill is high or low. The reason is that retailer has to share part of the cost of PBE under the CS scenario. In order to make up for the decline of its own profit, retailer order more products from supplier with lower wholesale prices. In addition, the products ordered from supplier in the QC scenario are less than that in the DD scenario when γ is small; otherwise, the situation is the opposite. This is because that when goodwill has little impact on the price, the price of products ordered from other channels is lower so that the retailer's profit gained from other channels is relatively low. With the incentive of the reward and punishment mechanism, retailer is more inclined to order products from the platform. While when goodwill has a greater impact on the price, the price of products ordered from other channels is higher, which increase the profit of retailer and retailer prefer to order products from other channels.

Both Corollaries 6 and 7 illustrate that QC can effectively motivate the retailer to order more products from platform, especially, it can reduce the product quantity ordered from other channels when goodwill has less of an effect on price, which is more effective than CS in reducing channel conflicts and ensuring product quality.

Corollary 8. *The optimal ratio of product quantities ordered by retailer from e-commerce platform to total quantity under different scenarios satisfy the relationship as* $r_p^{CD*} > r_p^{QC*} > r_p^{CS*} > r_p^{DD*}$.

Through Corollary 8, we find that QC and CS can effectively encourage retailer to increase the proportion of product quantity ordered from the platform and reduce that from other channels. Especially, the ratio under the QC scenario is greater than that under CS and DD scenarios.

Thus, e-commerce platform can effectively stimulate retailer's order ratios from itself by adopting minimum order quantity contract. Although Corollary 7 shows that QC will increase the product quantity from other channels when $(\overline{\varphi}/\theta^2) < \gamma^2 < (k(\beta+1)/\theta^2)$, the retailer's order ratio from the platform is greater than that under the CS and DD scenarios. That is to say, the platform can stimulate the retailer's quantity order from itself by providing QC, which possibly reduce the loss of goodwill of the platform caused by counterfeit products.

What's more, although Corollaries 6 and 7 show that products ordered from other channels in the CS scenario is more than that in the DD scenario, the CS also increases product quantity ordered from the platform. It can be drawn from Corollary 8 that the ratio of product quantity from platform in CS scenario is higher than that of DD scenario, which means CS contract still has a positive effect on encouraging retailer to order more products from the platform on the whole.

5.2. Comparison of Profits. In order to better analyze the changes in the profits of e-commerce platform and retailer under different scenarios, we assumed that $w_s = 0$. That is to say, the wholesale price of products ordered from other channels is very low, which closes to zero.

Corollary 9. *Comparing the profit of platform and retailer under CS scenario and DD scenario, we have*

(1) $\Pi_P^{CS*} > \Pi_P^{DD*}$

(2) $\Pi_R^{CS*} > \Pi_R^{DD*}$

Obviously, the empowerment cost sharing contract can alleviate the channel conflicts between e-commerce platform and retailer and realize the Pareto improvement of their profits compared with DD scenario.

Corollary 10. *The optimal total profit of platform and retailer under different scenarios satisfies the relationship as* $\Pi_C^{CD*} > \Pi_C^{QC*} > \Pi_C^{CS*} > \Pi_C^{DD*}$.

Both Corollaries 8 and 10 show that compared with CS, QC can not only better motivate retailer to order more products from platform but also improve the profits of platform and retailer.

Theorem 1. *If platform provides an incentive mechanism as* $\tau(Q-T)$ *in a nonvertically integrated mixed channel scenario, where* $\tau = w_p + \beta\lambda c_f$. *There is* $T_1, T_2, T_3, T_4, T_5, T_6$ *and* $T_4 > T_3 > T_2 > T_1$, $T_6 > T_5$:

(1) When $T \in (T_1, T_2]$, *after coordination of QC, the retailer's profit is greater than that under the DD scenario, but e-commerce platform's profit is opposite*

(2) When $T \in (T_2, T_3)$, *both e-commerce platform and retailer's profits are greater than that under the DD scenario*

(3) When $T \in [T_3, T_4)$, *e-commerce platform's profit is greater than that under the DD scenario, while retailer's profit is the opposite*

(4) When $T \in (T_5, T_6)$, *both the profits of e-commerce platform and retailer are greater than that under the CS scenario*

The first three illustrate that the QC can effectively reduce vertical channel conflicts of supply chain with PBE in a suitable range, and Theorem 1 (4) shows that the QC has a better effect than the CS; thus, e-commerce platform can choose incentive and punishment standard acceptable to itself and retailer according to the actual situation, so that they can achieve a win-win situation.

Corollary 11. *When* $T = 0$, $\Pi_P^{QC*} < 0$, *the linear rebate with minimum order quantity cannot achieve the coordination of supply chain.*

Corollary 11 demonstrates that the minimum order quantity contract can degenerate into a linear rebate mechanism when $T = 0$, and retailer will receive rebate $\tau > 0$ when $q_p^{QC} > 0$ of each unit product. In order to encourage retailer to order the same product quantity from e-commerce platform as CD scenario, the rebate provided by the platform needs to satisfy $\tau = u + w_s + \beta\lambda c_f$, which makes e-commerce platform's profit negative. Therefore, the linear rebate mechanism does not coordinate the supply chain.

6. Numerical Simulation

This part further analyzes the impact of key variables on the supply chain. Especially, we assume that $w_s \neq 0$, and consider the numerical analysis of the sensitivity of the coefficients and variables. From the assumptions and practical situations, the basic parameter values are assigned as follows: $a = 1$, $\beta = 0.4$, $\theta = 0.2$, $\gamma = 0.15$, $\lambda = 0.2$, $c_f = 0.15$, $u = 0.2$, $k = 0.005$, $w_s = 0.2$.

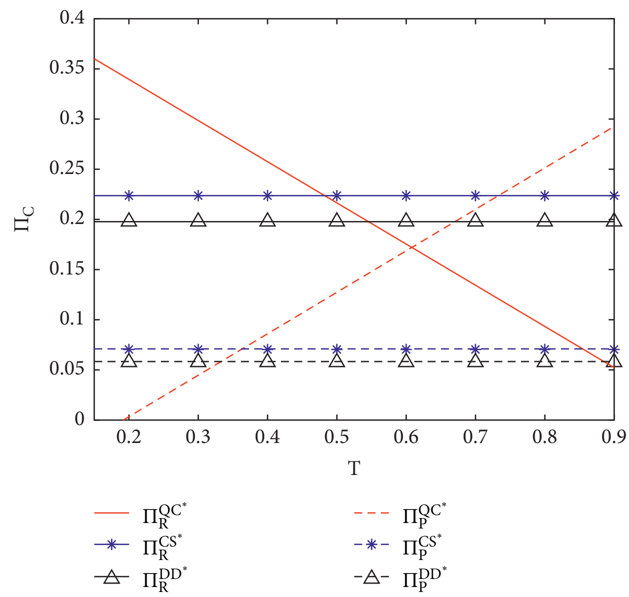


FIGURE 2: Impact of T on e-commerce platform and retailer's profits.

6.1. Impact of Reward and Punishment Standard T on the Profits of Platform and Retailer. Figure 2 shows the results of profit comparison under different scenarios. We find that both platform and retailer's profit under CS scenario are always higher than that under DD scenario. However, the situation under QC scenario is more complicated. When T is at (0.19, 0.33), the platform's profit is less than that under CS and DD scenarios, whereas the retailer's profit is highest of three scenarios, that is, the platform's profit is damaged and QC is beneficial to the retailer; as T increases in (0.33, 0.37), retailer's profit is still the highest under all three different scenarios; the platform's profit is higher than that under DD scenario but lower than that under CS scenario. As T increases in (0.37, 0.48), both the platform and retailer have the highest profit under the three scenarios and QC is the best choice for them. When T is at (0.48, 0.55), the platform's profit increases and it is the highest of the three scenarios, whereas retailer's profit decreases and it is greater than that under DD scenario but less than CS scenario, and QC is the best choice for platform. The reason is that retailer order expensive products from platform because of high standard, and platform's profit will be higher, whereas the retailer's profit will be lower with the increase of T.

In summary, platform and retailer have always embraced CS because it increases their profits. When T is small, the platform is not willing to provide QC for that the expected profit it brings will be lower than that under DD scenario. Figure 2 verifies Theorem 1 that only when T is within the appropriate range can QC be accepted by both retailer and platform, and there is Pareto improvements of platform and retailer's profits. For example, Alibaba requires that retailers must order more than 30% of products from Ali Retail Link every month, if the amount of money reaches 5000 yuan, the retailers can get a reward, which can greatly encourage them to order more products from platform, and the profits of both sides can be improved. Otherwise, QC will reduce the profit of one party and the channel conflicts cannot be alleviated.

6.2. The Impact of Reaction Coefficient of Price to Goodwill γ on the Total Profit of Platform and Retailer. It can be seen that as γ increases, the price of product increases and retailer will order more products from e-commerce platform, which increases the profits of both. However, due to the existence of channel conflicts, the profit growth is slow, and the profits under DD scenario are always the lowest.

Figure 3 shows that as γ increases, the effect of the two contracts on channel conflicts gradually increases; especially, QC's effect is always better. When γ is small, the effect of CS is not obvious, and it gradually increases only when γ is large. What's more, the gap between the two contracts' improvement effects gradually becomes smaller as γ increases. Therefore, for the platform, it is necessary to choose contract based on the degree of influence of the products it provides on the goodwill, minimum order quantity contract can be selected for products whose prices are less affected by goodwill, such as JD Convenience Store mainly deals in snacks and daily necessities, whose price is less affected by goodwill. With the incentive of minimum order quantity contract, retailers are more willing to order products from the platform in order to increase profits, and channel conflicts have been effectively reduced; otherwise, both QC and CS contracts can be selected. However, by comparing Figures 2 and 3, we can find that QC is not always the best choice for the platform, if the reward and punishment mechanism is not set properly, platform's profit in QC will decline, which is a problem it needs to be aware of.

6.3. Impact of Product Substitutability β on the Total Profit of Platform and Retailer. Figure 4 shows that with the increase of β, not only is profit decreases but also is their improvement on channel conflicts, especially the downward trend of QC slow down as the competition intensifies because the incentive and punishment coefficient set by the

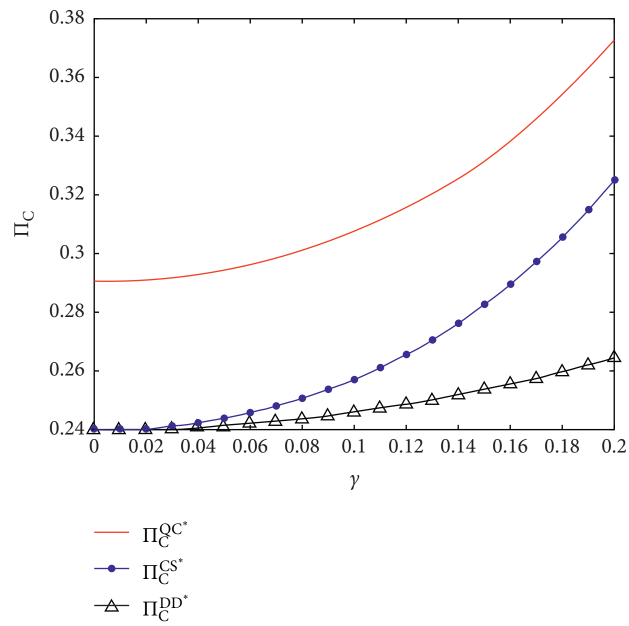


Figure 3: Impact of γ on the total profit of platform and retailer.

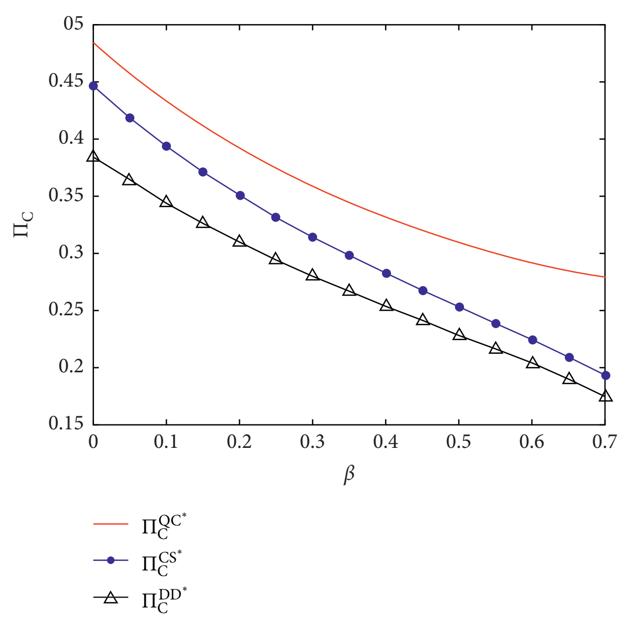


Figure 4: Impact of β on the profit of platform and retailer.

platform will increase, which encourages retailer to increase product quantity ordered from the platform, and the goodwill loss suffered by it will gradually reduce. With the intensification of product competition, QC's effect in improving profits is more obvious.

In order to improve the features and differences of products, e-commerce platform often takes product differentiation strategy, such as improve the product quality and style to reduce channel conflicts. For example, JD.com has established differentiated competitive advantage by making its own products to ensure their quality and cost performance. However, as can be seen from Figure 4, because the wholesale price of the products provided by the platform is higher, retailers will order cheaper but lower quality products from other suppliers in order to maximize their profits. Therefore, the platform will still suffer the loss of goodwill due to the potential quality problems of products. Thus, for commodities and other competitive products, e-commerce platform can adopt QC and CS to coordinate, while for some valuable products, e-commerce platform can reduce channel conflicts through QC.

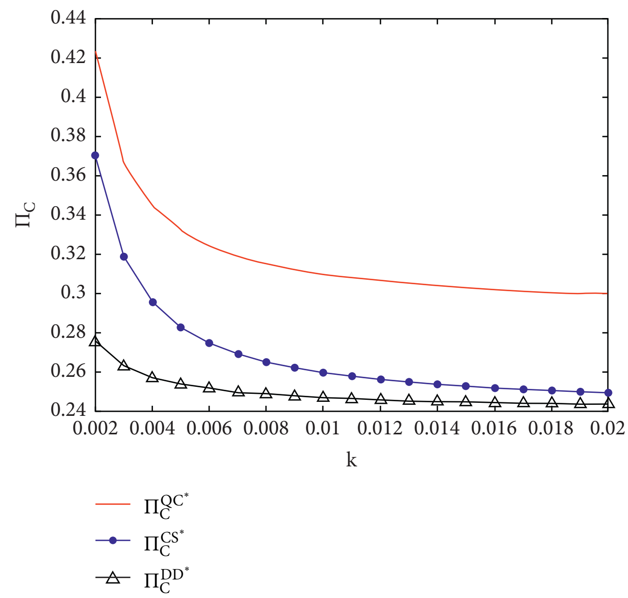


Figure 5: Impact of k on the profit of platform and retailer.

6.4. Impact of Empowerment Cost Coefficient k on the Total Profit of Platform and Retailer. Figure 5 indicates that the increase of k has a negative impact on the total profit of platform and retailer under three different scenarios, the reason is that platform will increase products' wholesale price because of high empowerment cost; thus, retailer will order more products from supplier to increase its profit, which may cause goodwill loss to platform. We find that the improvement effect of CS and QC both weaken with the increase in k in a small range, whereas when k increases in a large range, their decreasing trends gradually slow down. Especially, the improvement effect of QC is always better than that of CS in the part of analysis.

On the one hand, when the cost of PBE is high, the platform can use QC to reduce channel conflicts; otherwise, both CS and QC can be used. On the other hand, Figure 5 shows that the profit will increase rapidly as k decreases. Thus, how to reduce the cost coefficient of PBE is an important issue for the platform to consider. For example, platform can use event marketing, personalized customization, and other ways to improve the efficiency of empowerment to reduce the cost of empowerment. Take JD Convenience Store as an example, with the help of JD.com, it became the design theme of a design competition and successfully publicized its brand and concept to the society; also, JD.com uses its advantages to help it optimize the consumption scene and make it more in line with the shopping habits of target customers, thus stimulating consumers' demand.

7. Conclusion

In this article, we construct a two-echelon supply chain model, explore how to mitigate vertical channel conflicts caused by retailer's "free-rider" behavior chasing short-term interests and inconsistent business goals between the two parties, and further study the coordination effect of cost sharing contract and minimum order quantity contract on vertical channel conflicts.

First, channel conflicts can reduce product quantity ordered by retailer from platform and lead to a loss of goodwill of platform. Secondly, both cost sharing contract and minimum order quantity contract can effectively reduce channel conflicts, and the effect of minimum order quantity contract is greater than that of cost sharing contract. Minimum order quantity contract can not only help e-commerce platform and retailer obtain more profits but also achieve the optimal PBE level and product quantity from platform under centralized decision scenario. Thirdly, e-commerce platform through differentiated product strategy can reduce the degree of competition between the two products and reduce vertical channel conflicts, but it cannot achieve the coordination of the supply chain, and platform still needs to design proper contract to improve the product quantity ordered from itself. Finally, with the increase in the degree of competition between the two products and the cost of PBE, the improvement effect of empowerment cost sharing contract decreases, whereas the effect of minimum order quantity contract increase.

However, with the increase in the reaction coefficient of prices on goodwill, the effect of those two contracts increases.

There are some directions for future research:

(1) We only consider the problem under the context of information symmetry; however, the information between platform and retailer is probably asymmetric in reality. Taking JD.com as an example, in order to better meet the market demand, retailers are allowed to order products from other channels; however, JD.com does not know the quality of these products. High-quality products will not affect the goodwill of the platform, while unqualified products may hurt the goodwill of the platform. Therefore, the study of asymmetric information about product quality is very useful to reduce channel conflict.

(2) This article studies the coordination strategy of channel conflicts under the determined demand, future research can explore channel conflicts under the uncertain demand, which is a classic problem that has always been studied in supply chain management. Also, as competition intensifies, the business environment is undergoing dynamic changes, and market demand is in an uncertain state, so it is necessary to study the problem of channel conflict under the uncertain demand.

(3) This article examines the single-period model; however, PBE is a continuous process and the platform will decide the level of PBE according to its own interests, market environment, and other specific circumstances. Therefore, the study of channel conflict in multiperiod dynamic environment can indicate the dynamic decision-making process of the platform.

Appendix

Proof of Proposition 1. First, from function (3), the first-order partial derivative of q_s, q_p, and t_p for Π_{CD} are as follows:

$$\frac{\partial \Pi_{CD}}{\partial q_p} = a - 2q_p - 2q_s\beta + t_p\theta\gamma = 0, \tag{A.1}$$

$$\frac{\partial \Pi_{CD}}{\partial q_s} = a - 2q_s - c_f\lambda - 2q_p\beta + \theta t_p\gamma - w_s = 0, \tag{A.2}$$

$$\frac{\partial \Pi_{CD}}{\partial t_p} = q_p\gamma\theta - kt_p + q_s\gamma\theta = 0. \tag{A.3}$$

Then, we continue to calculate partial derivative, and the results are as follows:

$$\begin{aligned}
\frac{\partial^2 \Pi_C^{CD}}{\partial q_p^2} &= -2,\\
\frac{\partial^2 \Pi_C^{CD}}{\partial q_s^2} &= -2,\\
\frac{\partial^2 \Pi_C^{CD}}{\partial t_p^2} &= -k,\\
\frac{\partial^2 \Pi_C^{CD}}{\partial q_p \partial q_s} &= -2\beta,\\
\frac{\partial^2 \Pi_C^{CD}}{\partial q_p \partial t_p} &= \theta\gamma,\\
\frac{\partial^2 \Pi_C^{CD}}{\partial q_s \partial q_p} &= -2\beta,\\
\frac{\partial^2 \Pi_C^{CD}}{\partial q_s \partial t_p} &= \theta\gamma,\\
\frac{\partial^2 \Pi_C^{CD}}{\partial t_p \partial q_p} &= \theta\gamma,\\
\frac{\partial^2 \Pi_C^{CD}}{\partial t_p \partial q_s} &= \theta\gamma.
\end{aligned} \tag{A.4}$$

When $k(\beta + 1) - \theta^2\gamma^2 > 0$, the Hesse matrix on Π_{CD} is negative, so Π_{CD} can get the maximum under specific q_p, q_s, and t_p.

From the formulation (A.1)–(A.3), we get the equilibrium solution in the centralized scenario:

$$q_p^{CD*} = \frac{2ak(1-\beta) + (w_s + \lambda c_f)(2k\beta - \theta^2\gamma^2)}{4(1-\beta)K_1}, \tag{A.5}$$

$$q_s^{CD*} = \frac{2ak(1-\beta) - (w_s + \lambda c_f)(2k - \theta^2\gamma^2)}{4(1-\beta)K_1}, \tag{A.6}$$

$$t_p^{CD*} = \frac{\theta\gamma(2a - w_s - \lambda c_f)}{2K_1}. \tag{A.7}$$

According to $0 \le t_p \le 1$, we get $k(\beta + 1) > \theta^2\gamma^2$, and $k \ge (\theta\gamma(2a - w_s - \lambda c_f) + 2\theta^2\gamma^2)/(2(1+\beta))$. □

Proof of Corollary 1. (1) From equation (A.5)–(A.7), the first-order partial derivatives of c_f, γ, k, and β can be shown as follows:

$$\begin{aligned}
\frac{\partial q_p^{CD*}}{\partial c_f} &= \frac{\lambda\left(2k\beta-\theta^2\gamma^2\right)}{4(1-\beta)K_1^2}>0,\\
\frac{\partial q_s^{CD*}}{\partial c_f} &= -\frac{\lambda\left(2k\beta-\theta^2\gamma^2\right)}{4(1-\beta)K_1^2}<0,\\
\frac{\partial t_p^{CD*}}{\partial c_f} &= -\frac{\theta\lambda\gamma}{2K_1}<0,\\
\frac{\partial q_p^{CD*}}{\partial \gamma} &= \frac{\theta^2 k\gamma\left(2a-w_s-\lambda c_f\right)}{2K_1^2}>0,\\
\frac{\partial t_p^{CD*}}{\partial \gamma} &= \frac{\theta\left(k+\theta^2\gamma^2+k\beta\right)\left(2a-w_s-\lambda c_f\right)}{2K_1^2}>0;\\
\frac{\partial q_p^{CD*}}{\partial k} &= -\frac{\theta^2\gamma^2\left(2a-w_s-\lambda c_f\right)}{4K_1^2}<0,\\
\frac{\partial q_s^{CD*}}{\partial k} &= -\frac{\theta^2\gamma^2\left(2a-w_s-\lambda c_f\right)}{4K_1^2}<0,\\
\frac{\partial q_s^{CD*}}{\partial \gamma} &= \frac{\theta^2 k\gamma\left(2a-w_s-\lambda c_f\right)}{2K_1^2}>0,\\
\frac{\partial t_p^{CD*}}{\partial k} &= -\frac{\theta\gamma(\beta+1)\left(2a-w_s-\lambda c_f\right)}{2K_1^2}<0;\\
\frac{\partial q_p^{CD*}}{\partial \beta} &= -\frac{2ak^2(1-\beta)^2-\left(w_s+\lambda c_f\right)\left[\theta^2\gamma^2\left(-2k+\theta^2\gamma^2-2k\beta\right)+2k^2\left(\beta^2+1\right)\right]}{4(1-\beta)^2K_1^2}<0.
\end{aligned}\tag{A.8}$$

The reason is that $2ak^2(1-\beta)^2-(w_s+\lambda c_f)(\theta^4\gamma^4+2k^2\beta^2+2k^2-2k\theta^2\gamma^2-2k\beta\theta^2\gamma^2)>(k-\theta^2\gamma^2+k\beta)(w_s+\lambda c_f)>0$ due to $2a>w_s+\lambda c_f$.

$$\begin{aligned}
\frac{\partial q_s^{CD*}}{\partial \beta} &= -\frac{\left(w_s+\lambda c_f\right)\left(\theta^2\gamma^2+2k^2\beta^2+2k^2-2k\theta^2\gamma^2-2k\beta\theta^2\gamma^2\right)}{4K_1^2(1-\beta)^2}<0,\\
\frac{\partial t_p^{CD*}}{\partial \beta} &= -\frac{k\theta\gamma\left(w_s+\lambda c_f\right)}{2K_1^2}<0.
\end{aligned}\tag{A.9}$$

□

Proof of Proposition 2. From equation (1), the first-order partial derivatives of Π_R to q_p, q_s, we get

$$q_p^{\mathrm{DD}}(t_p) = \frac{(a - w_s + \theta t_p \gamma)(1-\beta) - u}{2(1-\beta^2)}, \tag{A.10}$$

$$q_s^{\mathrm{DD}}(t_p) = \frac{(a - w_s + \theta t_p \gamma)(1-\beta) + u\beta}{2(1-\beta^2)}. \tag{A.11}$$

Taking equation (A.10) and (A.11) into Π_p^{DD} and calculating its first-order partial derivative, we get ($\partial\Pi_p^{\mathrm{DD}}/\partial t_p) = ((-2kt_p(\beta+1) + \theta\gamma$ $(u + w_s - \lambda c_f))/2(\beta+1))$.

Let it equal to 0, we obtain the equilibrium solution,

$$t_p^{\mathrm{DD}^*} = \frac{\theta\gamma(u + w_s - \lambda c_f)}{2k(\beta+1)}. \tag{A.12}$$

In order to simplify the formula, let $K_2 = \theta\gamma/(2k(\beta+1))$, thus,

$$t_p^{\mathrm{DD}^*} = K_2(u + w_s - \lambda c_f). \tag{A.13}$$

Taking (A.13) into formula (1), we get

$$q_p^{\mathrm{DD}*} = \frac{(1-\beta)(a - w_s + \theta\gamma K_2(u + w_s - \lambda c_f)) - u}{2(1-\beta^2)}, \tag{A.14}$$

$$q_s^{\mathrm{DD}*} = \frac{(1-\beta)[a - w_s + \theta\gamma K_2(u + w_s - \lambda c_f)] + \beta u}{2(1-\beta^2)}, \tag{A.15}$$

and $(1-\beta)[a - w_s + \theta\gamma K_2(u + w_s - \lambda c_f)] \geq u$. □

Proof of Corollary 2. From equation (A.13)–(A.15), the first-order partial derivatives of γ, c_f, β, and k can be are as follows:

$$\begin{aligned}
\frac{\partial q_p^{\mathrm{DD}*}}{\partial\gamma} &= \frac{\theta^2\gamma(u + w_s - \lambda c_f)}{2k(\beta+1)^2} > 0,\\
\frac{\partial q_s^{\mathrm{DD}*}}{\partial\gamma} &= \frac{\theta^2\gamma(u + w_s - \lambda c_f)}{2k(\beta+1)^2} > 0,\\
\frac{\partial t_p^{\mathrm{DD}*}}{\partial\gamma} &= \frac{\theta(u + w_s - \lambda c_f)}{2k(\beta+1)} > 0;\\
\frac{\partial q_p^{\mathrm{DD}*}}{\partial c_f} &= -\frac{\theta^2\lambda\gamma^2}{4k(\beta+1)^2} < 0,\\
\frac{\partial t_p^{\mathrm{DD}*}}{\partial c_f} &= -\frac{\theta\lambda\gamma}{2k(\beta+1)} < 0,\\
\frac{\partial q_s^{\mathrm{DD}*}}{\partial c_f} &= -\frac{\theta^2\gamma^2\lambda}{4k(\beta+1)^2} < 0;\\
\frac{\partial q_p^{\mathrm{DD}*}}{\partial k} &= -\frac{\theta^2\gamma^2(u + w_s - \lambda c_f)}{4k^2(\beta+1)^2} < 0,\\
\frac{\partial q_s^{\mathrm{DD}*}}{\partial k} &= -\frac{\theta^2\gamma^2(u + w_s - \lambda c_f)}{k^2(\beta+1)^2} < 0,\\
\frac{\partial t_p^{\mathrm{DD}*}}{\partial k} &= -\frac{\theta\gamma(u + w_s - \lambda c_f)}{2k^2(\beta+1)} < 0;\\
\frac{\partial q_p^{\mathrm{DD}*}}{\partial\beta} &= -\frac{(\beta-1)^2(ak(\beta+1) - kw_s(\beta+1) + \theta^2\gamma^2(u + w_s - \lambda c_f)) + 2ku\beta(\beta+1)}{k(\beta-1)^2(\beta+1)^3} < 0,\\
\frac{\partial q_s^{\mathrm{DD}*}}{\partial\beta} &= -\frac{(\beta-1)^2(k(\beta+1)(a - w_s) + \theta^2\gamma^2(u + w_s - \lambda c_f)) - ku(\beta+1)(\beta^2+1)}{2k(\beta-1)^2(\beta+1)^3},\\
\frac{\partial t_p^{\mathrm{DD}*}}{\partial\beta} &= -\frac{\theta\gamma(u + w_s - \lambda c_f)}{2k(\beta+1)^2} < 0.
\end{aligned} \tag{A.16}$$

Let $f(k) = (\beta-1)^2 (k(\beta+1)(a-w_s) + \theta^2\gamma^2 (u+w_s-\lambda c_f)) - ku(\beta+1)(\beta^2+1)$, $f'(k) = (\beta+1)((\beta-1)^2(a-w_s) - u(\beta^2-1)) > 0$, we have $f(k) > 0$ because $k > 0$ and $f(k=0) = \theta^2\gamma^2(\beta-1)^2(u+w_s-\lambda c_f) > 0$; thus, $(\partial q_s^{DD*}/\partial\beta) = -(((\beta-1)^2 (k(\beta+1)(a-w_s) + \theta^2\gamma^2(u+w_s-\lambda c_f)) - ku(\beta+1)(\beta^2+1))/(2k(\beta-1)^2(\beta+1)^3)) < 0$. □

Proof of Proposition 3. From equation (6), the first-order partial derivatives of q_p^{CS} and q_s^{CS} can be shown as

$$\frac{\partial \Pi_R^{CS}}{\partial q_p} = a - 2q_p - u - w_s - 2\beta q_s + \theta\gamma t_p = 0, \quad (A.17)$$

$$\frac{\partial \Pi_R^{CS}}{\partial q_s} = a - 2q_s - w_s - 2\beta q_p + \theta\gamma t_p = 0. \quad (A.18)$$

From the formula (A.17) and (A.18), we get

$$q_p^{CS} = \frac{-a + u + w_s + a\beta - \beta w_s - \theta\gamma t_p + \theta\beta\gamma t_p}{2(\beta^2-1)}, \quad (A.19)$$

$$q_s^{CS} = \frac{a\beta - a + u + \theta\beta\gamma t_p - \beta w_s - \beta u - \theta\gamma t_p}{2(\beta^2-1)}. \quad (A.20)$$

Taking (A.19) into equation 5 and calculating the partial derivatives of t_p, we have

$$t_p^{CS} = \frac{\theta\gamma(u + w_s - \lambda c_f)}{-2k(\mu-1)(\beta+1)}. \quad (A.21)$$

Similarly, taking (A.19) and (A.21) into Π_C and calculating the partial derivatives of u, we have

$$\mu^{CS^*} = \frac{\theta^2\gamma^2(u + w_s - \lambda c_f) + k(\beta+1)(2a - u - 2w_s)}{k(\beta+1)(2a - w_s - \lambda c_f)}, \quad (A.22)$$

where $\mu^{CS^*} \in [0,1]$.

Taking (A.22) into (A.19) and (A.21), we get

$$t_p^{CS*} = \frac{\theta\gamma(2a - w_s - \lambda c_f)}{2K_1}, \quad (A.23)$$

$$q_p^{CS*} = \frac{2k(\beta+1)((1-\beta)(a-w_s) - u) + \theta^2\gamma^2(2u + (1-\beta)(w_s - \lambda c_f))}{4K_1(1-\beta^2)}, \quad (A.24)$$

$$q_s^{CS*} = \frac{2k(\beta+1)((1-\beta)(a-w_s) + u\beta) - \theta^2\gamma^2((\beta-1)(w_s - \lambda c_f) + 2u\beta)}{4K_1(1-\beta^2)}. \quad (A.25)$$

□

Proof of Corollary 3. From the formula (A.22)–(A.25), the first-order partial derivative of c_f, γ, k, and β can be shown as

$$\frac{\partial q_P^{CS*}}{\partial c_f} = -\frac{\theta^2\gamma^2\lambda}{4K_1(\beta+1)} < 0,$$

$$\frac{\partial q_s^{CS*}}{\partial c_f} = -\frac{\theta^2\gamma^2\lambda}{4K_1(\beta+1)} < 0,$$

$$\frac{\partial t_p^{CS*}}{\partial c_f} = -\frac{\theta\gamma\lambda}{2K_1} < 0,$$

$$\begin{aligned}
\frac{\partial \mu^{CS*}}{\partial c_f} &= \frac{\lambda K_1 (2a-u-2w_s)}{k(\beta+1)\left(2a-w_s-\lambda c_f\right)^2} > 0,\\
\frac{\partial q_p^{CS*}}{\partial \gamma} &= \frac{\theta^2 k\gamma\left(2a-w_s-\lambda c_f\right)}{2\left(-k+\theta^2\gamma^2-k\beta\right)^2} > 0,\\
\frac{\partial q_s^{CS*}}{\partial \gamma} &= \frac{\theta^2 k\gamma\left(2a-w_s-\lambda c_f\right)}{2\left(-k+\theta^2\gamma^2-k\beta\right)^2} > 0,\\
\frac{\partial t_p^{CS*}}{\partial \gamma} &= \frac{\theta\left(k+\theta^2\gamma^2+k\beta\right)\left(2a-w_s-\lambda c_f\right)}{2\left(-k+\theta^2\gamma^2-k\beta\right)^2} > 0,\\
\frac{\partial \mu^{CS*}}{\partial \gamma} &= \frac{2\theta^2\gamma\left(u+w_s-\lambda c_f\right)}{k(\beta+1)\left(2a-w_s-\lambda c_f\right)} > 0,\\
\frac{\partial q_p^{CS*}}{\partial k} &= -\frac{\theta^2\gamma^2\left(2a-w_s-\lambda c_f\right)}{4\left(-k+\theta^2\gamma^2-k\beta\right)^2} < 0,\\
\frac{\partial q_s^{CS*}}{\partial k} &= -\frac{\theta^2\gamma^2\left(2a-w_s-\lambda c_f\right)}{4\left(-k+\theta^2\gamma^2-k\beta\right)^2} < 0,\\
\frac{\partial t_p^{CS*}}{\partial k} &= -\frac{\theta\gamma(\beta+1)\left(2a-w_s-\lambda c_f\right)}{2\left(-k+\theta^2\gamma^2-k\beta\right)^2} < 0,\\
\frac{\partial \mu^{CS*}}{\partial k} &= \frac{\theta^2\gamma^2\left(u+w_s-\lambda c_f\right)}{k^2(\beta+1)\left(w_s+\lambda c_f-2a\right)} < 0;\\
\frac{\partial q_p^{CS*}}{\partial \beta} &= -\frac{2k^2(\beta+1)^2\left((\beta-1)^2(a-w_s)+2u\beta\right)+\theta^2\gamma^2\left(-2k-2k\beta+\theta^2\gamma^2\right)\left((\beta-1)^2\left(\lambda c_f-w_s\right)+4u\beta\right)}{4K_1^2\left(\beta^2-1\right)^2}.
\end{aligned} \tag{A.26}$$

Let $f(k) = 2k^2(\beta+1)^2\ ((\beta-1)^2(a-w_s)\ +2u\beta)+\theta^2\gamma^2\ (-2k\ -2k\beta+\theta^2\gamma^2)((\beta-1)^2(\lambda c_f-w_s)+4u\beta)$, $f'(k) = 4(\beta+1)^2\ ((\beta-1)^2(a-w_s)+2u\beta)>0$, $(1-\beta)(a-w_s +\theta\gamma K_2(u+w_s-\lambda c_f))\geq u$ and $f(k=(\theta^2\gamma^2\ /2\beta)) = \theta^4\gamma^4(\beta-1)^2\ (((\beta-1)(a-w_s\ +a\beta-\beta\lambda c_f) -2u\beta)/2\beta^2)>0$, thus $f(k)>0$.

Similarly,

$$\begin{aligned}
\frac{\partial q_s^{CS*}}{\partial \beta} &= \frac{-2k^2(\beta+1)^2\left((\beta-1)^2(a-w_s)-u\left(\beta^2+1\right)\right)+\theta^2\gamma^2\left(-2k+\theta^2\gamma^2-2k\beta\right)\left(2u\left(\beta^2+1\right)+(\beta-1)^2\left(w_s-\lambda c_f\right)\right)}{4\left(\beta^2-1\right)^2\left(k+k\beta-\theta^2\gamma^2\right)^2} < 0,\\
\frac{\partial t_p^{CS*}}{\partial \beta} &= \frac{\theta\gamma k\left(w_s-2a+\lambda c_f\right)}{2\left(k+k\beta-\theta^2\gamma^2\right)^2} < 0;\\
\frac{\partial \mu^{CS*}}{\partial \beta} &= \frac{\theta^2\gamma^2\left(u+w_s-\lambda c_f\right)}{k(\beta+1)^2\left(w_s+\lambda c_f-2a\right)} < 0.
\end{aligned} \tag{A.27}$$

□

Proof of Proposition 4. From the equation (10), the first-order partial derivatives of q_p and q_s can be shown as follows:

$$q_p^{\mathrm{QC}} = \frac{(\beta-1)\left(a - w_s + \theta\gamma t_p\right) + u - \tau}{2\left(\beta^2-1\right)}, \quad \text{(A.28)}$$

$$q_s^{\mathrm{QC}} = \frac{(\beta-1)\left(a - w_s + \theta\gamma t_p\right) + \beta(\tau - u)}{2\left(\beta^2-1\right)}. \quad \text{(A.29)}$$

Taking (A.28) and (A.29) into (9) and calculating the first-order partial derivatives of t_p, τ, we have

$$\tau = u + w_s + \beta\lambda c_f, \quad \text{(A.30)}$$

$$t_p^{\mathrm{QC}*} = \frac{\theta\gamma\left(2a - w_s - \lambda c_f\right)}{2K_1}. \quad \text{(A.31)}$$

We take (A.30) and (A.31) into (A.28) and (A.29) the equilibrium solutions are as follows:

$$q_p^{\mathrm{QC}*} = \frac{2ak(1-\beta) + \left(w_s + \lambda c_f\right)\left(2k\beta - \theta^2\gamma^2\right)}{4K_1(1-\beta)}, \quad \text{(A.32)}$$

$$q_s^{\mathrm{QC}*} = \frac{2ak(1-\beta) - \lambda c_f\left(2k\beta^2 - \theta^2\gamma^2(2\beta-1)\right) - w_s\left(2k - \theta^2\gamma^2\right)}{4K_1(1-\beta)}. \quad \text{(A.33)}$$

□

Proof of Corollary 4. From the equation (A.31)–(A.33), the partial derivative of c_f, γ, k, and β, we get

$$\begin{aligned}
\frac{\partial q_p^{\mathrm{QC}*}}{\partial c_f} &= -\frac{\lambda\left(\theta^2\gamma^2 - 2k\beta\right)}{4(\beta-1)\left(-k + \theta^2\gamma^2 - k\beta\right)} > 0,\\
\frac{\partial q_s^{\mathrm{QC}*}}{\partial c_f} &= \frac{\left(-\theta^2\gamma^2 - 2k\beta^2 + 2\theta^2\gamma^2\beta\right)\lambda}{4(\beta-1)\left(-k + \theta^2\gamma^2 - k\beta\right)} < 0,\\
\frac{\partial t_p^{\mathrm{QC}*}}{\partial c_f} &= \frac{\theta\lambda\gamma}{2\left(-k + \theta^2\gamma^2 - k\beta\right)} < 0;\\
\frac{\partial q_p^{\mathrm{QC}*}}{\partial \gamma} &= \frac{\theta^2\gamma k\left(2a - w_s - \lambda c_f\right)}{2\left(-k + \theta^2\gamma^2 - k\beta\right)^2} > 0,\\
\frac{\partial q_s^{\mathrm{QC}*}}{\partial \gamma} &= \frac{\theta^2\gamma k\left(2a - w_s - \lambda c_f\right)}{2\left(-k + \theta^2\gamma^2 - k\beta\right)^2} > 0,\\
\frac{\partial t_p^{\mathrm{QC}*}}{\partial \gamma} &= \frac{\theta\left(k + \theta^2\gamma^2 + k\beta\right)\left(2a - w_s - \lambda c_f\right)}{2\left(-k + \theta^2\gamma^2 - k\beta\right)^2} > 0;\\
\frac{\partial q_p^{\mathrm{QC}*}}{\partial k} &= -\frac{\theta^2\gamma^2\left(2a - w_s - \lambda c_f\right)}{4\left(-k + \theta^2\gamma^2 - k\beta\right)^2} < 0,\\
\frac{\partial q_s^{\mathrm{QC}*}}{\partial k} &= -\frac{\theta^2\gamma^2\left(2a - w_s - \lambda c_f\right)}{4\left(-k + \theta^2\gamma^2 - k\beta\right)^2} < 0,\\
\frac{\partial t_p^{\mathrm{QC}*}}{\partial k} &= -\frac{\theta\gamma(\beta+1)\left(2a - w_s - \lambda c_f\right)}{2\left(-k + \theta^2\gamma^2 - k\beta\right)} < 0;\\
\frac{\partial q_p^{\mathrm{QC}*}}{\partial \beta} &= -\frac{2ak^2(\beta-1)^2 - \left(2k^2\left(\beta^2+1\right) + \theta^2\gamma^2\left(-2k + \theta^2\gamma^2 - 2k\beta\right)\right)\left(w_s + \lambda c_f\right)}{4(\beta-1)^2\left(-k + \theta^2\gamma^2 - k\beta\right)^2}.
\end{aligned} \quad \text{(A.34)}$$

Let $f(k) = 2ak^2(\beta-1)^2 - (2k^2(\beta^2+1) + \theta^2\gamma^2(-2k+\theta^2\gamma^2 - 2k\beta))(w_s + \lambda c_f)$, we have $f''(k) = 4(a(\beta-1)^2 - (\beta^2+1)(w_s+\lambda c_f)) > 0$,

$$f'\left(k = \frac{\theta^2\gamma^2}{\beta+1}\right) = 2\theta^2\gamma^2(\beta-1)^2 \frac{2a - w_s - \lambda c_f}{\beta+1} > 0,$$
$$f\left(k = \frac{\theta^2\gamma^2}{\beta+1}\right) = \theta^4\gamma^4(\beta-1)^2 \frac{2a - w_s - \lambda c_f}{(\beta+1)^2} > 0, \tag{A.35}$$

and thus, $(\partial q_p^{QC*}/\partial\beta) < 0$.

$$\frac{\partial q_s^{QC*}}{\partial\beta} = -\frac{2ak^2(\beta-1)^2 + (\theta^2\gamma^2 - 2k)(\theta^2\gamma^2 - 2k\beta)(w_s+\lambda c_f)}{4(\beta-1)^2(-k+\theta^2\gamma^2 - k\beta)^2} < 0,$$
$$\frac{\partial t_p^{QC*}}{\partial\beta} = -\theta k\gamma \frac{2a - w_s - \lambda c_f}{2(-k+\theta^2\gamma^2 - k\beta)^2} < 0. \tag{A.36}$$

□

Proof of Corollary 5. From (A.7), (A.13), (A.23) and (A.31), we get $t_p^{QC*} = t_p^{CS*} = t_p^{CD*}$,

$$t_p^{CD*} - t_p^{DD*} = \theta\gamma \frac{k(\beta+1)(2a-u-2w_s) + \theta^2\gamma^2(u + w_s - \lambda c_f)}{2kK_1(\beta+1)} > 0, \tag{A.37}$$

and thus, $t_p^{CD*} = t_p^{QC*} = t_p^{CS*} > t_p^{DD*}$. □

Proof of Corollary 6. From (A.5), (A.14), (A.24) and (A.32), we have $q_p^{CS*} - q_p^{QC*} = ((u + w_s + \beta\lambda c_f)/2(\beta^2-1)) < 0$ and $q_p^{CS*} - q_p^{DD*} = \theta^2\gamma^2((k(\beta+1)(2a-u-2w_s) + \theta^2\gamma^2(u + w_s - \lambda c_f))/(4K_1(\beta+1)^2)) > 0$.

Hence, $q_p^{QC*} > q_p^{DD*}$ and $q_p^{QC*} = q_p^{CD*}$. □

Proof of Corollary 7. From (A.6), (A.15), (A.25), and (A.33), we have $q_s^{CS*} - q_s^{QC*} = \beta((u + w_s + \beta\lambda c_f)/2(1-\beta^2)) > 0$,

$$q_s^{CS*} - q_s^{DD*} = \theta^2\gamma^2 \frac{k(\beta+1)(2a-u-2w_s) + \theta^2\gamma^2(u + w_s - \lambda c_f)}{4kK_1(\beta+1)^2} > 0, \tag{A.38}$$

and $q_s^{QC*} - q_s^{DD*} = ((2k^2\beta(\beta+1)^2(u + w_s + \beta\lambda c_f) + f(\theta^2\gamma^2))/(4K_1 k(\beta-1)(\beta+1)^2))$.

Let $f(\theta^2\gamma^2) = \theta^2\gamma^2[k(\beta+1)(2a(\beta-1) - u(3\beta-1) - 2w_s(2\beta-1) - 2\beta^2\lambda c_f) + \theta^2\gamma^2(u + w_s - \lambda c_f)(\beta-1)]$ and $g(\theta^2\gamma^2) = 2k^2\beta(\beta+1)^2(u + w_s + \beta\lambda c_f) + f(\theta^2\gamma^2)$, $g'(\theta^2\gamma^2) < 0$, $g(0) = 2k^2\beta(\beta+1)^2(u + w_s + \beta\lambda c_f) > 0$ and $g(k(\beta+1)) = k^2(\beta-1)(\beta+1)^2(2a - w_s - \lambda c_f) < 0$.

Hence, there must be $\overline{\varphi} > 0$ that makes $g(\overline{\varphi}) = 0$ and $\overline{\varphi} = g^{-1}(0)$. Especially, $g(\theta^2\gamma^2) > 0$ when $\theta^2\gamma^2 \in (0, \overline{\varphi}]$ and $g(\theta^2\gamma^2) < 0$ when $\theta^2\gamma^2 \in (\overline{\varphi}, k(\beta+1)]$.

Therefore, when $0<\theta^2\gamma^2<\overline{\varphi}$, $q_s^{\mathrm{QC}*}<q_s^{\mathrm{DD}*}$; when $\overline{\varphi}<\theta^2\gamma^2<k(\beta+1)$. $q_s^{\mathrm{DD}*}<q_s^{\mathrm{QC}*}$. □

Proof of Corollary 8. From (A.6), (A.15), (A.25), and (A.33), we have

$$
\begin{aligned}
r_s^{\mathrm{CD}*} &= \frac{q_s^{\mathrm{CD}*}}{q_s^{\mathrm{CD}*}+q_p^{\mathrm{CD}*}} = \frac{2ak(\beta-1)+(w_s+\lambda c_f)(2k-\theta^2\gamma^2)}{2k(\beta-1)(2a-w_s-\lambda c_f)},\\
r_s^{\mathrm{DD}*} &= \frac{q_s^{\mathrm{DD}*}}{q_s^{\mathrm{DD}*}+q_p^{\mathrm{DD}*}} = \frac{2ak(1-\beta^2)+2k(\beta+1)(-w_s+u\beta+\beta w_s)+\theta^2\gamma^2(1-\beta)(u+w_s-\lambda c_f)}{2(1-\beta)(k(\beta+1)(2a-u-2w_s)+\theta^2\gamma^2(u+w_s-\lambda c_f))},\\
r_s^{\mathrm{CS}*} &= \frac{q_s^{\mathrm{CS}*}}{q_s^{\mathrm{CS}*}+q_p^{\mathrm{CS}*}} = \frac{2ak(1-\beta^2)+2k(\beta+1)(-w_s+u\beta+\beta w_s)-\theta^2\gamma^2(2u\beta+(\beta-1)(w_s-\lambda c_f))}{(1-\beta)(k(\beta+1)(2a-u-2w_s)+\theta^2\gamma^2(u+w_s-\lambda c_f))},\\
r_s^{\mathrm{QC}*} &= \frac{q_s^{\mathrm{QC}*}}{q_s^{\mathrm{QC}*}+q_p^{\mathrm{QC}*}} = \frac{2ak(1-\beta)-2k(w_s+\beta^2\lambda c_f)+\theta^2\gamma^2(w_s-\lambda c_f+2\beta\lambda c_f)}{2(1-\beta)(-kw_s+2ak-\theta^2\gamma^2\lambda c_f+k\beta\lambda c_f)},\\
r_s^{\mathrm{DD}*}-r_s^{\mathrm{CS}*} &= \theta^2\gamma^2 u\frac{\beta+1}{2(1-\beta)(k(\beta+1)(2a-u-2w_s)+\theta^2\gamma^2(u+w_s-\lambda c_f))}>0,\\
r_s^{\mathrm{CS}*}-r_s^{\mathrm{QC}*} &= \frac{K_1(2k(\beta+1)(a-w_s)+\theta^2\gamma^2(w_s-\lambda c_f))(u+w_s+\beta\lambda c_f)}{(1-\beta)(k(2a-w_s)+\lambda c_f(k\beta-\theta^2\gamma^2))(k(\beta+1)(2a-u-2w_s)+\theta^2\gamma^2(u+w_s-\lambda c_f))}>0.
\end{aligned}
\tag{A.39}
$$

Thus, we get $r_s^{\mathrm{DD}*}>r_s^{\mathrm{CS}*}>r_s^{\mathrm{QC}*}>r_s^{\mathrm{CD}*}$ and similarly, $r_p^{\mathrm{SC}*}>r_p^{\mathrm{QC}*}>r_p^{\mathrm{CS}*}>r_p^{\mathrm{DD}*}$. □

Proof of Corollary 9. From the above proposition, we have

$$
\begin{aligned}
\Pi_p^{\mathrm{CS}*} &= \frac{4k(\beta+1)[a(u-\lambda c_f)(\beta-1)+u(u+\beta\lambda c_f)]}{8K_1(\beta^2-1)},\\
\Pi_R^{\mathrm{CS}*} &= \frac{2k(\beta+1)[2a(\beta-1)(a-u)-u^2]+\theta^2\gamma^2[2u(a\beta-a+u)+\lambda c_f(\beta-1)(u-2a)]}{8K_1(\beta^2-1)},\\
\Pi_p^{\mathrm{DD}*} &= \frac{au(1-\beta)-u^2-\lambda c_f(a-a\beta+\beta u)-K_2(\lambda c_f-u)^2(\beta-1)(\theta\gamma-kK_2-k\beta K_2)}{2(1-\beta^2)},\\
\Pi_R^{\mathrm{DD}*} &= \frac{2a(\beta-1)(a-u)-u^2-2\theta\gamma K_2(\beta-1)(\lambda c_f-u)[2a-u+\theta\gamma K_2(u-\lambda c_f)]}{4(\beta^2-1)}.
\end{aligned}
\tag{A.40}
$$

$\Pi_p^{\mathrm{CS}*}-\Pi_p^{\mathrm{DD}*} = ((\theta^2\gamma^2(u-\lambda c_f)[k(\beta+1)(2a-u)+\theta^2\gamma^2(u-\lambda c_f)])/(8kK_1(\beta+1)^2))>0$ because of $2k\beta>\theta^2\gamma^2$ and $k\geq((\theta\gamma(2a-w_s-\lambda c_f)+2\theta^2\gamma^2)/2(1+\beta))$. Similarly, $\Pi_R^{\mathrm{CS}*}-\Pi_R^{\mathrm{DD}*} = (\theta^2\gamma^2[k(\beta+1)(2a-2u+\lambda c_f)+\theta^2\gamma^2(w_p-\lambda c_f)][k(\beta+1)(2a-u)+\theta^2\gamma^2(u-\lambda c_f)])/(8K_1k^2(\beta+1)^3)$.

Let $f(k)=k(\beta+1)(2a-u)+\theta^2\gamma^2(u-\lambda c_f)$, $(\partial f(k)/\partial k) = (\beta+1)(2a-u)>0$ and $k>(\theta^2\gamma^2/(\beta+1))$, which show that $f(k=(\theta^2\gamma^2/(\beta+1)))=\theta^2\gamma^2(2a-\lambda c_f)>0$ and $f(k)=k(\beta+1)(2a-u)+\theta^2\gamma^2(u-\lambda c_f)>0$, thus $\Pi_R^{\mathrm{CS}*}>\Pi_R^{\mathrm{DD}*}$. □

Proof of Corollary 10. From profit of platform and retailer, we have

$$
\begin{aligned}
\Pi_C^{\mathrm{CD}*} &= \frac{4ak(1-\beta)\left(a-\lambda c_f\right)+\lambda^2 c_f^2\left(2k-\theta^2\gamma^2\right)}{8(1-\beta)K_1},\\
\Pi_C^{\mathrm{CD}*} &= \frac{u\left(u+2\beta\lambda c_f\right)+2a(\beta-1)\left(a-\lambda c_f\right)-2K_2\left(\lambda c_f-u\right)(\beta-1)\left[\theta\gamma\left(2a-\lambda c_f\right)+K_2K_1\left(\lambda c_f-u\right)\right]}{4\left(\beta^2-1\right)},\\
\Pi_C^{\mathrm{CS}*} &= \frac{2K_1u\left(u+2\beta\lambda c_f\right)+\theta^2\lambda^2\gamma^2c_f^2(\beta-1)+4ak\left(\beta^2-1\right)\left(a-\lambda c_f\right)}{8K_1\left(\beta^2-1\right)},\\
\Pi_C^{\mathrm{QC}*} &= \frac{4ak(\beta-1)\left(a-\lambda c_f\right)+\lambda^2 c_f^2\left(-2k\beta^2+\theta^2\gamma^2(2\beta-1)\right)}{8(\beta-1)K_1},\\
\Pi_C^{CS*}-\Pi_C^{D\,D\,*} &= \frac{\theta^2\gamma^2\left(-ku+2ak+2ak\beta+\theta^2\gamma^2u-k\beta u-\theta^2\gamma^2\lambda c_f\right)^2}{8k^2(\beta+1)^3K_1}>0,\\
\Pi_C^{\mathrm{CS}*}-\Pi_C^{\mathrm{QC}*} &= \frac{\left(u+\beta\lambda c_f\right)^2}{4\left(\beta^2-1\right)}<0,\\
\Pi_C^{\mathrm{QC}*}-\Pi_C^{\mathrm{CD}*} &= -\frac{1}{4}\lambda^2c_f^2<0.
\end{aligned}
\tag{A.41}
$$

Hence, $\Pi_C^{\mathrm{CD}*}>\Pi_C^{\mathrm{QC}*}>\Pi_C^{\mathrm{CS}*}>\Pi_C^{\mathrm{DD}*}$. □

Proof of Theorem 1. From equations (9) and (10), we have

$$
\prod_p^{\mathrm{QC}*} = \frac{8TuK_1^2-4a^2\theta^2k\gamma^2+\lambda c_f\left[\theta^2\gamma^2\left[4ak(\beta+2)+\lambda c_f\left(K_1+k\beta-\theta^2\gamma^2\right)\right]-4ak^2(\beta+1)^2+8T\beta K_1^2\right]}{8K_1^2},
\tag{A.42}
$$

$$
\prod_R^{\mathrm{QC}*} = \frac{4(\beta-1)\left(-a^2k^2(\beta+1)+2TuK_1^2\right)+\lambda c_f\left[8T\beta(\beta-1)K_1^2+4ak\left(\beta^2-1\right)\left(\theta^2\gamma^2-k\beta\right)+\lambda c_f\left[\theta^4\gamma^4-2k\beta(\beta+1)\left(\theta^2\gamma^2-k\beta\right)\right]\right]}{8K_1^2(1-\beta)}.
\tag{A.43}
$$

From (A.42) and (A.43), the partial derivative of equilibrium with T can be shown as $(\partial\Pi_p^{\mathrm{QC}*}/\partial T)=u+\beta\lambda c_f>0$, and $(\partial\Pi_R^{\mathrm{QC}*}/\partial T)=-(u+\beta\lambda c_f)<0$.

Let $\Pi_p^{\mathrm{QC}}(T=T_1)=0$, $\Pi_R^{\mathrm{QC}}(T=T_4)=0$, we have

$$
\begin{aligned}
T_1 &= \frac{4ka^2\theta^2\gamma^2-\lambda c_f\left(-4ak^2(\beta+1)^2+\theta^2\gamma^2\left(\lambda c_f\left(K_1-\theta^2\gamma^2+k\beta\right)+4ak(\beta+2)\right)\right)}{8K_1^2\left(u+\beta\lambda c_f\right)},\\
T_4 &= \frac{4a^2k^2\left(\beta^2-1\right)-\lambda c_f\left[\lambda c_f\left(\theta^4\gamma^4-2k\beta\left(\theta^2\gamma^2-k\beta\right)(\beta+1)\right)+4ak\left(\theta^2\gamma^2-k\beta\right)\left(\beta^2-1\right)\right]}{8K_1^2(\beta-1)\left(u+\beta\lambda c_f\right)}.
\end{aligned}
\tag{A.44}
$$

$T_4 - T_1 = -((4ak(\beta-1)(a-\lambda c_f) + \lambda^2 c_f^2 [2\beta(\theta^2\gamma^2 - k\beta) - \theta^2\gamma^2])/(8K_1(1-\beta)(u+\beta\lambda c_f))) > 0$, where $T \in (T_1, T_4)$ because of $\Pi_P^{QC*} > 0$ and $\Pi_R^{QC*} > 0$.

Similarly, let $\Pi_P^{QC}(T = T_2) = \Pi_P^{DC}$, $\Pi_R^{QC}(T = T_3) = \Pi_R^{DC}$, we have $T_2 = ((R_2 + R_3)/(w_p + \beta\lambda c_f))$, $T_3 = -((R_4 + R_5)/(u + \beta\lambda c_f))$, where

$$R_2 = \frac{1}{2(\beta^2-1)}\left(u\left(u+\beta\lambda c_f\right) + \frac{1}{2}\left(u-\lambda c_f\right)(\beta-1)\left(2a+\theta\gamma\varphi\left(u-\lambda c_f\right)\right)\right),$$

$$R_3 = -\frac{1}{8K_1^2}\left(\lambda c_f\left(-4ak^2(\beta+1)^2 + \theta^2\gamma^2\left(\lambda c_f\left(K_1 - \theta^2\gamma^2 + k\beta\right) + 4ak(\beta+2)\right)\right) - 4a^2\theta^2\gamma^2 k\right),$$

$$R_4 = \frac{1}{4(\beta^2-1)}\left(-u^2 + 2a(\beta-1)(a-u) + 2\theta\gamma\varphi(\beta-1)\left(u-\lambda c_f\right)\left(2a-u+\theta\gamma\varphi\left(u-\lambda c_f\right)\right)\right), \tag{A.45}$$

$$R_5 = \frac{\lambda c_f\left(\lambda c_f\left(\theta^4\gamma^4 - 2k\beta\left(\theta^2\gamma^2-k\beta\right)(\beta+1)\right) + 4ak\left(\theta^2\gamma^2-k\beta\right)\left(\beta^2-1\right)\right) - 4a^2k^2\left(\beta^2-1\right)}{8K_1^2(\beta-1)},$$

$$T_2 - T_3 = \frac{8a\theta\gamma\varphi K_1(\beta-1)\left(uK_1 - ak(\beta+1) + \theta^2\gamma^2\lambda c_f\right) + 2K_1^2u^2(\theta\gamma\varphi(\beta-1)(2\theta\gamma\varphi-1)+1) + R_6}{8K_1^2\left(\beta^2-1\right)\left(u+\beta\lambda c_f\right)},$$

where $R_6 = \lambda c_f(-4K_1^2 u(2\theta^2\gamma^2\varphi^2(\beta-1)-\beta) + \lambda c_f K_1(2\beta^2 K_1 - 4\varphi^2\theta^4\gamma^4(\beta-1))) > 0$. We have $T_2 < T_3$.

Let $f(T) = \Pi_R^{CS*} - \Pi_R^{QC*}$, we have $f'(T) = u + \beta\lambda c_f > 0$, and when $f(T) = 0$, $T_6 = -((R_7 + R_8)/(u + \beta\lambda c_f))$, where

$$R_7 = \frac{4ak\left(\beta^2-1\right)\left(ak - \lambda c_f\left(\theta^2\gamma^2-k\beta\right)\right) - \lambda^2c_f^2\left(\theta^4\gamma^4 - 2k\beta(\beta+1)\left(\theta^2\gamma^2-k\beta\right)\right)}{8(1-\beta)K_1^2},$$

$$R_8 = \frac{2u^2K_1 - (\beta-1)\left(4ak(\beta+1)(a-u) + \theta^2\gamma^2\left(2a\left(u-\lambda c_f\right) + \lambda u c_f\right)\right)}{8\left(1-\beta^2\right)K_1}. \tag{A.46}$$

Similarly, let $g(T) = \Pi_P^{CS*} - \Pi_P^{QC*}$, we have $g'(T) = -(u + \beta\lambda c_f) < 0$, and when $g(T) = 0$, $T_5 = -((R_9 + R_{10})/(u + \beta\lambda c_f))$, where

$$R_9 = \frac{\lambda c_f\left(\theta^2\gamma^2\left(\lambda c_f\left(k(2\beta+1) - 2\theta^2\gamma^2\right) + 4ak(2+\beta)\right) - 4ak^2(\beta+1)^2\right) - 4a^2\theta^2\gamma^2 k}{8K_1^2},$$

$$R_{10} = \frac{ak\left(u-\lambda c_f\right)(\beta-1) + uk\left(u+\beta\lambda c_f\right)}{2K_1(1-\beta)},$$

$$T_5 - T_6 = \frac{2a\theta^2\gamma^2\left(\lambda c_f - u\right)(\beta-1) - u\left(2k(\beta+1)\left(u+2\beta\lambda c_f\right) + \theta^2\gamma^2\left(2u+\lambda c_f(\beta-1)\right)\right) + \lambda^2c_f^2(\beta+1)\left(\theta^2\gamma^2(2\beta-1) - 2k\beta^2\right)}{\left(1-\beta^2\right)\left(u+\beta\lambda c_f\right)K_1} < 0. \tag{A.47}$$

□

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Di Xiao acknowledges funding from the Philosophy and Social Science Program of Zhejiang Province (Grant 21NDJC079YB) and the Natural Science Foundation of Zhejiang Province (Grant LY18G020002). Qi Sun thank the financial support of Humanity and Social Science Foundation from the Ministry of Education of China (20YJC630130).

References

[1] P. Kotler and G. Armstrong, *Principles of Marketing*, Pearson, London, UK, 15 edition, 2014.

[2] A. J. H. Thorlakson and R. P. Murray, "An empirical study of empowerment in the workplace," *Group and Organization Management*, vol. 21, no. 1, pp. 67–83, 1996.

[3] S. Lenka, V. Parida, and J. Wincent, "Digitalization capabilities as enablers of value co-creation in servitizing Firms," *Psychology and Marketing*, vol. 34, no. 1, pp. 92–100, 2016.

[4] M. Corstjens and R. Lal, "Building store loyalty through store brands," *Journal of Marketing Research*, vol. 37, no. 3, pp. 281–291, 2000.

[5] S. Fiona, F. Morton, and Zettelmeyer, "The strategic positioning of store brands in retailer-manufacturer negotiations," *Review of Industrial Organization*, 2004.

[6] D. T. Mai, T. Liu, M. D. S. Morris, and S. Sun, "Quality coordination with extended warranty for store-brand products," *European Journal of Operational Research*, vol. 256, no. 2, pp. 524–532, 2017.

[7] W. Li, J. Chen, and B. Chen, "Supply chain coordination with customer returns and retailer's store brand product," *International Journal of Production Economics*, vol. 203, pp. 69–82, 2018.

[8] G. M. Grossman and C. Shapiro, "Comparative advantage in a multi-product and multi-factor model of multilateral trade," *Contributions to Economic Analysis*, vol. 78, no. 1, pp. 59–75, 1988.

[9] G. M. Grossman and C. Shapiro, "Foreign counterfeiting of status goods," *The Quarterly Journal of Economics*, vol. 103, no. 1, pp. 79–100, 1988.

[10] S.-H. Cho, X. Fang, and S. Tayur, "Combating strategic counterfeiters in licit and illicit supply chains," *Manufacturing and Service Operations Management*, vol. 17, no. 3, pp. 273–289, 2015.

[11] C. Fershtman, "Identification of classes of differential games for which the open loop is a degenerate feedback Nash equilibrium," *Journal of Optimization Theory and Applications*, vol. 55, no. 2, pp. 217–231, 1987.

[12] V. P. Rindova, I. O. Williamson, A. P. Petkova, and J. M. Sever, "Being good or being known: an empirical examination of the dimensions, antecedents, and consequences of organizational reputation," *Academy of Management Journal*, vol. 48, no. 6, pp. 1033–1049, 2005.

[13] G. Feichtinger, R. F. Hartl, and S. P. Sethi, "Dynamic optimal control models in advertising: recent developments," *Management Science*, vol. 40, no. 2, pp. 195–226, 1994.

[14] N. Amrouche, G. Martín-Herrán, and G. Zaccour, "Feedback Stackelberg equilibrium strategies when the private label competes with the national brand," *Annals of Operations Research*, vol. 164, no. 1, pp. 79–95, 2008.

[15] P. De Giovanni, "Quality improvement vs. advertising support: which strategy works better for a manufacturer?" *European Journal of Operational Research*, vol. 208, no. 2, pp. 119–130, 2011.

[16] A. Nair and R. Narasimhan, "Dynamics of competing with quality- and advertising-based goodwill," *European Journal of Operational Research*, vol. 175, no. 1, pp. 462–474, 2006.

[17] F. El Ouardighi and F. Pasin, "Quality improvement and goodwill accumulation in a dynamic duopoly," *European Journal of Operational Research*, vol. 175, no. 2, pp. 1021–1032, 2006.

[18] S. L. Ting and A. H. C. Tsang, "Using social network analysis to combat counterfeiting," *International Journal of Production Research*, vol. 52, no. 15, pp. 4456–4468, 2013.

[19] Y. Qian, "Counterfeiters: foes or friends? how counterfeits affect sales by product quality tier," *Management Science*, vol. 60, no. 10, pp. 2381–2400, 2014.

[20] Y. Qian, Q. Gong, and Y. Chen, "Untangling searchable and experiential quality responses to counterfeits," *Marketing Science*, vol. 34, no. 4, pp. 522–538, 2015.

[21] K. Liu, J.-A. Li, Y. Wu, and K. K. Lai, "Analysis of monitoring and limiting of commercial cheating: a newsvendor model," *Journal of the Operational Research Society*, vol. 56, no. 7, pp. 844–854, 2005.

[22] P. E. Chaudhry, A. Zimmerman, J. R. Peters, and V. V. Cordell, "Preserving intellectual property rights: managerial insight into the escalating counterfeit market quandary," *Business Horizons*, vol. 52, no. 1, pp. 57–66, 2009.

[23] G. Häubl and V. Trifts, "Consumer decision making in online shopping environments: the effects of interactive decision aids," *Marketing Science*, vol. 19, no. 1, pp. 4–21, 2000.

[24] M. Q. Saeed, Z. Bilal, and C. D. Walter, "An NFC based consumer-level counterfeit detection framework," in *Proceedings of the Eleventh International Conference on Privacy*, IEEE, Tarragona, Catalonia, Spain, July 2013.

[25] K. L. Webb and J. E. Hogan, "Hybrid channel conflict: causes and effects on channel performance," *Journal of Business and Industrial Marketing*, vol. 17, no. 5, pp. 338–356, 2002.

[26] A. T. Coughlan, E. Anderson, L. W. Stern, and A. I. El-Ansary, *Marketing Channels*, Prentice Hall lnc, Hoboken, NJ, USA, 2001.

[27] L. J. Rosenberg and L. W. Stern, "Toward the analysis of conflict in distribution channels: a descriptive model," *Journal of Marketing*, vol. 34, no. 4, pp. 40–46, 1970.

[28] M. Etgar, "Sources and types of intrachannel conflict," *Journal of Retailing*, vol. 55, no. 01, pp. 61–78, 1976.

[29] J. R. Brown and R. L. Day, "Measures of manifest conflict in distribution channels," *Journal of Marketing Research*, vol. 18, no. 3, pp. 263–274, 1981.

[30] M. Duarte and G. Davies, "Testing the conflict-performance assumption in business-to-business relationships," *Industrial Marketing Management*, vol. 32, no. 2, pp. 91–99, 2003.

[31] S. A. Samaha, R. W. Palmatier and R. P. Dant, Poisoning relationships: perceived unfairness in channels of distribution," *Journal of Marketing*, vol. 75, no. 3, pp. 99–117, 2011.

[32] S. J. Skinner, J. B. Gassenheimer, and S. W. Kelley, "Cooperation in supplier-dealer relations," *Journal of Retailing*, vol. 68, no. 2, pp. 174–193, 1992.

[33] J. R. Brown, R. F. Lusch, and L. P. Smith, "Conflict and satisfaction in an industrial channel of distribution," *International Journal of Physical Distribution and Logistics Management*, vol. 21, no. 6, pp. 15–26, 2016.

[34] W. Baker, M. Marn, and C. Zawada, "Price smarter on the net," *Harvard Business Review*, vol. 79, no. 2, pp. 122–157, 2001.

[35] K. L. Webb, "Managing channels of distribution in the age of electronic commerce," *Industrial Marketing Management*, vol. 31, no. 2, pp. 95–102, 2002.

[36] E. Brynjolfsson, Y. J. Hu, and M. D. Smith, "From niches to riches: anatomy of the long tail," *Social Science Electronic Publishing*, vol. 47, no. 4, pp. 67–71, 2006.

[37] D.-Q. Yao, X. Yue, X. Wang, and J. J. Liu, "The impact of information sharing on a returns policy with the addition of a direct channel," *International Journal of Production Economics*, vol. 97, no. 2, pp. 196–209, 2005.

[38] Y. Liang and X. Sun, "Product green degree, service free-riding, strategic price difference in a dual-channel supply chain based on dynamic game," *Optimization*, vol. 1, pp. 1–42, 2020.

[39] T.-H. Chen, "Effects of the pricing and cooperative advertising policies in a two-echelon dual-channel supply chain," *Computers and Industrial Engineering*, vol. 87, pp. 250–259, 2015.

[40] S. K. Mukhopadhyay, D.-Q. Yao, and X. Yue, "Information sharing of value-adding retailer in a mixed channel hi-tech supply chain," *Journal of Business Research*, vol. 61, no. 9, pp. 950–958, 2008.

[41] L. Zhang and J. Wang, "Coordination of the traditional and the online channels for a short-life-cycle product," *European Journal of Operational Research*, vol. 258, pp. 639–651, 2017.

[42] Y. Du, M. Cui, and J. Su, "Implementation processes of online and offline channel conflict management strategies in manufacturing enterprises: a resource orchestration perspective," *International Journal of Information Management*, vol. 39, pp. 136–145, 2018.

[43] A. A. Tsay and N. Agrawal, "Channel conflict and coordination in the e-commerce age," *Production and Operations Management*, vol. 13, no. 1, pp. 93–110, 2009.

[44] G. Xu, B. Dan, X. Zhang, and C. Liu, "Coordinating a dual-channel supply chain with risk-averse under a two-way revenue sharing contract," *International Journal of Production Economics*, vol. 147, pp. 171–179, 2014.

[45] J. Xie, W. Zhang, L. Liang, Y. Xia, J. Yin, and G. Yang, "The revenue and cost sharing contract of pricing and servicing policies in a dual-channel closed-loop supply chain," *Journal of Cleaner Production*, vol. 191, pp. 361–383, 2018.

[46] I. Biswas and B. Avittathur, "Channel coordination using options contract under simultaneous price and inventory competition," *Transportation Research Part E: Logistics and Transportation Review*, vol. 123, pp. 45–60, 2019.

[47] A. Ahmadi-Javid and P. Hoseinpour, "On a cooperative advertising model for a supply chain with one manufacturer and one retailer," *European Journal of Operational Research*, vol. 219, no. 2, pp. 458–466, 2012.

[48] Y.-C. Tsao and G.-J. Sheen, "Effects of promotion cost sharing policy with the sales learning curve on supply chain coordination," *Computers and Operations Research*, vol. 39, no. 8, pp. 1872–1878, 2012.

[49] M. Leng and M. Parlar, "Game-theoretic analyses of decentralized assembly supply chains: non-cooperative equilibria vs. coordination with cost-sharing contracts," *European Journal of Operational Research*, vol. 204, no. 1, pp. 96–104, 2010.

[50] T. Chakraborty, S. S. Chauhan, and M. Ouhimmou, "Cost-sharing mechanism for product quality improvement in a supply chain under competition," *International Journal of Production Economics*, vol. 208, pp. 566–587, 2019.

[51] Y.-W. Zhou, J. Guo, and W. Zhou, "Pricing/service strategies for a dual-channel supply chain with free riding and service-cost sharing," *International Journal of Production Economics*, vol. 196, pp. 198–210, 2018.

[52] G. P. Cachon and M. A. Lariviere, "Capacity choice and allocation: strategic behavior and supply chain performance," *Management Science*, vol. 45, no. 8, pp. 1091–1108, 1999.

[53] Y. Zhao and M. N. Katehakis, "On the structure of optimal ordering policies for stochastic inventory systems with minimum order quantity," *Probability in the Engineering and Informational Sciences*, vol. 20, no. 2, pp. 257–270, 2006.

[54] J. Hou, A. Z. Zeng, and L. Sun, "Backup sourcing with capacity reservation under uncertain disruption risk and minimum order quantity," *Computers and Industrial Engineering*, vol. 103, pp. 216–226, 2016.

[55] E. Porras and R. Dekker, "An efficient optimal solution method for the joint replenishment problem with minimum order quantities," *European Journal of Operational Research*, vol. 174, no. 3, pp. 1595–1615, 2006.

[56] W. A. Feng, C. Adb, and D. Lw, "Supply chain coordination with competing suppliers under price-sensitive stochastic demand," *International Journal of Production Economics*, vol. 234, 2021.

[57] A. A. Tsay, "The quantity flexibility contract and supplier-customer incentives," *Management Science*, vol. 45, no. 10, pp. 1339–1358, 1999.

[58] B. Tomlin, "Capacity investments in supply chains: sharing the gain rather than sharing the pain," *Manufacturing and Service Operations Management*, vol. 5, no. 4, pp. 317–333, 2003.

[59] A. Halati and Y. He, "Analysis of supply chains with quantity based fixed incentives," *European Journal of Operational Research*, vol. 202, no. 1, pp. 214–222, 2010.

[60] J. Zhang and R. Q. Zhang, "Supply chain structure in a market with deceptive counterfeits," *European Journal of Operational Research*, vol. 240, no. 1, pp. 84–97, 2015.

[61] Y. Liu, F. Qin, M. J. Fry, and A. S. Raturi, "Multi-period modeling of two-way price commitment under price-dependent demand," *European Journal of Operational Research*, vol. 221, no. 3, pp. 546–556, 2012.

[62] J. Zhang, Q. Gou, L. Liang, and Z. Huang, "Supply chain coordination through cooperative advertising with reference price effect," *Omega*, vol. 41, no. 2, pp. 345–353, 2013.

[63] T. Xiao, Y. Xia, and G. P. Zhang, "Strategic outsourcing decisions for manufacturers that produce partially substitutable products in a quantity-setting duopoly situation," *Decision Sciences*, vol. 38, no. 1, pp. 81–106, 2007.

[64] G. P. Cachon and A. G. Kök, "Competing manufacturers in a retail supply chain: on contractual form and coordination," *Management Science*, vol. 56, no. 3, pp. 571–589, 2010.

[65] Y. H. Albert, S. Tong, and H. Zhang, "Sharing demand information in competing supply chains with production diseconomies," *Management Science*, vol. 57, no. 3, pp. 566–581, 2011.

[66] E. Adida and G. Perakis, "The effect of supplier capacity on the supply chain profit," *Annals of Operations Research*, vol. 223, no. 1, pp. 1–52, 2014.

[67] D. Ghosh and J. Shah, "Supply chain analysis under green sensitive consumer demand and cost sharing contract," *International Journal of Production Economics*, vol. 164, pp. 319–329, 2015.

[68] Y. Zha, J. Zhang, X. Yue, and Z. Hua, "Service supply chain coordination with platform effort-induced demand," *Annals of Operations Research*, vol. 235, no. 1, pp. 785–806, 2015.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6541437, 13 pages
https://doi.org/10.1155/2021/6541437

Research Article

Digital Inclusive Finance, Multidimensional Education, and Farmers' Entrepreneurial Behavior

Ziqiang Liu,[1] Yihao Zhang ,[1] and Hongyi Li[2]

[1]*School of Economics, North Minzu University, Yinchuan, China*
[2]*Business School, Chinese University of Hong Kong, Central Ave, Hong Kong*

Correspondence should be addressed to Yihao Zhang; 914833342@qq.com

Received 22 July 2021; Accepted 6 September 2021; Published 5 October 2021

Academic Editor: Weilin Xiao

Scarce financial supply and low education level are important factors that inhibit the entrepreneurial behavior of rural residents in China. Based on the Static Career Choice Model, this paper matches the 2016 China Household Finance Survey (CHFS) microdata with the Peking University Digital Financial Inclusive Index of the previous years to empirically test the impact of digital financial inclusion and academic education, tacit knowledge, and Internet learning on farmers. The direct influence and linkage effect of entrepreneurial behaviors revealed that digital financial inclusion and multidimensional education could significantly promote farmers' entrepreneurial choices. Digital financial inclusion can relieve the constraints of insufficient academic education on farmers' entrepreneurial choices, partially replace the tacit knowledge for rural residents, and improve the efficiency of Internet learning, which ultimately enhances the entrepreneurial behavior choices of the farmers. Our results are still significant and robust with respect to the sample data, explained variables, and estimation methods. We also consider the use of instrumental variables to overcome the potential endogeneity issues. Through comparative analysis of different regions, it is found that the performance is particularly obvious in the eastern region.

1. Introduction

Entrepreneurship of farmers can help improve rural employment and farmers' incomes, which is an important means to implement the Rural Revitalization Strategy of the Central Government of China. The research on the factors affecting farmers' entrepreneurship and policies to promote entrepreneurship has aroused extensive discussions in the academic community. The factors affecting farmers' entrepreneurship, individual education, and the rapid development of digital financial inclusion have played an important supporting role.

Education is closely related to the entrepreneurial behavior of farmers. In addition to the traditional academic education, other types of education will also become a key factor influencing rural residents' entrepreneurial choices. Since academic education has a certain lag in the impact of farmers' entrepreneurial behavior, for the promotion of farmers' entrepreneurship at this stage, broadening of multidimensional education channels will enable farmers to enhance their entrepreneurial capabilities in practice and alleviate the constraints of insufficient academic qualifications on entrepreneurship. The economic environment is also an important factor affecting entrepreneurial behavior. In recent years, with the continuous enhancement of the integration of digital technology and traditional finance, digital finance has become an emerging format of inclusive financial services. Through the strong geographic penetration and the information advantage of capturing long-tail customers, digital financial inclusion has become a key financial tool on entrepreneurship for serving small private businesses, small and micro enterprises, and especially the more disadvantaged rural households. For farmers, the use of digital financial inclusion can alleviate credit constraints and is an important way to achieve multidimensional education and improve financial literacy and entrepreneurial capabilities.

This article focuses on answering the following questions. How do multidimensional education, academic education, tacit knowledge, Internet learning, and digital financial inclusion affect farmers' entrepreneurial responses, respectively? How does multidimensional education affect farmers' entrepreneurial behavior in the digital financial inclusion environment? What is the interaction between academic education and digital financial inclusion on entrepreneurial performance? Can farmers in different regions benefit from digital financial inclusion relatively equally in terms of entrepreneurship?

The rest of the paper is organized as follows: In Section 2, we review the related literature and discuss the theoretical justification of our research. The theoretical analysis framework and proposed hypotheses for empirical testing are also discussed in this section. In Section 3, regression models, variables, and data are discussed. In Section 4, the empirical results are presented. In Section 5, we perform robustness checking and also test the endogeneity. Regional analysis by subsamples is presented in Section 6. Finally, Section 7 concludes the paper.

2. Literature Review and Theoretical Basis

2.1. Literature Review. At present, academic research on digital inclusive finance and multidimensional education on residents' entrepreneurial decision-making mainly focuses on the effects of digital inclusive finance on residents' entrepreneurship and the impact of multidimensional education on residents' entrepreneurship, respectively.

As far as the role of digital financial inclusion in residents' entrepreneurship is concerned, the existing literature mainly focuses on the following two aspects. First, from the perspective of practical application and financial product innovation, it illustrates the support path of digital financial inclusion for entrepreneurial behavior through cases, such as the "Inclusive Finance and Smart County" project of the Internet Commercial Bank in Wugong County, Shaanxi Province, and the "Youth Entrepreneurship Loan" model developed by the Yingquan Rural Commercial Bank of Fuyang City, which provide strong support for local farmers to start businesses [1]. Second, it is to use different levels of macro- and micro-data to study the impact of digital financial inclusion on the entrepreneurial behavior of farmers. For example, Luo and Zhang [2] measured the entrepreneurial participation rate of residents in each city by (number of individual employees + number of employees engaged by private enterprise investors)/resident population. They believe that when the level of development of digital financial inclusion, insurance coverage, traditional finance, and industrial structure crosses the corresponding thresholds, the role of digital financial inclusion in promoting residents' entrepreneurship will continue to increase as the threshold variables increase. In contrast, they believe that when human capital crosses the corresponding threshold, the role of digital financial inclusion in promoting residents' entrepreneurship will continue to weaken as the level of threshold variables increases. Tao et al. [3] used provincial panel data and CFPS data to show that digital finance is inclusive, promoting family entrepreneurial decision-making. At this stage, the effect of digital finance on rural household entrepreneurship is greater than that of urban households. Based on the Peking University Digital Financial Inclusive Index and the Sun Yat-sen University China Labor Force Dynamics Survey (CLDS) Database, Zhang and Huang [4] found that the level of development of digital financial inclusion and its various dimensions have a significant positive effect on the self-employment of rural labor, especially for rural labor to become self-employed. In addition, digital financial inclusion has a stronger role in promoting self-employment activities of rural laborers over the age of 35 in underdeveloped areas with lower education levels. He and Li [5] used the rural inclusive finance survey conducted by China Agricultural University. They believed that digital finance eases the credit constraints of farmers, increases the availability of information for farmers, enhances farmers' social trust, and ultimately promotes farmers' entrepreneurial decision-making. At the same time, it is believed that the impact of digital finance on nonagricultural and survival entrepreneurship is very significant. Still, the impact on agricultural and development entrepreneurship is not obvious.

In addition to traditional academic education [6, 7], tacit knowledge [8, 9], "learning by doing" entrepreneurial practice [10, 11], financial literacy [12–14], Internet learning [15, 16], and so on will also become important factors influencing the entrepreneurial choices of rural residents. Wang and Li [22] proposed multidimensional education, specifically academic education, tacit knowledge, and Internet learning. From these three dimensions, the impact of multidimensional education on farmers' entrepreneurship can be sorted out, including the following aspects. First, academic qualifications have a significant positive impact on farmers' entrepreneurship [18, 19]. Zhou [20] argued that the level of artistic quality of farmers directly affects the choice of occupation, location, and time for migrant workers to go out to work, and indirectly affects the decision of migrant workers to return home to start a business. Huang et al. [21] explained that rural students involving formal full-time education can positively promote entrepreneurship intentions exposed to campus entrepreneurship culture and practice courses. Second, tacit knowledge has a significant role in promoting entrepreneurship. Tacit knowledge was originally referred to as knowledge that people know but is indescribable [22], and later it was widely regarded as the part of knowledge that is generated in the organization and owned by individuals, which can be obtained and transferred through observation and interaction [23]. Cao and Luo [6] demonstrated that farmers who have close contacts with relatives and friends, and those whose family members or relatives are village cadres, government staff, or bank clerk, are more inclined to participate in entrepreneurial financing. In Hu and Zhang [8], the notable role of social networks in promoting entrepreneurship in urban and rural families is discussed. But, it has a greater effect on the latter. Liu et al. [24] believed that the kinship network can improve the entrepreneurial performance of farmers, and compared with returning farmers, the kinship network has a greater effect on the farmers who have not gone out. Third, Internet learning can also positively promote farmers' entrepreneurial behavior, including Internet embedding [25], Internet procurement and sales [16], mobile payment [26], and similar factors. Internet learning is conducive to farmers' exploratory and utilization entrepreneurial learning, thereby promoting their entrepreneurial decision-making and business performance [27].

Combining the literature mentioned above, although current studies have focused on the impact of education level and digital financial inclusion on rural residents' entrepreneurial choices from multiple perspectives, there is a lack of integration of these three into the same theoretical framework for research. The financing environment and education level are the two key factors for farmers' entrepreneurial behavior. Digital financial inclusion is a new type of financial format based on Internet technology, which can help farmers expand their knowledge in their use. In this process, the financial environment and multi-dimensional education interact, and ultimately the two have a cross-influence on the entrepreneurial decision-making of farmers. This article establishes an analytical framework that includes three elements: digital financial inclusion, multidimensional education, and farmers' entrepreneurship. We consider data selected from "2016 China Family Tracking Survey (CFPS)," "2017 China Household Finance Survey (CHFS)," "China Economic Statistics Yearbook (2017 and 2018)," and "Digital Inclusive Finance Index (2015 and 2016)," and use regression models such as binary Probit, Logit, IV-Probit, and Tobit to focus on the cross-influence of the digital inclusive financial environment and multidimensional education on the entrepreneurial behavior of the farmers.

This study makes three contributions to the existing literature. Firstly, based on the static career choice model, it analyzes the linkage mechanism of the digital inclusive financial environment and multidimensional education on the entrepreneurial behavior of farmers. By analyzing the external environment, farmers' entrepreneurial characteristics, and their mechanism of action, we seek to stimulate the sustainable vitality of entrepreneurship and improve the choice of entrepreneurial paths. Secondly, we aim to find ways that actively respond to development concepts such as "village revitalization strategy" and "digital empowerment," and provide a theoretical reference for improving the endogenous power of rural economic development. Finally, we comprehensively consider the regional heterogeneity of digital inclusive finance, multidimensional education, and farmer entrepreneurship, which is conducive to understanding the actual value of digital financial inclusion in various regions.

2.2. Theoretical Analysis Framework. The static career choice model shows that after the individual's multidimensional education level increases, the same financial capital investment will produce different entrepreneurial benefits [28]. The basic premise assumes that rational individuals follow the principle of maximization of utility. They will choose to start a business only when their income from becoming an entrepreneur is higher than that from becoming an employee. This model assumes that the resources owned by farmers can be divided into two types: financial capital represented by digital financial inclusion K and educational capital represented by multidimensional education A. Assume that all farmers are risk-neutral. They can freely choose to be hired or start their businesses.

If a farmer is employed, the wage income function obtained is affected by educational capital:

$$w(A) = \eta A^{\gamma}, \tag{1}$$

where η is a constant, $\eta > 0$, and $\gamma \in (0, 1)$, representing the level of capital of farm households.

Idle financial assets of farmers can also generate certain rewards. Assuming that the rate of return on assets is i, the total output of the employed farmers y_w is

$$y_w(K, A) = w(A) + (1 + i)K. \tag{2}$$

Assuming that the comprehensive contribution coefficient of capital is μ and the total output of its entrepreneurial activity production function y_c is only determined by financial capital K and educational capital A; then,

$$y_c(K, A) = \mu K^{\alpha} A^{\beta}. \tag{3}$$

Among them, $\mu > 0$ and $\alpha, \beta \in (0, 1)$.

Furthermore, the net income of farmers' entrepreneurship π_c is the total output of entrepreneurship minus the financing principal and interest, and we obtain

$$\pi_c(K, A) = y_c(K, A) - (1 + i)K. \tag{4}$$

For rational farmers with entrepreneurial willingness, the necessary condition for choosing entrepreneurship is that the total output of entrepreneurship is greater than the total output of being employed. At this time,

$$\begin{aligned} y_c(K, A) &> y_w(K, A), \\ \pi_c(K, A) &> w(A). \end{aligned} \tag{5}$$

In other words, the necessary condition for farmers to engage in entrepreneurial behavior is that the family's net income level is greater than the operating income at the time of employment. Assuming that the financial capital investment K is constant, the first-order partial derivatives of net entrepreneurial income and wage income concerning multidimensional education levels are both greater than zero, namely,

$$\begin{aligned} \frac{\partial \pi_c}{\partial A} &= \mu \beta K^{\alpha} A^{\beta - 1} > 0, \\ \frac{\partial \mathrm{w}}{\partial A} &= \eta \gamma A^{\gamma - 1} > 0. \end{aligned} \tag{6}$$

When $\beta > \gamma$ and $\mu \beta K^{\alpha} > \eta \gamma$, we have $(\partial \pi_c / \partial A) > (\partial \mathrm{w} / \partial A)$, $\pi_c(0) > 0$, and $W(0) = 0$. The function of $\pi_c(K, A)$ and $W(A)$ is shown in Figure 1, the abscissa is education level A, and the ordinate is net entrepreneurial income or wage income. The only focus is A^*.

Figure 1 analyzes the impact of educational capital on career choices, ignoring financial capital, and calculating the partial derivatives of net entrepreneurial income and wage income with respect to academic education. Figure 2 analyzes the impact of financial capital on career choices, without considering academic education, and calculating the

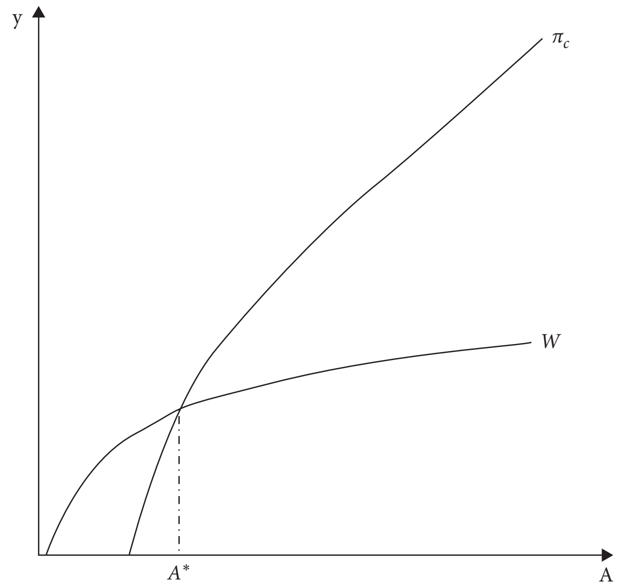


Figure 1: Farmers' multidimensional education and entrepreneurial decision-making.

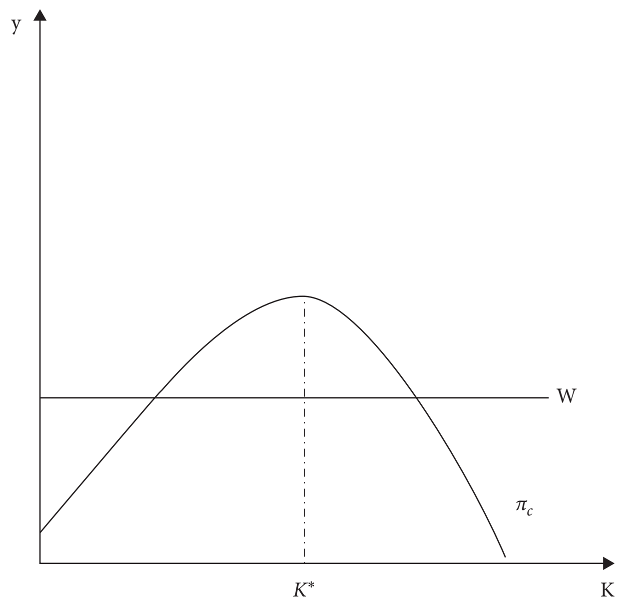


Figure 2: Digital inclusive finance and rural household entrepreneurial decision.

partial derivatives of net entrepreneurial income and wage income with respect to financial capital. Figure 3 comprehensively analyzes the impact of educational capital and financial capital on career choices. With the increase of financial capital, the wage income of rural households remains unchanged, but the total output of entrepreneurship presents an inverted U-shaped curve, as shown in Figure 2. The net income of entrepreneurship finds the first-order derivative of financial capital. Then, we set the derivative equal to 0, and find the optimal borrowing scale of farmers when A is constant, namely,

$$\frac{\partial \pi_c}{\partial K} = \alpha\mu K^{\alpha-1}A^{\beta} - (1+i) = 0. \tag{7}$$

That is,

$$K^* = \left(\frac{\alpha\mu}{1+i}\right)^{1/(1-\alpha)} A^{\beta/(1-\alpha)} > 0. \tag{8}$$

Substituting into equation (4), we get

$$\pi_c^*(A) = \mu(1-\alpha)\left(\frac{\alpha\mu}{1+i}\right)^{\alpha/(1-\alpha)} A^{\beta/(1-\alpha)}. \tag{9}$$

The inevitable condition for choosing entrepreneurship is $\pi_c^*(A) > w(A)$. Dong et al. [29] pointed out that the multidimensional education level in rural areas is lower than the equilibrium point A^*. In reality, farmers are generally subject to credit constraints. Therefore, farmers' venture capital K is lower than the optimal capital scale K^*, indicating that both digital financial inclusion development and multidimensional education will increase the probability of entrepreneurship.

When $K < K^*$, the development of digital financial inclusion can alleviate the adverse effects of the lack of multidimensional education of farmers to a certain extent, increase the multidimensional education level of farmers, and change the entrepreneurial curve of farmers from π_{c1} to π_{c2} in Figure 3, thus increasing the optimal financial capital scale of farmers from K_1^* to K_2^*. When the same ΔK (productive lending directly provided by digital financial inclusion) is provided, farmers who use digital financial inclusion are more inclined to start businesses and get more benefits than those who do not. Digital financial inclusion improves the probability of residents' entrepreneurship from three aspects of multidimensional education.

2.3. Proposed Hypotheses for Empirical Analysis.

2.3. Proposed Hypotheses for Empirical Analysis. Based on the above literature review and our research purpose discussion, we consider the following hypotheses for our empirical analysis.

Hypothesis 1. Digital financial inclusion will alleviate the constraints of insufficient academic education on farmers' entrepreneurial choices.

The average level of education of rural households in China is not high. When starting a business, they encounter prejudices in the financial market. The problems of expensive and difficult financing coexist, and access to funds is restricted. In 2016, the State Council promulgated the "Promoting Inclusive Finance Development Plan (2016–2020)," which proposed encouraging the provision of financial services to entrepreneurial farmers and other long-tail customers at affordable costs through Internet financial service platforms. Digital inclusive finance uses the Internet and the technology of information and communication to complete a series of financial activities such as third-party payment, online lending, funding direct sales, crowdfunding, and online insurance, alleviating the low efficiency, high threshold, and professionalism in the process of traditional inclusive financial services. It broadens the financing channels of farmers, brings about

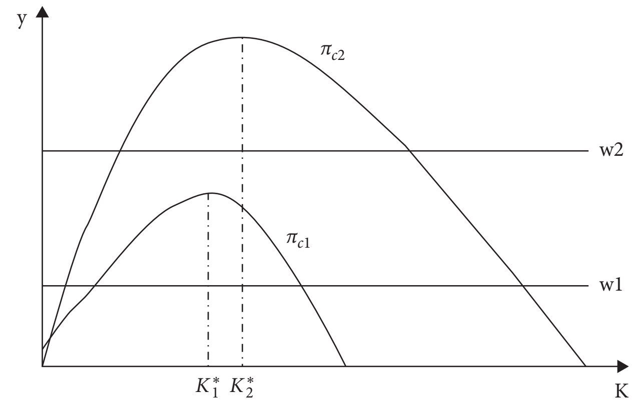


FIGURE 3: Influence of multidimensional education on farmers' entrepreneurial choice under the environment of digital inclusive finance.

an increase in the family income, and provides financial support for entrepreneurs in poor and vulnerable areas [30]. At the same time, in the process of exposure to digital financial inclusion, farmers have subtly improved financial literacy and information screening ability. They are more inclined to participate in e-commerce or other digital economic activities, thereby regulating the impact of insufficient academic qualifications on identifying entrepreneurial opportunities.

Hypothesis 2. Digital financial inclusion will partially replace the influence of tacit knowledge on rural residents' entrepreneurial choices.

Rural China is an "acquaintance society" based on blood, kinship, and geographic locations, wherein most of the tacit knowledge is acquired by farmers. Although it is beneficial to the actual entrepreneurial behavior of some farmers, there are temporal and spatial boundaries that are likely to consolidate their knowledge and experience. Digital inclusive finance relies on digital technology and is incorporated into commercial banks, financial technology companies, technology associations, public welfare teams, and other organizations to form a relatively rich and complete social circle. It indirectly reduces tacit knowledge on farmers' capital and information acquisition, thereby promoting their entrepreneurial behavior. Specifically, digital financial inclusion may adjust tacit knowledge to stimulate farmers' entrepreneurship in two ways: on the one hand, the use of digital finance by farmers has weakened the mutual guarantee or lending between acquaintances, thereby reducing the financing constraints of entrepreneurs [31]; on the other hand, digital finance uses technology dissemination and information sharing to explore business opportunities and extend the social network radius of microeconomic entities. For example, Alipay's interfaces such as Ant Forest, Ant Manor, and Baba Farm integrate daily life, online interaction, public welfare, and environmental protection into a powerful social network system, expanding demand for agricultural products such as fruits and saplings, ultimately partially replacing tacit knowledge and enhancing entrepreneurial response.

Hypothesis 3. Digital financial inclusion improves the efficiency of Internet learning and enhances farmers' entrepreneurial behavior choices.

The improvement of the rural network infrastructure and the long-term use of the Internet can strengthen the entrepreneur's knowledge category and accept external information [32]. Compared with traditional Internet learning, digital financial inclusion aims to deeply integrate technology represented by artificial intelligence, big data, cloud computing, and the Internet of Things with the traditional financial industry, which encourages financial technology companies to guide large state-owned banks, joint-stock commercial banks, rural commercial banks, and rural banks to conduct their businesses into rural financial markets, and then reduces information asymmetry and transaction costs. It has professional technical guidance and management training functions, which can provide an open online learning path, strong oriented and professional operation reducing the cost of user information search, and promoting farmers to understand, recognize, compare, and imitate information more purposefully. An efficient entrepreneurial information acquisition process replaces the Internet's relatively common knowledge retrieval process, reducing the time cost of farmers' Internet search-based learning and enhancing farmers' entrepreneurial choices. Simply put, it replaces the relatively common knowledge retrieval process on the Internet with a more efficient process, so it reduces the time cost of farmers' Internet search-based learning and enhances farmers' entrepreneurial choices. Digital financial inclusion in financial institutions is often a combination of policies that benefit the people and digital technology. The application of digital technology usually includes information collection, credit rating assessment, automatic calculation of credit lines and loan interest rates. Farmers often only need to download relevant banking apps to their mobile phones in order to independently involve in comprehensive financial services such as loan processing, payment, transfer, deposit, wealth management, social security card activation, collecting for another agency, pay for another, wage payment, and so on. In addition to the bank's corresponding publicity and some training activities, farmers can take questions to the bank to consult the staff on how to use mobile banking. In this process, the efficiency of Internet learning has been improved, more financial knowledge and application capabilities have been accumulated, and ultimately it enhances the possibility of farmers' entrepreneurial choices.

Concisely, as an entrepreneurial environment, digital inclusive finance and multidimensional education have an interactive impact, which plays a role in the entrepreneurial choices of farmers, as shown in Figure 4.

3. Models, Variables, and Data

3.1. Models. First, we study the impact of digital financial inclusion and multidimensional education on the entrepreneurial behavior of rural residents in China. Considering that the individual characteristics, family characteristics, and regional characteristics will affect their entrepreneurial

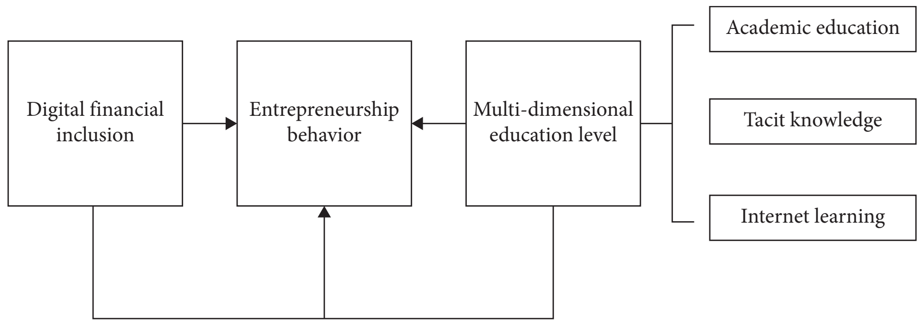


Figure 4: Theoretical framework of the influence of multidimensional education on farmers' entrepreneurial choice in the context of digital inclusive finance.

choices, the specific measurement model is designed as follows:

$$Entre_{i,j} = \alpha + \beta_0 In\ de\ x_j + \beta_1 E\ DU_{i,j} + \beta_2 X_{i,j} + \beta_3 Z_j + \varepsilon_{i,j}. \tag{10}$$

$Entre_{i,j}$ represents the binary variable of whether the i-th rural resident in the j-th province chooses to start a business; $In\ de\ x_j$ represents the independent variable, which is the level of digital financial inclusion in the j-th province; $E\ DU_{i,j}$ represents all education-related variables of the i-th farmer household surveyed in the j-th province; $X_{i,j}$ represents the set of individual characteristic variables; Z_j represents the set of regional characteristic variables; $\varepsilon_{i,j}$ represents a random error.

To further test the impact of multidimensional education on farmers' entrepreneurial choices in the digital financial inclusion environment, the interactive term of digital financial inclusion and education is added to the above model. Thus, we have

$$\begin{aligned} Entre_{i,j} = \alpha + \beta_0 In\ de\ x_j + \beta_1 E\ DU_{i,j} + \beta_2 X_{i,j} \\ + \beta_3 In\ de\ x_j \times E\ DU_{i,j} + \beta_4 Z_j + \varepsilon_{i,j}. \end{aligned} \tag{11}$$

3.2. Variables

3.2.1. Dependent Variables. The dependent variable is: whether rural residents start a business ($Entre_{i,j}$). According to Wang and Huang [10], we choose the CFPS2016, "Is your job working for yourself/your family or being employed by another person/another family/organization/unit/company?" and "Is your job agricultural or nonagricultural work?" to represent the type of job generated. If the respondent chooses "private enterprise/individual business/other self-employed," it means starting a business and a value of 1 is assigned. On the contrary, if the respondent chooses "own agricultural production and operation," "agricultural employment," and "nonagricultural employment," it means not starting a business, a value of 0 is assigned. Statistics show that there are 1,147 entrepreneurs, accounting for 7.47% of the total sample.

3.2.2. Core Independent Variables. We consider the following variables: multidimensional education which includes academic education, tacit knowledge, Internet learning. First, educational qualifications (*eduC*) are judged by the question of "the highest degree in the most recent survey." The results show that no schooling (34.40%), elementary school (26.26%), junior high school (24.36%), high school (7.79%), junior college (2.25%)), undergraduate (0.72%), master (0.02%), respectively. Second, social networks are closely related to traditional tacit knowledge. Referring to Han and Zhang [33], who used human relations as an indicator variable to measure social relations, tacit knowledge (*eduN*) is expressed by the question "How much cash or in-kind help has your family received from other relatives living in different places in the past 12 months? (ten thousand yuan)." Third, the indicator of farmers' self-learning through the Internet (*eduW*) chooses the question "Are you online" to measure. Last but not the least, we consider the interaction terms of multidimensional education and index to observe the changes in the impact of three dimensions of education on the entrepreneurial choices of rural residents under digital financial inclusion.

3.2.3. Moderator. According to Fu and Huang [34], considering that there may be a reverse causal relationship between digital financial inclusion and farmers' entrepreneurship, digital financial inclusion is measured by the provincial digital financial inclusion index (*Index*) of the previous year.

3.2.4. Control Variables. First, for personal characteristics, we use age, gender, and "frequency of using Internet business activities (times)" (trade) in CFPS2016; second, for regional characteristics, we use the logarithm of the local GDP in 2016 (lnGDP) in the "China Statistical Yearbook 2017" to represent the local economy development level; third, because various tax preferential policies have an incentive effect on innovative services [35], we choose the rate of various taxes to GDP (rate) to describe the local entrepreneurial environment. The statistical description of the variables is shown in Table 1.

Table 1: Statistical description of variables.

Variables	Obs	Mean	Std. dev.	Min	Median	Max
Whether rural residents start a business (*Entre*)	14501	0.07	0.26	0	0	1
Academic education (*eduC*)	17574	2.16	1.13	1	2	7
Tacit knowledge (*eduN*)	17574	0.10	0.76	0	0	40
Internet learning (*eduW*)	15653	0.32	0.47	0	0	1
Digital financial inclusion index (*Index*)	17574	213.50	17.28	193.30	206.30	278.10
Respondent's age (*age*)	17574	46.90	17.45	16	48	99
Respondent's gender (*gender*)	17574	0.51	0.50	0	1	1
Frequency of internet business activities (*trade*)	15653	0.33	0.85	0	0	4
The level of economic development (*lnGDP*)	17574	0.07	0.02	0.05	0.07	0.20
The proportion of various taxes in GDP (*rate*)	17574	10.02	0.76	7.85	10.01	11.30

3.3. Data. First, the micro-data of farmers comes from the 2016 China Family Panel Studies, referred to as CFPS. The survey was organized and implemented by the Chinese Social Science Research Center of Peking University. It aims to track and collect data at three levels of individuals, families, and communities to reflect the changes in China's society, economy, population, education, and health. The CFPS sample covers 25 provinces/municipalities/autonomous regions, the target sample size is 16,000 households, and the survey objects include all family members in the sample households. Second, the data related to digital financial inclusion comes from the China Digital Financial Inclusive Development Index. The index is released after a follow-up survey conducted by the Peking University Digital Finance Research Center and Ant Financial Services Corporation since 2011. Third, the provincial environmental data come from the China Statistical Yearbook (2016). According to the province to which the individual micro-data belongs in CFPS (2016), the relevant data are connected with the Digital Financial Inclusion Index (2016). After excluding invalid samples, 18,740 samples from rural areas are extracted for our research and analysis.

4. Regression Results and Analysis

We discuss the impact of multidimensional education on farmers' entrepreneurial choices in the context of digital financial inclusion. Table 2 reports the regression results of the benchmark model (11 (10) and (11)). First, in column (1), only independent variables such as academic education and digital financial inclusion are added. The results show that academic education positively affects the entrepreneurial behavior of the farmers, that is, the increase of academic education will affect the increase in the probability of entrepreneurship. Second, in column (2), we further add the cross-terms. The regression results show that the coefficients of the cross-terms between the index and academic education are negative, that is, digital financial inclusion has a certain supplementary effect on academic education and alleviates the lack of academic education for rural residents' entrepreneurial choices. The possible explanation is that digital financial inclusion allows rural residents to accumulate more specialized information and knowledge, which will partially supplement the impact of academic education on rural residents' entrepreneurial choices. Third, in column (3), we add control variables to the model, and the results remain robust. The results support Hypothesis 1.

Fourth, the model in column (4) only involves tacit knowledge, total index, and control variables, and we find that tacit knowledge has a significant positive impact on the entrepreneurial choices of rural residents. Fifth, the cross-term of tacit knowledge and index is added to the model in column (5). It is found that the influence of tacit knowledge on entrepreneurial choices of rural residents is significantly negative, and the cross-term of the index and tacit knowledge is opposite to the coefficient of tacit knowledge. It shows that digital finance reduces the impact of tacit knowledge on entrepreneurship. The possible explanation is that with the development of digital financial inclusion, the impact of relying on the tacit knowledge of social relationships on farmers' entrepreneurship has begun to diminish. As a result, farmers have more standardized channels to obtain information. They may gradually get rid of the influence of the "acquaintance society" and are accustomed to accepting the consciousness about the modern commercial society. The results support Hypothesis 2.

Sixth, in the model in column (6), we add Internet learning, index, and control variables, and the results show that using the Internet will increase the probability of starting a business. Seventh, when the interaction term between the index and Internet learning is added to the model in column (7), its coefficient is significantly negative, which shows that under the digital financial inclusion environment, the degree of dependence of rural residents' entrepreneurial choices on Internet learning level is decreasing. The possible reason is that digital financial inclusion provides professional, entrepreneurial information and knowledge, making knowledge retrieval more concentrated in entrepreneurial credit, which in turn has a positive impact on rural residents' entrepreneurship. Hypothesis 3 is supported. By observing the above models, it can be seen that the positive impact of academic education, Internet learning, and tacit knowledge on entrepreneurship has begun to diminish. In contrast, digital financial inclusion has complementary effects on multidimensional education.

In addition, the control variables also show the following unique characteristics: (1) The more frequently the farmers use the Internet for business activities, the greater the chance of choosing entrepreneurship. (2) The age of the farmers and

Table 2: Baseline regression results of the impact of multidimensional education on rural residents' entrepreneurial choice under the context of digital financial inclusion.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Academic education			Tacit knowledge		Internet learning	
Index	0.005***	0.008***	0.011***	0.01***	0.01***	0.01***	0.01***
	(5.11)	(3.50)	(3.60)	(3.56)	(3.38)	(3.26)	(3.48)
eduC	0.177***	0.903***	0.93***	0.922***	0.931***	0.851***	0.588***
	(14.48)	(5.15)	(4.62)	(4.59)	(4.72)	(4.22)	(2.68)
eduC×eduC		−0.08***	−0.07***	−0.07***	−0.07***	−0.07***	−0.07***
		(−7.87)	(−6.69)	(−6.72)	(−6.83)	(−6.19)	(−6.25)
index×eduC		−0.001*	−0.002**	−0.002**	−0.002**	−0.002**	−0.0006
		(−1.68)	(−2.19)	(−2.15)	(−2.19)	(−2.01)	(−0.58)
eduN				0.034**	−0.829***		
				(2.15)	(−3.29)		
index×eduN					0.004***		
					(3.40)		
eduW						0.305***	1.813***
						(6.16)	(3.86)
index×eduW							−0.007***
							(−3.23)
Trade			0.12***	0.119***	0.119***	0.076***	0.08***
			(5.48)	(5.45)	(5.45)	(3.38)	(3.61)
Age			0.024***	0.024***	0.024***	0.031***	0.033***
			(2.88)	(2.87)	(2.84)	(3.72)	(3.84)
age×age			−0.0004***	−0.0004***	−0.0004***	−0.0004***	−0.0004***
			(−3.76)	(−3.75)	(−3.73)	(−4.10)	(−4.21)
gender			0.163***	0.164***	0.164***	0.154***	0.157***
			(4.59)	(4.62)	(4.61)	(4.33)	(4.42)
Rate			−2.278*	−2.244*	−2.224*	−2.058*	−2.220*
			(−1.83)	(−1.80)	(−1.77)	(−1.66)	(−1.80)
lnGDP			0.047	0.046	0.047	0.051	0.048
			(1.48)	(1.45)	(1.47)	(1.60)	(1.51)
Intercept	−2.852***	−4.120***	−5.162***	−5.127***	−5.024***	−5.298***	−5.377***
	(−15.08)	(−8.33)	(−8.62)	(−8.58)	(−8.43)	(−8.86)	(−9.13)
Pseudo R2	0.0280	0.0379	0.0608	0.0616	0.0637	0.0667	0.0685
Obs	14501	14501	13302	13302	13302	13302	13302

Note. Robust standard errors are in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

the entrepreneurial choice are in an inverted U-shaped relationship, which means the probability of farmers starting a business first increases and then decreases with age. The possible reason is that age is closely related to risk appetite, and there is an inverted U-shaped relationship between risk aversion and age. (3) Female farmers are more likely to choose to start a business. (4) The higher the proportion of tax in GDP, the stricter the entrepreneurial environment, and the lower the probability of farmers choosing to start a business. (5) The local economic development level, LnGDP, has no significant impact on farmers' choice of entrepreneurship.

In general, through the following three mechanisms, digital financial inclusion has improved multidimensional education and increased the probability of rural residents' entrepreneurial choices, which further responds to the theoretical analysis in the previous research. First, digital financial inclusion partially compensates for academic education by accumulating new knowledge. Second, digital financial inclusion can effectively expand standardized information acquisition channels and reduce the influence of tacit knowledge on the entrepreneurial choices of rural residents. Third, digital financial inclusion directly provides a large number of specialized information channels for rural residents' entrepreneurial choices, which improves the efficiency of Internet learning to a certain extent.

5. Robustness Test

5.1. Replacing Sample Data for the Robustness Test. First, we consider using the nationwide household finance survey (CHFS) conducted by the Southwestern University of Finance and Economics in 2017 for robustness testing. The database covers 29 provinces, 363 counties, and more than 1,400 villages. After excluding problematic samples, there are a total of 43229 individual observations. The CHFS micro-data are merged with the Digital Financial Inclusion Index (2017) and China Economic Statistics Yearbook 2018 according to the province where the individual is located. Then, we select variables as follows. Select the question "Are you engaged in industrial and commercial production and operation?" as the entrepreneurial variable (*Entre*). Select the question of "after-tax income from industrial and commercial projects and income" as the entrepreneurial performance variable (*income*). Select the number of years of education as a variable to

Table 3: Descriptive statistics of variables based on CHFS2017.

Variables	Obs	Mean	Std. dev.	Min	Max
Entre	43229	0.12	0.32	0	1
Income	529	180000	480000	0	3.000e+06
Index	43229	271.5	19.4	240.2	336.6
eduC	37900	2.72	1.4	1	9
Age	43229	38.86	21.21	0	97
Gender	43229	0.52	0.5	0	1
Married	43229	0.64	0.48	0	1
Health	43198	0.78	0.41	0	1
Insure	42628	0.94	0.24	0	1
Party	23916	0.0500	0.21	0	1
lnGDP	43229	8.780	0.33	8.14	9.66

measure the level of education (*eduC*). Select the digital financial inclusion index to measure the development level of digital financial inclusion (*Index*) and use the interactive item of education and digital financial inclusion. Choose gender, age, marital status, physical status, party membership, and participation in medical insurance as individual and family control variables (*gender, age, married, health, party, insure*) to observe their changes. The logarithm of the province's per capita GDP in 2017 (*lnGDP*) in the "China Statistical Yearbook 2018" is selected as the regional control variable. The statistical description of the variables is shown in Table 3.

The regression results are reported in Table 4. Columns (1) and (3) of Table 4 use the Logit model to verify Hypothesis 1 again. The results show that academic education has a significant positive impact on rural residents' choice of entrepreneurship. At the same time, digital financial inclusion reduces the dependence of entrepreneurial choices on academic education. After adding the control variables, the results are the same.

5.2. Replacing the Dependent Variable and Model for the Robustness Test. The dependent variable income data (*income*) are partly continuous and partly discrete, so the Tobit model is used to verify the impact of academic education on the entrepreneurial performance of rural residents under the digital financial inclusion environment. Column (2) shows that academic education has a positive impact on the entrepreneurial performance of rural residents, and digital financial inclusion replaces the increase in entrepreneurial performance of academic education. After adding the control variable in column (4), the basic result remains unchanged. The results show that the key variables are highly robust.

5.3. Endogeneity Test. Although the independent variable uses the previous period's macro index, it can partially alleviate the endogeneity problem caused by reverse causality. However, due to the variables that affect education, many unobservable missing variables may also affect entrepreneurial choices, such as individual ability, personality, etc., leading to the correlation between educational variables and error terms, making the Probit model contain endogenous variables. To solve the estimation error caused by endogeneity, the control function method is selected to carry out the "two-step method" estimation. We choose parents' education level as an instrumental variable [36] and add quadratic terms and control variables for regression analysis. This is because the education level of rural residents' parents is less likely to have a direct impact on children's entrepreneurship.

Table 4: Robustness test based on CHFS2017 data.

	(1) Logit 1	(2) Tobit 1	(3) Logit 2	(4) Tobit 2
Index	0.015***	0.013***	0.026***	0.021***
	(−0.002)	(−0.002)	(−0.003)	(−0.003)
eduC	1.033***	0.849***	1.085***	0.825***
	(−0.15)	(−0.133)	(−0.264)	(−0.233)
eduC×eduC	−0.049***	−0.041***	−0.037***	−0.028***
	(−0.006)	(−0.005)	(−0.011)	(−0.009)
Index×eduC	−0.002***	−0.001***	−0.002**	−0.002**
	(−0.0005)	(−0.0005)	(−0.0009)	(−0.0008)
Age			−0.028***	−0.023***
			(−0.002)	(−0.002)
Gender			−0.098**	−0.093**
			(−0.05)	(−0.04)
Married			0.553***	0.385***
			(−0.095)	(−0.076)
Health			0.562***	0.471***
			(−0.060)	(−0.049)
Insure			0.171	0.177*
			(−0.108)	(−0.091)
Party			0.284***	0.238***
			(−0.091)	(−0.082)
lnGDP			−0.545***	−0.445***
			(−0.133)	(−0.113)
Intercept	−7.441***	−6.552***	−4.896***	−4.207***
	(−0.494)	(−0.434)	(−0.99)	(−0.846)
Pseudo R2	0.0238	0.0570	0.0186	0.0442
Obs	37,900	37,900	23,717	23,717

Note. Robust standard errors are in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

In contrast, the parent's education level significantly impacts the children's education level. The results in columns (1), (2), and (3) in Table 5 show that the instrumental variables of the first stage are significant at the 1% level, and the F value is greater than 10, indicating that there is no problem of weak instrumental variables; meanwhile, in the Wald endogeneity of the second stage, the p-values of the test in columns (4), (5), and (6) show that academic education is an endogenous variable. The IV-Probit model shows that academic education and entrepreneurial behavior have a significant positive impact, indicating that for the farmers who rationally choose to start a business (excluding those who must and must not), the improvement of academic education can promote the entrepreneurial choice of rural residents.

6. Analysis of the Heterogeneity of Samples by Region

In the benchmark analysis, it is observed that the local economic development level GDP has no significant impact on the entrepreneurial choices of rural residents. Thus,

Table 5: Influence of education on the entrepreneurial choice of rural residents: instrumental variable regression.

	(1)	(2) Step one	(3)	(4)	(5) Step two	(6)
	eduC	*eduC*	*eduC*	*Entre*	*Entre*	*Entre*
edu	0.586 ***	0.053 ***	0.036***			
	(0.014)	(0.004)	(0.005)			
eduC				0.325***	2.522***	2.074**
				(0.04)	(0.492)	(0.813)
eduC×eduC					−0.409***	−0.343**
					(0.086)	(0.14)
Controls	No	No	Yes	No	No	Yes
Obs	11,031	11,031	11,031	11,031	11,031	11,031
F statistic value	1762	74943	16436			
Wald test				15.83	15.12	3.45
p value				$p \leq 0.001$	$p \leq 0.001$	$p \leq 0.1$

Note. Robust standard errors are in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

whether the development of digital financial inclusion can break through the original traditional financial development situation in the region remains to be studied. To analyze in different regions what impact multidimensional education has on farmers' entrepreneurial choices under the digital financial inclusion, the 31 provinces were divided into eastern, central, and western regions for subsample testing (The eastern region includes 11 provinces including Liaoning, Beijing, Tianjin, Shanghai, Hebei, Shandong, Jiangsu, Zhejiang, Fujian, Guangdong, and Hainan; the central region includes 8 provinces including Heilongjiang, Jilin, Shanxi, Anhui, Henan, Hubei, Hunan, and Jiangxi Provinces; the western region includes 12 provinces including Chongqing, Sichuan, Yunnan, Guizhou, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang, Inner Mongolia, and Guangxi). The subsample regression results in Table 6 show that the economic significance and statistical significance of the eastern region are basically consistent with the benchmark regression results, but the counterparts are not significant. A possible explanation is that on the one hand, the extensive and in-depth use of digital financial inclusion requires a process. The coverage of traditional financial outlets is wider in economically developed areas. Rural residents in the east use traditional finance and the Internet more frequently than its counterparts. On the other hand, it is in the east that there are smaller, medium, and micro enterprises, and farmers are more accepting in using digital financial inclusion. Therefore, in noncentral and western provinces, digital financial inclusion has a greater substitution effect on academic education and Internet learning, and the possibility of entrepreneurship is greater. In general, digital financial inclusion cannot break through the traditional financial development at this stage to regulate multidimensional education and entrepreneurial choices.

7. Results and Discussion

This paper incorporates the new financial format of digital financial inclusion into the consideration of the entrepreneurial environment. It attempts to analyze the combined effect of multidimensional education and digital financial inclusion on the entrepreneurial choices of farmers from the perspective of the interaction of independent variables with empirical methods. First, based on the static career choice model, it explains that digital financial inclusion interacts with rural residents' academic education, tacit knowledge, and Internet learning to improve the multidimensional education level and influence their entrepreneurial choices. Three hypotheses were subsequently proposed. Then, CFPS2016 individual micro-survey data and the digital financial inclusion index are used to test the above three hypotheses empirically. The results show that: first, the expansion of digital financial inclusion supplements the impact of academic education on entrepreneurial choices of rural residents, thereby relaxing the constraints of academic education on entrepreneurial choices. Second, the popularization of digital financial inclusion partially replaces and gradually changes the influence of traditional tacit knowledge on farmers' entrepreneurship. Third, digital financial inclusion has effectively improved the efficiency of Internet learning. Finally, further analysis found that the role of multidimensional education on rural residents' entrepreneurial choices in the digital financial inclusion environment is affected by regional heterogeneity. Rural residents in the eastern region have a higher acceptance and use of digital financial inclusion. At the same time, compared with tacit knowledge, digital financial inclusion has a greater substitution effect on academic education and Internet learning.

The conclusion has important policy implications. First, strengthening the promotion of digital financial inclusion, including vigorously implementing network environment infrastructure construction in remote areas, providing material guarantees for encouraging farmers to use digital financial inclusion, comprehensively breaking up the technical barriers between financial services and users, actively guiding the conducting of digital services in commercial banks to rural areas, and promptly launching digital inclusive financial products to meet the needs of the "three rural" groups. Second, encouraging the coordinated development of digital inclusive finance and

Table 6: Influence of multidimensional education on rural residents' choice of entrepreneurship in different regions under the environment of digital inclusive finance.

	(1) Eastern	(2) Eastern	(3) Eastern	(4) Central	(5) Central	(6) Central	(7) Western	(8) Western	(9) Western
Index	0.02***	0.01***	0.02***	−0.033	−0.018**	−0.027**	0.024**	0.016*	0.018**
	(4.85)	(3.78)	(5.40)	(−1.07)	(−2.17)	(−2.18)	(2.09)	(1.93)	(2.09)
eduC	1.272***	0.07**	0.07**	−0.491	0.094***	0.075**	1.163	0.128***	0.096***
	(3.90)	(2.57)	(2.52)	(−0.24)	(3.30)	(2.54)	(1.42)	(4.62)	(3.29)
eduC×eduC	−0.045***			−0.09***			−0.1***		
	(−2.65)			(−4.10)			(−4.56)		
EduC×index	−0.004***			0.005			−0.002		
	(−3.10)			(0.51)			(−0.64)		
eduN		−0.129			4.180			−3.193	
		(−0.28)			(0.82)			(−1.53)	
eduN×index		0.001			−0.020			0.015	
		(0.60)			(−0.81)			(1.54)	
EduW			2.723***			−2.458			1.322
			(4.00)			(−0.72)			(0.79)
Eduw×index			−0.011***			0.014			−0.004
			(−3.82)			(0.82)			(−0.50)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.0649	0.0646	0.0652	0.0575	0.0494	0.0569	0.0644	0.0530	0.0678
Obs	4533	4533	4533	3731	3731	3731	5038	5038	5038

Note. Robust standard errors are in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

multidimensional education, including attaching importance to the new vocational education system for farmers, improving the financial literacy of rural young people and their acceptance of digital inclusive finance, providing targeted assistance to rural labor with entrepreneurship needs, for those who have low academic qualifications and lack tacit knowledge or have low frequency of Internet use, organizing learning lectures and exchange meetings, and inviting model entrepreneurs with experience and digital financial inclusion professionals to share their experiences. Third, alleviate regional differences step by step. That is, paying attention to the gaps in the development of digital financial inclusion in the eastern, central, and western regions, and taking relevant measures to increase the acceptance and use of digital financial inclusion by farmers in fragile areas, thereby effectively spreading the marginal role of digital financial inclusion.

Data Availability

The CHFS Stata Dataset data used to support the findings of this study may be released upon application to the SurveyAnd Research Center for China Household Finance, who can be contacted at https://chfs.swufe.edu.cn/. The CFPS Stata Dataset data used to support the findings of this study may be released upon application to the Institute of Social Science Survey, Peking University, who can be contacted at http://isss.pku.edu.cn/cfps/download/login. The Peking University Digital Financial Inclusion Index (2011–2020) PDF data used to support the findings of this study may be released upon application to the Institute of Digital Finance Peking University, who can be contacted at https://idf.pku.edu.cn/yjcg/zsbg/index.htm. The 2017 and 2018 China Economic Statistical Yearbooks Excel data used to support the findings of this study may be released upon application to the National Bureau of Statistics, who can be contacted at http://data.stats.gov.cn/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank Weidong Xu for valuable and constructive comments. This work was supported by the Fundamental Research Funds for the Central Universities "Study on the Synergy Mechanism of Rural Revitalization and the Development of Digital Inclusive Finance" (ZDZX201903).

References

[1] X. Ma, "The practical dilemma and countermeasures of financial support for rural households' entrepreneurship under the background of rural revitalization," *Southwest Finance*, no. 10, pp. 36–46, 2020, in Chinese with English abstract.

[2] X. Luo and L. Zhang, "The entrepreneurial effect of digital financial inclusion: mechanism, threshold and policy value," *Finance Theory and Practice*, vol. 2, pp. 17–26, 2021, in Chinese.

[3] Y. Tao, Y. Cao, J. Zhang, and K. Zou, "The Impact of Digital Finance on Entrepreneurship: Evidences from Regional and National CFPS," *Journal of Zhejiang University (Humanities and Social Sciences)*, vol. 51, no. 01, pp. 129–144, 2021, in Chinese with English abstract.

[4] Z. Zhang and F. Huang, "Influence of digital inclusive finance on self-employment of rural labor force," *Financial Forum*, vol. 26, no. 4, pp. 58–68, 2021, in Chinese with English abstract.

[5] J. He and Q. Li, "Digital finance and farmers' entrepreneurship," *Chinese Rural Economy*, vol. 1, pp. 112–126, 2019, in Chinese with English abstract.
[6] L. Cao and J. Luo, "Social capital, financial literacy and farmers' entrepreneurial financing behavior," *Journal of Zhongnan University of Economics and Law*, vol. 3, pp. 3–158, 2019, in Chinese with English abstract.
[7] L. Liu, J. Zhao, and H. Zhong, "Study on intergenerational income mobility from the perspective of entrepreneurship: stratum differences and influencing factors," *Nankai Economic Studies*, vol. 5, pp. 163–179, 2019, in Chinese with English abstract.
[8] J. Hu and B. Zhang, "Social networks, informal financing and household entrepreneurship: empirical analysis based on the urban-rural differences in China," *Journal of Financial Research*, vol. 10, pp. 148–163, 2014, in Chinese with English abstract.
[9] C. Guo and A. He, "Social capital, entrepreneurship environment and peasants' entrepreneurship performance in agriculture-related businesses," *Journal of Shanghai University of Finance and Economics*, vol. 19, no. 2, pp. 76–85, 2017, in Chinese with English abstract.
[10] Q. Wang and W. Huang, "Does entrepreneurship make farmers happier? Empirical evidence from China's labor-force dynamic survey," *Journal of Nanjing Agricultural University (Social Science Edition)*, vol. 20, no. 3, pp. 99–110, 2020, in Chinese with English abstract.
[11] Z. Wang, "Entrepreneurship education family, entrepreneurship resources and quality of entrepreneurship education: An empirical study based on the data from the universities in Zhejiang province," *Journal of Nanjing Normal University (Social Science Edition)*, vol. 3, pp. 39–52, 2020, in Chinese with English abstract.
[12] Z. Yin, Q. Song, Y. Wu, and C. Peng, "Financial knowledge, entrepreneurial decision and entrepreneurial motivation," *Management World*, vol. 1, pp. 87–98, 2015, in Chinese.
[13] G. Sun, Q. Li, and Y. Yang, "The impact of financial knowledge on the entrepreneurship of landless rural households: an empirical analysis based on IV-Heckman model," *China Rural Survey*, vol. 3, pp. 124–144, 2019, in Chinese with English abstract.
[14] Y. Zhang, Y. Liu, and G. Han, "Financial literacy and family entrepreneurial returns," *Journal of Finance and Economics*, vol. 8, pp. 4–12, 2020, in Chinese.
[15] Z. Wang and T. Wu, "entrepreneurship in the Internet context," *Journal of Zhejiang University (Humanities and Social Sciences)*, vol. 46, no. 1, pp. 131–141, 2016, in Chinese with English abstract.
[16] L. Su and R. Kong, "Does Internet use improve rural households' entrepreneurial performance? An empirical analysis based on endogenous switching regression model," *Chinese Rural Economy*, vol. 2, pp. 62–80, 2020, in Chinese with English abstract.
[17] J. Wang and Q. Li, "Research on the relationship between dimensional education and rural residents′ entrepreneurship under the perspective of E-commerce environment: based on the data of CHIPS2013 and CFPS2014," *Nankai Economic Studies*, vol. 6, pp. 75–92, 2017, in Chinese with English abstract.
[18] Z. Hao, X. Ping, H. Zhang, and S. Liang, "Liang S. Rural households' entrepreneurial choice and its influencing factors: evidence from the "Rural Finance Survey," *Chinese Rural Economy*, vol. 4, pp. 57–95, 2012, in Chinese.
[19] J. Wang, S. Mou, and Y. Sheng, "Can E-commerce help rural residents' entrepreneurship? Based on the perspective of social capital," *Research on Economics and Management*, vol. 40, no. 02, pp. 95–110, 2019, in Chinese with English abstract.
[20] Y. Zhou, "Rural culture in the new era and the willingness of migrant workers to return home and start businesses," *Seeker*, vol. 12, pp. 122–130, 2017, in Chinese.
[21] M. Huang, Y. Zhang, and X. Tu, "The influence of entrepreneurial education on entrepreneurial intention: the moderating effect of entrepreneurial culture," *Journal of Xidian University (Social Science Edition)*, vol. 29, no. 4, pp. 96–103, 2019, in Chinese with English abstract.
[22] P. Michael, *Study of Man*, Chicago University Press, Chicago, IL, USA, 1958.
[23] C. Dhanaraj, M. Lyles, H. Steensma, and L. Tihanyi, "Managing tacit and explicit knowledge transfer in IJVs: the role of relational embeddedness and the impact on performance," *Journal of International Business Studies*, vol. 35, no. 5, pp. 428–442, 2004.
[24] J. Liu, X. He, and Y. Zhang, "How do Farmers Succeed in Entrepreneurship? an Empirical Study Based on the Dual Perspectives of Human Capital and Social Capital," *Sociological Review of China*, vol. 8, no. 3, pp. 105–117, 2020, in Chinese with English abstract.
[25] Z. Yao, J. Luo, X. Zhang, and L. Zhang, "Internet embedding, ambidextrous entrepreneurship learning and farmers' entrepreneurial performance," *Studies in Science of Science*, vol. 38, no. 4, pp. 685–695, 2020, in Chinese with English abstract.
[26] Z. Yin, X. Gong, and B. Guo, "The Effect of Mobile Payments on Household Money Demand: Micro Evidence from the China Household Finance Survey," *China Industrial Economics*, vol. 3, pp. 119–137, 2019, in Chinese with English abstract.
[27] G. Zhou and G. Fang, "Internet Usage and Household Entrepreneurship: Evidence from CFPS," *Economic Review*, vol. 5, pp. 134–147, 2018, in Chinese with English abstract.
[28] D. Evans and B. Jovanovic, "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints," *Journal of Politics*, vol. 97, no. 4, pp. 808–827, 1989.
[29] X. Dong, N. Sun, and W. Wu, "Human capital, household financing and entrepreneurial choice of farmer households: an analysis based on examination of 7981 valid samples in CFPS," *China Rural Survey*, vol. 03, pp. 109–123, 2019, in Chinese with English abstract.
[30] X. Xu, J. Li, and L. Jin, "The alleviation effect of inclusive finance on rural education poverty," *Chinese Rural Economy*, vol. 9, pp. 41–64, 2020, in Chinese with English abstract.
[31] Y. Wu, C. Li, X. Li, and D. Yi, "The impact of digital finance development on traditional private credit market," *Management World*, vol. 36, no. 10, pp. 53–65, 2020, in Chinese with English abstract.
[32] X. Yu and Q. Zhou, "Impact of external search tactics and knowledge absorptive capacity on technology innovation performance," *Science Research Management*, vol. 39, no. 08, pp. 11–18, 2018, in Chinese with English abstract.
[33] C. Han and L. Zhang, "Has the Internet improved the income mobility of residents in China? An empirical study based on CFPS data from 2010 to 2016," *Journal of Yunnan University of Finance and Economics*, vol. 35, no. 1, pp. 86–99, 2019, in Chinese with English abstract.
[34] Q. Fu and Y. Huang, "Digital finance's heterogeneous effects on rural financial demand: evidence from China household

finance survey and inclusive digital finance index," *Journal of Financial Research*, vol. 11, pp. 68–84, 2018, in Chinese with English abstract.

[35] J. Cui and Y. Cheng, "The incentive effect of incubator tax preferential policy on innovation service," *Research in Science of Science*, vol. 34, no. 1, pp. 30–39, 2016, in Chinese with English abstract.

[36] D. Shan, "Financial literacy and urban poverty," *China Industrial Economics*, vol. 4, pp. 136–154, 2019, in Chinese with English abstract.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5683525, 12 pages
https://doi.org/10.1155/2021/5683525

Research Article

The Nearer, the Better? The Impact of Cultural and Geographic Distance on Crowdfunding Project Attractiveness

Hui Fan,[1] **Teng Gao**,[2,3] **and Shuman Liu**[4]

[1]*School of Economics and Management, Tongji University, Shanghai, China*
[2]*School of Tourism and Exhibition, Hefei University, Hefei, China*
[3]*Chaohu Research Center for Culture, Economic and Social Development, Hefei, China*
[4]*Faculty of Business, Lingnan University, Hong Kong, China*

Correspondence should be addressed to Teng Gao; gaoteng1963@163.com

Received 21 July 2021; Accepted 4 September 2021; Published 26 September 2021

Academic Editor: Wen-Tsao Pan

Information asymmetry between backers and project creators impedes the crowdfunding success. Consequently, creators usually rely on various information to alleviate information asymmetry. Particularly, the location information of both backers and creators embodies their geographic and cultural distance, which may affect crowdfunding project attractiveness. Whereas current literature almost ignores the role cultural distance in crowdfunding, this research focuses on the reward-based crowdfunding, so that it becomes salient to form the appreciation and judgment of the innovative, creative, or artistic nature of projects. Meanwhile, geographic distance is examined to join the debates between flat world hypothesis and home bias proposition. A series of econometric models are examined based on a sample of 264 fundraising projects collected from Kitckstarter.com through Python program. Results show that cultural distance exerts a U-shape effect, which initially impedes the crowdfunding performance but promote projects when large enough. Geographic distance generally exerts insignificant impact on crowdfunding performance. Furthermore, cultural and geographic distance exerts the asymmetric effects on experienced versus new backers. This article underscores the important implications of cultural distance on reward-based crowdfunding. By showing the differential effects of cultural and geographic distance on experience versus new backers, it empirically infers the social capital as the underlying mechanism.

1. Introduction

Crowdfunding refers to the efforts by entrepreneurial individuals and groups using the Internet to fund their cultural, social, and for-profit ventures by drawing on relatively small contributions from a relatively large number of individuals, without standard financial intermediaries [1]. Online crowdfunding platforms have become appealing channels for fundraising, and markets for crowdfunding have rapidly and significantly grown into a multibillion-dollar industry worldwide [2]. Although seemingly promising, not all crowdfunding projects can attract the desired amount of funding [3]. Information asymmetry in crowdfunding is pervasive, so that potential backers lack substantive knowledge on both the capabilities or trustworthiness and the characteristics of the proposed initiations [4–6].

Accordingly, current literature has documented various signals that crowdfunding creators can adopt to alleviate the information asymmetry and finally attract potential backers. Some work on the project quality signals embodied in the project descriptions [1, 6–11], creator-trustworthiness signals [8, 12–15], timing signals [16], founders' social capital [4], and product quality signals [14, 17]. Some other researchers focus on the implicit information during the dynamic fundraising cycle, such as herding information among backers [5], creator-backer interaction [18], and information implied by contributing patterns of previous backers [16, 19, 20].

This study aims to add to the pertinent literature by understanding whether and how the geographic and cultural distance between a creator and a backer influences the performance of crowdfunding projects. Geographic distance is salient due to the fact that, as the striking features, crowdfunding platforms remove geographic limitations but great geographic distance always exists between the creator and backers [19]. In offline investment context, home bias exists because geographic proximity reduces the cost associated with information acquisition, transaction, or monitoring (rational/economic causes) and engenders trust and overoptimism toward transaction partners or opportunities (behavioral causes) [19, 21–23]. The ubiquity of the Internet promotes a "flat world" hypothesis [24] because the aforementioned "home bias" [21] seems to become irrelevant in crowdfunding. Nevertheless, subsequent research on geographic distance reveals controversial findings, advocating either a negative effect [21, 25] or a positive effect [1, 26].

The cultural distance information contained in crowdfunding projects is also an important factor affecting the success or failure of crowdfunding projects. Burtch et al. [25] underscored the cultural fraction caused by cultural distance and revealed its detrimental effect on crowdfunding performance. However, every cloud has a silver lining. Cultural distance helps backers "go out of the box" and provides opportunities for individuals to learn about disparate concepts and ideas from different cultures. This is similar to what Lu et al. [27] posit that intercultural social interaction stimulates cultural learning to acquire new information and understand about the assumptions, beliefs, customs, norms, values, or language of another culture. Mollick [1] provides implicit evidence on a positive effect of geographic distance on crowdfunding projects that reflect the underlying cultural products of particular geographic areas.

To fully understand the effect of geographic and cultural distance on crowdfunding project performance, we choose Kickstarter.com for the empirical investigations, which is the leading reward-based crowdfunding platform in the United States and has provided more funding for artists than the National Endowment for the Arts [28]. Based on a sample of 264 fundraising projects, this article reveals the more complex effects of cultural and geographic distance than those identified in the previous research. Cultural distance has a U-shape relationship with crowdfunding project attractiveness. It initially discourages backers from supporting crowdfunding projects when relatively small; however, the larger cultural distance makes the creative ideas more appealing when large enough. Surprisingly, geographic distance does not play a significant role in attracting backers to contribute. Moreover, we empirically infer the social capital acquired and accumulated from the platform community as one plausible mechanism. Whereas the U-shape effect of cultural distance only exists for experienced backers, large geographic distance encourages experienced backers but discourages new backers.

This article provides three-folded contributions. First, it understands more thoroughly the role of cultural distance on crowdfunding performance, which has been mostly ignored by the literature [25]. Increasing cultural distance initially impedes the crowdfunding success and funding willingness, which is consistent with what Burtch et al. [25] advocate; however, when it enlarges to some extent, its impact turns to be positive. The U-shape relationship between the cultural distance and crowdfunding performance we have identified are innovative to the crowdfunding literature. By highlighting the salience of cultural background in evaluating creative ideas, we contend that backers make a complex tradeoff between the uncertainty and perceived creativity of projects.

Second, we show that the effects of geographic distance do not seem to be straightforward. Though it does not significantly impact the crowdfunding success or backers' contribution amount, it implicitly influences the backer composition. It encourages experienced backers but discourages new backers to contribute. Our findings speak to the controversial literature on the effect of geographic distance (i.e., the negative effect by Burtch et al. [25] and Lin and Viswanathan [21] or the positive effect by Kang et al. [26] and Mollick [1]) and support the "flat world" hypothesis [24], instead of the "home bias" proposition [21].

Third, the difference between the experienced and naive backers reveals that crowdfunding platforms can cultivate the internal social capital in the community, which enables experienced backers to make the tradeoff between the uncertainty and perceived creativity. Also, the insights on how cultural distance and geographic distance influence the backer composition are innovative; whereas, currently, literature mostly focuses on the crowdfunding success [1, 10], backers' contribution decisions [17, 20], or interest rate [8, 12].

2. Literature Review and Hypotheses

Usually, crowdfunding involves an open call on the Internet-based platform for financial resources in the form of donation, exchange for future products, or some other rewards [29]. Various crowdfunding platforms (such as Kickstarter, Kiva, or SellaBand) reduce market frictions associated with geographic distance [19]. Remarkably, people still confront with uncertainty and information asymmetry on the trustworthiness of fundraisers [5] or project quality [11, 30]. Thus, crowdfunding creators utilize various signals to alleviate information asymmetry and attract potential backers.

2.1. Various Signals to Enhance Crowdfunding Project Attractiveness. Current literature has investigated various signals that help reduce uncertainty and information asymmetry in the crowdfunding market. Researchers underscore the signaling value of informational and social cues that facilitate backers to judge the project quality or trustworthiness of borrowers. The information in the project description determines crowdfunding success, ranging from the voluntary self-disclosure to the more subtle such as spelling errors or linguistic style of the text [1, 6, 9, 10]. Meanwhile, backers attend to many cues to judge the credibility of a borrowers, such as his popularity in his social network as indicated by his friendship or online word-of-mouth [7, 12] or simply his appearance attractiveness [8, 13]. Moreover, crowdfunding success is vulnerable to social influence; thus, the information implied by the contributing patterns of previous backers' shapes subsequent backers' decisions [5, 16, 19, 20].

The broad geographic dispersion of backers supports a "flat world" hypothesis [24]. Existing findings on the effect of geographic distance are far from conclusive. Geographic proximity may limit backers' potential to leverage social networks. For instance, Mollick [1] and Kang et al. [26] disclosed a positive effect of geographic distance in that large geographic distance signals the widely recognized project quality or reflects the underlying cultural products. However, Lin and Viswanathan [21] and Burtch et al. [25] demonstrated that backers still favor geographically proximate projects, supporting a "home bias" proposition.

Among the four basic types of crowdfunding posited by Mollick [1], the reward-based model fits well with creative crowdfunding projects undertaken by artists, musicians, filmmakers, inventors, and social enterprise. First, individuals supporting such projects receive a reward but not any financial incentives, returns, or repayment [7]. Second, the biggest difference lies in information asymmetry and uncertainty [18]. In the all-or-nothing game, project creators set a funding goal and receive the donations only if the goal is reached, and only after the project is successfully funded and implemented, the product described in the creative crowdfunding project will exist [1] and can enlarge information asymmetry between creators and backers in the crowdfunding markets. Third, people make judgment and funding decisions based on their appreciation of the "innovative," "creative," or "artistic" nature of the products, for which the cultural background becomes salient. Thus, we consider whether and how cultural and geographic distance influences crowdfunding project attractiveness, aiming to contribute to the existing literature as follows.

First, we underscore the role of cultural distance in the reward-based crowdfunding, which has been largely neglected currently. The cultural background of both backers and fundraisers become salient for the reward-based crowdfunding for creative or artistic projects. Second, we investigate the role of geographic distance, aiming to resolve the inconsistent findings between the two schools. Third, by demonstrating the asymmetric effects of cultural and geographic distance on experiential and new backers, we empirically infer the social capital, in which the backers develop inside the crowdfunding community, as the plausible underlying mechanism. This differs but complements what Colombo et al. [31] postulated that the internal social capital project creators acquire in the crowdfunding platform helps attract contributions.

Notably, Burtch etal. [25] also examined the dual roles of geographic and cultural distance and evidence that backers do prefer culturally similar and geographically proximate project creators. We differ from it for at least two points. First, Burtch et al. [25] investigated the aggregated effects of geographic and cultural distance at the national level by aggregating all crowdfunding by country. We dig deeply at the individual project level and map out their effects on individual crowdfunding project attractiveness. Second, while Burtch et al. [25] proposed the IT-based trust to explain the adverse effects of geographic and cultural distance, we empirically show the social capital as one plausible explanation by uncovering the their asymmetric effects on experienced and new backers.

2.2. Crowdfunding Project Attractiveness. When confronting various crowdfunding projects on Kickstart.com, potential backers usually aim to tap the most attractive one and decide the contribution amount. Although crowdfunding seems to be a promising channel, not all projects are able to attract the desired amount of funding [3]. Thus, whether a project is successfully funded measures the overall attractiveness. Meanwhile, the average fund indicates backers' willingness to contribute.

2.3. The Effect of Cultural Distance on Crowdfunding Project Attractiveness. There are strong forces within nations to create and maintain a shared culture [32]. Cultural distance refers to the extent to which the shared norms and values in one country differ from those in another and is operationalized in terms of the six dimensions: power distance, avoidance of uncertainty, individualism vs. collectivism, masculinity vs. feminity, long-term orientation vs. short-term normative orientation, and indulgence vs. restraint [33]. Online crowdfunding usually involves cross-border business transactions or interactions with different societal value systems [34]. Furthermore, because we particularly focus on the reward-based crowdfunding platforms for creative or artistic ideas, a backer's motivation to help others realize their creative ideas, instead of financial incentives, is generally more significant [16]. The backer selects the most appealing ideas based on the appreciation of the "innovative," "creative," or "artistic" nature of the funded products, among other project signals. As a result, the cultural backgrounds of both the creator and backers become salient for funding decisions (whether and how much to fund a creative idea). This is quite consistent with what Chua et al.' [35] content that cultural distance is one of three cultural characteristics particularly relevant in understanding creative ideas or solutions in a global context.

Nevertheless, cultural distance has received very limited discussion in crowdfunding literature. Exceptionally, Burtch et al. [25] demonstrated that cultural differences play a significant impeding role in crowd funders' decision-making. The authors interpret this as an awareness effect, suggesting that cultural differences are only relevant insofar as backers are aware of them. Although Mollick [1] focuses on geographic distance, the implications he suggests for its positive effect on crowdfunding success propose that culture may play a pivotal role especially for those projects that reflect the underlying cultural components of particular geographic areas. These existing studies reveal either a positive or a negative effect, a situation similar to international business research.

On the one hand, cultural familiarity theory holds that firms are less likely to invest in culturally distant countries and that cultural difference hampers multinational enterprises' performance when investing in culturally distant countries (e.g., [36, 37]). On the other hand, some researchers provide evidence that cross-border acquisitions in culturally distant countries tend to be more valuable as more diversified cultural integration helps enhance post-acquisition performance [38] or that high cultural distance has been associated with low rates of joint venture failure [39]. In sum, pertinent literature has identified a double-edged sword effect of cultural distance in cross-border business [40].

This article proposes a nonlinear, U-shape relationship between cultural distance and crowdfunding project attractiveness, which indicates that cultural distance exerts a negative effect on project attractiveness when it is relatively small or moderate and a positive effect when large enough. This contention is based on the notion that, when deciding the attractiveness of a project located in a culturally different country, backers make the tradeoff between the uncertainty and perceived creativity. When relatively small or moderate, cultural distance produces frictions and poses difficulties to some extent in understanding and interpreting the creative ideas in the crowdfunding projects. Meanwhile, the small cultural distance between somewhat similar cultures engenders low perceived creativity for a creative idea described by a crowdfunding project, that is, backers' evaluation of a creative idea tends to be constrained by the conventions and routines of their home culture. Cultural proximity could reduce lenders' reach for wider and untapped potential [26]. Thus, cultural distance exerts a detrimental effect on the project attractiveness when it is low or moderate.

However, we postulate a positive effect of cultural difference when it is large enough. When people perceive an object, the distance increases the uncertainty and gives people a broader space for imagination, thus producing a kind of beauty of distance. A crowdfunding project needs to gain sufficient attention and recognition to encourage the public to contribute [26]. When evaluating an idea from a strikingly different culture, backers face large comprehensive complexity and difficulties, which is supposed to hamper its attractiveness. However, being exposed to more culturally diverse ideas can increase the creative content of the mind [41], motivate individuals to perform more adeptly in creative insight tasks [42], and stimulate cultural learning to acquire new information and understand about the assumptions, beliefs, customs, norms, values, or language of another culture [27]. Global crowdfunding projects may contain more novel information for backers [26] and receive more support due to a high level of novelty [43]. Thus, we propose that cultural distance that is large enough exerts a positive effect on backers' judgment of crowdfunding project attractiveness.

> H1: cultural distance has a U-shape relationship with the crowdfunding project attractiveness; it exerts a negative impact when it is relatively low but a positive impact when large enough on the success of crowdfunding project (H1a) and on the average contribution amount of backers (H1b).

2.4. The Effect of Geographic Distance on Crowdfunding Project Attractiveness. The "flat world" hypothesis that crowdfunding projects are usually supported by the broad geographic dispersion of investors [24] gives rise to an interesting question "do crowdfunding platforms reduce market frictions associated with geographic distance?" Several studies examine the effect of geographic distance on funding decisions and reveal mixed findings. The home bias that backers favor geographically approximate fundraisers exists for both the equity-based crowdfunding on Prospers [21] and reward-based crowdfunding on SellaBand [19]. Contrarily, Mollick [1] and Kang et al. [26] disclosed a positive effect of geographic distance, implying that crowdfunding not only relaxes geographic constraints but also activates other mechanisms. Mollick [1] demonstrates this positive effect when crowdfunding projects reflect the underlying cultural products of particular geographic areas. Moreover, Kang et al. [26] contend that larger geographic distance may enhance entrepreneurs' reach for wider and untapped potential, and their study reveals that further geographical distance led to higher funding. In sum, previous studies have not concluded whether and how geographic distance exerts a significant effect on crowdfunding success performance.

Since the Internet facilitates instant and inexpensive communication across large distances, the impact of geographical distance between project creators and backers on crowdfunding projects has tended to fade in the context of online e-commerce and a globalized economy. The Kickstarter platform has users all over the world and belongs to a group of investors interested in or relatively familiar with the cultural and artistic creative industries. Culture and art creative projects are based on their own cultural attributes, with a certain story and sentimentality, which easily attract the attention of backers. Therefore, a project that sounds attractive or creative will break through geographical distance and have a group of enthusiastic supporters worldwide. Therefore, we propose the hypothesis as follows:

> H2: geographic distance exerts an insignificant impact on crowdfunding performance. There will be no significant difference between large geographic distance and small geographic distance in crowdfunding project attractiveness.

2.5. Social Capital for Experienced (vs. New) Backers. We further contend that the social capital the backers acquire from their community participation on the crowdfunding platform may drive the complex effects of cultural and geographic distance. The social capital refers to "the sum of the actual and potential resources embedded within, available through, and derived from the social contacts of an individual or organization" [44]. Crowdfunding platforms are not only intermediaries of monetary transactions but also loci of social connections [31]. Online community members routinely help one another, often going to great lengths to volunteer and share their expertise and resources with other members even when there are no apparent benefits from doing so [45–47]. Consumers derive many benefits from online community participation, such as learning, problem-solving, and the opportunity to socialize and ward off loneliness [48]. Most of previous literature investigated social capital from the perspective of the creators and has found that creators' social capital (i.e., a creator's social network ties, obligations to fund other creators, and the shared meaning of the crowdfunding project between the creator and the backers) had significant effects on crowdfunding performance in both China and the US [49], but has underscored the role of backers' social capital [26, 31]. The computer-mediated interactions occur among creators and backers of projects [50] and entail behaviors specific to crowdfunding communities [31].

By participating in the crowdfunding platform, backers can accumulate social capital in terms of information source and social support [51, 52]. Thus, experienced backers are more able to make the complex tradeoff between the perceived uncertainty and creativity when evaluating the creative ideas. We empirically demonstrate this by the asymmetric effects of cultural and geographic distance on experienced vs. new backers. Cultural distance has a U-shape effect in attracting more experienced backers (but not new backers) to support the project. This is similar to research on the multicultural experience–creativity link that the more contacts among two intercultural individuals, the more likely they assimilate and draw upon ideas from both cultures to synthesize novel and useful insights [53]. Furthermore, we speculate that when the cultural distance is large enough, backers will obviously perceive that the project is different from the local culture of their own country. People with investment experience pursue novel psychological feelings and have the need to explore new things. Even if the cultural distance of the project is too large to lead to a certain degree of risk-taking, cultural distance will not be the reason to hinder their investment. Instead, the project uncertainty and risk-taking brought by cultural distance may become the attraction of their investment. Therefore, we propose the hypothesis as follows:

> H3a: cultural distance has a U-shape effect in attracting the experienced backers; it exerts a negative impact on the number of the experienced backers when relatively small, but a positive impact when large enough.

New backers lack enough social capital and are less able to make connections between disparate ideas originating from different cultures. Furthermore, they are less likely to inspire creative thinking or break away from structured and routine ways of approaching problems. In order to adapt to cultural differences, new backers need to pay more physical and mental costs. Because many investors will encounter culture shock, when facing strange stimulus, people are easy to lose the basic ability to understand problems and even distinguish things and choose to escape and return. Therefore, we propose the hypothesis as follows:

> H3b: cultural distance has no significant effect in attracting the new backers

The entry barrier for launching projects on crowdfunding platforms is very low, so most new backers are not competent to compare and optimize choices in so many projects. This challenge can enlarge information asymmetry between creators and backers in the crowdfunding markets [18]. Also, new backers lack the information resource or social support from the acquired social capital on the online platform, so they cannot accurately evaluate the potential benefits and risks of competing crowdfunding projects [11]. Therefore, preference for geographically proximate projects seems to be reasonable for naive backers. Therefore, we postulate that geographic distance exerts differential effects on experienced vs. new backers.

> H3c: whereas large geographic distance discourages the new backers from contributing, it encourages the experienced backers

3. Methodology

3.1. Sample. Kickstarter is the largest reward-based crowdfunding provider worldwide [28], and data from the platform have been used in several prior studies (e.g., [1, 6, 10, 14, 31]). Kickstarter is appropriate for this study because first, the "creative" or "artistic" nature of creative ideas described by Kickstarter projects makes different cultural backgrounds salient in appreciation and judgment. Second, Kickstarter presents all project information such as resident city, nationality, funding and fundraising history, the funds raised thus far, number of backers who have contributed, and the frequency distribution of the backers. Till August 17th 2021, its several million community members have pledged \$6,033,885,042 to fund 206,813 creative ideas. Among 20,068,256 total backers, 33.79% backers have backed two or more projects.

This study uses Python to collect the real data generated by creators and backers from kickstarter.com. Specifically, this study grabs the key fields of project which ended in 2017 and get the data of 264 fundraising projects, among which 175 succeeded and 89 failed.

3.2. Data Descriptions. We summarize in Table 1 the project characteristics for the whole sample and the successful and unsuccessful subsamples, respectively. The average fundraising cycle is 34 days, and those successful ones tend to have shorter cycles (32 vs. 37 days). On average, the projects receive funds (\$189,77) much more than they request (\$14,895). Those successful projects tend to receive more contributions (\$26,477) than the requested amount (\$11,027), while those unsuccessful ones usually set higher targets (\$22,501) but receive much fewer contributions (\$4,485).

For each project, the "community" section demonstrates the backers' composition, including the frequency distribution of backer origins and the number of new and experienced backers, respectively. The summary of cultural distance, geographic distance, and percentage of experienced backers are given in Table 1. The projects can appeal to the experienced backers a bit more (59%). The focal variable we are interested in is the cultural distance between the creator and backers. Following Cho and Kim [54], we gauge the cultural distance between the creator and the top 10 countries (along with the number of backers from each of the 10 countries) based upon the six-dimension national culture model by Hofstede [33]. We aggregate the distance scores, which is weighted by the frequency of each country,

Table 1: Statistical summary of the sample.

	Total sample (264)			Successful (175)			Unsuccessful (89)		
	Mean	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.
Fundraising cycle (day)	34	7	379	32	7	61	37	10	379
Target (USD)	14,895	1	165,000	11,027	1	160,000	22,501	400	165,000
Funded amount (USD)	189,77	0	506,351	26,477	67	506,351	4,485	0	50,010
% of experienced backers	0.59	0	1	0.59	0	1	0.57	0.05	1
Cultural distance	55.84	42.13	69.8	55.66	42.13	62.95	56.24	53.20	69.8
Geographic distance (km)	2,962	8	13,082	2,905	15	13,082	3,090	8	9,637

on six available dimensions (https://geert-hofstede.com/cultural-survey.html). To measure the geographic distance between the creator and the top 10 cities, we use the interface of Google Maps APIs to obtain the latitudes and longitudes for each city. We manually measure the straight-line geographic distance for each pair of cities.

Besides the cultural and geographic distance, we include several essential covariates. The project duration, which indicates the degree of project exposure, offers for awareness and attention-building and promote the project performance [20]. The fundraising target usually set up a shared group goal among backers. When group identification is relatively weak (as in crowdfunding community with anonymous members), individuals decide to pursue the shared group goal if believed to be worthwhile. Thus, both the fundraising cycle and project target should play a role in crowdfunding decisions. Additionally, researchers have illustrated that the personal characteristics of creators lead to discriminations in the crowdfunding market [12, 55]. The number of initiated projects and that of successfully funded ones of a particular creator usually indicate the sophistication of creators in the market and thus can influence the crowdfunding project attractiveness. All the relevant variables are defined in detail in Table 2.

3.3. Empirical Models and Results. To test the effect of cultural and geographic distance on crowding project attractiveness, we build a series of empirical models. Although we propose a quadratic relationship between the cultural distance and project attractiveness, we include the linear relationship as the baseline models.

3.3.1. Effect of Cultural and Geographic Distance on Crowdfunding Project Attractiveness. As Kickstarter adopts the rule of "All-or-Nothing" support, backers can successfully contribute to a project only when the requested amount has been reached. Thus, this article measures the crowdfunding project attractiveness by whether the project is successfully funded and average fund each backer invests. The regression models are shown in model^1 and model^2, respectively. Particularly, as the dependent variable in model^1 is a dummy to indicate when a project is successfully funded (i.e., 1) or not (i.e., 0), we adopt the probit regression models. Please note that we also examine the linear relationship between the cultural and geographic distance and project attractiveness as the baseline models in $\text{model}^1_{\text{baseline}}$ and $\text{model}^2_{\text{baseline}}$.

$$
\begin{aligned}
\text{Prob}\,(\text{success}) = {} & \alpha_0^1 + \beta_1^1 \text{culture distance} + \beta_2^1 \text{culture distance}^2 + \beta_3^1 \log\,(\text{geographic distance}) \\
& + \gamma_1^1 \text{requested amount} + \gamma_2^1 \text{fundraising cycle} + \gamma_3^1 \text{historical projects} \\
& + \gamma_4^1 \text{historical success} \\
& + \gamma_5^1 \text{total backers} + \gamma_6^1 \text{naive backers} + \varepsilon^1, \quad \left(\text{model}^1\right). \\
\text{Average fund} = {} & \alpha_0^2 + \beta_1^2 \text{culture distance} + \beta_2^2 \text{culture}^2 + \beta_3^2 \log\,(\text{geographic distance}) \\
& + \gamma_1^2 \text{requested amount} + \gamma_2^2 \text{fundraising cycle} + \gamma_3^2 \text{historical projects} + \gamma_4^2 \text{historical success} \\
& + \gamma_5^2 \text{total backers} + \gamma_6^2 \text{naive backers} + \varepsilon^2, \quad \left(\text{model}^2\right).
\end{aligned}
\tag{1}
$$

As displayed in the first column of Table 3, the cultural distance ($-.132^{**}$, $\text{SE} = 0.054$) does not have a significant linear effect on crowdfunding project success. However, the estimation results for the quadratic relationships shown in the second column reveal that the cultural distance has a significant U-shape relationship with the crowdfunding success (-1.453^{***}, $\text{SE} = 0.703$; 0.012^{**}, $\text{SE} = 0.006$). When the cultural distance is relatively small, the larger the average cultural distance between the backers and the project creator, the less attractive the artistic crowdfunding project, and thus the lower the success likelihood. When the cultural distance is large to some extent, the more culturally distant projects become more appealing and are more likely to succeed. Nevertheless, the geographic distance does not

Table 2: Definitions of the variables.

Variables	Type and unit	Definition
Dependent variables		
Success or failure	Dummy coded	1 if a project succeeds and 0 otherwise
Average fund	Ratio variable; USD	The average amount per backer contributes for a particular project, calculated by the total amount a project raises divided by the number of total backers
The number of experienced backers	Counting variable	The number of experienced backers funding a particular project
The number of new backers[a]	Counting variable	The number of new backers funding a particular project
Independent variables		
Cultural distance	Ratio variable	The average cultural distance between the creator and the funding backers based on the six-dimension culture model
Geographic distance	Ratio variable; kilometer.	The average geographic distance between the creator and the funding backers based on the Google Maps APIs
Covariates		
The number of total backers	Counting variable	The number of total backers funding a particular project
Requested amount	Ratio variable; USD	The specified amount a creator requests for the initiated project
Fundraising cycle	Ratio variable; day.	The specified days for fundraising
Historical projects	Counting variable	The number of projects a creator previous initiated
Historical success	Counting variable	The number of successful projects previously initiated by a creator

[a]The number of total backers and the number of naïve backers included as control variables in the regressions on project attractiveness.

Table 3: The effects of cultural and geographic distance on crowdfunding project attractiveness.

	$\text{Model}^1_{\text{baseline}}$	Model^1	$\text{Model}^2_{\text{baseline}}$	Model^2
	DV, success or failure		DV, the average fund per backer	
Cultural distance	−0.131** (0.054)	−1.446* (0.813)	−7.397 (5.781)	−241.640*** (67.189)
Cultural distance^2	—	0.0113** (0.562)	—	2.087*** (0.596)
Log (geographic distance)	−0.023 (0.099)	−0.030 (0.099)	3.871 (10.043)	6.326 (9.839)
Requested amount	$-3.93\acute{e}(-5)$*** ($7.10\acute{e}(-6)$)	$-4.33\acute{e}(-5)$*** ($8.25\acute{e}(-6)$)	0.004*** (0.0005)	0.004*** (0.0005)
Fundraising cycle	−0.006 (0.005)	−0.006 (0.005)	−0.001 (0.512)	−0.219 (0.500)
Historical projects	0.185* (0.086)	0.176** (0.086)	−0.322 (3.045)	−0.219 (2.975)
Historical success	0.011 (0.009)	0.011 (0.009)	−0.150 (0.325)	−0.109 (0.318)
Total backers	0.004** (0.002)	0.005** (0.002)	−0.085* (0.044)	−0.077* (0.043)
New backers	0.033*** (0.008)	0.032*** (0.008)	0.047 (0.113)	−0.030 (0.111)
Constant	7.356 (3.271)	45.606 (29.291)	459.033 (340.880)	7002.579*** (1899.579)
	Prob > chi^2 = 0.000 Pseudo R^2 = 0.385	Prob > chi^2 = 0.000 Pseudo R^2 = 0.390	Adjusted R^2 = 0.170	Adjusted R^2 = 0.2067

***Significant at 0.01. **Significant at 0.05. *Significant at 0.10.

exert any significant effect on the success probability of crowdfunding projects either in $\text{model}^1_{\text{baseline}}$ (−0.022, SE = 0.099) or in model^1 (−0.029, SE = 0.100).

The last two columns of Table 3 also reveal how the cultural and geographic distance influence the backers' decisions of contribution amount. Similarly, although the cultural distance does not show a significant effect (−7.402, SE = 5.777) on the average contributed amount in the baseline model ($\text{model}^2_{\text{baseline}}$), it demonstrates a significant U-shape impact in the quadratic relationship (−240.606***, SE = 67.149; 2.078***, SE = 0.596). When the cultural distance between the backers and project creator is moderate, the backers may perceive the project to be less creative and less attractive due to the somewhat similar cultures and thus be more reluctant to contribute. Therefore, the attractiveness of artistic crowdfunding project is negatively related to the cultural distance. The result was in accord with Kang et al.' [26] contention and support H1a. However, when the cultural distance is large enough, backers are exposed to the information which is brand new or strikingly different from the conventions and routines of their home culture. Accordingly, they may consider the crowdfunding project to be eye catching and innovative and pay more attention to the project, rendering them to contribute considerable funds. Therefore, H1b is support. Again, the geographic distance does not exert any significant effect on the average amount a backer is willing to contribute either in $\text{model}^2_{\text{baseline}}$ (4.367, SE = 10.018) or in model^2 (6.836, SE = 9.817).

3.3.2. Effect of Cultural and Geographic Distance on Experienced vs. New Backers. We postulate that the social capital the backers acquire from their community participation of crowdfunding platform is one of the plausible mechanisms that may explain the effects of cultural distance. We empirically disclose this potential by investigating the asymmetric effects of cultural distance on experienced vs. new backers. Meanwhile, we investigate whether geographic distance shows any significant effect on experienced vs. new backers. For comparison purpose, we add the model for their impact on the total number of backers as well. On Kickstarter, each project discloses not only the frequency distribution of backer origins but also the number of new backers and that of experienced backers. As the three dependent variables are counting variables, the negative binomial regression modes are employed in model^3, model^4, and model^5 (as well as their baseline models) as follows:

$$
\begin{aligned}
\text{The total number of backers} = {} & \alpha_0^3 + \beta_1^3\text{culture distance} + \beta_2^3\text{culture distance}^2 + \beta_3^3\log(\text{geographic distance}) \\
& + \gamma_1^3\text{requested amount} + \gamma_2^3\text{fundraising cycle} + \gamma_3^3\text{historical projects} + \gamma_4^3\text{historical success} \\
& + \gamma_5^3\text{total backers} + \gamma_6^3\text{naive backers} + \varepsilon^3, \quad \left(\text{model}^3\right). \\
\text{The number of experienced backers} = {} & \alpha_0^4 + \beta_1^4\text{culture distance} + \beta_2^4\text{culture distance}^2 \\
& + \beta_3^4\log(\text{geographic distance}) + \gamma_1^4\text{requested amount} + \gamma_2^4\text{fundraising cycle} \\
& + \gamma_3^4\text{historical projects} + \gamma_4^4\text{historical success} + \gamma_5^4\text{total backers} + \gamma_6^4\text{naive backers} + \varepsilon^4, \quad \left(\text{model}^4\right). \\
\text{The number of new backers} = {} & \alpha_0^5 + \beta_1^5\text{culture distance} + \beta_2^5\text{culture distance}^2 + \beta_3^5\log(\text{geographic distance}) \\
& + \gamma_1^5\text{requested amount} + \gamma_2^5\text{fundraising cycle} + \gamma_3^5\text{historical projects} + \gamma_4^5\text{historical success} \\
& + \gamma_5^5\text{total backers} + \gamma_6^5\text{naive backers} + \varepsilon^5, \quad \left(\text{model}^5\right).
\end{aligned}
\tag{2}
$$

The estimation results for model^3, model^4, and model^5 as well as their baseline models are given in Table 4. The cultural distance does not exert a significant effect on the total number of backers either in the linear form in $\text{model}^3_{\text{baseline}}$ (-0.063, SE = 0.042) or in the quadratic form in model^3 (-0.465, SE = 0.563; $-4.64\hat{e}(-3)$, SE = $4.95\hat{e}(-3)$). Surprisingly, the geographic distance manifests a significant positive effect in attracting more backers to contribute (0.226, SE = 0.059). This result is consistent with what Mollick [1] and Kang et al. [26] advocate but contradicts with the negative effect identified by Lin and Viswanathan [21] and Burtch et al. [25].

To empirically infer the social capital the backers acquire from community interactions as the underlying mechanism, the aforementioned effects of cultural distance should be further qualified by their asymmetric effects on the new and experienced backers. As indicated by the fifth and seventh columns, the U-shape relationship exists of the cultural distance only for the experienced backers (-0.991^{***}, SE = 0.341; $.009^{***}$, SE = 0.003) but does not for the new backers (-0.065, SE = 0.379; $-5.36\hat{e}(-4)$, SE = $3.43\hat{e}(-3)$). The asymmetric impacts of cultural distance on the experience vs. new backers are intriguing. It may imply that, for it to play a significant role in funding decisions, culture distance should become salient. Our findings disclose that, only if backers acquire social capital from the platform interactions and become experienced in evaluating the attractiveness of creative ideas, they are more likely to be able to make the complex tradeoff between the uncertainty and perceived creativity engendered by cultural distance.

Interestingly, geographic distance exerts a negative effect in attracting the naive backers (-0.323^{***}, SE = 0.054) but a positive effect in attracting experienced backers (0.137^{***}, SE = 0.045). Large geographic distance discourages the backers who seldom have contributing experience on Kickstarter, probably because the lack of a common set of beliefs and expression systems hampers the communication and understanding between each other, reducing the uncertainty of cooperation. Also, small distance indicates a low cost of project knowledge search and acquisition, but large distance could increase the cost of cooperation and innovation through increased communication time and traffic distance. Therefore, naive backers are more likely to contribute to the crowdfunding projects initiated by geographically proximate creators.

However, for those experienced backers, they have benefited from online community participation and accumulated social capital from their previous contributions and they believe online community members from different countries or areas would help each other. Therefore, trust and knowledge exchange of experienced backers are not constrained by geographic distance; they would not consider geographic distance convenience as the most important factor, but focus on the benefit of social capital brought by far project rather than nearby projects. Also, excessive proximity could reduce the learning range from each other, so they reduce innovation enthusiasm. Thus, those experienced backers are more attracted by the creative ideas

Table 4: The effect of cultural and geographic distance on backer composition.

	Model$^3_{\text{baseline}}$	Model^3	Model$^4_{\text{baseline}}$	Model^4	Model$^5_{\text{baseline}}$	Model^5
	DV, the total number of backers		DV, the number of experienced backers		DV, the number of new backers	
Cultural distance	−0.062 (0.042)	−0.459 (0.565)	−0.048 (0.041)	−0.994*** (0.341)	0.006 (0.029)	−0.063 (0.381)
Cultural distance^2	—	−4.58é(−3) (4.97é(−3))	—	0.009*** (0.003)	—	−5.14é(−4) (3.45é(−3))
Log (geographic distance)	0.235*** (0.059)	0.227*** (0.059)	0.147*** (0.046)	0.140***(0.045)	−0.319*** (0.053)	−0.321*** (0.054)
Requested amount	2.76é(−5)*** (5.09é(−6))	2.79é(−5)*** (5.14é(−6))	4.72é(−6)* (2.57é(−6))	5.04é(−6)* (2.59é(−6))	9.84é(−6)*** (3.52é(−6))	9.91é(−6)*** (3.56é(−6))
Fundraising cycle	−2.97é(−3) (4.79é(−3))	−2.62é(−3) (4.94é(−3))	−5.51é(−3) (4.12é(−3))	−5.63é(−3) (4.00é(−3))	3.14é(−3) (3.26é(−3))	3.15é(−3) (3.26é(−3))
Historical projects	8.66é(−4) (1.50é(−2))	7.24é(−4) (0.015)	0.052** (0.023)	0.052** (0.023)	−0.014 (0.011)	−0.014 (0.011)
Historical success	1.16é(−3) (2.39é(−3))	1.16é(−3) (2.38é(−3))	4.26é(−3)* (2.30é(−3))	4.09é(−3)* (2.24é(−3))	−4.10é(−3)*** (1.52é(−3))	−4.10é(−3)*** (1.52é(−3))
Total # of backers	—	—	2.77é(−3)*** (3.00é(−4))	2.74é(−3)*** (2.97é(−4))	1.96é(−3)*** (2.27é(−4))	1.96é(−3)*** (2.27é(−4))
Constant	6.430*** (2.512)	−8.320 (16.044)	5.112** (2.328)	−24.35** (9.71)	4.898*** (1.692)	3.351 (10.488)
	Prob > chi^2 = 0.000 Pseudo R^2 = 0.0291	Prob > chi^2 = 0.000 Pseudo R^2 = 0.0294	Prob > chi^2 = 0.000 Pseudo R^2 = 0.1037	Prob > chi^2 = 0.000 Pseudo R^2 = 0.1061	Prob > chi^2 = 0.000 Pseudo R^2 = 0.1032	Prob > chi^2 = 0.000 Pseudo R^2 = 0.1035

***Significant at 0.01. **Significant at 0.05. *Significant at 0.10.

proposed by geographically distant creators. Therefore, the discussions on the asymmetric effects of geographic distance enrich the understanding of the effect of geographic distance in the debating literature and empirically advocate the social capital as the plausible mechanism.

4. Discussion and Conclusions

By using a sample of 264 crowdfunding projects on Kickstarter.com, this study investigates the impact of cultural and geographic distance on crowdfunding performance and unveils several exciting findings. First, cultural distance exerts a U-shape effect, which initially impedes the crowdfunding performance but promotes projects when large enough. Specifically, cultural distance exerts a negative impact when it is relatively low but a positive impact when large enough on the success of crowdfunding project and the average contribution amount of backers. Second, this study reveals that geographic distance exerts an insignificant impact on either project success or the average contribution amount. Third, cultural and geographic distance exerts the asymmetric effects on experienced versus new backers. Cultural distance has a U-shape effect in attracting experienced backers, but no effect in attracting new backers. But projects with large geographic distance appeal to the experienced backers but discourage new backers from contributing.

4.1. Theoretical Contributions. This study differs from prior research along several vital dimensions. First, it is one of the scarce studies to understand the implications of cultural distance on reward-based crowdfunding thoroughly. In crowdfunding, transactions are mediated on platforms which increase information asymmetry between backers and the creators. Backers evaluate the information provided by the creators and contribute more if the creator delivers more valuable signals that alleviate information asymmetry. The primary literature focuses on the project quality signals [e.g., [7]], creator-trustworthiness signals [e.g., [8]] or social influence during the dynamic fundraising cycle [e.g., [5]], but ignores the role of cultural distance [25].

Different from the only two existing studies on cultural distance [i.e., [1, 25]], the U-shape relationship between the cultural distance and crowdfunding performance are original and creative to the crowdfunding literature, and backers' complicated tradeoff between uncertainty and perceived creativity drives the U-shape effect. When evaluating a project from a similar culture, backers tend to reduce the imagination of potential, pay more attention to the ambiguity, and thus feel not so attractive. When evaluating a project embodying some characteristics of different cultures, backers tend to pay particular attention to the project and inspire the creative content of the mind. Thus, their perceived creativity overcomes the uncertainty engendered.

Second, this study responds to previous disputes on geographic distance effect and reveals the complex effects of distance on crowdfunding. Rather than the "home bias" proposition [21], this study supports the "flat world" hypothesis [24] and shows that geographic distance does not significantly affect project success or average contribution amount. Noticeably, geographic distance implicitly influences the backer composition by successfully encouraging experienced backers' investment but discourages new backers.

Third, our findings offer new avenues for research toward understanding how cultural and geographic distance determines our choices and actions. This study innovatively and empirically infers the social capital as one plausible mechanism for the effects of cultural and geographic distance by showing the asymmetric effects on experienced versus new backers, while previous literature ignores how cultural distance and geographic distance influence the backer composition. The social capital that experienced backers acquired and accumulated from the Internet-mediated interactions [26, 31] helps reduce the uncertainty of the project and synthesize novel and useful insights. Therefore, experienced backers are more able to formulate funding decisions by balancing the uncertainty and perceived creativity. Contrarily, due to the lack of experience or social connections on the platform, new backers may feel difficult to communicate or assimilate essentially different ideas originating from different cultures. Furthermore, naïve bakers are less likely to get rid of rigid and fixed thinking way. Home bias exists in their decision-making process, making them show a local preference.

4.2. Practical Implications. From a practical point of view, creators can make use of the U-shape effect of cultural distance to adjust the pledging and project characteristics to increase the chance that a project is successfully funded. Creators should construct project descriptions of cultural distance to meet the needs of experienced backers. Based on the profile data (e.g., funding experience, background cultures, or geographic areas in this study), creators may disclose information (e.g., the number of Facebook connections, backers with similar interest), which help to influence the other signals sent in context of the project campaign. Creators could present the high degree of cultural distance by demonstrating the cultural diversity by photos or videos. The relevant and detailed textual descriptions or media content of projects can signal preparedness and seriousness to potential investors [56].

Second, both crowdfunding platforms and project creators should enhance Internet-mediated interactions to cultivate social capital for backers, which help them overcome the fractions caused by cultural and geographic distance. Social capital investment is a long-term process; therefore, creators may link the project to social media platforms such as Facebook or LinkedIn and actively interact with potential backers via two-way communication on Kickstarter to interact with potential backers to promote the project. The platform should encourage community participation, supporting backers in sharing, learning, problem-solving, and the opportunity to socialize. Besides, to attract experienced investors, creators may highlight the geographical distance of the project and set the distance unit to meters instead of kilometers when displaying the project description. However, to attract new backers, creators may weaken the geographical distance of the project when displaying the project, such as reducing the number used to indicate the distance.

Third, creators especially from start-up firms can use the Internet as a channel to promote their projects regardless of geographic distance. Internet gives potential backers the chance to process relevant information for the investment decision at a low cost, while also providing an opportunity to chat with creators. It is for this reason that creators should consider crowdfunding as a resource to cost-effectively bridge geographic boundaries and link investment opportunities.

4.3. Limitations and Future Research Directions. This study unavoidably suffers from several limitations that open avenues for further research. First, while the cultural background of bakers and creators becomes salient for reward-based crowdfunding for artistic projects, the main findings need caution when generalizing to other types of projects (e.g. fast-moving consumer goods and luxury goods). Second, using data from reward-based crowdfunding raises concerns about the generalizability of our findings; thus, future research may collect data from multiple platforms. Third, although we used econometric models to empirically infer that the social capital experienced backers acquire from their previous platform participation drives the asymmetric effects of both cultural and geographic distance; further research may replicate this study in a more controlled experiment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was funded by Humanities and Social Sciences Research Projects of Anhui Province (SK2019A0698) and Talent Research Projects of Hefei University (20RC74).

References

[1] E. Mollick, "The dynamics of crowdfunding: an exploratory study," *Journal of Business Venturing*, vol. 29, no. 1, pp. 1–16, 2014.

[2] T. Hildebrand, M. Puri, and J. Rocholl, "Adverse incentives in crowdfunding," *Management Science*, vol. 63, no. 3, pp. 587–900, 2016.

[3] J. Hollas, "Is crowdfunding now a threat to traditional finance?" *Corporate Finance Review*, vol. 18, no. 1, p. 27, 2013.

[4] C. Courtney, S. Dutta, and Y. Li, "Resolving information asymmetry: signaling, endorsement, and crowdfunding success," *Entrepreneurship: Theory and Practice*, vol. 41, no. 2, pp. 265–290, 2017.

[5] J. Zhang and P. Liu, "Rational herding in microloan markets," *Management Science*, vol. 58, no. 5, pp. 892–912, 2012.

[6] N. Steigenberger and H. Wilhelm, "Extending signaling theory to rhetorical signals: evidence from crowdfunding," *Organization Science*, vol. 29, no. 3, pp. 529–546, 2018.

[7] S. Bi, Z. Liu, and K. Usman, "The influence of online information on investing decisions of reward-based crowdfunding," *Journal of Business Research*, vol. 71, pp. 10–18, 2017.

[8] R. Iyer, A. I. Khwaja, E. F. P. Luttmer, and K. Shue, "Screening peers softly: inferring the quality of small borrowers," *Management Science*, vol. 62, no. 6, pp. 1554–1577, 2016.

[9] J. Michels, "Do unverifiable disclosures matter? evidence from peer-to-peer lending," *The Accounting Review*, vol. 87, no. 4, pp. 1385–1413, 2012.

[10] A. Parhankangas and M. Renko, "Linguistic style and crowdfunding success among social and commercial entrepreneurs," *Journal of Business Venturing*, vol. 32, no. 2, pp. 215–236, 2017.

[11] G. K. C. Ahlers, D. Cumming, C. Günther, and D. Schweizer, "Signaling in equity crowdfunding," *Entrepreneurship: Theory and Practice*, vol. 39, no. 4, pp. 955–980, 2015.

[12] M. Lin, N. R. Prabhala, and S. Viswanathan, "Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending," *Management Science*, vol. 59, no. 1, pp. 17–35, 2013.

[13] J. Duarte, S. Siegel, and L. Young, "Trust and credit: the role of appearance in peer-to-peer lending," *Review of Financial Studies*, vol. 25, no. 8, pp. 2455–2484, 2012.

[14] T. H. Allison, B. C. Davis, J. W. Webb, and J. C. Short, "Persuasion in crowdfunding: an elaboration likelihood model of crowdfunding performance," *Journal of Business Venturing*, vol. 32, no. 6, pp. 707–725, 2017.

[15] J. Cox, T. Nguyen, A. Thorpe, A. Ishizaka, S. Chakhar, and L. Meech, "Being seen to care: the relationship between self-presentation and contributions to online pro-social crowdfunding campaigns," *Computers in Human Behavior*, vol. 83, pp. 45–55, 2018.

[16] V. Kuppuswamy and B. L. Bayus, "Does my contribution to your crowdfunding project matter?" *Journal of Business Venturing*, vol. 32, no. 1, pp. 72–89, 2017.

[17] B. C. Davis, K. M. Hmieleski, J. W. Webb, and J. E. Coombs, "Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: the influence of perceived product creativity and entrepreneurial passion," *Journal of Business Venturing*, vol. 32, no. 1, pp. 90–106, 2017.

[18] N. Wang, Q. Li, H. Liang, T. Ye, and S. Ge, "Understanding the importance of interaction between creators and backers in crowdfunding success," *Electronic Commerce Research and Applications*, vol. 27, pp. 106–117, 2018.

[19] A. Agrawal, C. Catalini, and A. Goldfarb, "Crowdfunding: geography, social networks, and the timing of investment decisions," *Journal of Economics and Management Strategy*, vol. 24, no. 2, pp. 253–274, 2015.

[20] G. Burtch, A. Ghose, and S. Wattal, "An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets," *Information Systems Research*, vol. 24, no. 3, pp. 499–519, 2013.

[21] M. Lin and S. Viswanathan, "Home bias in online investments: an empirical study of an online crowdfunding market," *Management Science*, vol. 62, no. 5, pp. 1393–1414, 2016.

[22] S. Lai and M. Teo, "Home-biased analysts in emerging markets," *Journal of Financial and Quantitative Analysis*, vol. 43, no. 3, pp. 685–716, 2008.

[23] N. Strong and X. Xu, "Understanding the equity home bias: evidence from survey data," *The Review of Economics and Statistics*, vol. 85, no. 2, pp. 307–312, 2003.

[24] A. Agrawal, C. Catalini, and A. Goldfarb, "Friends, family, and the flat world: the geography of crowdfunding," *NBER Working Paper*, vol. 16820, pp. 1–61, 2011.

[25] G. Burtch, A. Ghose, A. Ghose, and S. Wattal, "Cultural differences and geography as determinants of online prosocial lending," *MIS Quarterly*, vol. 38, no. 3, pp. 773–794, 2014.

[26] L. Kang, Q. Jiang, and C.-H. Tan, "Remarkable advocates: an investigation of geographic distance and social capital for crowdfunding," *Information & Management*, vol. 54, no. 3, pp. 336–348, 2017.

[27] J. G. Lu, A. C. Hafenbrack, P. W. Eastwick, D. J. Wang, W. W. Maddux, and A. D. Galinsky, ""Going out" of the box: close intercultural friendships and romantic relationships spark creativity, workplace innovation, and entrepreneurship," *Journal of Applied Psychology*, vol. 102, no. 7, pp. 1091–1108, 2017.

[28] E. Mollick and R. Nanda, "Wisdom or madness? comparing crowds with expert evaluation in funding the arts," *Management Science*, vol. 62, no. 6, pp. 1533–1553, 2015.

[29] P. Belleflamme, T. Lambert, and A. Schwienbacher, "Crowdfunding: tapping the right crowd," *Journal of Business Venturing*, vol. 29, no. 5, pp. 585–609, 2014.

[30] J.-A. Koch and M. Siering, "The recipe of successful crowdfunding campaigns," *Electronic Markets*, vol. 29, no. 4, pp. 661–679, 2019.

[31] M. G. Colombo, C. Franzoni, and C. Rossi–Lamastra, "Internal social capital and the attraction of early contributions in crowdfunding," *Entrepreneurship: Theory and Practice*, vol. 39, no. 1, pp. 75–100, 2015.

[32] G. Hofstede, G. J. Hofstede, and M. Minkov, *Cultures and Organizations: Software of the Mind*, McGraw-Hill, New York, NY, USA, 2010.

[33] G. Hofstede, *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*, Sage Publications, Thousand Oaks, CA, USA, 2001.

[34] L. Tihanyi, D. A. Griffith, and C. J. Russell, "The effect of cultural distance on entry mode choice, international diversification, and MNE performance: a meta-analysis," *Journal of International Business Studies*, vol. 36, no. 3, pp. 270–283, 2005.

[35] R. Y. J. Chua, Y. Roth, and J.-F. Lemoine, "The impact of culture on creativity," *Administrative Science Quarterly*, vol. 60, no. 2, pp. 189–227, 2015.

[36] S.-H. Lee, O. Shenkar, and J. Li, "Cultural distance, investment flow, and control in cross-border cooperation," *Strategic Management Journal*, vol. 29, no. 10, pp. 1117–1125, 2008.

[37] Y. Luo and M. W. Peng, "Learning to compete in a transition economy: experience, environment, and performance," *Journal of International Business Studies*, vol. 30, no. 2, pp. 269–295, 1999.

[38] P. Morosini, S. Shane, and H. Singh, "National cultural distance and cross-border acquisition performance," *Journal of International Business Studies*, vol. 29, no. 1, pp. 137–158, 1998.

[39] S. H. Park and G. R. Ungson, "The effect of national culture, organizational complementarity, and economic motivation on joint venture dissolution," *Academy of Management Journal*, vol. 40, no. 2, pp. 279–307, 1997.

[40] T. H. Reus and B. T. Lamont, "The double-edged sword of cultural distance in international acquisitions," *Journal of International Business Studies*, vol. 40, no. 8, pp. 1298–1316, 2009.

[41] W. W. Maddux and A. D. Galinsky, "Cultural borders and mental barriers: the relationship between living abroad and creativity," *Journal of Personality and Social Psychology*, vol. 96, no. 5, pp. 1047–1061, 2009.

[42] C.-Y. Cheng and A. K.-Y. Leung, "Revisiting the multicultural experience-creativity link," *Social Psychological and Personality Science*, vol. 4, no. 4, pp. 475–482, 2013.

[43] M. Schulz, P. Haas, K. Schulthess, I. Blohm, and J. M. Leimeister, "How idea creativity and hedonic value influence project success in crowdfunding," in *Proceedings of the 12th International Conference On Wirtschaftsinformatik*, Osnabrück, Germany, March 2015.

[44] J. Nahapiet and S. Ghoshal, "Social capital, intellectual capital, and the organizational advantage," *Academy of Management Review*, vol. 23, no. 2, pp. 242–266, 1998.

[45] M. Ma and R. Agarwal, "Through a glass darkly: information technology design, identity verification, and knowledge contribution in online communities," *Information Systems Research*, vol. 18, no. 1, pp. 42–67, 2007.

[46] F. Olivera, P. S. Goodman, and S. S.-L. Tan, "Contribution behaviors in distributed environments," *MIS Quarterly*, vol. 32, no. 1, pp. 23–42, 2008.

[47] M. M. Wasko and S. Faraj, "Why should i share? examining social capital and knowledge contribution in electronic networks of practice," *MIS Quarterly*, vol. 29, no. 1, pp. 35–57, 2005.

[48] N. Hara, "Online communities of practice: beyond organizational boundaries," *Communities of Practice: Information Science and Knowledge Management*, Springer, Berlin, Germany, 2009.

[49] H. Zheng, D. Li, J. Wu, and Y. Xu, "The role of multidimensional social capital in crowdfunding: a comparative study in China and US," *Information & Management*, vol. 51, no. 4, pp. 488–496, 2014.

[50] S. Faraj and S. L. Johnson, "Network exchange patterns in online communities," *Organization Science*, vol. 22, no. 6, pp. 1464–1480, 2011.

[51] P. S. Adler and S.-W. Kwon, "Social capital: prospects for a new concept," *Academy of Management Review*, vol. 27, no. 1, pp. 17–40, 2002.

[52] C. Mathwick, C. Wiertz, and K. De Ruyter, "Social capital production in a virtual P3 community," *Journal of Consumer Research*, vol. 34, no. 6, pp. 832–849, 2008.

[53] A. K.-Y. Leung and C.-Y. Chiu, "Multicultural experience, idea receptiveness, and creativity," *Journal of Cross-Cultural Psychology*, vol. 41, no. 5-6, pp. 723–741, 2010.

[54] M. Cho and G. Kim, "A cross-cultural comparative analysis of crowdfunding projects in the United States and south Korea," *Computers in Human Behavior*, vol. 72, pp. 312–320, 2017.

[55] D. G. Pope and J. R. Sydnor, "What's in a picture?: evidence of discrimination from prosper.com," *Journal of Human Resources*, vol. 46, no. 1, pp. 53–92, 2011.

[56] A. Hoegen, D. M. Steininger, and D. Veit, "How do investors decide? an interdisciplinary review of decision-making in crowdfunding," *Electronic Markets*, vol. 28, no. 3, pp. 339–365, 2018.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 7915706, 8 pages
https://doi.org/10.1155/2021/7915706

Research Article

AIU-Net: An Efficient Deep Convolutional Neural Network for Brain Tumor Segmentation

Yongchao Jiang, [1,2] Mingquan Ye, [1,2] Daobin Huang, [1,2] and Xiaojie Lu [1,2]

[1] *School of Medical Information, Wannan Medical College, Wuhu 241002, China*
[2] *Research Center of Health Big Data Mining and Applications, Wannan Medical College, Wuhu 241002, China*

Correspondence should be addressed to Mingquan Ye; ymq@wnmc.edu.cn

Received 20 May 2021; Revised 23 July 2021; Accepted 29 July 2021; Published 4 August 2021

Academic Editor: Shianghau Wu

Automatic and accurate segmentation of brain tumors plays an important role in the diagnosis and treatment of brain tumors. In order to improve the accuracy of brain tumor segmentation, an improved multimodal MRI brain tumor segmentation algorithm based on U-net is proposed in this paper. In the original U-net, the contracting path uses the pooling layer to reduce the resolution of the feature image and increase the receptive field. In the expanding path, the up sampling is used to restore the size of the feature image. In this process, some details of the image will be lost, leading to low segmentation accuracy. This paper proposes an improved convolutional neural network named AIU-net (Atrous-Inception U-net). In the encoder of U-net, A-inception (Atrous-inception) module is introduced to replace the original convolution block. The A-inception module is an inception structure with atrous convolution, which increases the depth and width of the network and can expand the receptive field without adding additional parameters. In order to capture the multiscale features, the atrous spatial pyramid pooling module (ASPP) is introduced. The experimental results on the BraTS (the multimodal brain tumor segmentation challenge) dataset show that the dice score obtained by this method is 0.93 for the enhancing tumor region, 0.86 for the whole tumor region, and 0.92 for the tumor core region, and the segmentation accuracy is improved.

1. Introduction

Glioma is the most common brain tumor, and it is also the brain tumor with the highest mortality and morbidity. Accurate segmentation of gliomas is of great significance for the diagnosis and treatment of gliomas. Magnetic resonance imaging (MRI) is an important technical means to assist doctors in the diagnosis and treatment of brain tumors [1]. Various sequences of MRI can provide different brain tumor tissue structures, and it is usually combined with multimodal MRI of brain tumor to segment brain tumors. Because of the complexity of brain tumor structure, the fuzziness of tumor boundary, and the difference of different individuals, the accurate segmentation of brain tumor is a complicated and difficult task [2]. The traditional manual segmentation needs a lot of time for doctors to complete, and the segmentation accuracy is relatively rough. In recent years, the automatic segmentation method based on deep learning has achieved good results in medical image segmentation [3].

The deep learning method based on convolutional neural network performs well in computer vision tasks such as image classification, segmentation, and object detection. Convolutional neural network can automatically learn the complex features of data in the training process without relying on manual extraction of features, which can further improve the segmentation accuracy of brain tumors [4, 5]. Long et al. [6] proposed the full convolutional neural network (FCN), which transformed the full connection layer of the convolutional neural network into the convolutional layer and used up sampling to restore the output feature map to the same size as the input image, so as to achieve end-to-end semantic segmentation of the image. Ronneberger et al. [7] proposed U-net for biological cell segmentation; U-net is composed of contracting path and expanding path.

Contracting path includes convolution block for feature extraction and max-pooling for down sampling. Expanding path is composed of convolution block and up-sampling module. Features with the same resolution are fused by skip connection between contracting path and expanding path. U-net has a simple structure and can obtain better segmentation results in the case of small sample size of medical images. However, the application of U-net in brain tumor image segmentation still needs to be improved [8]. On the one hand, the contracting path of U-net uses the pooling layer to reduce the feature map to expand the receptive field. Continuous pooling operation may cause the loss of image details and affect the segmentation result. On the other hand, it is the difference of the size, shape, and location of brain tumor, and how to obtain more detailed features of segmentation target and how to obtain multiscale features are the important problems [9]. In order to solve the problem of gradient disappearance and network degradation with the increase of network depth, the residual network (ResNet) [10] is proposed. By adding identity mapping between the input and output of several convolutional layers, the network is easier to converge and prevent network degradation [11].

In order to obtain multiscale features, Chen et al. [12] proposed the DeepLab model. In this model, the last pooling layers were removed, and atrous convolutions were used to expand the receptive field. The atrous spatial pyramid pooling (ASPP) samples the given input in parallel with the atrous convolution of different dilation rates and then splices the results together. ASPP has a better effect on the extraction of multiscale features. Szegedy et al. [13] proposed inception network. Inception module can increase the width of the network. GoogLeNet, which is composed of inception module, obtained the best classification and detection performance in ILSVRC 2014 competition. This paper proposes to use atrous convolution to expand the receptive field and reduce the use of pooling layer, so as to reduce the loss of image details. Atrous convolution is to insert holes into the standard convolution kernel to expand the receptive field of feature extraction without additional parameters [14]. In this paper, atrous convolution and inception are combined to form a new structure named A-Inception module, and a new network architecture based on U-net is proposed. The encoder of this network adopts A-inception module to increase the depth and width of the network and obtain the different sizes of receptive field. At the same time, the atrous spatial pyramid pooling is added into the network to extract the multiscale features of the image.

2. Related Work

In recent years, methods based on convolutional neural networks have provided good performance in the field of computer vision. Compared with traditional methods, algorithms based on convolutional neural network can automatically learn the complex features of the original data without relying on manual extraction of features, which further improves the accuracy of image segmentation [15].

The framework of encoder-decoder is a common structure in image segmentation. In the encoding process, the pixels of the image are mapped to a high-dimensional distribution, and the decoding process is to gradually restore the details and spatial dimensions of the image. Therefore, the encoder-decoder structure can achieve the end-to-end semantic segmentation of the image [16]. SegNet [17] is a typical encoder-decoder network framework in image segmentation. The encoder network in SegNet has the same topology as the convolution layer in VGG16, but removes the fully connected layers. The network uses max-pooling to reduce the dimension of feature maps, and the decoder uses max-pooling indices received from the corresponding encoder to perform nonlinear up sampling of their input feature maps. U-net is also encoder-decoder structure and has been widely used in medical image segmentation. It adds skip connections between the encoder and the decoder, which are used to fuse the feature maps with the same resolution between the encoder and the decoder. Shaikh et al. [18] introduced dense connection and replaced the basic convolution module in U-net with dense connection module, which further improved the segmentation performance of the network. Oktay et al. [19] introduced the attention gates into the standard U-net architecture that automatically learns to focus on target structures of varying shapes and sizes. In the training process, the attention weight gradually tended to the target region, while the attention weight of the background region gradually decreased so that the segmentation accuracy was improved.

In the semantic segmentation of images, convolutional neural network uses pooling to realize down sampling, which reduces the image size and increases the receptive field and then uses up sampling to restore the original image size. In this process, some detail features of the image will be lost [20]. Atrous convolution can increase the receptive field without losing the image resolution, thus improving the accuracy of image semantic segmentation. Zhao et al. [21] proposed pyramid scene parsing network (PSPNET), which aggregates the context of different regions through pyramid pooling module and improves the ability of the network to obtain global information. In order to segment multiscale objects, DeepLabv3 [22] proposes to connect several atrous convolutions with different dilation rates in series and parallel, which can obtain larger receptive fields in cascade mode, and different receptive fields in the parallel mode for the same input, which can extract multiscale features better. DenseASPP [23] integrates atrous convolutions with different dilation rates through dense connection. Without the use of pooling operation, the receptive field of output neurons is expanded so that the output features cover a large range of semantic information and acquire multiscale features. DeepLabv3+ [24] is an extension of DeepLabv3, adding a simple decoder module to recover the object boundary details. The ASPP is an improvement on the basis of spatial pyramid pooling. For multiscale object segmentation, parallel pooling modules are designed to obtain multiscale features.

3. Methods

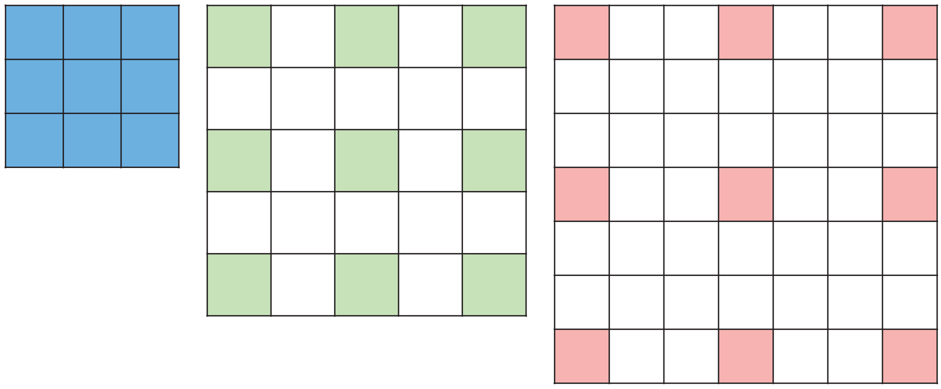
3.1. Atrous Convolution. When convolutional neural network is used for end-to-end semantic segmentation of images, down sampling will reduce the resolution of the feature maps, which can reduce the amount of computation and expand the receptive field. After that, the feature maps can be restored to the original image size through up sampling. In this process, some details related to the boundary of the segmentation object will be lost, resulting in the image segmentation results which are not accurate enough. Atrous convolution can control the resolution of features and adjust the size of receptive field to capture multiscale information. In fact, atrous convolution is to inject holes into the standard convolution kernel, and the dilation rate is used to define the interval of convolution kernel insertion [25]. The atrous convolution with the dilation rate of 1 is the same as the standard convolution. Figure 1 shows the atrous convolution with dilation rates of 1, 2, and 3, respectively. Compared with the ordinary convolution with convolution kernel of 3×3, the receptive field of atrous convolution is larger.

Increasing the depth and width of the network to improve network performance will bring a large number of parameters, which can easily lead to overfitting and increase the amount of calculation. The fundamental method to solve this problem is to keep the sparsity of neural network structure, but the computational efficiency of computer for nonuniform sparse data is very low. A large number of literatures show that the sparse matrix can be clustered into relatively dense submatrix to improve the computational performance. The main purpose of the inception structure is to use dense components to approximate the optimal local sparse structure. In this paper, a new module A-Inception is proposed. In this module, there are three branches in parallel, each branch uses different convolution kernel, instead of directly connecting convolution kernel in series, thus increasing the width of the network. In this module, atrous convolution is used to replace ordinary convolution. Different branches have different receptive fields. The convolution of different receptive fields is connected in parallel. Because the size of brain tumors is greatly different, receptive fields of different scales can reduce the fluctuation caused by the disturbance of brain tumor size, improve the robustness of neural networks, and obtain more detailed features at the same time. The BN layer is added after each convolution layer to avoid the gradient vanishing [26]. At the same time, inspired by the Inception-ResNet [27] module, the residual connection is added between the input and the output, which makes the network easier to learn and faster to converge. The specific model structure is shown in Figure 2.

3.2. ASPP. In Deeplabv2, atrous spatial pyramid pooling is proposed, which uses atrous convolutions with different dilation rates in parallel to obtain multiscale features of images. In DeepLabv3, the BN layer is added to the atrous spatial pyramid pooling, and global pooling is paralleled. Atrous convolutions with different dilation rates have different receptive fields for the same input, and these results can be stitched together to better capture the multiscale features of the image. In order to reduce the number of channels after splicing, the 1×1 convolution layer is connected. In [22], when the output-stride = 16 (output-stride is the ratio of input image spatial resolution to final output resolution), three 3×3 convolutions with rates = 6, 12, 18 are adopted in ASPP, while when the output stride = 8, the rates should be doubled. In this paper, three times downsampling is used in the encoder part, so the output stride is 8. The experiment proves that the segmentation effect is better when the atrous convolutions of 3×3 with rates = 12, 18, 24 are used. The ASPP module adopted in this paper is shown in Figure 3. In addition to three atrous convolutions with different rates, there is also a 1×1 convolution and a pooling layer in parallel.

3.3. Network Structure. In this paper, an improved brain tumor segmentation algorithm based on U-net is proposed. The encoder obtains the higher-level semantic information of the image, and the decoder gradually recovers the spatial information of the image. The encoder uses five A-inception modules. The first three A-inception modules use the atrous convolution with the dilation rate of 1, which is the standard 3×3 convolution, and then use the down-sampling module to reduce the feature resolution. The down-sampling modules use the max-pooling and the 3×3 convolution with the stride of 2 for down sampling the input, respectively, and then parallel the results of the two. In order to reduce the use of the pooling layer and prevent more loss of image details, the last two A-inception modules in the encoder use atrous convolution with larger dilation rates. The rate1, rate2, and rate3 are 2, 2, 4 and 4, 4, 8 in A-inception block 4 and A-inception block 5, respectively. It can not only expand the receptive field but also connect different receptive fields in parallel, which can better capture multiscale features and obtain more image details [28]. The ASPP module is used between the encoder and decoder. The decoder uses three residual blocks and bilinear interpolation up sampling to restore the feature maps to the same size as the input image. At the same time, the feature maps of the same resolution in the encoder and decoder are combined, and low-level features are introduced to increase the segmentation accuracy of spatial information features. The specific network structure is shown in Figure 4:

The optimization algorithm adopted in this paper is adaptive moment estimation (Adam) [29], which has the advantages of simple implementation, low memory requirement, and high computational efficiency. The loss function used in this paper is a linear combination of cross-entropy loss function and Dice loss function. The Dice loss function is suitable for the situation where the positive and negative samples are not balanced, and it focuses more on the prospects. However, if there are many small targets in the training data during the experiment, the loss curve is likely to oscillate. Therefore, this paper adopts the loss function of the Dice loss function combined with the cross-entropy loss



(a) (b) (c)

Figure 1: Atrous convolution with different dilation rates.

function, which can alleviate the problem of sample imbalance and obtain a smoother loss curve. The loss function is defined as follows:

$$\mathrm{Loss}_{\mathrm{CE}} = -\sum_{i}^{N} g_i \log\left(p_i\right),$$
$$\mathrm{Loss}_{\mathrm{Dice}} = 1 - \frac{2 \times \sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i}, \tag{1}$$
$$\mathrm{Loss} = 0.5\mathrm{Loss}_{\mathrm{CE}} + \mathrm{Loss}_{\mathrm{Dice}}.$$

The set N of all samples is calculated, where g_i is the thermal code (0 or 1) of the ith sample tag and p_i is the prediction probability of the ith sample tag.

3.4. Data Processing. The experimental data used in this paper are brain tumor challenge datasets BraTS2018 and BraTS2019. BraTS2018 dataset includes 210 HGG patients and 75 LGG patients, each of which includes T1 (T1 weighted), T2 (T2 weighted), T1c (contrast enhanced T1 weighted), and flair (fluid attenuated inversion recovery) four MRI sequences and ground truth labels [30]. These data were used as a training set for the experiment. BraTS2019 dataset added 49 HGG patients and 1 LGG patient on the basis of BraTS2018 dataset, and these data were used as the testing set of the experiment. The size of each modal MR image is $240 \times 240 \times 155$. The ground truth labels are the result of tumor manually labeled by 1 to 4 experts according to the same annotation protocol, including normal tissue (label 0), necrotic and nonenhancing tumor (label 1), edema (label 2), and enhancing tumor (label 4) [31].

In the data preprocessing method, firstly, the data are standardized, that is, subtracting the mean value and dividing by the standard deviation. Then, the redundant background in the original data is cropped to alleviate the problem of data imbalance [32]. Then, the three-dimensional images were sliced to obtain two-dimensional images, and the slices without lesions in the training set were discarded to alleviate the category imbalance. The slices of four MRI scan modalities are combined into four channel samples for training and testing data.

4. Results and Discussion

The experimental environment is Intel Xeon Silver 4116 CPU@2.10 GHz, GPU NVIDIA GeForce RTX2080Ti. After preprocessing the experimental data, a two-dimensional image sample with the size of 160×160 is obtained. 80% of the training set is used for model training, and 20% is used as the validation set to adjust parameters to monitor whether the model is overfitting. The testing set is used to verify the segmentation effect.

4.1. Evaluation Metrics. In order to quantitatively evaluate the segmentation performance of the proposed algorithm, the evaluation metrics used in this paper include the Dice similarity coefficient (DSC), intersection over union (IOU), and positive predictive value (PPV). These indicators were used to evaluate the experimental results [33]. The Dice similarity coefficient represents the similarity between the experimental segmentation results and the ground truth labels. DSC, PPV, and IOU are commonly used in image segmentation. The definitions are as follows:

$$\mathrm{DSC} = \frac{2\mathrm{TP}}{\mathrm{FP} + 2\mathrm{TP} + \mathrm{FN}},$$
$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{2}$$
$$\mathrm{IOU} = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{TP} + \mathrm{FN}},$$

where TP is true positive, FP is false positive, and FN is false negative. The range of the result is 0 to 1. The closer the test result is to 1, the more accurate the segmentation result is.

4.2. Experimental Results. In this paper, the results of glioma segmentation include the whole tumor region and tumor subregion, which are whole tumor region (WT), tumor core (TC), and enhancing tumor region (ET). ET is enhancing tumor, TC includes enhancing tumor and necrotic, and WT includes regions of enhancing tumor, necrotic, and edema. In this paper, an improved network architecture AIU-net

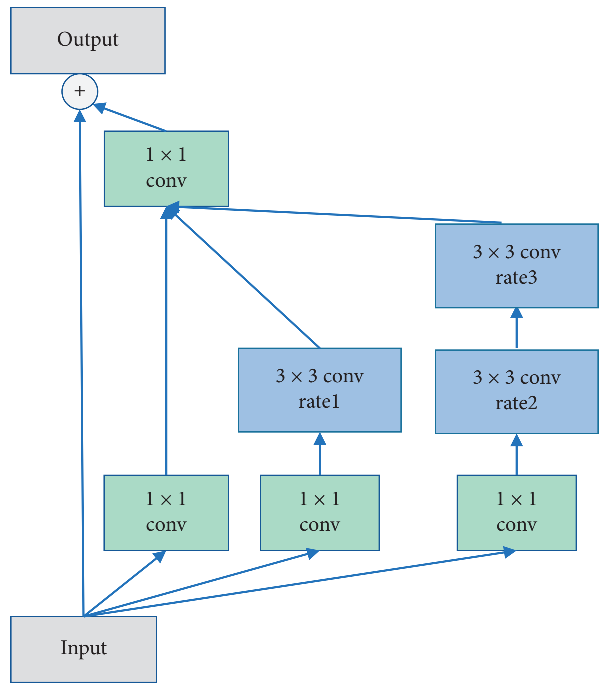


Figure 2: The structure of A-inception module.

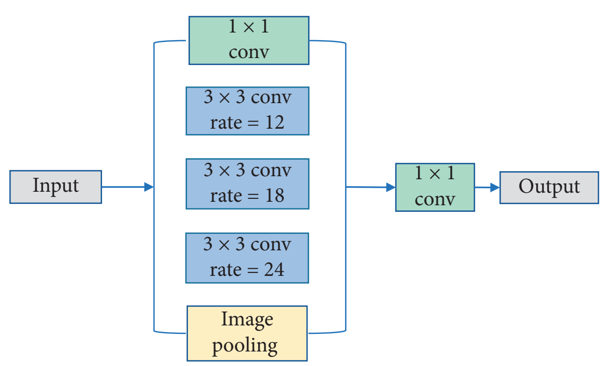


Figure 3: The structure of ASPP module.

based on U-net is proposed. In order to test the segmentation effect of this network on brain tumors and compare it with U-net and DeepLabv3+ network structure, the same dataset and parameters are used for training and testing on each network. The experimental results are shown in Figure 5. From the segmentation results shown in the figure, it can be seen that the method proposed in this paper is better than the other two methods in the segmentation of tumor details, and the segmentation results are closer to the ground truth. Figure 6(a) shows the change of the loss of the three networks with epoch in the training process. It can be seen from the figure that the loss value of AIU-net proposed in this paper is already less than 0.1 when the epoch was 50, and the convergence speed is faster than that of U-net and DeepLabv3+ networks. Figure 6(b) shows the change of IOU with epoch in the training process. It can be seen from the figure that when epoch was 50, the IOU value of AIU-net had exceeded 0.9, while the IOU value of the other two networks were all less than 0.9.

Table 1 shows the evaluation results of DSC, IOU, and PPV of several methods. The results show that the values of DSC, IOU, and PPV of the proposed method are higher than those of U-net and DeepLabv3+, and the values of DSC, IOU, and PPV of the whole tumor region are greatly improved compared with those of the other two methods. The segmentation results of the whole tumor region, enhancing tumor region, and tumor core region are better than those of the other two methods, and the segmentation performance is improved.

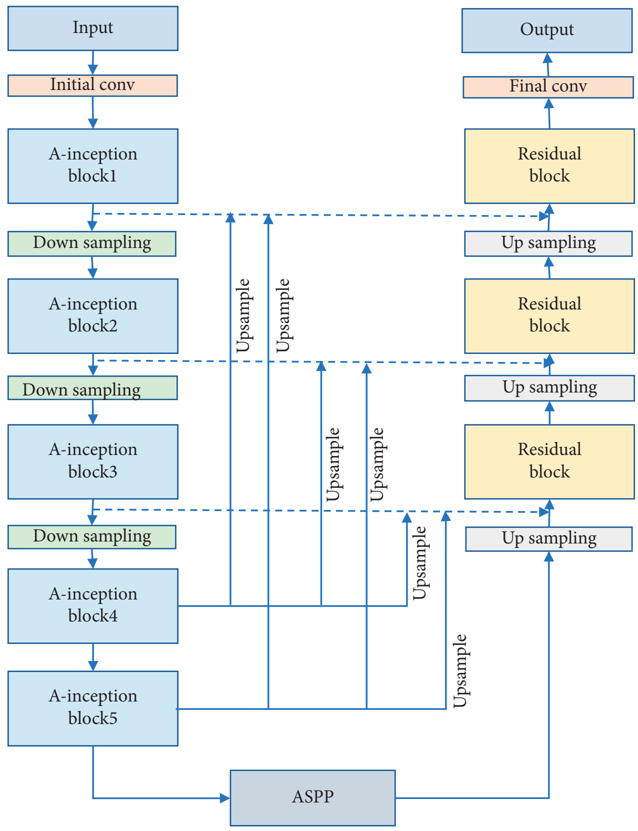


Figure 4: Overall architecture of the AIU-net model.

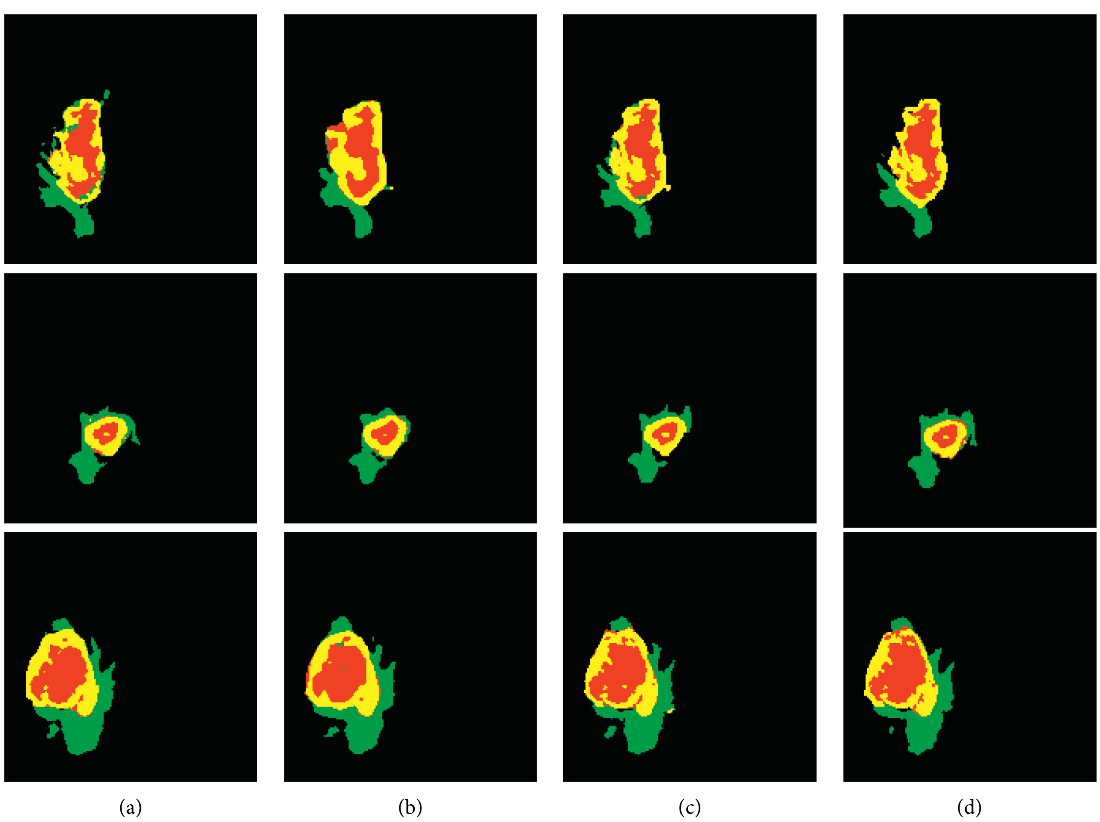


Figure 5: Example segmentation results on the BRATS dataset. From left to right are the segmentation results of U-net, deepLabv3 +, AIU-net, and ground truth. The whole tumor (WT) class includes all visible labels (a union of green, yellow, and red labels), the tumor core (TC) class is a union of red and yellow, and the enhancing tumor core (ET) class is shown in yellow. (a) U-net. (b) DeppLabv3+. (c) AIU-net. (d) GT.

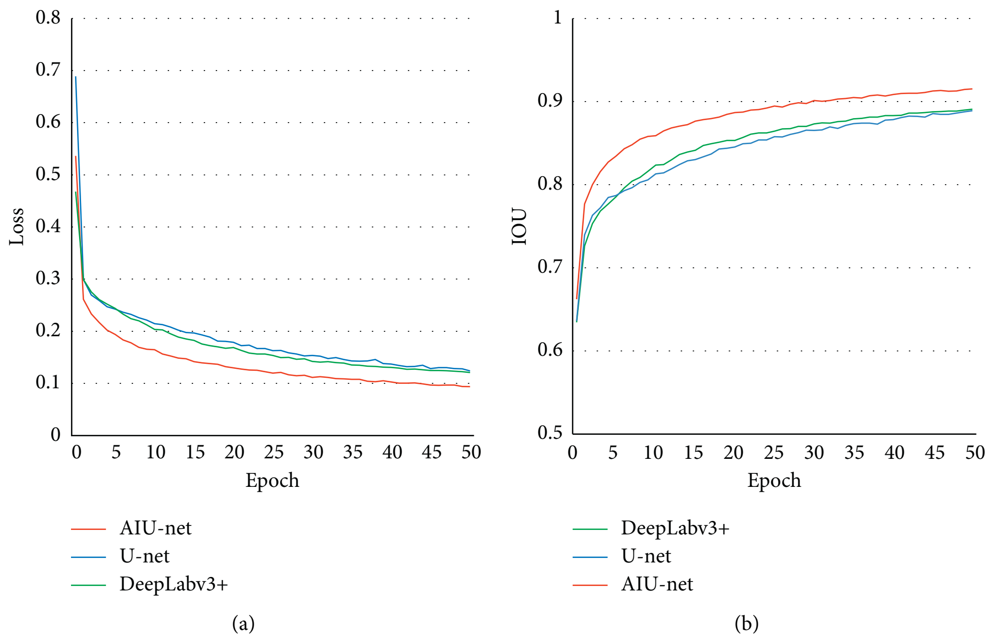


FIGURE 6: Training processes performance. (a) Loss of training process. (b) IOU of the training process.

TABLE 1: Segmentation performance of different models.

Model	Dice			IOU			PPV		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
U-net	0.9222	0.7696	0.9004	0.8998	0.7412	0.8805	0.9371	0.7828	0.9106
DeepLabv3+	0.8986	0.6819	0.8745	0.8697	0.6504	0.8521	0.9215	0.6922	0.8839
AIU-net	**0.9395**	**0.8696**	**0.9276**	**0.9170**	**0.8425**	**0.9079**	**0.9631**	**0.8927**	**0.9392**

5. Conclusion

In order to improve the accuracy of brain tumor automatic segmentation, this paper proposes a network architecture AIU-net based on U-net, which uses a new module combining inception and atrous convolution as encoder, and introduces ASPP module to obtain multiscale features. Experiments show that the new architecture AIU-net can effectively improve the accuracy of brain tumor segmentation and is conducive to multiscale information extraction. The segmentation of tumor details has been improved. In comparison with U-net and DeepLabv3+, the results of DSC, IOU, and PPV of the proposed method are better than those of other methods, and better segmentation performance is obtained. However, compared with U-net and DeepLabv3+, the training time and test time of the proposed method are longer mainly because the network architecture is more complex. The further work in the future is to obtain higher segmentation accuracy and better efficiency at the same time.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded and supported in part by the National Natural Science Foundation of China, under Grant 61672386, and the Anhui Provincial Natural Science Foundation of China, under Grant 1708085MF142.

References

[1] D. Ivana, G. Bart, and P. Wilfried, "MRI segmentation of the human brain: challenges, methods, and applications," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 450341, 23 pages, 2015.

[2] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 8, pp. 582–596, 2019.

[3] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Procedia Computer Science*, vol. 102, pp. 317–324, 2016.

[4] P. Moeskops, M. A. Viergever, A. M. Mendrik et al., "Automatic segmentation of MR brain Images with a

convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, 2019.

[5] Y. Zhang, C. Yin, Q. Wu, Q. He, and H. Zhu, "Location-aware deep collaborative filtering for service recommendation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3796–3807, 2021.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2015.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, Munich, Germany, 2015.

[8] C. G. B. Yogananda, B. R. Shah, M. Vejdani-Jahromi et al., "A fully automated deep learning network for brain tumor segmentation," *Tomography*, vol. 6, no. 2, pp. 186–193, 2020.

[9] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[10] K. He, X. Zhang, S. Ren et al., "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proceedings of the 2016 European Conference on Computer Vision ECCV*, vol. 9980, pp. 630–645, Amsterdam, Netherlands, 2016.

[12] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[13] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, 2015.

[14] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *International Conference on Learning Representations (ICLR)*, San Juan, PR, USA, 2016.

[15] F. Hoseini, A. Shahbahrami, and P. Bayat, "An efficient implementation of deep convolutional neural networks for MRI segmentation," *Journal of Digital Imaging*, vol. 31, no. 5, pp. 738–747, 2018.

[16] M. B. Naceur, R. Saouli, M. Akil, and R. Kachouri, "Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images," *Computer Methods and Programs in Biomedicine*, vol. 166, pp. 39–49, 2018.

[17] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, pp. 2481–2495, 2017.

[18] M. Shaikh, G. Anand, G. Acharya, A. Amrutkar, V. Alex, and G. Krishnamurthi, "Brain tumor segmentation using dense fully convolutional neural network," in *Proceedings of the 3rd International Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 309–319, Springer, Granada, Spain, 2018.

[19] O. Oktay, J. Schlemper, L. L. Folgoc et al., "Attention U-net: learning where to look for the pancreas," 2018, https://arxiv.org/abs/1804.03999.

[20] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[21] H. Zhao, J. Shi, X. Qi et al., "Pyramid scene parsing network," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017.

[22] L. C. Chen, G. Papandreou, F. Schroff et al., "Rethinking atrous convolution for semantic image segmentation," 2017, https://arxiv.org/abs/1706.05587.

[23] M. Yang, K. Yu, Z. Chi et al., "DenseASPP for semantic segmentation in street scenes," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3684–3692, Salt Lake, UT, USA, 2018.

[24] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*, Springer, Cham, Switzerland, 2018.

[25] H. Guo and D. Yang, "PRDNet: medical image segmentation based on parallel residual and dilated network," *Measurement*, vol. 173, no. 1, Article ID 108661, 2020.

[26] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, pp. 448–456, Lille, France, 2015.

[27] C. Szegedy, S. Ioffe, V. Vanhoucke et al., "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, CA, USA, 2016.

[28] A. Kermi, I. Mahmoudi, and M. T. Khadir, "Deep convolutional neural networks using U-net for automatic brain tumor segmentation in multimodal MRI volumes," *Lecture Notes in Computer Science*, p. 11384, Springer, Berlin, Germany, 2019.

[29] D. Kingma and J. Ba, "ADAM: a method for stochastic optimization," in *Proceedings of the 2015 International Conference on Learning Representations*, San Diego, CA, USA, 2015.

[30] B. H. Menze, A. Jakab, S. Bauer et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[31] G. Urban, M. Bendszus, F. A. Hamprecht et al., "Multi-modal brain tumor segmentation using deep convolutional neural networks," in *Proceedings of the MICCAI—BraTSs*, Boston, MA, USA, 2014.

[32] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation," *Frontiers in Computational Neuroscience*, vol. 13, p. 56, 2019.

[33] I. S. Muhammad and P. L. Jiang, "Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images," *Pattern Recognition Letters*, vol. 129, pp. 181–189, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 7502676, 11 pages
https://doi.org/10.1155/2021/7502676

Research Article

The Influence of E-Marketing on Performance of Real Estate Enterprises: Based on Super-Efficiency DEA and Grey Entropy Methods

Zhong-Huan Wu [1] **and Hong-jie Chen** [2]

[1] *School of Management, Guangzhou Huashang College, Guangzhou, China*
[2] *School of Business Administration, South China University of Technology, Guangzhou, China*

Correspondence should be addressed to Zhong-Huan Wu; m18819453269_1@163.com

Received 17 June 2021; Accepted 24 July 2021; Published 3 August 2021

Academic Editor: Shianghau Wu

E-marketing is an important tool for real estate enterprises. We evaluate 3 online marketing channels of 44 Chinese real estate companies. Super-efficiency DEA and grey entropy methods are applied to analyse the influence of E-marketing on the performance of real estate enterprises. We find that E-marketing will affect the business performance of real estate companies. Real estate company managers should adopt more strategies to improve corporate performance.

1. Introduction

The Internet is a basic marketing tool, which has brought many new opportunities. To gain more competitive advantage in the market, many enterprises have established their websites and APPs and promoted it through social networks. In the online market, E-marketing is very helpful to the company's performance growth and development [1]. An efficient operating model [2] absorbing external information and knowledge [3], managing customer relationships [4, 5], and exploring other market information [6] can improve the performance of enterprises. Merono-Cerdan and Soto-Acosta [7] evaluated 228 Spanish firms and found a positive relationship between web content and firm performance. Despite these findings, other research has produced mixed results concerning the relationship between Internet marketing and firm performance. Shang et al. [8] found that there are no significant differences in efficiency owing to different e-commerce adoption statuses.

As a main industry in China, real estate enterprises have employed the online marketing and spend millions on E-marketing. It is increasingly important to the operation and development of enterprises, and the evaluation of E-marketing performance also attaches great practical significance to companies. What is the relationship between E-marketing and real estate enterprises performance? What are the effects of online marketing between different channels? Whether the E-marketing can be the strategic resource of the enterprises? In this paper, we will try to explore whether real estate companies' E-marketing can influence operation performance using the data envelopment analysis (DEA). The study first constructs a new variable IMT to evaluate Internet marketing for Chinese real estate enterprises. Next, we study the impact of IMT on the performance of real estate enterprises in two scenarios. Thirdly, we use the grey entropy method to sort the four inputs. Finally, to further study the influencing factors of IMT, we bring the six attributes of IMT into the DEA model.

The paper is organized as follows. In the next section, we motivate our work by reviewing some related works in the literature. In Section 3, we present the methodology we use in this paper. Section 4 will design the research and set out the data collected related to the real estate companies we study. We will present the results obtained in Section 5. Finally, we draw some conclusions and put some future research in Section 6.

2. Literature Review

2.1. Internet Marketing. In recent years, E-marketing is becoming popular in academia. The study mostly focuses on the following three areas: (1) the subject of Internet marketing; (2) Internet marketing focuses on industry research; (3) the elements of Internet marketing. Firstly, the subject of Internet marketing mostly focused on consumers, especially the searching behaviour of consumers. Information search plays a very critical role in consumer decision-making. A large amount of information can be obtained online. Optimizing search becomes a hot topic. For example, Ghose et al. [9] proposed a structured econometric model to understand consumer preferences and then improve the user experience of social media. Du et al. [10] established a hierarchical Bayesian model to study how keyword categories and matching types affect customer. Secondly, Internet marketing research on the industry mainly focuses on online word-of-mouth. For example, in book industry, Chevalier and Mayzlin [11] found an improvement in book's reviews that lead to an increase in relative sales at that site. Jeong and Chung [12] have proved that the film has a greater sense of confidence in word of mouth information and influences consumers' choice. And, in the restaurant industry, the financial impact of online customer reviews in the restaurant industry and the restaurant eWOM (review volume and review rating) contributed to restaurant profitability [13]. Thirdly, there are more literature on the attributes of online marketing research focusing on user satisfaction and emotional analysis [14, 15].

Above all, research subject of Internet marketing is mostly related to consumer and rarely from the E-marketing implementer. And, there are few literature on elements of Internet marketing, such as the online information, online transactions, and online interaction. We will study Internet marketing from the provider of network marketing and put the network information, online transactions, and other network interaction factors into the evaluation system.

2.2. Real Estate Performance Measurement. Performance measurement refers to the relationship between inputs and outputs. Evaluating organizational efficiency units is usually difficult, especially when the inputs (resources and costs) and outputs (services and products) are multiple variables [16]. Data envelopment analysis (DEA) is a nonparametric method to empirically measure the relative efficiency of multiple decision-making units (DMU). It is a valuable analysis tool for performance evaluation. DEA can analyse and quantify the efficiency of each DMU without a specific function. El-Mashaleh et al. [17] used data envelopment analysis (DEA) to establish a benchmark for evaluating the construction company performance. Some scholars also have modified previously established benchmark models, such as Fisher et al., Hudson, and Construction Industry Association (2000). Horta et al. [18] used DEA to evaluate a web-based job performance. The performance indicators are organizational performance indicators (productivity, profitability, accident rate, and unchecked invoices) and operational performance indicators (contractor-customer cooperation, contractor's satisfaction with payment, and contractor satisfaction with cooperation and predictability of cost). These may be proved to be a benchmark to improve their organization management. Tsolas [19] also integrated DEA and ratios to evaluate construction companies' performance in profitability and effectiveness. Jia-Jane et al. (2011) used DEA and grey entropy to study the influence of E-marketing on hotel performance. There are few literature studies on the impact of Internet marketing on the performance of real estate enterprises by DEA.

3. Research Methods

3.1. DEA. The DEA model (also known as the CCR model) is first proposed by Charnes, Cooper, and Rhodes [20]. By using the linear planning method, the relative efficiency of the same type (DMU) is measured. Each DMU has multiple inputs and outputs. The unit of a high-efficiency DMU is 1.0, and the low-efficiency DMU is less than 1.0. DEA is a nonparametric analysing method and need not make any assumptions.

Suppose that we want to calculate a set of n decision-making units (DMUs); the DMUs may be hospitals, government departments, enterprises, or schools. The technical efficiency of n DMUs is recorded as DMU_j ($j = 1, 2, 3, \ldots, n$), and each DMU has m inputs, denoted as x_j ($i = 1, 2, 3, \ldots, m$), and the weight of the input is expressed as v_i ($i = 1, 2, 3, \ldots, m$); each DMU has q outputs, denoted as y_i ($i = 1, 2, 3, \ldots, q$), and the weight of the output is expressed as u_i ($i = 1, 2, 3, \ldots, q$). The DMU measured is denoted as DMU_k, and its output-input ratio is

$$h_k = \frac{u_1 y_{1k} + u_2 y_{2k} + \ldots + u_q y_{qk}}{v_1 x_{1k} + k x_{2k} + \ldots + v_q x_{qk}} = \frac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}}, \quad (u \geq 0; v \geq 0). \tag{1}$$

The efficiency value θ_j obtained by the weight (1) is limited to the interval [0, 1]:

$$\frac{\sum_{r=1}^{q} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1. \tag{2}$$

Formula (2) is the input-oriented CCR model, and the nonlinear planning

$$\max \frac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}}$$
$$\text{s.t. } \frac{\sum_{r=1}^{n} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1,$$
$$\sum_{i=1}^{m} v_i x_{ik} = 1, \tag{3}$$
$$u \geq 0, v \geq 0,$$
$$i = 1, 2, \ldots, m,$$
$$j = 1, 2, \ldots, n.$$
$$r = 1, 2, \ldots, q,$$

The purpose of formula (3) is to maximize the efficiency of DMU, and all efficiency values do not exceed 1. The nonlinear planning formula with infinite optimal solutions is a problem. Therefore, formula (3) is transformed into an equivalent linear planning formula:

$$\max \sum_{r=1}^{q} u_r y_{rk}$$
$$\text{s.t. } \sum_{r=1}^{q} u_r y_{rj} \leq \sum_{i=1}^{m} v_i x_{ij},$$
$$\sum_{i=1}^{m} v_i x_{ik} = 1, \tag{4}$$
$$u \geq 0, v \geq 0,$$
$$i = 1, 2, \ldots, m,$$
$$r = 1, 2, \ldots, q,$$
$$j = 1, 2, \ldots, n.$$

Formula (4) is the input-oriented CCR model to find the solution of DMU_k.

3.2. Super-Efficiency DEA. Most DEA models (such as CCR and BCC) [20] cannot provide efficient DMU [21]. And, Banker and Gifford removed the DMU, evaluated from the reference, and developed a super-efficient DEA model to find the efficient DMU.

In the super-efficiency model, the value can be greater than or equal to 1, and the model is not affected by the marginal effects produced by other enterprises. It can measure the input-output ratio well. Therefore, we will use the super-efficiency model to analyse the technical efficiency of real estate companies. The super-efficiency DEA model of DMU_s is

$$\max \sum_{r=1}^{q} u_{rk} y_{rk}$$
$$\text{s.t. } \sum_{\substack{r=1 \\ r\neq k}}^{q} u_{rj} y_{rj} \leq \sum_{i=1}^{m} v_{ij} x_{ij},$$
$$\sum_{i=1}^{m} v_{ik} x_{ik} = 1, \tag{5}$$
$$i = 1, 2, \ldots, m,$$
$$r = 1, 2, \ldots, q,$$
$$j = 1, 2, \ldots, n.$$

3.3. Grey Entropy. Grey entropy analysis is one of the good tools for factor analysis, especially suitable for multifactor analysis where the data distribution cannot be described by common probability distributions. There are many improved versions, but the common feature of these versions is used to calculate the average point-by-point grey correlation coefficient. The grey entropy was proposed by Wen et al. [22] and Wang et al. [23]. It introduces the nature of entropy to improve the lack of grey relation.

Let X be a factor set of grey relation, and one sequence can be denoted as

$$x_i = \left(x_i(1), x_i(2), x_i(3), \ldots, x_i(k)\right), \tag{6}$$

where $i = 0, 1, 2, \ldots, m$ and $k = 1, 2, \ldots, n$.

Compute the summation of each attribute's value for all sequences (AGO):

$$D_k\colon D_k = \sum_{i=1}^{m} x_i(k). \tag{7}$$

Compute the normalization coefficient K:

$$K = \frac{1}{\left(e^{0.5} - 1\right)n}, \tag{8}$$

where n represents the number of attributes.

Find the entropy for the specific attribute e_k:

$$e_k = K \sum_{i=1}^{m} W_e(z_i), \tag{9}$$

where,

$$W_e(z_i) = z_i e^{(1-z_i)} + (1 - z_i) e^{z_1} - 1,$$
$$z_i = \frac{x_i(k)}{D_k}. \tag{10}$$

Compute the total entropy value E:

$$E = \sum_{k=1}^{n} e_k. \tag{11}$$

Determine the relative weighting factor λ_k:

$$\lambda_k = \frac{1}{n - E}\left(1 - e_k\right). \tag{12}$$

The normalized weight of each attribute can be calculated as

$$\beta_k = \frac{\lambda_k}{\sum_{i-1}^{n} \lambda_i}. \tag{13}$$

4. Performance of Internet Marketing and Research Design

4.1. The Conceptual Framework of Internet Marketing Tools. The inputs and outputs of DEA are all quantitative variables, and the impact of the qualitative variable of Internet marketing is unknown. Therefore, we will add Internet marketing tools (IMT) as a qualitative variable to evaluate performance. Angel and Nath [24] developed a conceptual framework to evaluate Internet marketing strategies with three dimensions: electronic information, communication, and transaction. Our framework was built upon the work of Angel and Nath [24] and Cherif and Grant [25], covering effective real estate company website design and E-marketing [26, 27].

IMT is a qualitative variable. There are three first-level indicators, 6 second-level indicators, and 26 third-level indicators. The information includes 3 second-level indicators (1. company information. 2. Real estate information. 3. Promotions), communication includes 2 second-level indicators (1. interaction. 2. Multilingual capabilities), and transaction includes a second-level indicator (online trading). And, the second-level indicators are divided into 6 groups:

(1) Company information includes 6 third-level evaluation indicators: company introduction; photos of company features; financial statements; employment opportunities; relevant business information; new media links

(2) Real estate information includes 6 third-level evaluation indicators: availability price information; 3D tour; real estate introduction; real estate address; surrounding facilities; transportation

(3) Promotion includes 3 third-level evaluation indicators: any promotion mentioned; up-to-date information; banner advertisement

(4) Interaction includes 5 third-level evaluation indicators: contact details; customer service centre; online comments; suggestion feedback; search capabilities

(5) Multilingual capabilities includes 3 third-level evaluation indicators: simplified Chinese; English; others

(6) Online transactions includes 3 third-level evaluation indicators: online appointment; online shopping mall; online payment

4.2. Data Collection. The E-marketing evaluation, a structured form, consisting of 44 companies, was developed to access the contents of the three marketing channels for real estate companies in China.

The data is divided into two parts: (1) IMT data comes from the public annual reports or the websites and new media platforms (Table 1 shows the content evaluation of Internet marketing tools); (2) input-output data of 44 companies comes from the CREIS in 2019. And, we also use special reports and annual data from companies to ensure the confidence of data (the data are shown in Table 2).

4.3. Research Steps. Firstly, we develop a structured evaluation form (IMT) with 26 evaluation indicators in 3 online marketing channels (website, WeChat official account, and Applet). There are 78 elements in the table, and each element is measured by a binary variable (0 or 1), which represents whether or not a company has the particular marketing feature. "IMT" contains 78 indicators to check whether the real estate websites or new media provide certain features or services, and each item is 0 or 1. The statistical results are to assess the current usage of network marketing tools. However, there are no new media links on WeChat official account and Applet, and the total score is 76 points.

Second, construct the super-efficiency DEA model. The three basic inputs that generate profit are the area of land acquired, the amount of land acquired, and the number of employees, and the two outputs, sales area and sales amount, represent most of company revenue. To test the relationship between Internet marketing and the performance, we add "Internet marketing tools" (IMT) as the fourth input to the super-efficiency DEA model. Among these inputs and outputs, only the IMT is a composed variable and a qualitative variable. The super-efficiency DEA evaluation model is shown in Figure 1.

At last, analyse the data obtained with three methods: super-efficiency DEA, paired-sample t-test, and grey entropy method. First, the super-efficiency DEA model is employed to calculate the relative efficiency of 44 real estate companies for two scenarios (A and B). The outputs of two scenes are the same (sales area and sales amount), but the inputs of scenarios A and B are different, and the difference is that IMT is added to scenario B. Next, the paired-sample t-test is used to test the difference between two scenarios. Then, grey entropy is employed to calculate the weighting of four inputs, and it represents the relations between enterprises performance and E-marketing characteristics. Finally, to explore the impact of various attributes of IMT, we put 6 attributes (company information, real estate information, promotion, interaction, multilingual capabilities, and online transactions) to the super-efficiency DEA model.

The research steps are shown in Figure 2.

Table 1: Conceptual framework of Internet marketing tools for real estate companies.

Marketing features	Category	Items	References
Information	Company information	Company introduction; photos of company features; financial statements; employment opportunities; relevant business information; new media links	Meroño-Cerdan and Soto-Acosta [7]; Cherif & Grant. [25]; Ullah et al. [26]; Ipoo et al. [28]; Shuai and Wu [27]
	Real estate information	Availability price information; 3D tour; real estate introduction; real estate address; surrounding facilities; transportation	
	Promotion	Any promotion mentioned; up-to-date information; banner advertisement	
Communication	Interaction	Contact details; customer service center; online comments; suggestion feedback; search capabilities	
	Multilingual capabilities	Simplified Chinese; English; others	
Transaction	Online transactions	Online appointment; online shopping mall; online payment	

Table 2: The input-output data of real estate enterprises in 2019.

DMU	Amount of land acquired	Area of land acquired	The number of employees	IMT	Sales amount	Sales area
DMU1	1610.00	2996.00	131505	47	6260.30	4035.90
DMU2	1303.00	4253.00	101784	41	7715.00	8630.30
DMU3	1034.00	813.00	6200	26	3371.20	1787.10
DMU4	1000.00	1965.00	50834	30	5556.00	3808.50
DMU5	813.00	1328.00	46518	13	2425.00	1367.00
DMU6	804.00	3577.00	39091	32	3880.00	3420.00
DMU7	759.00	1011.00	24107	21	2425.00	1432.00
DMU8	644.00	1055.00	34227	28	2106.00	1077.50
DMU9	569.00	495.00	7418	40	2001.00	1055.70
DMU10	521.00	768.00	49014	23	2205.50	1130.00
DMU11	517.00	803.00	11370	32	1608.10	748.40
DMU12	493.00	1688.00	28058	33	2747.80	2436.90
DMU13	472.00	1132.00	13693	31	1526.00	1582.30
DMU14	459.00	847.00	7870	41	2090.00	1261.00
DMU15	437.00	2138.00	131694	40	6205.60	5990.90
DMU16	411.00	1307.00	22812	26	1803.40	1968.90
DMU17	405.00	828.00	16420	20	1980.20	550.40
DMU18	377.00	806.00	9414	26	870.80	681.60
DMU19	343.00	492.00	1092	20	434.50	245.60
DMU20	338.00	581.00	286242	31	1002.00	742.20
DMU21	327.00	293.00	13658	33	347.60	185.40
DMU22	323.00	654.00	14413	38	1011.40	1011.70
DMU23	313.00	391.00	1902	41	1607.00	1058.60
DMU24	312.00	768.00	38313	38	1185.10	893.70
DMU25	309.00	428.00	18540	40	678.20	363.20
DMU26	303.00	1097.00	17378	14	1219.00	1285.70
DMU27	292.00	663.00	26779	20	1510.00	1366.80
DMU28	289.00	367.00	2999	30	1037.80	696.60
DMU29	268.00	825.00	24054	20	1207.50	1149.90
DMU30	257.00	453.00	10854	22	2607.80	1470.60
DMU31	253.00	400.00	7566	8	922.70	401.40
DMU32	247.00	310.00	6698	33	783.60	532.40
DMU33	200.00	201.00	7537	35	751.20	488.70
DMU34	191.00	355.00	9462	23	758.00	274.10
DMU35	184.00	248.00	11400	42	861.00	492.40
DMU36	181.00	182.00	62305	28	808.00	316.30
DMU37	180.00	646.00	62305	39	1381.90	1360.90
DMU38	170.00	401.00	3488	25	1174.70	1076.30
DMU39	166.00	381.00	11631	20	610.40	381.30

Table 2: Continued.

DMU	Amount of land acquired	Area of land acquired	The number of employees	IMT	Sales amount	Sales area
DMU40	154.00	411.50	2059	33	1020.00	376.70
DMU41	152.00	210.00	17100	28	923.50	648.90
DMU42	144.00	177.00	11571	19	468.00	254.00
DMU43	87.00	219.00	16723	27	1180.60	1178.50
DMU44	338.45	480.00	3527	19	921.20	91.50

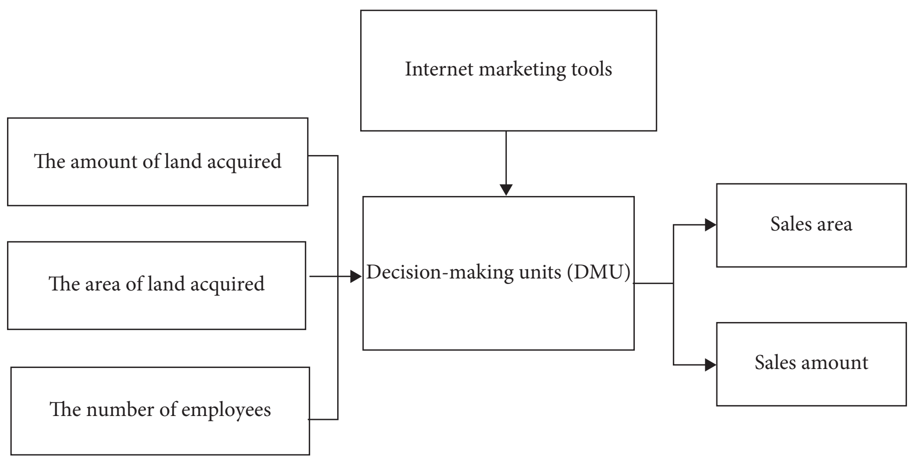


Figure 1: Super-efficiency DEA evaluation model.

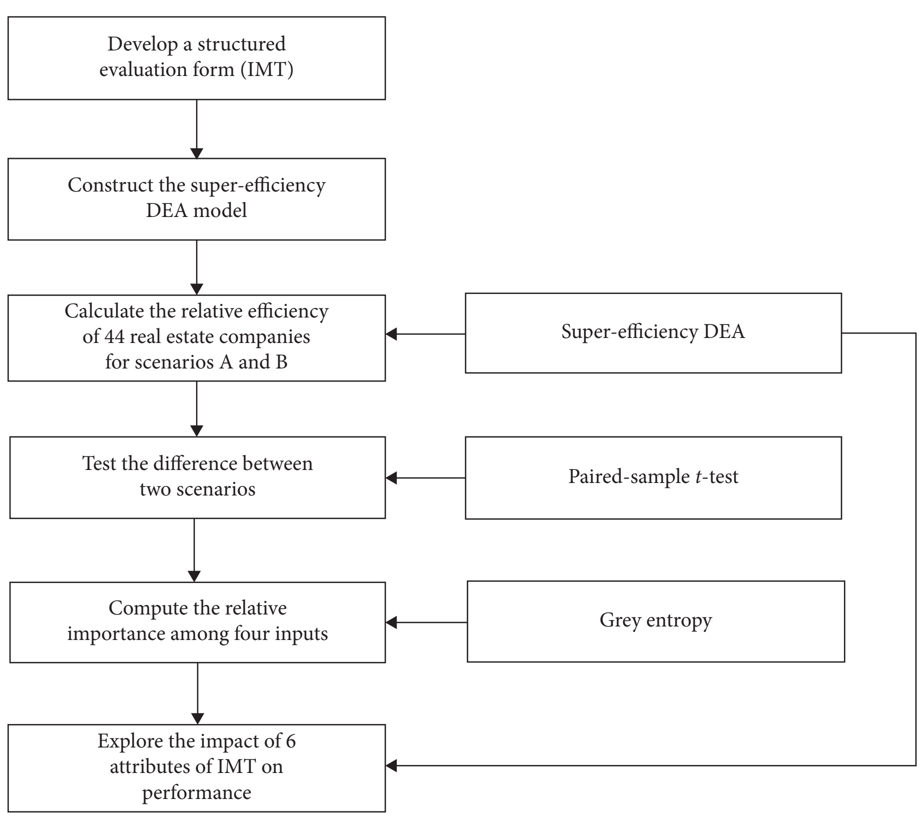


Figure 2: Research steps.

5. Results and Discussion

5.1. Internet Marketing Tools Analysis. Internet marketing tools (IMT) evaluate the online marketing content, services, and features of 44 real estate enterprises (as shown in Table 3). There are three marketing channels for evaluation: websites, WeChat official accounts, and Applets. In 44 real estate companies, 42 enterprises own websites, 41 enterprises have WeChat official accounts, 35 enterprises have Applets, and 32 enterprises have three marketing channels.

Website marketing is mature, focusing on company information, promotion, and interaction; more than 95% of companies on the site will introduce the company and related business. Perhaps, a self-operated website provides a complete introduction to improve the corporate brand image. The maturity of WeChat official account and Applet is similar, but their marketing focuses are different. WeChat official account, which focuses on company information, promotion, and interaction, is a platform for displaying the overview of the company, introducing the main products and communicating with consumers. The Applet focuses on buildings information, promotion, and online transaction. Traditionally, real estate sales are offline, but Applets broke this marketing model. It can show various listings information online, including price, surrounding environment, traffic, and location. When the clients see the property through Applets, then they can make an appointment and pay a deposit to see the house. The applet is a good promotional channel for companies.

The 3D tour indicator and the online trading indicator in the "real estate information" have a lower score. Due to the particularity of real estate products, it is difficult to achieve transactions online. With the development of virtual technology, it can simulate or restore "real" houses and build a virtual 3D stereoscopic environment to help buyers see the overall environment for real time. This function deserves to be paid attention by the manager.

5.2. The Super-Efficiency DEA and Grey Entropy Analysis. To evaluate whether IMTs improve the performance, the data analysis is performed through DEA and grey entropy. First, the super-efficiency DEA model calculates the relative efficiency of 44 real estate companies in the two cases (A and B). Scenario A consists of three inputs (amount of land acquired, area of land acquired, and number of employees) and two outputs (sales area and sales amount), while scenario B consists of four inputs (amount of land acquired, area of land acquired, number of employees, and IMTs) and same outputs. As shown in Table 4, Score 1 is the relative efficiency of scenario A and Score 2 is the relative efficiency of scenario B. The average value of Score 2 is 80.68%, which is higher than the average value of Score 1 of 64.10%. The difference is 16.58%, a big difference. The efficiency difference between Scores 1 and Scores 2 is tested by the paired-sample t-test, and the result of the paired-sample t-test reached a significant level (t value $= -0.4966$, d$f = 44$, and two-tailed significance < 0.001). We believe that there is a positive correlation between Internet marketing tools and real estate company performance; in other words, the E-marketing can advance real estate company's performance. Given the increasing popularity of new media, managers should take full advantage of it. For large-scale companies, they should increase investment in E-marketing and develop new online marketing models. For small and medium enterprises, they can choose low-cost channels or third-party platforms for E-marketing.

Secondly, we use grey entropy to find the weighting of four inputs. For the grey entropy is an objective weighting technique with strict rules and requirement, we employ it to calculate the relative weighting of four inputs, and the results show that the IMT is 0.9959, which is slightly lower than the area of land acquired (0.9984), followed by the amount of land acquired (0.9956) and the number of employees (0.9905). Although the best performer to real estate performance is the area of land acquired rather than IMT, the weight of IMT is not too different from these three inputs. In fact, the network marketing has become the essential investment for real estate companies. Real estate company managers should adopt more strategies to improve corporate performance.

5.3. Attributes in Internet Marketing Tools Analysis. To explore the impact of various attributes of IMT (shown as Table 1), we put 6 attributes (company information, real estate information, promotion, interaction, multilingual capabilities, and online transactions) to the super-efficiency DEA model, and the results are shown in Table 5. There are 7 real estate companies whose efficiency of company information, real estate information, promotion, interaction, and multilingual capabilities are greater than 1, including DMU3, DMU15, DMU23, DMU30, DMU38, DMU40, and DMU43. The average efficiency value of six attributes is higher than Score 1 (Table 4), so we believe that six attributes have a positive impact on the performance of real estate companies. The various attributes affecting the performance of companies in the descending order are online transactions, real estate information, interaction, company information, and multilingual capabilities and promotions.

"Online transaction" shows that a customer has a certain degree of loyalty to the real estate company. It can promote user's willingness to buy and is benefitted from output. From the three dimensions of online transaction, we make the following recommendations: (1) real estate enterprises pay attention to this function and make its identification to be clear, while simplifying the process; (2) online shopping malls can put on shelves some characteristic or peripheral products, increase consumers' online experience, and thus change consumers' stereotyped image of the company; (3) expand the channels of payment while ensuring payment security.

"Interactive function" reflects the interaction with users. There are five subattributes: contact details, online comments, suggestion feedback, keyword search, and customer service centre. In Table 3, to facilitate users to contact the company, most companies have contact details in each channel. And, the popularity of keyword search is high, for

Table 3: Content evaluation of Internet marketing tools.

First-level indicators	Second-level indicators	Evaluation content	Website		WeChat official account		Applet	
			N (42)	Percentage (%)	N (41)	Percentage (%)	N (35)	Percentage (%)
Information	Company information	Company introduction	42	95.45	30	68.18	14	31.82
		Photos of company features	34	77.27	23	52.27	13	29.55
		Financial statements	27	61.36	27	61.36	2	4.55
		Employment opportunities	40	90.91	21	47.73	3	6.82
		Relevant business information	42	95.45	32	72.73	13	29.55
		New media links	38	86.36	—	—	—	—
	Buildings' information	Availability price information	6	13.64	3	6.82	28	63.64
		3D tour	0	0.00	4	9.09	17	38.64
		Building introduction	32	72.73	24	54.55	26	59.09
		Surrounding facilities	15	34.09	9	20.45	28	63.64
		Real estate address	31	70.45	18	40.91	30	68.18
		Transportation	16	36.36	5	11.36	25	56.82
	Promotion	Any promotion mentioned	0	0.00	9	20.45	22	50.00
		Up-to-date information	40	90.91	36	81.82	25	56.82
		Banner advertisement	34	77.27	5	11.36	30	68.18
Communication	Interaction	Contact details	40	90.91	20	45.45	35	79.55
		Online comments	1	2.27	41	93.18	3	6.82
		Suggestion feedback	15	34.09	4	9.09	5	11.36
		Search capabilities	27	61.36	41	93.18	24	54.55
		Customer service center	13	29.55	8	18.18	7	15.91
	Multilingual capabilities	Simplified Chinese	23	52.27	0	0.00	0	0.00
		English	27	61.36	1	2.27	0	0.00
		Others	0	0.00	0	0.00	0	0.00
Transaction	Online transactions	Online appointment	4	9.09	0	0.00	19	43.18
		Online shopping mall	0	0.00	8	18.18	11	25.00
		Online payment	0	0.00	6	13.64	13	29.55

Table 4: DEA model results.

DMU	Input 1	Input 2	Input 3	Output 1	Output 2	Input 4 (IMT)	Scores 1 (%)	Scores 2 (%)
DMU 1	1610.00	2996.00	131505	6260.30	4035.90	47	36.76%	72.95%
DMU2	1303.00	4253.00	101784	7715.00	8630.30	41	75.68%	152.31%
DMU 3	1034.00	813.00	6200	3371.20	1787.10	26	95.37%	176.92%
DMU 4	1000.00	1965.00	50834	5556.00	3808.50	30	57.94%	116.80%
DMU 5	813.00	1328.00	46518	2425.00	1367.00	13	31.72%	99.20%
DMU 6	804.00	3577.00	39091	3880.00	3420.00	32	58.12%	86.33%
DMU 7	759.00	1011.00	24107	2425.00	1432.00	21	43.30%	72.76%
DMU 8	644.00	1055.00	34227	2106.00	1077.50	28	34.68%	52.27%
DMU 9	569.00	495.00	7418	2001.00	1055.70	40	81.13%	81.13%
DMU10	521.00	768.00	49014	2205.50	1130.00	23	49.88%	69.72%
DMU11	517.00	803.00	11370	1608.10	748.40	32	42.21%	48.02%
DMU12	493.00	1688.00	28058	2747.80	2436.90	33	64.53%	78.96%
DMU13	472.00	1132.00	13693	1526.00	1582.30	31	50.12%	75.41%
DMU14	459.00	847.00	7870	2090.00	1261.00	41	67.50%	73.99%
DMU15	437.00	2138.00	131694	6205.60	5990.90	40	104.65%	177.04%
DMU16	411.00	1307.00	22812	1803.40	1968.90	26	61.41%	78.52%
DMU17	405.00	828.00	16420	1980.20	550.40	20	49.19%	71.34%
DMU18	377.00	806.00	9414	870.80	681.60	26	30.37%	42.42%
DMU19	343.00	492.00	1092	434.50	245.60	20	47.09%	50.45%
DMU20	338.00	581.00	286242	1002.00	742.20	31	30.79%	36.32%

Table 4: Continued.

DMU	Input 1	Input 2	Input 3	Output 1	Output 2	Input 4 (IMT)	Scores 1 (%)	Scores 2 (%)
DMU21	327.00	293.00	13658	347.60	185.40	33	20.61%	20.61%
DMU22	323.00	654.00	14413	1011.40	1011.70	38	47.58%	49.00%
DMU23	313.00	391.00	1902	1607.00	1058.6	41	187.20%	187.20%
DMU24	312.00	768.00	38313	1185.10	893.70	38	31.69%	33.87%
DMU25	309.00	428.00	18540	678.20	363.20	40	27.53%	27.53%
DMU26	303.00	1097.00	17378	1219.00	1285.70	14	53.86%	74.30%
DMU27	292.00	663.00	26779	1510.00	1366.80	20	52.20%	71.46%
DMU28	289.00	367.00	2999	1037.80	696.60	30	67.17%	69.96%
DMU29	268.00	825.00	24054	1207.50	1149.90	20	46.64%	57.90%
DMU30	257.00	453.00	10854	2607.80	1470.60	22	143.54%	170.25%
DMU31	253.00	400.00	7566	922.70	401.40	8	43.35%	75.50%
DMU32	247.00	310.00	6698	783.60	532.4	33	52.39%	52.39%
DMU33	200.00	201.00	7537	751.20	488.70	35	65.80%	65.80%
DMU34	191.00	355.00	9462	758.00	274.10	23	38.48%	38.48%
DMU35	184.00	248.00	11400	861.00	492.40	42	60.38%	60.38%
DMU36	181.00	182.00	62305	808.00	316.30	28	77.12%	77.12%
DMU37	180.00	646.00	62305	1381.90	1360.90	39	55.51%	55.51%
DMU38	170.00	401.00	3488	1174.70	1076.30	25	146.67%	146.67%
DMU39	166.00	381.00	11631	610.40	381.30	20	34.10%	34.21%
DMU40	154.00	411.50	2059	1020.00	376.70	33	107.78%	107.78%
DMU41	152.00	210.00	17100	923.50	648.90	28	78.05%	78.05%
DMU42	144.00	177.00	11571	468.00	254.00	19	45.93%	45.93%
DMU43	87.00	219.00	16723	1180.60	1178.50	27	177.12%	177.12%
DMU44	338.45	480.00	3527	921.20	91.50	19	47.13%	60.27%
Mean	430.67	896.44	32310	1890.67	1357.20	28	64.10%	80.68%

Table 5: Evaluate the effect of attributes.

DMU	Company information	Real estate information	Promotions	Interaction	Multilingual language	Online transactions
DMU 1	93.10%	62.00%	61.99%	61.12%	39.99%	48.58%
DMU2	152.51%	141.22%	163.87%	126.01%	93.41%	95.92%
DMU 3	150.70%	186.25%	202.57%	169.81%	197.92%	157.13%
DMU 4	101.64%	108.56%	77.68%	136.29%	60.94%	79.87%
DMU 5	57.88%	80.02%	46.75%	100.00%	31.72%	66.80%
DMU 6	83.15%	95.15%	71.87%	96.85%	69.89%	71.51%
DMU 7	63.97%	72.02%	67.48%	63.82%	43.31%	62.48%
DMU 8	51.79%	49.91%	34.68%	48.50%	34.68%	73.22%
DMU 9	81.66%	81.87%	81.13%	81.13%	81.13%	81.13%
DMU10	78.44%	66.14%	49.88%	53.69%	49.88%	96.53%
DMU11	50.57%	46.96%	45.36%	49.29%	42.21%	76.14%
DMU12	78.88%	82.52%	76.47%	83.08%	76.60%	65.99%
DMU13	71.76%	77.32%	65.62%	85.45%	69.69%	58.29%
DMU14	73.24%	76.05%	71.34%	74.51%	68.48%	74.53%
DMU15	204.32%	125.86%	142.20%	185.70%	186.21%	236.75%
DMU16	67.92%	85.65%	75.14%	88.77%	74.15%	117.74%
DMU17	69.53%	79.30%	50.13%	56.16%	49.76%	117.05%
DMU18	34.79%	53.33%	37.61%	39.76%	34.92%	59.40%
DMU19	59.46%	47.09%	49.64%	47.20%	78.03%	49.37%
DMU20	37.89%	34.48%	30.79%	38.64%	31.53%	43.44%
DMU21	20.61%	20.61%	20.61%	20.61%	20.61%	29.90%
DMU22	48.43%	54.12%	48.33%	48.63%	48.39%	47.58%
DMU23	187.20%	187.20%	187.20%	187.20%	187.20%	187.20%
DMU24	35.09%	32.09%	32.10%	36.05%	32.59%	36.79%
DMU25	27.53%	27.53%	27.53%	27.53%	27.53%	27.53%
DMU26	64.22%	77.47%	69.03%	70.41%	74.15%	90.41%
DMU27	62.19%	99.26%	64.92%	65.00%	81.80%	124.18%
DMU28	69.37%	77.25%	67.17%	67.76%	77.71%	120.56%
DMU29	52.13%	65.20%	55.53%	56.60%	74.98%	91.06%
DMU30	169.31%	157.01%	185.96%	168.87%	232.93%	148.93%

Table 5: Continued.

DMU	Company information	Real estate information	Promotions	Interaction	Multilingual language	Online transactions
DMU31	45.89%	100.00%	100.00%	59.28%	46.73%	85.82%
DMU32	52.39%	52.71%	52.39%	52.39%	55.48%	52.39%
DMU33	65.80%	65.80%	65.80%	65.80%	65.80%	112.95%
DMU34	38.48%	61.68%	38.48%	38.48%	38.48%	46.13%
DMU35	60.38%	60.38%	60.38%	60.38%	60.38%	60.38%
DMU36	77.12%	77.12%	77.12%	77.12%	77.12%	118.83%
DMU37	62.19%	55.51%	55.51%	58.35%	59.73%	58.77%
DMU38	146.67%	146.67%	146.67%	146.67%	146.67%	153.90%
DMU39	34.61%	34.24%	34.10%	34.58%	34.10%	72.79%
DMU40	107.78%	107.78%	107.78%	107.78%	107.78%	108.44%
DMU41	78.05%	78.05%	78.05%	78.05%	95.18%	84.37%
DMU42	45.93%	45.93%	45.93%	45.93%	45.93%	79.36%
DMU43	177.12%	177.12%	177.12%	177.12%	177.12%	177.12%
DMU44	56.25%	61.50%	57.04%	67.98%	66.37%	56.33%
Mean	78.36%	81.23%	76.29%	79.64%	76.12%	88.72%

the users can find information quickly. The other three have lower penetration rates, with the lowest being online comment. Online comment is a communication channel before forming a community. In all the enterprise-owned websites, there is almost no online comment, and the enterprise should add this function.

Real estate information, company information, multilingual capabilities, and promotional activities display the relevant information of enterprises so that users have a basic understanding of enterprises. Table 3 shows that company information is comprehensive, but the indicators of 3D tour and promotion are bad. Real estate enterprises should use 3D tour online because compared to text and pictures, 3D tour is convenient for users to receive information, and they can truly "experience" the environment. Moreover, real estate companies should increase new promotional activities because users or potential consumers can understand companies' characteristics, achievements, and core products through these activities.

6. Conclusion

We have studied the relationship between Internet marketing and operational performance of real estate companies in China and developed real estate Internet marketing tools (IMT) so that the evaluation becomes more comprehensive. And, IMT as a qualitative input was added to the super-efficient DEA model. The results suggest that the Internet marketing tool positively affect firm performance. Furthermore, not only we study the correlation between the Internet marketing and the performance but also study the impact of six attributes of IMT on performance. The results show that six attributes can influence the enterprise performance. Thus, managers should take advantage of full range of features of the Internet to interact with customers rather than simply appearing on the Internet (only information). The companies should increase the investment in online marketing, for example, provide AR guide, increase promotional activities, and build user community, or they can develop new online marketing models. However, the best performer to market real estate performance is the area of land acquired rather than IMT. Although E-marketing has a positive impact on firm performance, the business focus for real estate companies is to take land.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Additional Points

The Limitation. This study only selects the input-output data in 2019. A longitudinal study could enrich the findings. And, the channels are home-owned channels rather than third-party platforms. We hope that researchers will consider the third-party platforms in the future.

Conflicts of Interest

The authors declare there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was financially supported by Scientific Research Special Fund of Guangzhou Huashang College (Grant: 2021HSDS02), Guangdong Planning Office of Philosophy and Social Sciences Project (Youth) (Grant: GD20YGL09), and Department of Education of Guangdong Province "Innovative Projects with Characteristics of Ordinary Universities" Project (Grant: 2019WTSCX158).

References

[1] W. K. Chong, D. Bian, and N. Zhang, "E-marketing services and e-marketing performance: the roles of innovation, knowledge complexity and environmental turbulence in influencing the relationship," *Journal of Marketing Management*, vol. 32, no. 1-2, pp. 149–178, 2016.

[2] C. W. Y. Wong, K.-H. Lai, and T. C. E. Cheng, "Value of information integration to supply chain management: roles of internal and external contingencies," *Journal of Management Information Systems*, vol. 28, no. 3, pp. 161–200, 2011.

[3] J. Bockstedt and K. H. Goh, "Seller strategies for differentiation in highly competitive online auction markets," *Journal of Management Information Systems*, vol. 28, no. 3, pp. 235–268, 2011.

[4] P. Harrigan, E. Ramsey, and P. Ibbotson, "Critical factors underpinning the e-CRM activities of SMEs," *Journal of Marketing Management*, vol. 27, no. 5-6, pp. 503–529, 2011.

[5] C. O'Leary, S. Rao, and C. Perry, "Improving customer relationship management through database/Internet marketing," *European Journal of Marketing*, vol. 38, no. 3/4, pp. 338–354, 2004.

[6] C. Soh, M. L. Markus, and K. H. Goh, "Electronic marketplaces and price transparency: strategy, information technology, and success," *MIS Quarterly*, vol. 30, no. 3, pp. 705–723, 2006.

[7] A. L. Meroño-Cerdan and P. Soto-Acosta, "External Web content and its influence on organizational performance," *European Journal of Information Systems*, vol. 16, no. 1, pp. 66–80, 2007.

[8] J.-K. Shang, W.-T. Hung, C.-F. Lo, and F.-C. Wang, "Ecommerce and hotel performance: three-stage DEA analysis," *The Service Industries Journal*, vol. 28, no. 4, pp. 529–540, 2008.

[9] A. Ghose, P. G. Ipeirotis, and B. Li, "Modeling consumer footprints on search engines: an interplay with social media," *Management Science*, vol. 65, no. 3, pp. 1363–1385, 2019.

[10] X. Du, M. Su, X. Zhang, and X. Zheng, "Bidding for multiple keywords in sponsored search advertising: keyword categories and match types," *Information Systems Research*, vol. 28, no. 4, pp. 711–722, 2017.

[11] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: online book reviews," *Journal of Marketing Research*, vol. 43, no. 3, pp. 345–354, 2006.

[12] Y.-J. Jeong and D.-s. Chung, "A study on the effect of knowledge-based resources and word-of-mouth effects on Korean movies," *The Journal of Internet Electronic Commerce Resarch*, vol. 19, no. 1, pp. 255–271, 2019.

[13] Y. Wang, J. Kim, and J. Kim, "The financial impact of online customer reviews in the restaurant industry: a moderating effect of brand equity," *International Journal of Hospitality Management*, vol. 95, Article ID 102895, 2021.

[14] G. J. Tellis, D. J. Macinnis, S. Tirunillai, and Y. Zhang, "What drives virality (sharing) of online digital content? the critical role of information, emotion, and brand prominence," *Journal of Marketing*, vol. 83, no. 4, 2019.

[15] E. W. Anderson and M. W. Sullivan, "The antecedents and consequences of customer satisfaction for firms," *Marketing Science*, vol. 12, no. 2, pp. 125–143, 1993.

[16] W. D. Cook and L. M. Seiford, "Data envelopment analysis (DEA)-t," *European Journal of Operational Research*, vol. 192, no. 1, pp. 1–17, 2009.

[17] M. S. El-Mashaleh, R. Edward Minchin, and W. J. O'Brien, "Management of construction firm performance using benchmarking," *Journal of Management in Engineering*, vol. 23, no. 1, pp. 10–17, 2007.

[18] I. M. Horta, A. S. Camanho, and J. M. Da Costa, "Performance assessment of construction companies integrating key performance indicators and data envelopment analysis," *Journal of Construction Engineering and Management*, vol. 136, no. 5, pp. 581–594, 2010.

[19] I. E. Tsolas, "Modelling profitability and effectiveness of Greek-listed construction firms: an integrated DEA and ratio analysis," *Construction Management and Economics*, vol. 29, no. 8, pp. 795–807, 2011.

[20] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[21] S. Li, G. R. Jahanshahloo, and M. Khodabakhshi, "A super-efficiency model for ranking efficient units in data envelopment analysis," *Applied Mathematics and Computation*, vol. 184, no. 2, pp. 638–648, 2007.

[22] K. L. Wen, T. C. Chang, and M. L. You, "The grey entropy and its application in weighting analysis," *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2, pp. 1842–1844, 1998.

[23] C. Wang, T. W. Lin, and S. S. Hu, "Optimizing the rapid prototyping process by integrating the Taguchi method with the Gray relational analysis," *Rapid Prototyping Journal*, vol. 13, no. 5, pp. 304–315, 2007.

[24] R. Angeles and R. Nath, "Business-to-business e-procurement: success factors and challenges to implementation," *Supply Chain Management: An International Journal*, vol. 12, no. 2, pp. 104–115, 2007.

[25] E. Cherif and D. Grant, "Analysis of e-business models in real estate," *Electronic Commerce Research*, vol. 14, no. 1, pp. 25–50, 2014.

[26] F. Ullah, S. Sepasgozar, and C. Wang, "A systematic review of smart real estate technology: drivers of, and barriers to, the use of digital disruptive technologies and online platforms," *Sustainability*, vol. 10, no. 9, p. 3142, 2018.

[27] J. J. Shuai and W. W. Wu, "Evaluating the influence of E-marketing on hotel performance by DEA and grey entropy," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8763–8769, 2011.

[28] C. Ipoo, R. Law, and H. A. Lee, "Research on network marketing performance evaluation of tourism destination," *Business Management*, vol. 25, 2013.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 3043797, 13 pages
https://doi.org/10.1155/2021/3043797

Research Article

Novel Network Public Opinion Prediction and Guidance Model Based on "S-Curve": Taking the Loss of Contact with "Malaysia Airlines"

Xiangdong Liu,[1] **Axiao Cao**,[1] **and Chuyang Li**[2]

[1]*School of Economics, Jinan University, Guangzhou 510632, China*
[2]*College of Computation Science, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China*

Correspondence should be addressed to Xiangdong Liu; tliuxd@jnu.edu.cn

Received 28 May 2021; Accepted 13 July 2021; Published 22 July 2021

Academic Editor: Weilin Xiao

It is of great significance for the government to control the network public opinion in time and maintain social stability to predict the network public opinion in emergency. This paper proposes a novel improvement method to "S-curve" theory in the context of big data and establishes three novel network public opinion prediction models. These models take into account the proliferation trend of initial and follow-up network public opinion over a long period of time when emergencies are formed and the objective environment suddenly changes, based on the information diffusion model conforming to the traditional "S-curve" theory. The novel improvement and establishment allow our model to have more accurate predictions than other scholars' models that mainly study the first network public opinion in a shorter period of time. And it is more applicable to real social conditions, in line with the public's cognition of reality, and provides more reference for the government to take preventive and corresponding positive guiding measures in advance. To better establish the model, we obtained the 24-day Weibo data associated with the incident of "Malaysia Airlines" loss of contact from big data for model establishment, public opinion prediction, and comprehensive evaluation. The result innovatively shows that, in addition to the initial public opinion that is worthy of attention, the follow-up public opinion is also noteworthy and proves that our model has more practical value.

1. Introduction

Public opinion is the abbreviation of "the situation of public opinion." It is the summation of the beliefs, attitudes, opinions, and emotions expressed by many people about various phenomena and problems in the society. "From the perspective of traditional sociological theory, public opinion is a comprehensive reflection of public opinion and belongs to the category of public opinion theory. From the theory of public opinion in modern society, in addition to the simple generalization of the law of public opinion, public opinion is more focused on the description of "public opinion and its role and the rule of governing and political orientation"" [1]. Network public opinion is a special form of public opinion; that is, network public opinion is a collection of feelings, attitudes, opinions, and opinions of the majority of netizens, with the network as the carrier and events as the core. The expression of network public opinion is fast, the information is diverse, and the way is interactive. The openness and virtuality of the network also determine the characteristics of directness, arbitrariness, diversification, abruptness, concealment, and deviation of the network public opinion. Emergency is the inducing factor of network public opinion, as a kind of unconventional event, and it has the characteristics of instantaneous, accidental burst point, instant gathering, harm to the subject and society, and the crisis of the development trend. The sum of related comments, reports, opinions, and sentiments and attitudes formed in cyberspace after an emergency occurs is the network public opinion of the emergency, which is a special form of network public opinion [2].

Weibo, or microblog, is a broadcast-style social network platform that shares brief and real-time information through a following mechanism. In China, it is difficult for any public opinion that may spread across the whole society to circumvent Weibo as a communication carrier. Sina Weibo is currently one of the largest online platforms in China that can satisfy ordinary users to deliver information to the country. According to the annual financial results released by Weibo, as of December 2020, the number of monthly active users reached 521 million, an increase of about 5 million over the same period last year; the average number of daily active users reached 225 million, a net increase of about 3 million users over the same period last year. Sina Weibo has a large number of active users, a rapid growth in monthly active users, and a high degree of openness on the platform. Now, Weibo has become an important platform for ordinary people to express their views on news hotspots and has strong information dissemination and emotional contagion capabilities [3]. Therefore, Sina Weibo plays an important role in public opinion monitoring of emergencies.

Public opinion research is based on the research of information diffusion mode. Foreign scholar Hagerstrand was the first one to study the traditional information diffusion mode of geographical entity space. He pointed out that the accumulation and change of information adoption over time conformed to the S-shaped logistic curve. He also pointed out that the information diffusion of traditional geographic entity space has multiple effects and will be affected by communication and interaction [4]. Rogers elaborated on the “S-curve” theory in his book “Diffusion of Innovation,” combined with this theory, to point out that the innovation and diffusion of new things and new ideas require the use of certain social networks to achieve [5]. Fiona Duggan proposed the “crisis information dissemination mode” that combined the “S-curve” theory [6]. The existing domestic research on the network public opinion model for emergencies in China is mainly concerned with the initial public opinion in a short period of time. There is a lack of models on the emergence of unclear or undisclosed emergencies as the follow-up network public opinion of the event. After an emergency breaks out, the unknown nature of the matter may put people in a state of excessive speculation and psychological panic. The release of updated news may push network public opinion to a climax, and the scale of public opinion on the Internet at this time will even exceed the public opinion when the emergency breaks out. The fierce follow-up network public opinion is actually a reflection of the public’s attitude towards emergencies, which may be accompanied by or lead to extreme behavioral reactions, which seriously affect social stability. Therefore, the government’s supervision of follow-up public opinion is also crucial. According to the established novel model, the government can take corresponding measures to guide people’s thoughts and behaviors to avoid derivative emergencies.

Many studies have shown that the trend of information diffusion caused by the formation of the initial public opinion of an emergency is in line with the “S-curve” theory, but it is rarely discussed that the emergence of subsequent public opinion leads to the diffusion of information. Therefore, this article mainly discusses the incident of the “Malaysia Airlines” loss of contact based on the corresponding 24-day Weibo data and defines the network public opinion of emergency, including the initial network public opinion and the follow-up network public opinion. Based on the traditional information diffusion model that conforms to the “S-curve” theory, this article further discusses the information diffusion model of the initial and subsequent network public opinion by adding factors that occur when the objective environment changes suddenly in cyberspace and emergency formation. That is, three public opinion prediction models are established based on the traditional curve fitting method to perform data fitting, and the original model is reasonably explained and improved based on the characteristics of the network public opinion and the actual situation during the fitting process. Then, we innovatively put forward the influence of follow-up public opinion on the network public opinion of emergent events. Then, the mean absolute percentage error (MAPE) was used to evaluate and compare three different models simultaneously, and MAPE is the average percentage of the sum of each absolute error (the absolute value of the deviation between the observed value and the true value) and the true value. And finally, it was concluded that, with the subsequent sudden changes in the objective environment, the traditional “S-curve” theory cannot be mechanically applied in the discussion of network public opinion. The network public opinion in the new environment should conform to the public’s perception of reality to analyze and predict objective development. Therefore, when predicting the network public opinion of emergencies, in addition to the initial public opinion, the follow-up public opinion is also worthy of attention. Moreover, by predicting network public opinion, the government can guide positive public opinion forecasting and analysis when emergencies occur to promote harmonious and healthy development and take timely measures against negative public opinion analysis and forecasting to control network public opinion and maintain social stability.

The research content of this paper is mainly divided into three parts. The first part introduces public opinion and Weibo big data platform and then determines the research direction and methods by combining the relevant literature of domestic and foreign scholars on public opinion information. The second part mainly explains the selection and processing of data, and this paper mainly analyzes the information diffusion model of network public opinion through the relevant public opinion data of the loss of contact with “Malaysia Airlines.” The third section is also the most important part, which introduces the theoretical improvement and the establishment of three types of models on the basis of “S-curve” and then makes a comprehensive analysis and evaluation of the three models.

2. Data Selection and Specific Processing

2.1. Selection of Data. The research data in this article comes from the official big data platform of Sina Weibo (see micro index: https://data.weibo.com/index). To better combine the actual situation to discuss the network public opinion. This article selects the platform's data related to the loss of contact with "Malaysia Airlines" from March 8, 2014, to March 31, 2014 (24 days in total), as the data basis for the establishment of the model. At 2:40 am on March 8, 2014, Malaysia Airlines flight number is MH370. The plane that was originally scheduled to fly from Kuala Lumpur to Beijing lost contact. That day, it caused heated discussions with Weibo. At 10 p.m on March 24, 2014, Malaysian Prime Minister Najib announced in Kuala Lumpur that Malaysia Airlines' missing flight MH370 had crashed in the southern Indian Ocean. No one survived on the plane. This caused a Weibo heat on the same day; Figure 1 shows the single-day word frequency curve of the keywords "Malaysia Airlines" and "Malaysia" during the observation period. The single-day word frequency of the keywords "Malaysia Airlines" and "Malaysia" during the observation period increased rapidly on the day of the sudden incident of "Malaysia Airlines" loss of connection (March 8). Taking the emergency of March 8 as the formation event, a large-scale change in the frequency of Japanese words was triggered when the news of the crash was officially confirmed (March 24). Taking the emergency of March 8 as the formation event, the official confirmation of the crash (March 24) triggered a large-scale change in the word frequency on a single day. The date of the obvious change in the data coincides with the date of the key event described in the background introduction of the loss of contact with "Malaysia Airlines."

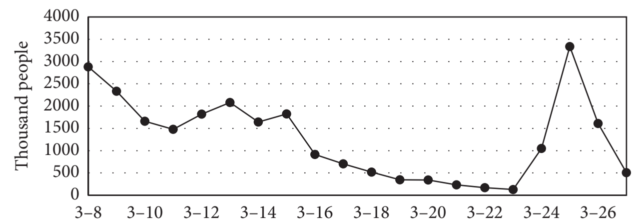


Figure 1: The single-day word frequency curve of the keywords "Malaysia Airlines" and "Malaysia" during the observation period.

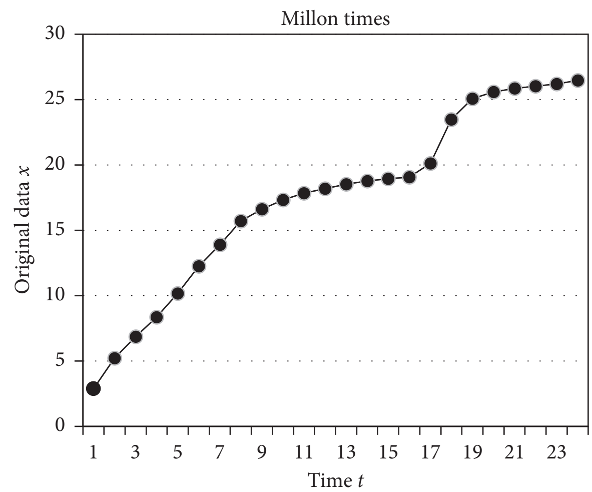


Figure 2: x and t data image.

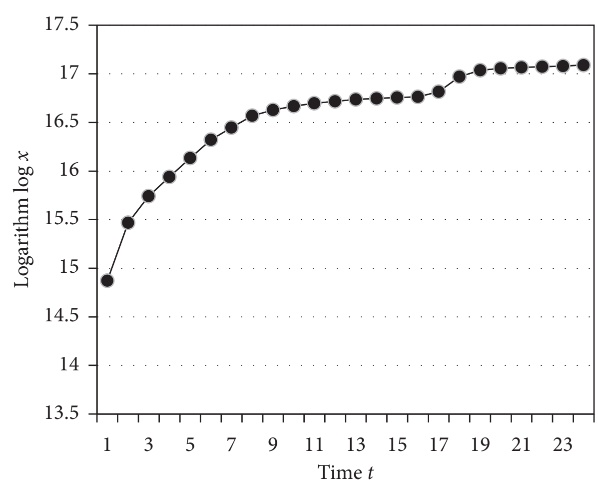


Figure 3: $\log(x)$ and t data image.

2.2. Processing of Data. First, we use numbers 1 to 24 to represent each piece of data between March 8, 2014, and March 31, 2014, expressed in a as the original daily word frequency. To weaken the randomness of the original data and highlight the law of the original data. We accumulate the original data a and store the accumulated data as the new variable x (the relationship between x and t is shown in Figure 2). Logarithmically transform the variable x, and store the data in the variable $\log(x)$ (the relationship between $\log(x)$ and t is shown in Figure 3). The logarithmic transformation is the Box-cox transformation when $\lambda = 0$. There are two reasons for the logarithmic transformation of the original data: ① the scale of variables is compressed, and the absolute value of data is reduced to make it change between relatively small data ranges, which is convenient for calculation; ② the nature and correlation of the data will not be changed after the logarithm transformation.

After the data is preprocessed, it is summarized in Table 1. Table 1 summarizes the variables a, x, and $\log(x)$. Three numbers are selected randomly from the column of $\log(x)$ to form a test set to test the fitting effect of the model. The remaining 21 numbers constitute the training set to build the model. The three samples randomly selected in this paper are the corresponding data at $t = 10$, 22, and 12.

Table 1: Data frame.

t	a	x	$\log(x)$
1	2878813	2878813	14.87
2	2334038	5212851	15.47
3	1658542	6871393	15.74
4	1478279	8349672	15.94
5	1818663	10168335	16.13
6	2081481	12249816	16.32
7	1641772	13891588	16.45
8	1821670	15713258	16.57
9	912939	16626197	16.63
10	701404	17327601	16.67
11	518800	17846401	16.70
12	346422	18192823	16.72
13	343224	18536047	16.74
14	232773	18768820	16.75
15	168282	18937102	16.76
16	128418	19065520	16.76
17	1047551	20128509	16.82
18	3336808	23465317	16.97
19	1610509	25075826	17.04
20	505760	25581586	17.06
21	275780	25857366	17.07
22	166097	26023463	17.07
23	162715	26186178	17.08
24	278691	26464869	17.09

3. Model Establishment and Theoretical Improvement Based on "S-Curve"

The research of public opinion is based on the research of information diffusion mode. The change of information diffusion mode in physical space over time conforms to the trend of "S-curve." Therefore, based on the "S-curve" theory, this paper makes theoretical improvements, establishes three models on the basis of information diffusion, studies the prediction of network public opinion of emergencies, and evaluates and compares the fitting effects of different models on data.

3.1. Introduction to "S-Curve". The Logistic curve (also known as the "S-curve") was first proposed by Belgian biomathematician Pierre Francois Veluler when he was studying the subject of population growth. The basic assumption is that there is no difference among individuals in the population, they have the same growth rate, and the relative growth rate is proportional to the remaining space resources; space resources are limited, with a certain saturation value, and the saturation value has remained unchanged. The characteristic of the curve change is as follows: the initial stage is roughly exponential growth; then, as it becomes saturated, the increase slows down; finally, the increase stops when it reaches maturity (see Figure 4).

This law is abstracted as a mathematical expression:

$$x(t) = \frac{k}{1 + ae^{-rt}} \quad (k, a > 0). \tag{1}$$

The population growth rate expression is

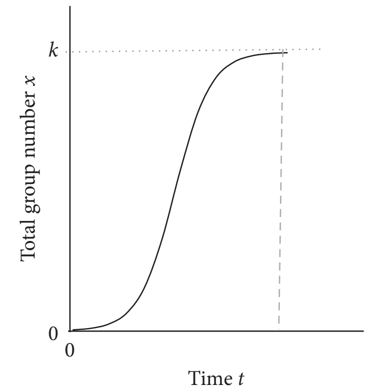


Figure 4: Logistic model.

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{k}\right). \tag{2}$$

The relative population growth rate expression iswhere k is the space saturation capacity, r is the innate rate of increase, and $r > 0$ represents the maximum increase of each individual in the population when it is not suppressed and reflects the inherent characteristics of the species [7]. The growth rate has a quadratic relationship with the population size x (see Figure 5).

$$\frac{dx}{xdt} = r\left(1 - \frac{x}{k}\right), \tag{3}$$

3.2. "S-Curve" and Information Dissemination in Physical Space. In the well-known spatial diffusion "average information domain" model [8], Hagerstrand simulates the probability distribution and empirical study of information diffusion by Monte Carlo method and draws an important conclusion that the accumulation of information users over time accords with the S-shaped logistic curve. The image of information diffusion rate varies with time as a whole, showing a "low-high-low" parabola shape "from low to high and then from high to low" (see Figure 5). Hagerstrand studies the information diffusion mode in the field of physical space; that is, in the physical diffusion space and environment, information diffusion depends on the ancient channels and ways of face-to-face communication. That is to say, the information diffusion of an event is accompanied by the formation of public opinion, and the object described by the information diffusion model in physical space is defined as the captured traditional social public opinion. However, generally speaking, in the physical space, the formation of traditional social public opinion is relatively slow.

3.3. Research on the Applicability of the "S-Curve" Network Public Opinion Prediction Model for Emergencies. In the following content, we apply the "S-curve" theory to the microblog public opinion analysis of the emergent formation event of the loss of contact with "Malaysia Airlines" and establish model 1 (m1) with the following expression:

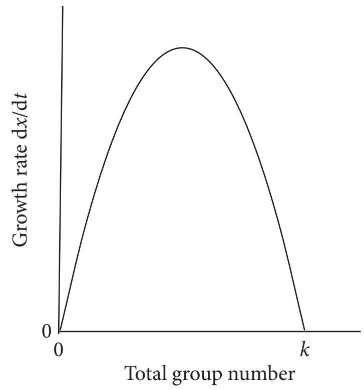


Figure 5: Logistic model (dx/dt) and x curve.

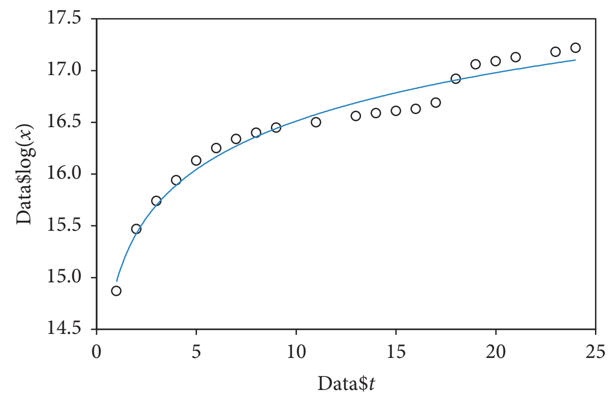


Figure 6: Fitted image of model 1 $\log(x)$ and t.

$$\log x_i = \frac{k}{1 + ae^{-rt_i}} + \varepsilon_i \quad (k, a > 0,\ i = 1, 2, \ldots, 24), \tag{4}$$

where k, a, and r are unknown parameters and ε_i is a random error variable and satisfies

$$\begin{aligned} E(\varepsilon_i) &= 0, \\ \operatorname{Var}(\varepsilon_i) &= \sigma_j^2, \\ \operatorname{Cov}(\varepsilon_i, \varepsilon_j) &= 0 \quad (i \neq j), \end{aligned} \tag{5}$$

where $E(\varepsilon_i)$ is the expected value of ε_i and $\operatorname{Var}(\varepsilon_i)$ is the variance of ε_i. In general, it is usually assumed that $\sigma_j^2 = \sigma^2$ (constant). $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = 0$ means that there is no correlation between observation errors; that is, the variation of errors is random and completely accidental. Under the assumption that $\sigma_j^2 = \sigma^2$ (constant), the ordinary least squares method can be used to estimate the parameters, but the remainder will be ignored when the model is linearized, and the estimation error will be large. Therefore, the nonlinear least squares method [9, 10] (unconstrained optimization) is used to estimate the parameters to make the error smaller; that is, $Q(k, a, r) = \sum_{i=1}^{n} (\log x_i - (k/(1 + ae^{-rt_i})))^2$ has a minimum value. The corresponding R language function is nls().

From the results of the program, model 1 and its various parameters have passed the significance test. The Residual Standard Error (RSE1) is 0.09769. We save the prediction result of model 1 on the test set in the variable test 1. The expression of Model 1 is

$$\log x(t) = \frac{16.991837}{1 + 0.164634e^{-0.223025t}}. \tag{6}$$

A graph of the relationship between the fitted line of model 1 and the original data is shown in Figure 6.

It can be roughly seen from Figure 6 that the fitting effect of model 1 is general, the data points from $t = 12$ to $t = 17$ are not well fitted, and there is a significant deviation from the original data, and after $t = 18$, the fitting of the final saturation value is also not ideal. Therefore, to make the fitting effect more ideal and more in line with the actual situation of the original data, this paper improves the theory of the "S-curve" and establishes a new model to explore the fitting effect of the network public opinion data.

3.4. Theoretical Improvement of "S-Curve". The traditional "S-curve" theory is suitable for the analysis and study of social public opinion of information diffusion in physical space. With the development of the Internet, social democracy is constantly dependent on the Internet. There is the occurrence of emergency, then inducing the network public opinion. At this time, the traditional "S-curve" theory can no longer accurately analyze and predict the network public opinion, so it is necessary to improve and innovate the "S-curve" theory.

3.4.1. Cyberspace Information Diffusion and Network Public Opinion. Traditional social public opinion is fleeting or difficult to capture. Some of them are daily discussions on the streets and alleys, while others only exist in the ideas of people. Therefore, the object described by the information diffusion model of physical space is only the captured traditional social public opinion. With the popularization of the Internet, information diffusion in cyberspace has gradually become an important research object. Information diffusion in cyberspace is a process, in which information continues to expand its scope of adoption and application in cyberspace over time. It is based on proliferating behavioral participants (diffusion nodes), supported by the network environment, and driven by information potential. Conditioned by information circulation and transmission, it is a complex process involving many factors [11].

The spread of network information for a certain event can also promote the formation of social public opinion, and the formation speed is faster than the physical space. The Internet breaks through the space constraints and reduces the time cost of information transmission, thereby reducing intermediate links [12]. Modern news media, enterprises, and the government also make full use of this feature of the Internet to publish as soon as a major event occurs, so that the masses can get news in a short time. The self-media era has also made information dissemination popular and autonomous, and the contents with dissemination values published by personal accounts will also be quickly disseminated. At the same time, as an equal, free, and open information communication platform, the Internet makes the flow and interaction of information possible, and both sides of information communication can interact with the other party equally. While people can quickly obtain information on the Internet, it also means that people can instantly express or share opinions, express attitudes, and vent their emotions on the Internet, that is, network public opinion; or they transfer the received network information to the physical space in a traditional way to promote the formation of public opinion in the physical space.

The formation of network public opinion and the promotion of public opinion in physical space are realized in the process of network information diffusion. Both are public opinions that can be captured and are part of modern social public opinion (see Figure 7). Therefore, the monitoring of the network public opinion has attracted increasing attention from the government.

3.4.2. "S-Curve" Applied to the Theoretical Improvement of Network Public Opinion in Emergency. The network information diffusion mode still conforms to the S-shaped logistic curve. However, in the "low-high-low" parabolic mode, where the information diffusion rate changes over time from low to high and then from high to low (see Figure 5), the transition of "low-high" from "low to high" is relatively rapid. For emergencies, the rapid transition of the "low-high" process of the network information diffusion model is more obvious. In the public opinion monitoring of the day, the data may only reflect the "high-low" pattern of "high to low." This is because, in the information dissemination of physical space or the network information dissemination of ordinary events, it may take three to five days or longer to form public opinion, while the network public opinion of emergencies can be formed within two days or even one day [13]. The "high-low" mode is determined by the momentary gathering of people and the destructiveness of behavior. Due to the sudden occurrence of emergencies and a higher degree of damage than conventional events, public opinion on emergencies can be formed extremely quickly with the help of the online platform. However, the dissemination of Internet public opinion in emergencies essentially still undergoes a process of "from low to high and then from high to low" with the time-varying information diffusion rate. The formation and extinction of public opinion still follows the mechanism of the formation and extinction of public opinion in physical space, that is, conforming to the logistic curve.

After the accumulated daily data, the images of the information diffusion mode of the physical space and the information diffusion mode of the emergency event are shown in Figures 8 and 9, respectively, where $k1 < k2$, $t1 > t2$; this is because, in the information network society, the Internet has a wide range of audiences, and knowledge and information are important sources of power. After emergencies, with the help of the Internet as a medium, the original accumulation of information with the help of the power of the network has reached an unprecedented degree of concentration, and the speed of information diffusion is faster than that of the physical space, so it takes a short time for the total cumulative information to reach the saturation value.

4. Theoretical Improvement and Model Establishment of Double "S-Curve"

4.1. Raising the Question. The "S-curve" is based on the assumption that the space saturation capacity k remains unchanged and does not take into account the sudden changes in the objective environment of public opinion. If the saturation capacity changes, which is reflected in the actual problem as a sudden change in the objective environment, the "S-curve" model cannot accurately describe the actual problem. However, in the real world, the constant change of the space saturation capacity is a frequent phenomenon. For example, technological innovation, productivity improvement, and even the release of updated news will bring about major structural changes [14]. At this time, it is more appropriate to use the double "S-curve" model to describe the reality.

4.2. Double "S-Curve" and Network Public Opinion of General Events. There are two situations in which external influencing factors lead to changes in the space saturation capacity: one is a jump-type mutation, that is, when the original curve is saturated or close to the saturation value due to the sudden addition of new influencing factors such as technological innovation. This causes the curve to reenter a new growth development period starting from the change point in time (see Figure 10). The other is a continuous and gradual change in the saturation capacity of the space; that is, a certain influencing factor has appeared in the early and mid-term of the curve development and continues to affect the development of the curve over time, resulting in the continuous expansion of the saturation capacity of the curve (see Figure 11). The above two situations describe the development law of the jumping double "S-curve" and the gradual double "S-curve," respectively. For the former, the curves before and after the change of the space saturation capacity are relatively independent and can be studied separately and can be merged into the original curve until the new growth reaches the new space capacity. For the latter, the curve change is an overall process, and changes at various moments are closely related and difficult to separate. Generally speaking, from the image, the two types of double "S-curves" have obvious jumps compared with the "S-curves."

After certain events in the physical space occur, if the reason for their occurrence is unclear or undisclosed, the original event is likely to exist in the form of an event when the official announcement of the reason or confirmation of the occurrence of the event occurs. And this original event will trigger a new round of public opinion, that is, follow-up public opinion. The formation of follow-up public opinion will change the saturation capacity of the original public opinion. Therefore, public opinion images that generally form events can be described by two types of double "S-curves."

4.3. Improving the Double "S-Curve" and Network Public Opinion of Emergencies. In cyberspace, when the cause of an event is unclear or undisclosed as a formative event, or the official announcement of the cause of the event or

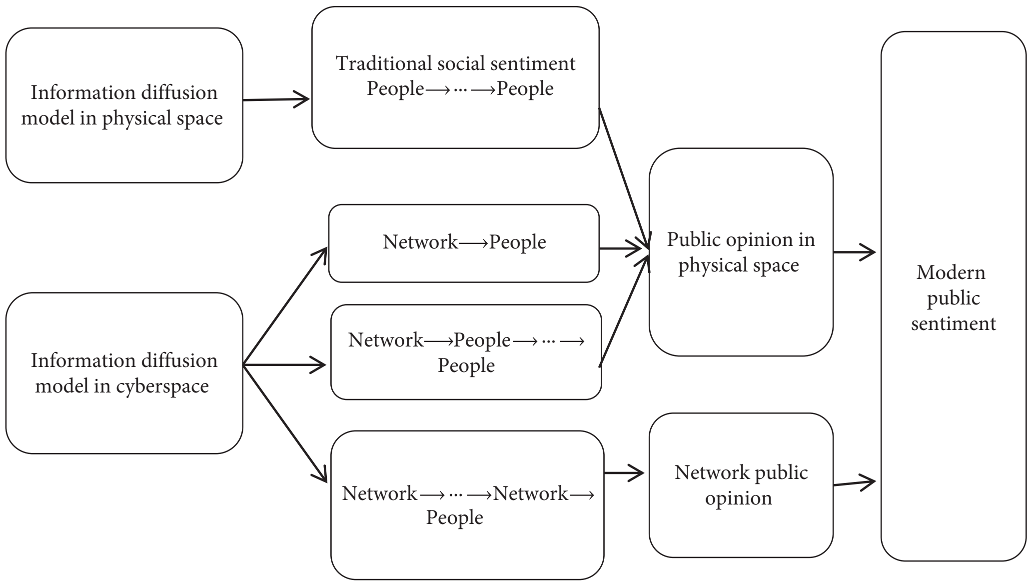


Figure 7: The relationship between traditional public opinion and modern public opinion.

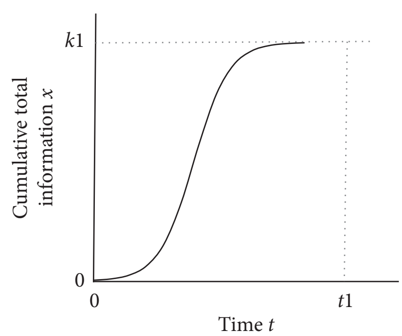


Figure 8: Information diffusion mode in physical space.

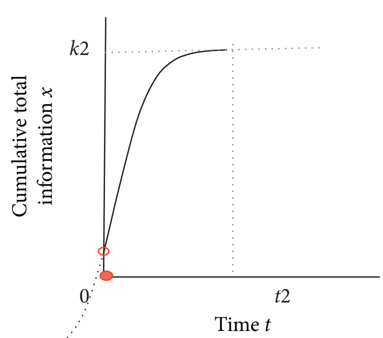


Figure 9: Network information diffusion mode of emergencies.

confirmation of the occurrence of the event may again cause major public opinion, that is, the overall network public opinion will undergo major structural changes, the public opinion, when public opinion is raised again, may be affected by an emergency that has occurred, or it may treat new news as a new event, and not be affected by an emergency that has occurred. Therefore, the network public opinion of emergencies can also be described by two types of double "S-curves."

However, in cyberspace, public opinion is formed faster than in physical space. Network public opinion for emergencies with unclear or undisclosed causes is formed faster than public opinion in physical space (initial public opinion formation), and such emergencies, such as forming events, once again form public opinion in cyberspace (follow-up public opinion formation). Therefore, in general, the formation of public opinion on the Internet for emergent events is faster than that for general events in physical space.

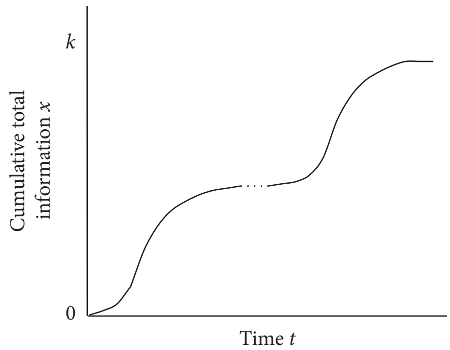


FIGURE 10: Information diffusion mode of general formation events (jump type).

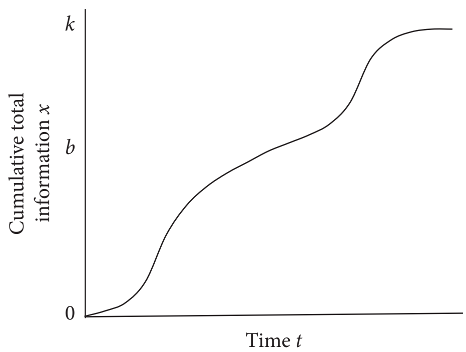


FIGURE 11: Information diffusion mode of general formation events (gradual type).

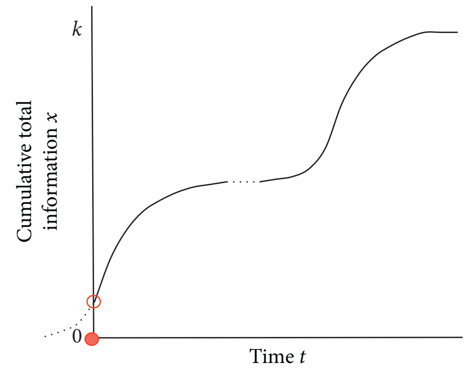


FIGURE 12: Network information diffusion mode of emergencies (jump type).

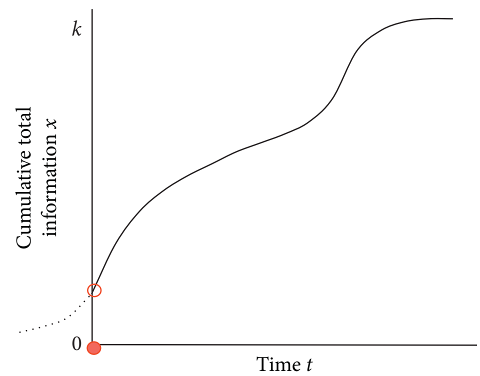


FIGURE 13: Network information diffusion mode of emergencies (gradual type).

The theoretical improvement in applying the double "S-curve" to the network public opinion of emergencies (Figures 10–12, Figures 11–13) is similar to the theoretical improvement in applying the "S-curve" to the network public opinion of emergencies(Figures 8 and 9).

The world is changing rapidly, and when emergencies with unclear or undisclosed causes are used as forming events, there may be many public opinions afterwards, the space capacity of the total amount of information will continue to increase, and the images will show more jumps. The development of network public opinion in emergencies may present a nonlinear development model of multiperiod, multiparameter, and dynamic multi-"S" shape development mechanisms. This article takes the "Malaysia Airlines" loss of connection in March 2014 as an example and establishes a double "S-curve" emergency event prediction model.

4.4. Establishing a Model Based on the Improved Jump-Type Double "S-Curve" Theory. The network information diffusion mode of emergency with jump change, that is, the improved jump double "S-curve," can be considered formed by the combination of two independent improved "S-curves" (see Figure 12). By using the improved jump double "S-curve" theory, we can establish model 2 (m2):

$$\log x_i = \sum_{h=1}^{2}\left(I\left(t_i \in T_h\right)\frac{k_h}{1 + a_h e^{-r_h t_i}}\right) + \varepsilon_i \quad \left(k_h,\ a_h > 0,\ h = 1, 2,\ i = 1, 2, \ldots, 24,\right). \tag{7}$$

$H = 1,\ 2$ represent the first and second independent "S-curves," respectively. And use nonlinear least squares to estimate the parameter $k_h, a_h, r_h (h = 1, 2)$ of model 2, so that $Q(k_1,\ a_1,\ r_1,\ k_2,\ a_2,\ r_2) = \sum_{i=1}^{n} (\log\ x_i - \sum_{h=1}^{2} (I(t_i\ \ \in T_h)$ $(k_h/(1 + a_h e^{-r_h t_i}))))^2$ has a minimum value.

From the x and t image in Figure 2, and the$\log(x)$ and t image in Figure 3, it can be seen that the "Mahan Airlines" event has a major jump at $t = 17$. Therefore, the split point is $t = 17$ to divide the data into two parts $t = 1$ to 16 and $t = 17$ to 24 for fitting. The corresponding R language function is nls(). From the results of program operation, Model 2 and its various parameters have passed the significance test. The Residual Standard Error (RSE2) is 0.03591. Store the prediction result of model 2 on the test set in the variable test 2. The expression of Model 2 is shown in the following way:

$$\log x(t) = I(t \le 16)\frac{16.791925}{1 + 0.168621e^{-0.300460t}} + I(t \ge 17)\frac{17.082978}{1 + 0.036851e^{-0.848520t}}. \quad (8)$$

$T \le 16$ means that, after the date is processed according to the number 1–24, the data of the first 16 days are displayed, and $t \ge 17$ can be expressed in the same way. Make a graph of the relationship between the fitted line of model 2 and the original data (see Figure 14). Figure 14 shows that the fitting effect of model 2 is very good, and the RSE2 is smaller than RSE1. This means that the fitting effect of model 2 is better than that of model 1. It indicates that the traditional information diffusion model of physical space is not so good in describing the network public opinion of the sudden formation of events. It means that when the objective environment of network public opinion in emergency changes suddenly, the traditional "S-curve" theory cannot be mechanically applied. Instead, the analysis and prediction of network public opinion under new circumstances should be more in line with objective development.

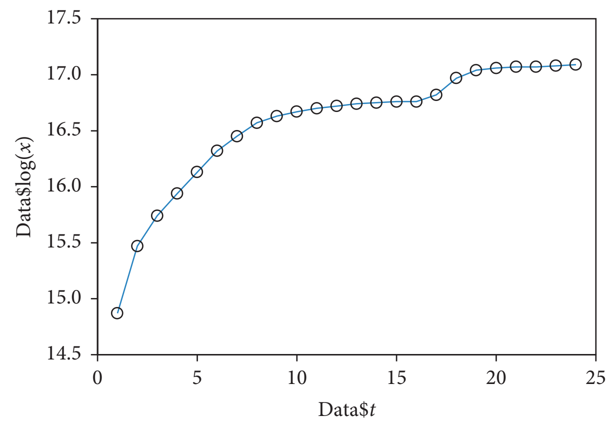


Figure 14: Fitting diagram of model 2 $\log(x)$ and t.

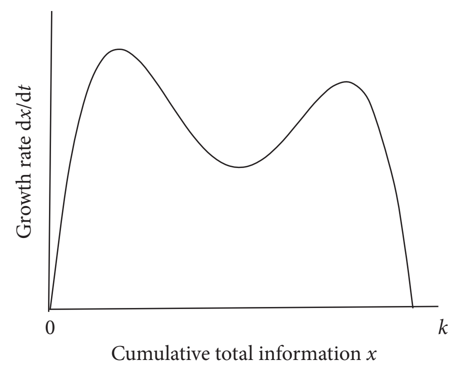


Figure 15: The $(dx/dt)$and x curve of the gradual double "S-curve" model.

4.5. Establishing a Model Based on the Improved Gradual Double "S-Curve" Theory

4.5.1. Gradual Double "S-Curve" Is Applied to the Public Opinion of General Events. The model we begin to build in this chapter is based on a differential equation model of our predecessor [6]. This kind of differential equation model can be used to describe the public opinion of general forming events in physical space (see Figure 10). This section will briefly introduce the process of establishing the model.

In practical applications, "S-curve" describes the information diffusion mode of physical space. If the gradual double "S-curve" expansion is carried out in physical space, firstly, according to the differential form of the "S-curve," it is a parabolic function with an opening downward (see Figure 5), and its relative growth rate image is a descending straight line. From this analogy, the differential curve of the gradual double "S-curve" should be "M-shaped" (see Figure 15); that is, there are two maximum values and one minimum value. It can be seen from the figure that the relationship of $(dx/dt)$and x is approximately a fourth-degree polynomial, and the differential form of the double "S-curve" satisfies the following equation:

$$\frac{dx}{dt} = rx\left|1 - \frac{x}{k}\right|\left|a\left(1 - \frac{x}{b}\right)^2 + c\right| \quad (k, b \ne 0), \quad (9)$$

where b is the space saturation capacity when the curve grows from the first stage to the second stage, and a and $c$are the parameters that control the shape of the curve.

Then, divide each point on the "M-shaped" growth rate curve by the value of x to get the equation that the relative growth rate dx/xdt satisfies:

$$\frac{dx}{xdt} = r\left|1 - \frac{x}{k}\right|\left|a\left(1 - \frac{x}{b}\right)^2 + c\right| \quad (k, b \ne 0). \quad (10)$$

We use approximate methods to replace dx/xdt with actual data. Use the values of dx/xdt and x, then use the least square method to estimate the values of the parameters r, k, a, b and c of equation (10), and solve the differential equation (9) to obtain x at each time.

4.5.2. The Theoretical Improvement of the Application of Gradual Double "S-Curve" in the Network Public Opinion of Emergencies. The mode of the gradual double "S-curve" to form the network public opinion prediction model of unexpected events should be similar to Figure 13. Therefore, the image of $(dx/dt)$and x should be as shown in Figures 16 and 17. As for Figure 16, the differential curve of the improved gradual double "S-curve" is still "M-type" (quartic function), and the differential form dx/dt of the double "S-curve" satisfies equation (9). The relative growth rate dx/xdt satisfies equation (10). As for Figure 17, the differential curve of the improved gradual double "S-curve" is more suitable for fitting with cubic functions, so the differential form dx/dt satisfies equation (11); the relative growth rate dx/xdt satisfies equation (12).

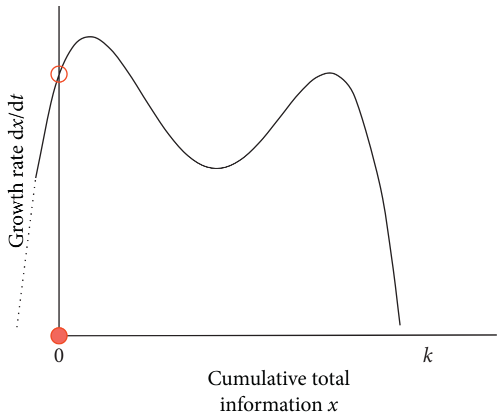


FIGURE 16: The $(dx/dt)$and x relationship of the improved gradual double "S-curve" model is a fourth-order image.

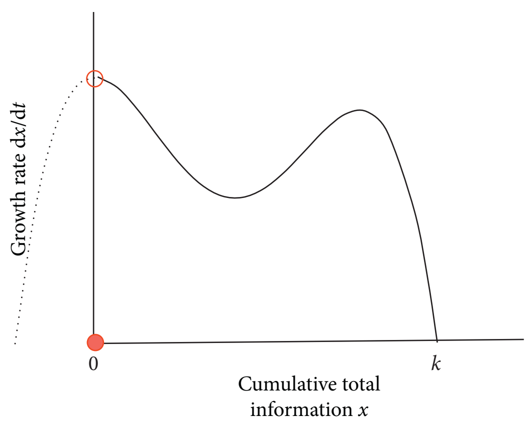


FIGURE 17: The $(dx/dt)$and x relationship of the improved gradual double "S-curve" model is a cubic image.

$$\frac{dx}{dt} = rx\left|1-\frac{x}{k}\right|\left|1-\frac{x}{b}\right| \quad (k, b \neq 0), \tag{11}$$

$$\frac{dx}{xdt} = r\left|1-\frac{x}{k}\right|\left|1-\frac{x}{b}\right| \quad (k, b \neq 0). \tag{12}$$

4.5.3. Improving the Gradual Double "S-Curve" and Network Public Opinion Prediction Model for Emergencies. Use the improved gradual double "S-curve" theory to establish a network public opinion prediction model for emergencies, which is recorded as Model 3. The detailed steps are as follows:

(i) Differential processing

The $\log(x)$ in the original data is differentially processed, and the derivative is approximately replaced by the central difference quotient to obtain the approximate value of the differential dx/dt and the relative growth rate dx/xdt (13), where x_t is the accumulated information data at a time t, and $dx/xdt|_{t=i}$ is the relative growth rate of the total amount of information at time t.

$$\frac{dx}{xdt}\Big|_{t=i} \approx \frac{x_{t=i+1} - x_{t=i-1}}{2\, x_{t=i}} \quad (i = 2, \ldots, 23). \tag{13}$$

Store the approximately calculated dx/dt and dx/xdt together with the data in rows 2–23 of the original data in the new data frame and put all data belonging to the training set ($i \in$ train) in data 2, used for the next parameter estimation.

(ii) Parameter Estimation

For different emergencies, when using curves to fit public opinion, a suitable function model should be selected according to the shape of the differential curve. From the training set $dx/dt|_{t=i}$ and $\log(x)$ (see Figure 18), it can be seen that the differential curve of the "Mahan" event is more suitable to be described by a cubic function. In this paper, the cubic function is used for fitting, and the four-degree fitting can be obtained by analogy.

Equation (12) is a cubic polynomial, converting it to a mathematical formula:

$$\frac{dx}{xdt} = Ax^2 + Bx + C. \tag{14}$$

To facilitate parameter estimation, the parameters A, B, C have a certain quantitative relationship with r, k, b in equation (12) and can be converted. Use the values of dx/xdt and x to establish a transition model (m3):

$$\frac{dx}{xdt}\Big|_{t=i} = Ax_i^2 + Bx_i + C + \varepsilon_i \quad (A, B, C \neq 0,\ i \in \text{train}). \tag{15}$$

The least square method is used to estimate the values of the parameters A, B, and C of equation (15). From the results of program operation, the transition model (m3) and its various parameters have passed the significance test.

(iii) Data Fitting

Substitute the estimated value of each parameter $A = 0.009175$, $B = -0.313680$, $C = 2.682796$ into equation (14), using MATLAB to solve the numerical solution of ordinary differential equations based on the Runge–Kutta method, and to obtain the estimated value of $\log(x)$ at each time ($t = 1$ to 24), namely, $\widehat{\log(x)}$, we store the predicted data belonging to the test set in the first 24 values solved in the variable test 3.

Draw the relationship between the fitted line of model 3 and the original data (see Figure 19). It can be seen initially from the figure that the fitting result of model 3 is better than that of model 1 in the period $t = 12$ to $t = 17$, and after $t = 22$, the final saturation value fits better than model 1. However, from the image point of view, compared with model 2, the fitting effect of model 3 is not ideal. To some extent, this is due to the irrational group behavior of

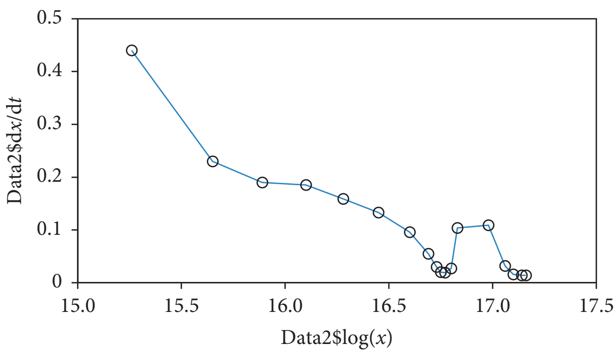


Figure 18: dx/dt and $\log(x)$ relationship diagram.

the sudden formation of network public opinion of events driven by the Internet. However, the description of model 3 is more consistent with the public's cognition of reality, reflecting the internal relations and internal laws of things; that is, the change of public opinion is an integral process, and the changes of each moment are closely related and difficult to be separated.

5. Comprehensive Evaluation of the Models

This article is based on predecessors' researches on the network public opinion of emergencies and obtains Weibo data related to the loss of contact with the emergent "Malaysia Airlines" from the big data platform. It innovatively puts forward the novel concept of network public opinion in emergency, that is, the sum of network public opinion caused by emergencies whose cause is unclear or undisclosed, and the follow-up network public opinion caused by the official announcement of the cause of the incident or the confirmation of the incident with this type of emergency as the formative event. Three types of prediction models are established for this network public opinion, and a comprehensive evaluation of the models is carried out at the same time. We use a more novel curve fitting method to establish a public opinion forecasting model. In the fitting process, the original model is reasonably explained and novel improved based on the characteristics of network public opinion and the actual situation. The three models fitted in the article are model 1 based on the "S-curve" theory, model 2 of the improved jump double "S-curve" theory, and model 3 of the improved gradual double "S-curve" theory.

Since the transition model (m3) was used when model 3 was established, the transition model (m3) is used to estimate that the intermediate parameters cannot represent model 3. Therefore, the RSE of the transition model (m3) is not comparable to the RSE of model 1 (m1) and model 2 (m2), so RSE cannot be used as a common standard for comprehensive model evaluation. To solve this problem, this paper uses the MAPE to evaluate different models. MAPE is expressed as a percentage and is dimensionless, eliminating the influence of the level of time series data and the unit of measurement. Therefore, it can be used to evaluate the same set of data in different models and measure the deviation between the observed value and the true value.

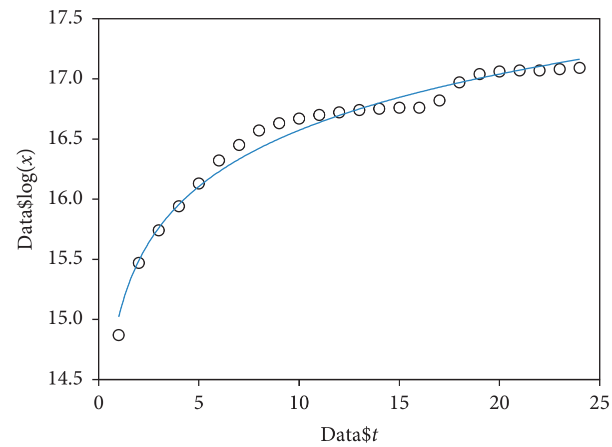


Figure 19: Fitted image of model 3 $\log(x)$ and t.

The expression is as follows:

$$\text{MAPE} = \frac{\sum_{t=1}^{n}\left|(\text{observed}_t - \text{predicted}_t)/\text{observed}_t\right|}{n} \times 100\%. \tag{16}$$

The original $\log(x)$of the test set and the prediction results of each model on the test set are gathered before the evaluation and comparison of the prediction effects of models 1, 2, and 3. Then, calculate each MAPE.

We obtained MAPE1 = 0.42703%, MAPE2 = 0.04263%, MAPE3 = 0.15782%. From the perspective of MAPE, the prediction situation from good to bad is model 2 > model 3 > model 1. And from the fitting curve of each model (see Figures 6, 14, and 19), the fitting situation from good to bad is model 2 > model 3 > model 1. Finally, we summarize the advantages and disadvantages of the three models, as shown in Table 2.

The results of the comprehensive evaluation and analysis of the three models from the perspective of fitting and prediction are of great significance to reality. First, the fitting and prediction effect of model 1 is inferior to models 2 and 3. Although the "S-curve" theory is suitable for forecasting in any period, its nonlinear fitting ability is weak, and the fitting effect is not ideal; this indicates that the traditional physical space information diffusion model is not suitable for describing the emergencies in the network public opinion. It means that when the objective environment of network public opinion in emergency changes suddenly, the traditional "S-curve" theory cannot be mechanically applied. Instead, the analysis and prediction of network public opinion under new circumstances should be more in line with objective development. Second, the important basic assumption for the establishment of model 2 is that the curve changes before and after the space saturation capacity changes are relatively independent and can be studied separately before the new growth reaches a stable level. The fitting effect and prediction effect of this model are very good, but the segmentation point needs to be established in advance; third, model 3 uses differential processing, which reduces the fitting progress to a certain extent; when mid- and long-term predictions are made, the deviation is large;

Table 2: Advantages and disadvantages of the three models.

Model	Principle	Advantage	Disadvantage
Model 1 (m1)	"S-curve"	Suitable for forecasting in any period	The nonlinear fitting ability is weak, and the fitting effect is not ideal
Model 2 (m2)	Improved jump double "S-curve"	Fitting effect and prediction effect are very good	The establishment of the model is based on the independence of local laws, and the segmentation point needs to be established in advance
Model 3 (m3)	Improved gradual double "S-curve"	It is suitable for forecasting in any period and can reflect the inner relationship and inner law of things	The use of differential processing reduces the fitting progress to a certain extent; the deviation is large when mid- and long-term prediction is made; the analytical formula of the differential equation is not easy to obtain, and the numerical solution should be used instead

the analytical formula of the differential equation is not easy to obtain, so it needs to be replaced by a numerical solution. Although the fitting effect of model 3 is not as good as that of model 2, the description of model 3 is more in line with people's perception of reality, reflecting the internal relations and internal laws of things; that is, the change of public opinion is an overall process, and the changes at each moment are closely related and difficult to separate. However, the fitting and prediction effect of model 2 is better than that of model 3, to a certain extent, indicating that the network public opinion of emergencies is irrational under the drive of the Internet. Under this circumstance, when the official announcement of the cause of the incident or confirmation of the occurrence of the incident is done, the network public opinion caused by the emergent event as a forming event can be treated as a new round of network public opinion that is inevitable, but the time of occurrence is uncertain. After an emergency occurs, the government has to not only deal with the panic caused by the current network public opinion, but also prepare for more intense network public opinion that may occur in the future. Before the release of updated news related to emergencies or the announcement of the truth of the matter, the government should use appropriate means to actively guide people's emotions in advance to reduce the peak of secondary public opinion on the Internet and to avoid excessive ideological reactions and extreme behavior of the public when the truth is restored, and to avoid emergencies derived from this. Therefore, when predicting the network public opinion of emergencies, in addition to the initial public opinion, the follow-up public opinion is also worthy of attention. This provides more practical value for subsequent supervision and guidance of public opinion.

6. Concluding Remarks

In this paper, we obtained 24-day data related to the loss of contact with "Malaysia Airlines" from the official Weibo platform to study the information diffusion mode under the network public opinion. According to the difference between network public opinion diffusion and physical space public opinion diffusion, as well as the occurrence of emergencies, on the basis of the traditional S-curve model, three novel public opinion prediction models are innovatively determining the specific calculation procedures to carry out data fitting analysis and prediction: model 1 based on the "S-curve" theory, model 2 of the improved jump double "S-curve" theory, and model 3 of the improved gradual double "S-curve" theory. The main contribution of this paper is that, through the novel explanation and improvement of the model, it provides a better analysis and prediction method for the information diffusion mode of network public opinion. Considering the formation of emergencies, changes in the objective environment, and rapid transmission of information in the context of cyberspace, not only the initial public opinion within a short period of time but also subsequent public opinion within a long period of time is taken into account. It solves the problem of public opinion prediction in the big data network. Then, specific calculation programs are given in these models, and the trend of information diffusion is predicted. We overcome the spread of network public opinion under the cognitive feedback of the public to the reality under the sudden change of objective environment to deal with the problem. Finally, through data fitting, MAPE is used to eliminate the influence of data level and units of measurement in time series, and the comprehensive evaluation shows that the novel improved model is representative of the information diffusion mode under network public opinion, with more effectiveness and higher practicability, in line with the development of modern society. At the same time, the government can use the novel improved "S-curve" model to timely and effectively predict and analyze public opinion. With positive public opinion forecasts, the government can step up publicity and send signals to the public in all aspects to guide them. For negative public opinion forecasts, the government can properly release relevant signals to the public in advance, so that the public can be psychologically prepared to receive them to prevent the sudden outbreak of the public after the incident, in case it is out of control.

Data Availability

The data used to support the findings of this study are available from the micro index (https://data.weibo.com/index).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was financially supported by the National Natural Science Foundation of China (No. 71471075), Fundamental Research Funds for the Central University (No. 19JNLH09), Innovation Team Project in Guangdong Province, P.R. China (No. 2016WCXTD004) and Industry-University-Research Innovation Fund of Science and Technology Development Center of Ministry of Education, P.R. China (No. 2019J01017), and the authors gratefully acknowledge this support.

References

[1] L.-h. Wang, "A brief discussion on the study of "public opinion"," *Tianjin Social Sciences*, vol. 24, no. 2, pp. 78–81, 2004, in Chinese.

[2] C.-h. Yi and Z. He, "Study on the regularities of formation and evolution of internet public opinions caused by emergent events," *Journal of Xiangtan University (Philosophy and Social Sciences)*, vol. 38, no. 2, pp. 74–78, 2014, in Chinese.

[3] Y. He, Y. Zhang, and M. Xiao, "Etc. Infection relationship research on the emotions of different microblogging community: with the topic of "air asia losing contact" as an example," *Statistics & Information Forum*, vol. 32, no. 8, pp. 110–116, 2017, in Chinese.

[4] T. Hagerstrand, *Innovation Diffusion as a Spatial Process*, The University Chicago Press, Chicago, IL, USA, 1967.

[5] E. M. Rogers, *Diffusion of Innovations*, Simon and Schuster, New York, NY, USA, 2003.

[6] W.-z. Zhu, "New characteristic of crisis information communication model and its management strategies in digital age," *Journal of Modern Information*, vol. 29, no. 2, pp. 60–63, 2009, in Chinese.

[7] J.-q. Liu and W.-y. Yuan, "Updated logistic model: double S-shape curve model," *Mathematics in Practice and Theory*, vol. 38, no. 17, pp. 60–65, 2008, in Chinese.

[8] Y.-r. Deng, *Research on Information Diffusion Based on Network Dimension-Force*, Harbin Engineering University, Harbin, China, 2008, in Chinese.

[9] X.-b. Wang, "The method of nonlinear least square fitting logistic curve and its application in diffusion of computer technology," *Journal of Shanghai University (Natural Science Edition)*, vol. 1, no. 4, pp. 369–375, 1995, in Chinese.

[10] M.-l. Cheng, "Extension and application of logistic curve," *Operations Research and Management*, vol. 12, no. 3, pp. 85–88, 2003, in Chinese.

[11] Y.-r. Deng and X. f. Xu, "Analysis of information diffusion under the network environment," *Modern Management Science*, vol. 27, no. 2, pp. 59–61, 2008, in Chinese.

[12] X.-d. Liu, Y.-t. Cao, and L.-m. Li, "Predicting model in network impact: a monitoring and warning system for public opinion in universities under big data framework," *Journal of Shenzhen University (Humanities & Social Sciences)*, vol. 32, no. 4, pp. 156–160, 2015, in Chinese.

[13] X.-d. Liu and J.-l. Chen, "A differentiated strategy of tourism e-commerce product promotion based on provincial tourism development difference: based on large data and data visualization method," *Statistics and Information Forum*, vol. 32, no. 8, pp. 31–38, 2017, in Chinese.

[14] X.-d. Liu and W.-j. Li, "Study on jump volatility of financial high-frequency data: based on the method of big-data kernel functions SVM," *Statistics & Information Forum*, vol. 33, no. 9, pp. 23–30, 2018, in Chinese.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5401499, 14 pages
https://doi.org/10.1155/2021/5401499

Research Article

Research on Freight Development of Guangdong Province Based on Grey Theory Model

Lianhua Liu ,[1,2] Aili Xie ,[1] and Hai Ping[1]

[1]*Guangzhou Huashang College, Guangzhou 511300, Guangdong, China*
[2]*Huashang Business Economic and Social Research Institute, Guangzhou 511300, Guangdong, China*

Correspondence should be addressed to Aili Xie; ailixie16@126.com

Received 3 May 2021; Revised 1 June 2021; Accepted 19 June 2021; Published 21 July 2021

Academic Editor: Yiwen Zhang

Logistics and economic development complement each other. The comprehensive competitiveness of Guangdong provincial economy ranks first in China. Under the influence of COVID-19, the freight development of Guangdong Province has been affected, but there is still lack of quantitative research. It is significant to explore the trend of economic development through the freight development of Guangdong Province. Based on the grey theory model, this paper uses six freight indexes to research freight development of Guangdong Province. Under the assumption that COVID-19 did not happen, we predicted the development value of freight index of Guangdong Province from January to December in 2020 and studied the influence based on the comparison between the predicted value and actual value. The empirical study shows three impact characteristics: stage characteristics, structural characteristics, and entity transmission characteristics. COVID-19 has a negative impact on the development of total freight volume, highway freight volume, waterway freight volume, and air freight volume in Guangdong Province. The influence values were −23.001%, −29.344%, −11.296%, and −3.838%. But, the freight volumes of railway and pipeline were positively affected by 14.343% and 13.057%, respectively, due to their continuity and substitution to other transportation modes. To further explore the abate measures of COVID-19 impact on Guangdong's freight development, the grey correlation model is introduced to study the correlation factors of freight development of Guangdong Province. Through the research of the related factors, this paper puts forward some measures to promote the freight development of Guangdong Province in the postepidemic era.

1. Introduction

Guangdong Province is the South Gate of China, located in the South China Sea shipping hub position, since the Qin and Han Dynasties have become the starting point of the maritime Silk Road. The comprehensive competitiveness of Guangdong provincial economy ranks first in China. Since 1989, Guangdong's GDP has ranked first in China for 32 consecutive years, becoming the largest economic province in China. In 2020, Guangdong Province's GDP reached 11.08 trillion yuan, accounting for one-eighth of China's economy. The nine cities in the Pearl River Delta of Guangdong Province join hands with Hong Kong and Macao to build the bay area of Guangdong, Hong Kong, and Macao and become one of the four bay areas in the world side by side with New York Bay area, San Francisco Bay area, and Tokyo Bay area. The developments of economy and logistics complement each other. The rapid economic development of Guangdong Province promotes the improvement of logistics development level [1–4], which further supports the economic development of Guangdong Province.

The COVID-19 outbreak as a public health emergency has a negative impact on the economy from multiple channels. It also has a direct impact on the logistics economy. The epidemic situation affects the demand and supply of the logistics market. It affects the development of logistics enterprises through the operation of logistics and indirectly transmits to the field of logistics through the impact on domestic real economy and foreign trade. In the postepidemic period, the relevant supporting policies need a certain industry and time transmission, so COVID-19's impact on the logistics industry is characterized by multiple channels and multiple spatial and temporal superposition. Guangdong Province as an important node of the global supply chain, the total freight volume, road freight volume, railway freight volume, waterway freight volume, air freight volume, and pipeline freight volume has been affected by COVID-19.

In order to study the impact of COVID-19 on freight transportation in Guangdong Province and explore mitigation measures, this paper intends to use the grey prediction model GM (1, 1) to analyze the impact and introduce the grey correlation model to conduct a quantitative study on the correlation degree of the factors affecting the freight development of Guangdong Province. According to the correlation factors of the freight development of Guangdong Province, this paper puts forward some measures to promote the freight development of Guangdong Province in the postepidemic era.

2. Literature Review

2.1. Research Progress on Forecasting Freight Development. In recent years, logistics development prediction has become a hot issue for scholars. The methods commonly used by scholars include exponential smoothing [5–7], linear model [8], BP neural network method [5–9], multiple regression analysis [5, 9], seasonal autoregressive model [10–13], discrete wavelet technology [14], vector autoregressive method [15], and Markov chain theory [16], which are widely used in the prediction of logistics and freight development. In addition to conventional prediction methods, scholars also innovate prediction methods, such as a genetic algorithm and backpropagation (GA-BP) prediction model (optimized backpropagation neural network model using genetic algorithm), which are used to predict freight volume demand with small error [17]. Scholars use L-OD logistics demand forecasting method and construct a new model of double constraint gravity model to forecast logistics distribution, which achieves good forecasting effect. The state travel demand model (STDM) [18] is introduced to forecasting freight development, and a new hybrid multicriteria decision-making model combining Delphi, analytic network process (ANP), and quality function deployment (QFD) in fuzzy environment is applied to freight forecasting [19].

2.2. Application of GM in Forecasting Freight. Grey system theory is an effective method of studying and modeling systems consisting of small sample sizes that contain a limited amount of information and is widely used in many fields. The valuable information is extracted by processing the known information. This is further used to explore the evolution laws of the system and thus establish a prediction model. As there are many factors influencing the freight development of Guangdong Province, e.g., environmental factors of transportation and logistics, regional economic environment, government policy, and science and technology environment, it can be regarded as a grey system. Thus, it can be described using a grey model (GM). The GM (1, 1) is the most generally used grey model. GM (1, 1) used to predict the freight volume achieved satisfactory results [5, 9, 20–24].

To sum up, the grey prediction model is widely used in logistics development and freight volume prediction. Therefore, it is scientific and feasible to use the grey prediction model in the quantitative research of Guangdong freight development:

(1) The innovation of this paper is the prediction of impact value refined to month which is a more precise observation of COVID-19's impact on Guangdong's freight development.
(2) Practical significance: through quantitative and accurate research on the monthly impact value of freight development in Guangdong Province, we can grasp the situation of economic operation in Guangdong Province from the side and promote the formulation and implementation of relevant economic stimulus policies, which has certain management practical significance.

3. Research Methods

3.1. Introduction of GM (1, 1). The differential equation of grey system theory is called GM. G stands for grey, M stands for model, and GM (1, 1) is a one-order and one-variable differential equation model. The modeling process and mechanism of GM (1, 1) are as follows:

$$X^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), x(3), \ldots, x^{(0)}(n)\right\}. \tag{1}$$

(1) Record the original data sequence as a nonnegative sequence, where $X^{(0)}$ is a nonnegative sequence, $X^{(0)}(k) \geq 0, k = 1, 2, \ldots, n$.
(2) Generate a cumulative data sequence $X^{(1)}$:

$$X^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \ldots, x^{(1)}(n)\right\}, \tag{2}$$

where $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)$, $k = 1, 2, \ldots, n$. $Z^{(1)}$is the adjacent mean generating sequence of $X^{(1)}$:

$$Z^{(1)} = \left\{z^{(1)}(1), z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n)\right\}, \tag{3}$$

where $Z^{(1)}$ $(k) = 0.5x^{(1)}(k) +$ $0.5x^{(1)}(k-1)$, $k = 1, 2, \ldots, n$.
(3) Establish GM(1,1):

$$x^{(0)}(k) + aZ^{(1)}(k) = b. \tag{4}$$

in which "a" and "b" are parameters, which are development grey number and endogenous control grey number, respectively.
(4) Solve the parameters "a" and "b":
If $a = \begin{pmatrix} a & b \end{pmatrix}^{T}$ is a parameter column and

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(N) \end{bmatrix},$$
$$B = \begin{bmatrix} z^{(1)}(2) & 1 \\ z^{(1)}(3) & 1 \\ \vdots & \vdots \\ z^{(1)}(n) & 1 \end{bmatrix}, \quad (5)$$

then find the least square estimation coefficient sequence of differential equation $x^{(0)}(k)+az^{(1)}(k) = b$, satisfying condition $\hat{a} = (B^T B)^{-1} B^T Y$.

(5) Establish a prediction model.

Establish the whitening equation corresponding to the grey differential equation, as the following formula:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b. \quad (6)$$

As mentioned above, the time response sequence of GM (1, 1) grey differential equation $x^{(0)}(k) + az^{(1)}(k) = b$ is the following formula:

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}, \quad k = 1, 2, 3, \ldots, n. \quad (7)$$

The predicted values generated by restoration are as follows:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \quad k = 1, 2, 3, \ldots, n. \quad (8)$$

3.2. Grey Prediction Model (GM (1, 1)) Test

3.2.1. Modeling Rationality Test. Firstly, the modeling rationality of the original sequence $x^{(0)}$ is tested, and the grade ratio$\lambda(k)$ is used to represent

$$\lambda(k) = \frac{x^{(0)}(k+1)}{x^{(0)}(k)}, \quad (9)$$

where $k = 1, 2, 3, \ldots, n$. When all $\lambda(k) \in (e^{-(2/n+1)}, e^{(2/n+1,)})$, the GM (1,1) can be used in the $x^{(0)}$ series for satisfactory prediction modeling.

3.2.2. Grey System Model Prediction Accuracy Test

(1) Residual Error Test. After modeling with GM (1, 1), the predicted value sequence is as follows:

$$\hat{X}^{(0)}(K) = \left\{\hat{x}^{(0)}(1), \hat{x}^{(0)}(1), \ldots, \hat{x}^{(0)}(n)\right\}. \quad (10)$$

By calculating the series $X^{(0)}(K)$ and $\hat{x}^{(0)}(K)$, the following GM (1, 1) modeling rationality residual test index is obtained.

① Absolute residual is

$$\varepsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k). \quad (11)$$

② Relative residual is

$$\Delta(k) = \left|\frac{\varepsilon(k)}{x^{(0)}(k)}\right|. \quad (12)$$

③ Average residual is

$$\overline{\Delta(k)} = \frac{1}{n}\sum_{i=1}^{n}\Delta(k). \quad (13)$$

④ Average accuracy is

$$\rho^0 = (1 - \overline{\Delta(k)}) \times 100\%. \quad (14)$$

Residual test standards are as shown in Table 1.

(2) Posterior Variance Test. Suppose $X^{(0)}$is the original sequence, $\hat{X}^{(0)}$is the simulation error sequence, and $\varepsilon^0$is the absolute residual sequence. Test standards are shown in Table 1.

① Mean value of $X^{(0)}$ is

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x^{(0)}(k). \quad (15)$$

② Variance of $X^{(0)}$ is

$$S_1 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(x^{(0)}(k) - \overline{x}\right)^2}. \quad (16)$$

③ Mean of absolute residuals is

$$\overline{\varepsilon} = \frac{1}{n}\sum_{k=1}^{n}\varepsilon(k). \quad (17)$$

④ Absolute residual variance is

$$S_2 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\varepsilon(k) - \overline{\varepsilon})^2}. \quad (18)$$

⑤ Ratio of variance is

$$C = \frac{S_2}{S_1}. \quad (19)$$

⑥ Probability of small error is

$$P = p\left(|\varepsilon(k) - \overline{\varepsilon}| < 0.6745 S_1\right), \quad (20)$$

$$S_0 = 0.6745 S_1. \quad (21)$$

3.3. Grey Correlation Degree. Correlation degree refers to the measurement of the correlation between the factors of the

Table 1: Grade reference table of grey system model parameters.

Accuracy class	Relative residuals ($\Delta(k)$)	Mean variance ratio (C)	Probability of small error (P)	Correlation degree (γ_{0i})
Good	0.01	0.35	0.95	0.90
Qualification	0.05	0.45	0.85	0.80
Barely qualified	0.1	0.50	0.70	0.70
Disqualification	0.2	0.65	0.70	0.60

two systems that change with time or different objects. Grey system theory puts forward the concept of grey correlation analysis for each subsystem and intends to seek the numerical relationship between each subsystem in the system through certain methods or measure the influence of each subfactor in the system on the main factor. Grey correlation analysis provides a quantitative measurement for the development and change trend of a system, which is suitable for dynamic process analysis.

The calculation steps of grey correlation degree are as follows:

(1) Establish a raw data matrix X_i for each index:

$$X_i = (x_i(1), x_i(2), \ldots, x_i(k)). \tag{22}$$

Here, $x_i(k)$ represents the original data of i factor in year k.

(2) Calculate initialization transformation of matrix X_i':

$$X_i' = \left(\frac{x_i(1)}{x_i(1)}, \frac{x_i(2)}{x_i(1)}, \ldots, \frac{x_i(k)}{x_i(1)}\right) = (x_i'(1), x_i'(2), \ldots, x_i'(k)). \tag{23}$$

(3) Calculate the difference sequence $\Delta_{0i}(k)$. The difference sequence occurs between the main factor sequence data and the measure factors:

$$\begin{aligned}\Delta_{0i}(k) &= \left|x_0(k) - x_i'k\right|, \quad k = 1, 2, \ldots, n, i = 1, 2, \ldots, m,\\ \Delta_{0i}(k) &= (\Delta_{0i}(1), \Delta_{0i}(2), \ldots, \Delta_{0i}(k)).\end{aligned} \tag{24}$$

(4) The correlation coefficient $\xi_{0i}(k)$ and grey correlation degree γ_{0i} are calculated:

$$\xi_{0i}(k) = \frac{\min_i \min_k \Delta_{0i}(k) + \phi \max_i \max_k \Delta_{0i}(k)}{\Delta_{0i}(k) + \phi \max_i \max_k \Delta_{0i}(k)}, \tag{25}$$

where$\varphi$is the resolution coefficient, which improves the significance of the difference between the correlation coefficients, $\varphi \in (0, 1)$, and the general value is 0.5. The grey correlation degree is shown as follows:

$$\gamma_{0i} = \frac{1}{n}\sum_{k=1}^{n} \xi_{0i}(k). \tag{26}$$

In summary, the relevant parameters of the grey system model level index [20] reference table are shown in Table 1.

4. Construction of GM (1, 1)

4.1. Data Source and Description. The data are from the monthly open statistics of Guangdong Provincial Bureau of Statistics from 2013 to 2020. The data content includes six groups of data: total freight volume, truck freight volume, railway freight volume, waterway freight volume, air freight volume, and pipeline freight volume of Guangdong Province. For the convenience of calculation, total freight volume, railway freight volume, truck freight volume, water freight volume, air cargo volume, and pipeline freight volume are expressed. The annual freight development of Guangdong Province is counted, covering the six freight indexes. The original data to predict the development of freight of Guangdong Province in 2020 are shown in Table 2.

4.2. Rationality of GM (1, 1) Construction. The rationality of model construction was tested in advance. On the basis of the original data sequence, the stage ratio $\lambda(k)$ is calculated according to formula (10).

4.2.1. Calculation of the Grade Ratio $\lambda(k)$. The grade ratio is calculated as follows:

$$\begin{aligned}\lambda(k)F &= (0.8486, 0.9526, 1.0130, 0.9395, 0.9341, 0.9565),\\ \lambda(k)R &= (1.0918, 1.0776, 0.9792, 1.4029, 0.9753, 0.9478),\\ \lambda(k)T &= (0.8256, 0.9319, 1.0372, 0.9450, 0.9423, 0.9595),\\ \lambda(k)W &= (0.8813, 1.0004, 0.9421, 0.8827, 0.9175, 0.9445),\\ \lambda(k)A &= (0.9167, 0.9730, 0.9250, 0.9697, 0.7466, 0.9325),\\ \lambda(k)P &= (0.9420, 1.0599, 0.9699, 0.9663, 0.8258, 1.0042).\end{aligned} \tag{27}$$

4.2.2. Judgment of the Grade Ratio. $\lambda(k) \in (e^{-(2/n+1)}, e^{(2/n+1)})$, when $n = 7$,$\lambda(k) \in (0.7788, 1.2840)$. If λ(k) in the above range, $x^{(0)}(k)$ is suitable for GM (1,1). $\lambda(k)F \in (0.8486, 1.0130)$, $\lambda(k)T \in (0.8256, 1.0372)$, $\lambda(k)\ W \in (0.8813, 1.0004)$, and $\lambda(k)\ P \in (0.8258, 1.0559)$, where $k = 2, 3, 4, 5, 6, 7$, and the grade ratios are in $(0.7788, 1.2840)$; hence,$X_F^0 X_T^0$,X_W^0, and X_P^0 can be modeled by GM (1,1). The values of grey development coefficient "-a" are 0.0429, 0.0364, 0.0752, and 0.0539, respectively, i.e., less than 0.3, which are suitable for medium- and long-term prediction.

However, $\lambda(k)R \in (0.9478, 1.4029)$ and $\lambda(k)A \in (0.7446, 1.0442)$. There is a grade ratio not in the region, which is not available

Table 2: Monthly data of freight indicators of Guangdong Province from 2013 to 2020 (unit: million tons) (%).

Year	Total freight volume (F)	Railway freight volume (R)		Truck freight volume (T)		Water freight volume (W)		Air cargo volume (A)		Pipeline freight volume (P)	
		Amount	Ratio	Amount	Ratio	Amount	Ratio	Amount	Ratio	Amount	Ratio
2013	3065.44	122.38	3.992	2176.43	70.999	688.51	22.460	1.32	0.043	76.81	2.506
2014	3612.49	112.09	3.103	2636.19	72.974	781.25	21.626	1.44	0.040	81.54	2.257
2015	3792.21	104.02	2.743	2828.87	74.597	780.93	20.593	1.48	0.039	76.93	2.029
2016	3743.48	106.23	2.838	2727.37	72.857	828.93	22.143	1.60	0.043	79.32	2.119
2017	3984.69	75.72	1.900	2886.19	72.432	939.04	23.566	1.65	0.041	82.09	2.060
2018	4265.78	77.64	1.820	3063.04	71.805	1023.51	23.994	2.21	0.052	99.41	2.330
2019	4459.75	81.92	1.837	3192.31	71.580	1083.71	24.300	2.37	0.053	98.99	2.220
2020	3547.51	77.67	2.189	2313.56	65.216	1036.57	29.220	2.38	0.067	117.35	3.308

Data source: according to the website data of Guangdong Provincial Bureau of Statistics.

for $X_R^{(0)}$and $X_A^{(0)}$modeled in GM (1, 1) directly. Therefore, it is necessary to do translation transformation on the data, and the translated data are $X_R^{(0)'}$ and $X_A^{(0)'}$, as follows:

$$X_R^{(0)'} = (202.38, 192.09, 184.02, 186.23, 155.72, 157.64, 161.92),$$
$$X_A^{(0)'} = (2.12, 2.24, 2.28, 2.40, 2.45, 3.01, 3.17). \tag{28}$$

The grade ratios of $X_R^{(0)'}$and $X_A^{(0)'}$are calculated:

$$\begin{aligned}
&\lambda' (k)R = (1.0536, 1.0439, 0.9881, 1.1959, 0.9878, 0.9736),\\
&\lambda'^{(k)}R \in (0.9736, 1.1959), \quad K = 2, 3, 4, 5, 6, 7,\\
&\lambda' (k)A = (0.9464, 0.9825, 0.9500, 0.9796, 0.8140, 0.9495),\\
&\lambda' (k)A \in (0.8140, 0.9825), \quad K = 2, 3, 4, 5, 6, 7.
\end{aligned} \tag{29}$$

The grade ratios of $X_R^0$and X_A^0 are in the rank of $\lambda(k)$ and $\in$ (0.7788, 1.2840). The values of grey development coefficient "$-a$" are −0.0435 and −0.0784, which are less than 0.3. So $X_R^0$and X_A^0 are suitable for medium- and long-term prediction.

4.3. Construction of GM (1, 1). According to the original sequence, the accumulated sequence is generated and the differential equation is constructed. On the basis of solving the values of "a" and "b," the time response sequence is obtained, as formulas (30)–(35). The predicted values are calculated according to the time response sequence, shown in Tables 3 and 4.

$$\widehat{X}^{(1)} (k + 1)_F = \left(X^{(0)} (1)_F - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} = 81237.44e^{0.0429k} - 78172.00, \tag{30}$$

$$\widehat{X}^{(1)} (k + 1)_R = \left(X^{(0)\prime} (1)_R - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} = 4713.89 - 4511.51e^{-0.0435k}, \tag{31}$$

$$\widehat{X}^{(1)} (k + 1)_T = \left(X^{(0)} (1)_T - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} = 71059.66e^{0.03637k} - 6888311.00, \tag{32}$$

$$\widehat{X}^{(1)} (k + 1)_W = \left(X^{(0)} (1)_W - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} = 9522.20e^{0.0752k} - 8833.69, \tag{33}$$

$$\widehat{X}^{(1)} (k + 1)_A = \left(X^{(0)\prime} (1)_A - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} = 25.8654e^{0.0784k} - 23.7454, \tag{34}$$

$$\widehat{X}^{(1)} (k + 1)_P = \left(X^{(0)} (1)_P - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} = 1357.82e^{0.0539k} - 1281.01. \tag{35}$$

4.4. Validation of Modeling Accuracy of GM (1, 1) of Freight Volume in Guangdong Province. According to the modeling results of GM (1, 1), we will test the accuracy of the model by the residual test and posterior test.

Table 3: Actual value and forecast value of the freight development index in Guangdong Province (unit: million tons).

Year	Total freight volume (F)		Railway freight volume (R)		Truck freight volume (T)	
	X_F^0	$\widehat{X}_F$	X_R^0	$\widehat{X}_R$	X_T^0	$\widehat{X}_T$
2013	3065.44	3065.44	202.38	202.38	2176.43	2176.55
2014	3612.49	3561.51	192.09	192.04	2636.19	2632.37
2015	3792.21	3717.65	184.02	183.87	2828.87	2729.88
2016	3743.48	3880.63	186.23	176.04	2727.37	2831.01
2017	3984.69	4050.76	155.72	168.55	2886.19	2935.88
2018	4265.78	4228.35	157.64	161.37	3063.04	3044.64
2019	4459.75	4413.72	161.92	154.50	3192.31	3157.43

Source: according to the calculation criteria.

4.4.1. Residual Test. The residual test is carried out to test the accuracy of GM (1, 1) in forecasting the freight volume of Guangdong Province. The predicted value of total freight volume, railway freight volume, truck freight volume, water freight volume, air cargo volume, and pipeline freight volume is compared with the original data. $\varepsilon(k)_F$, $\varepsilon(k)_R$, $\varepsilon(k)_T$, $\varepsilon(k)_W$, $\varepsilon(k)_A$, and $\varepsilon(k)_P$ and $\Delta(k)_F$, $\Delta(k)_R$, $\Delta(k)_T$, $\Delta(k)_W$, $\Delta(k)_A$, and $\Delta(k)_P$are calculated according to formulas (11)–(14), and the results are shown in Tables 5 and 6.

$$\begin{aligned}
&\Delta(k)_F \le 0.05,\\
&\overline{\Delta(k)_F} = 0.0152 < 0.05,\\
&\quad \rho_F^0 = \left(1 - \overline{\Delta(k)_F}\right) * 100\% = 98.48\% > 90\%,\\
&\Delta(k)_R \le 0.05,\\
&\overline{\Delta(k)_R} = 0.0297 < 0.05,\\
&\quad \rho_R^0 = \left(1 - \overline{\Delta(k)_R}\right) * 100\% = 97.03\% > 90\%,\\
&\Delta(k)_T \le 0.05,\\
&\overline{\Delta(k)_T} = 0.0155 < 0.05,\\
&\quad \rho_T^0 = \left(1 - \overline{\Delta(k)_T}\right) * 100\% = 98.45\% > 90\%,\\
&\Delta(k)_W \le 0.05,\\
&\overline{\Delta(k)_W} = 0.02548 < 0.05,\\
&\quad \rho_W^0 = \left(1 - \overline{\Delta(k)_W}\right) * 100\% = 97.96\% > 90\%,\\
&\Delta(k)_A \le 0.05,\\
&\overline{\Delta(k)_A} = 0.0359 < 0.05,\\
&\quad \rho_A^0 = \left(1 - \overline{\Delta(k)_A}\right) * 100\% = 96.41\% > 90\%,\\
&\Delta(k)_P \le 0.05,\\
&\overline{\Delta(k)_P} = 0.0441 < 0.05,\\
&\quad \rho_P^0 = \left(1 - \overline{\Delta(k)_C}\right) * 100\% = 95.59\% > 90\%.
\end{aligned} \tag{36}$$

To sum up, GM (1, 1) has a good accuracy in modeling and forecasting the freight development index of Guangdong Province and can pass the residual test.

4.4.2. Posterior Error Test. In order to further test the accuracy of GM (1,1) in forecasting the development of freight volume in Guangdong Province, a posteriori test is carried out. According to formulas (15)–(18), the original sequence's variance of X_F^0, X_R^0, X_T^0, X_W^0, X_A^0, and X_P^0 is calculated, and the results are as follows:

$$\begin{aligned}
S_{1F} &= 456.316,\\
S_{1R} &= 18.535,\\
S_{1T} &= 329.434,\\
S_{1W} &= 143.907,\\
S_{1A} &= 0.365,\\
S_{1P} &= 9.902.
\end{aligned} \tag{37}$$

Table 4: Actual value and forecast value of the freight development index in Guangdong Province (unit: million tons).

Year	Water freight volume (W)		Air cargo volume (A)		Pipeline freight volume (P)	
	X_W^0	$\widehat{X}_W$	X_A^0	$\widehat{X}_A$	X_P^0	$\widehat{X}_P$
2013	688.51	688.51	2.12	2.12	76.81	76.81
2014	781.25	744.06	2.24	2.11	81.54	75.14
2015	780.93	802.20	2.28	2.28	76.93	79.29
2016	828.93	864.88	2.40	2.47	79.32	83.68
2017	939.04	932.46	2.45	2.67	82.09	88.31
2018	1023.51	1005.32	3.01	2.89	99.41	93.20
2019	1083.71	1083.88	3.017	3.12	98.99	98.35

Source: according to the calculation criteria.

Table 5: Residual test results of GM (1,1) of freight volume in Guangdong Province.

Year	Total freight volume (F)		Railway freight volume (R)		Truck freight volume (T)	
	$\varepsilon(k)_F$	$\Delta(k)_F$	$\varepsilon(k)_R$	$\Delta(k)_R$	$\varepsilon(k)_T$	$\Delta(k)_T$
2013	0.000	0.000	0.000	0.000	0.120	0.000
2014	−50.980	0.014	−0.050	0.000	−3.820	0.001
2015	−74.560	0.020	−0.150	0.001	−98.990	0.035
2016	137.150	0.037	−10.190	0.055	103.640	0.038
2017	66.070	0.017	12.830	0.082	49.690	0.017
2018	−37.430	0.009	3.730	0.024	−18.400	0.006
2019	−46.030	0.010	−7.420	0.046	−34.880	0.011

Source: according to the calculation criteria.

Table 6: Residual test results of GM (1,1) of freight volume in Guangdong Province.

Year	Water freight volume (W)		Air cargo volume (A)		Pipeline freight volume (P)	
	$\varepsilon(k)_W$	$\Delta(k)_W$	$\varepsilon(k)_A$	$\Delta(k)_A$	$\varepsilon(k)_P$	$\Delta(k)_P$
2013	0.000	0.000	0.000	0.000	0.000	0.000
2014	−37.190	0.048	−0.130	0.058	−0.078	0.078
2015	21.270	0.027	0.000	0.000	0.031	0.031
2016	35.950	0.043	0.070	0.029	0.055	0.055
2017	−6.580	0.007	0.220	0.090	0.076	0.076
2018	−18.190	0.018	−0.120	0.040	−0.062	0.062
2019	0.170	0.000	0.103	0.034	−0.006	0.006

Source: according to the calculation criteria.

The absolute residual sequence's variance of $\varepsilon(k)_F$, $\varepsilon(k)_R$, $\varepsilon(k)_T$, $\varepsilon(k)_W$, $\varepsilon(k)_A$, and $\varepsilon(k)_p$ is calculated, and the results are as follows:

$$\begin{aligned} S_{2F} &= 76.211,\\ S_{2R} &= 7.497,\\ S_{2T} &= 64.003,\\ S_{2W} &= 24.194,\\ S_{2A} &= 0.124,\\ S_{2P} &= 4.885. \end{aligned} \tag{38}$$

According to formula (19), calculate the ratio of variance of the freight index. The results are as follows:

$$\begin{aligned} C_F &= 0.167,\\ C_R &= 0.404,\\ C_T &= 0.194,\\ C_W &= 0.168,\\ C_A &= 0.365,\\ C_P &= 0.493. \end{aligned} \tag{39}$$

C_P is less than 0.5, so the posterior error test is barely qualified. C_R and C_A are greater than 0.35 and less than 0.45, respectively, and the posterior error test result is qualified. C_F, C_T, and C_W are less than 0.35, and hence, the posterior error test result is good.

Calculate the probability of small error according to formulas (20) and (21). The results are as follows:

$$\begin{aligned} S_{0F} &= 307.785,\\ S_{0R} &= 12.502,\\ S_{0T} &= 222.203,\\ S_{0W} &= 97.065,\\ S_{0A} &= 0.246,\\ S_{0P} &= 6.679,\\ \left|\varepsilon(k)_F - \bar{\varepsilon}_F\right| &= (0.826, 50.154, 73.734, 137.976, 66.896, 36.604, 45.204), \quad P = 1 > 0.95,\\ \left|\varepsilon(k)_R - \bar{\varepsilon}_R\right| &= (0.179, 0.129, 0.029, 10.011, 13.009, 3.909, 7.241), \quad P = 85.7 > 0.8,\\ \left|\varepsilon(k)_T - \bar{\varepsilon}_T\right| &= (0.497, 3.443, 98.613, 104.017, 50.067, 18.023, 34.503), \quad P = 1 > 0.95,\\ \left|\varepsilon(k)_W - \bar{\varepsilon}_W\right| &= (0.653, 36.537, 21.923, 36.603, 5.927, 17.537, 0.823), \quad P = 1 > 0.95\\ \left|\varepsilon(k)_A - \bar{\varepsilon}_A\right| &= (0.020, 0.150, 0.020, 0.050, 0.200, 0.140, 0.083), \quad P = 1 > 0.95,\\ \left|\varepsilon(k)_P - \bar{\varepsilon}_P\right| &= (0.044, 6.356, 2.404, 4.404, 6.264, 6.166, 0.596), \quad P = 1 > 0.95. \end{aligned} \tag{40}$$

To sum up, the six indicators of freight development in Guangdong Province, total freight volume, railway freight volume, truck freight volume, water freight volume, air cargo volume, and pipeline freight volume, all meet the requirements of posterior error test standard $C < 0.5$, $0.8 < P$, which indicate that the accuracy of the model is qualified.

5. Impact of COVID-19 on Freight Development of Guangdong Province

The results of the rationality test of model construction, residual test, and posteriori test proved that GM (1, 1) is feasible to forecast the freight development of Guangdong Province in 2020. Due to the particularity of railway transportation and air transportation, after eliminating the influence of C value, the prediction results are shown in Table 7.

In order to reflect COVID-19's monthly impact on Guangdong Province's freight development, further explore the monthly forecast value of freight volume in Guangdong Province of 2020:

(1) Calculate the ratio of each month in different years to the freight volume of that year.

Record the freight development data of Guangdong Province from 2013 to 2019 as matrix of $A_{F7\times12}$,

Table 7: Forecast freight volume value of Guangdong Province in 2020 (unit: million ton).

Year	Total freight volume	Railway freight volume	Truck, freight volume	Water freight volume	Air cargo volume	Pipeline freight volume
2020	4607.222	67.927	3274.392	1168.572	2.475	103.797

Source: according to the calculation criteria.

$A_{R7\times12}$, $A_{T7\times12}$, $A_{W7\times12}$, $A_{A7\times12}$, and$A_{P7\times12}$, and calculate the ratio of monthly freight volume (a_{kj}) to the total freight volume of the year ($x^{(0)}(k)$) and record it as R_{kj}, where R is ratio, k is year, and j is month:

$$R_{kj} = \frac{a_{kj}}{x^{(0)}(k)}, \quad k = 1, 2, 3, \ldots, 7, j = 1, 2, 3, \ldots, 12. \tag{41}$$

(2) Calculate the average freight ratio of the same month in different years:

$$\overline{R_j} = \frac{1}{7}\sum_{k=1}^{7} r_{kj}, \quad k = 1, 2, 3, \ldots, 7, j = 1, 2, 3, \ldots, 12. \tag{42}$$

According to formulas (35)–(41), we can calculate$\overline{R_{jF}}$, $\overline{R_{jR}}$, $\overline{R_{jT}}$, $\overline{R_{jW}}$, $\overline{R_{jA}}$, and $\overline{R_{jP}}$ from January to December, as shown in Table 8.

(3) Calculate monthly forecast value as follows:

$$\widehat{a}_{kj} = \overline{R_j}\widehat{x}^{(0)}k \quad k = 1, 2, 3, \ldots, 9, j = 1, 2, 3, \ldots, 12. \tag{43}$$

(4) Calculate the influence value I as follows:

$$I_{kj} = \frac{\left(a_{kj} - \widehat{a}_{ij}\right)}{\widehat{a}_{ij}} \times 100\%, \tag{44}$$

$$I_k = \frac{\left(x^{(0)}(k) - \widehat{x}^{(0)}k\right)}{\widehat{x}^{(0)}k} \times 100\%. \tag{45}$$

According to formulas (41)–(45), the monthly forecast value and impact of freight development in Guangdong Province from January to December 2020 are calculated, as shown in Tables 9 and 10 and Figure 1.

According to Tables 9 and 10 and Figure 1, reflecting the influence of COVID-19, there are three characteristics of freight transport development of Guangdong Province in 2020, which are stage characteristics, structural characteristics, and entity conduction.

5.1. Stage Characteristics. The first stage is the direct impact stage. Freight transport in Guangdong Province is stagnant during January and February. Affected by the epidemic situation, most logistics enterprises shut down except emergency logistics, and some logistics transportation channels were interrupted. The risk mainly comes from the disconnection of logistics supply and demand and the interruption of logistics channel. COVID-19 had the greatest impact on freight development in Guangdong at that time. The negative impact value of freight volume of various forms of transportation reached the peak in February. The impact value of total freight volume is −53.726%, that of truck freight volume is −65.155%, that of water freight volume is −31.232%, that of air cargo volume is −32.011%, and that of pipeline freight volume is −19.533%. However, the impact of railway is positive, and the stagnation of other transportation promotes the actual value of railway freight volume to be higher than the predicted value, and the impact value is 15.829%.

The second stage is the stage of entity conduction. The impact value of each subindex of Guangdong freight volume from March to May is reduced. This stage mainly focuses on the period from March to May 2020. The main impact of this stage is the uneven structure of the type and quantity of freight demand business in Guangdong Province and the short-term mismatch between supply and demand. The development of Guangdong's port freight industry has changed from explosive logistics demand in the early stage of the full resumption of manufacturing and circulation industries to the gradual emergence of real economic difficulties and the decline of logistics demand with the spread of the global epidemic.

The third stage is the policy digestion stage. This stage is mainly after May 2020; the state has issued relevant policies to support the recovery and development of the real economy and logistics industry, such as providing freeway and other measures to promote further recovery of freight. The role of policy has a certain lag and conductivity. The development of freight transportation in Guangdong Province is affected by the favorable policies and the recovery of logistics demand of the upstream manufacturing industry, which can be shown in turn in the postepidemic period. COVID-19's impact on freight development in Guangdong showed a gradual reduction. The risk points in this period mainly focus on the recovery of logistics industry, the scale of logistics investment, and the choice of new development mode.

5.2. Structural Characteristics. The impact of the epidemic on the development of freight transportation in Guangdong Province has structural characteristics. In the same period, the impact of the epidemic on the internal subbusiness sectors of freight transportation in Guangdong Province is different. The overall impact value of the epidemic on freight transportation was −23.001%, and the most affected mode was truck transportation (−29.344%), followed by waterway transportation (−11.296%) and air transportation

Table 8: Average ratio of monthly freight (unit: %).

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
$\overline{R_{jF}}$	7.378	6.364	7.560	8.032	8.251	8.357	8.529	8.571	8.794	9.095	9.304	9.766
$\overline{R_{jR}}$	8.379	6.965	8.429	8.073	8.596	8.089	8.367	8.299	8.358	8.510	8.709	9.225
$\overline{R_{jT}}$	7.117	6.221	7.484	7.989	8.229	8.341	8.594	8.634	8.837	9.220	9.369	9.965
$\overline{R_{jW}}$	7.986	6.598	7.576	8.109	8.282	8.458	8.389	8.375	8.722	8.862	9.271	9.372
$\overline{R_{jA}}$	7.715	6.537	7.762	8.051	8.340	8.311	8.470	8.470	8.678	8.922	9.163	9.582
$\overline{R_{jP}}$	8.055	7.280	8.157	8.236	8.262	8.312	8.359	8.765	8.609	8.567	8.677	8.720

Source: according to the calculation criteria.

Table 9: COVID-19's impact of freight development in Guangdong Province (unit, million ton; %).

Month	Total freight volume			Railway freight volume			Truck freight volume		
	Actual value	Predictive value	Impact degree	Actual value	Predictive value	Impact degree	Actual value	Predictive value	Impact degree
Jan	250.750	339.903	−26.229	5.540	5.692	−2.668	158.130	233.029	−32.142
Feb	135.680	293.208	−53.726	5.480	4.731	15.829	70.980	203.705	−65.155
Mar	223.620	348.316	−35.800	6.270	5.725	9.511	132.740	245.071	−45.836
Apr	267.320	370.044	−27.760	5.980	5.484	9.048	170.170	261.582	−34.946
May	298.390	380.134	−21.504	6.910	5.839	18.336	191.570	269.464	−28.907
Jun	308.350	385.032	−19.916	6.780	5.495	23.387	204.060	273.113	−25.284
Jul	314.180	392.930	−20.042	7.260	5.683	27.744	211.690	281.402	−24.773
Aug	329.530	394.878	−16.549	6.650	5.637	17.964	215.540	282.716	−23.761
Sep	332.890	405.139	−17.833	6.960	5.678	22.586	220.050	289.356	−23.952
Oct	349.210	419.023	−16.661	6.360	5.780	10.029	237.200	301.893	−21.429
Nov	364.490	428.661	−14.970	6.990	5.916	18.163	245.020	306.762	−20.127
Dec	373.100	449.955	−17.081	6.490	6.266	3.568	256.410	326.298	−21.418
Total impact	4607.222	3547.510	−23.001	77.67	67.927	14.343	2313.560	3274.392	−29.344

Source: according to the calculation criteria.

Table 10: COVID-19's impact of freight development in Guangdong Province (unit: million ton; %).

Month	Water freight volume			Air cargo volume			Pipeline freight volume		
	Actual value	Predictive value	Impact degree	Actual value	Predictive value	Impact degree	Actual value	Predictive value	Impact degree
Jan	78.540	93.318	−15.836	0.190	0.191	−0.493	8.340	8.361	−0.248
Feb	53.020	77.100	−31.232	0.110	0.162	−32.011	6.080	7.556	−19.533
Mar	76.200	88.529	−13.926	0.180	0.192	−6.308	8.230	8.467	−2.798
Apr	82.360	94.757	−13.083	0.180	0.199	−9.662	8.640	8.549	1.065
May	91.330	96.785	−5.636	0.200	0.206	−3.105	8.390	8.575	−2.161
Jun	87.890	98.840	−11.079	0.210	0.206	2.087	9.410	8.628	9.068
Jul	86.630	98.033	−11.632	0.200	0.210	−4.591	8.410	8.677	−3.075
Aug	95.000	97.869	−2.932	0.200	0.210	−4.593	12.140	9.098	33.432
Sep	94.440	101.925	−7.344	0.240	0.215	11.745	11.200	8.936	25.332
Oct	93.680	103.560	−9.540	0.220	0.221	−0.367	11.750	8.893	32.131
Nov	99.080	108.343	−8.550	0.220	0.227	−2.994	13.180	9.007	46.332
Dec	98.400	109.513	−10.148	0.230	0.237	−3.017	11.580	9.051	27.948
Total impact	1036.570	1168.572	−11.296	2.380	2.475	−3.838	117.350	103.797	13.057

Source: according to the calculation criteria.

(−3.838%). At the beginning of the outbreak, the railway operation remained sustainable. Due to the suspension of road transport enterprises, a large number of goods were transferred to railway transportation. Pipeline transportation can be carried out without people. The epidemic has little impact on its production and operation. Therefore, the impact of COVID-19 on the railway transportation and pipeline transportation is positive. The influence values were 14.343 and 13.057, respectively. In summary, the development structure of freight transport in Guangdong has been characterized by structural development under the influence of COVID-19.

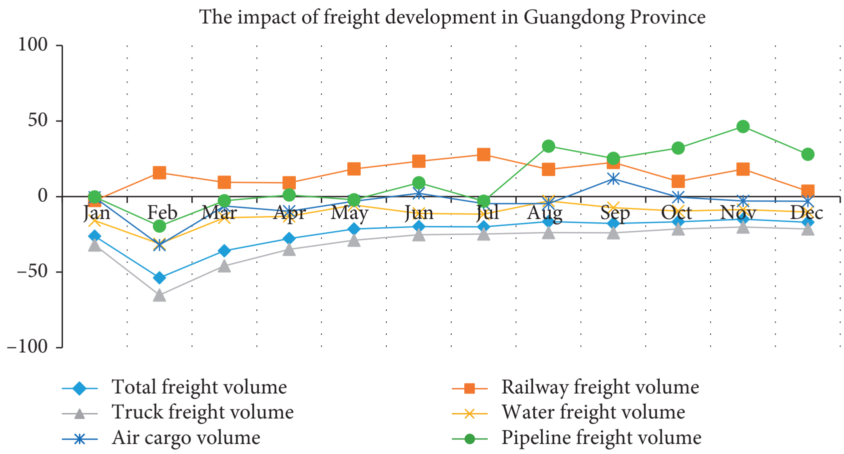


Figure 1: The impact of freight development in Guangdong Province.

5.3. Characteristics of Solid Conduction. Freight industry is a derivative industry. The demand of freight industry is mainly affected by the secondary industry and commodity circulation industry and is mainly affected by manufacturing industry, import and export trade, domestic trade, and manufacturing industry. As an auxiliary industry of the national economy, freight industry is an important support for the development of industrial economy. The crisis of the freight industry will be transmitted in reverse, resulting in damage to the real economy. During the epidemic period, due to traffic control, road restrictions, blocked logistics channels, staff unable to rework and resume production on time, and unsatisfied conditions for resumption of production, some freight enterprises were unable to provide normal freight service, which affected the resumption of some industrial enterprises and could not guarantee the logistics of raw materials and finished products.

6. Research on the Related Factors of Freight Development in Guangdong Province

The grey prediction model GM (1, 1) can scientifically measure the impact of COVID-19 on the development of freight transport in Guangdong. From the perspective of system theory, the development of freight transportation in Guangdong Province is the output of logistics system and the product of adapting to the economic environment and regional development of Guangdong Province. Therefore, in order to explore the countermeasures of the development of freight transportation in Guangdong Province in the post-epidemic period, we should also analyze the environmental factors. To sum up, this paper will use the grey correlation model to study the related factors of freight development in Guangdong Province.

6.1. Construction of the Index System. Scholars in the study of freight development factors divide them mainly into the regional economic factors, logistics environmental factors, and industrial factors for quantitative index selection; this paper, considering the scientificity, rationality, comprehensiveness, and comparability of the index construction, divides the related factors of the freight development of Guangdong Province into economic factors, transportation factors, science and technology factors, and policy factors, as shown in Table 11.

6.2. Calculation of Grey Correlation Degree of Influencing Factors. Based on the freight development of Guangdong Province from 2011 to 2019, total freight volume, railway freight volume, truck freight volume, water freight volume, air cargo volume, and pipeline freight volume are the main factor sequence data of the system. The historical data of $X_{11} - X_{46}$ index factors are regarded as subfactor series data, and the grey correlation resolution coefficient is 0.5 based on experience. The data involved are from the public data of the Guangdong Statistical Yearbook in 2013–2019. The grey correlation degree of influencing factors of freight development in Guangdong Province is calculated as given in formulas (24)–(26) and $\mathrm{AVG}\gamma_{0i} = \mathrm{avg}(\gamma_{0iF} + \gamma_{0iR} + \gamma_{0iT} + \gamma_{0iW} + \gamma_{0iA} + \gamma_{0iP})$, and the calculation results are shown in Tables 12–14.

The closer the value to 1, the higher the correlation between the self-factor sequence data and the main factor sequence data. According to the calculation results of correlation degree of Tables 12–14, $\mathrm{AVG}R_{042} > \mathrm{AVG}R_{044} > \mathrm{AVG}R_{014} > \mathrm{AVG}R_{041} = \mathrm{AVG}R_{046} > \mathrm{AVG}R_{023} > \mathrm{AVG}R_{011} > \mathrm{AVG}R_{045} > \mathrm{AVG}R_{015} > \mathrm{AVG}R_{024} > \mathrm{AVG}R_{021} > \mathrm{AVG}R_{031} > \mathrm{AVG}R_{043} > \mathrm{AVG}R_{043} > \mathrm{AVG}R_{016} > \mathrm{AVG}R_{017} > \mathrm{AVG}R_{012} = \mathrm{AVG}R_{022} > \mathrm{AVG}R_{032} > \mathrm{AVG}R_{013} > \mathrm{AVG}R_{025}$; there are 15 factors with $\mathrm{AVG}R_{0i}$; there are 15 factors with $\mathrm{AVG}R_{0i}$ greater than or equal to 0.8, indicating that the vast majority of factors are highly related to the development of freight transportation in Guangdong Province, including operating mileage of railway, tonnage of trucks, disposable income, deposit balance of financial institutions, number of employees in transportation, warehousing and postal industry, transport aircraft, R&D personnel, GDP, mileage of oil pipeline, total retail sales of consumer goods, financial expenditure on education, Internet broadband access users, general budget revenue, tonnage of cargo ships, and total

Table 11: Index system of influencing factors of freight development in Guangdong Province.

First level indicators	Second level indicators	Variable	Characteristic
Regional economy factors	Gross domestic product	X_{11}	Positive correlation
	Proportion of added value of secondary industry	X_{12}	Positive correlation
	Total investment in fixed assets	X_{13}	Positive correlation
	Per capita disposable personal income	X_{14}	Positive correlation
	Total retail sales of social consumption	X_{15}	Positive correlation
	Total imports and exports	X_{16}	Positive correlation
	Balance of domestic and foreign currency deposits of financial institutions	X_{17}	Positive correlation
Science and technology factors	Number of Internet broadband access users	X_{21}	Positive correlation
	Internal expenditure of R&D funds	X_{22}	Positive correlation
	Number of employees of R&D	X_{23}	Positive correlation
	Education expenditure	X_{24}	Positive correlation
	Expenditure on science and technology	X_{25}	Positive correlation
Government policy factors	Revenue	X_{31}	Positive correlation
	Transportation expenditure	X_{32}	Positive correlation
Transportation environment factors	Number of employees in transportation, storage, and post industry	X_{41}	Positive correlation
	Length of railroad lines in service	X_{42}	Positive correlation
	Net tonnage of carrying vessel	X_{43}	Positive correlation
	Tonnage of truck	X_{44}	Positive correlation
	Oil pipeline mileage	X_{45}	Positive correlation
	Transport plane	X_{46}	Positive correlation

Table 12: Grey correlation results of freight development in Guangdong Province (1).

Variable	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}
γ_{0iF}	0.896	0.762	0.807	0.925	0.900	0.774	0.818
γ_{0iR}	0.747	0.916	0.689	0.806	0.737	0.903	0.708
γ_{0iT}	0.894	0.747	0.816	0.904	0.897	0.758	0.829
γ_{0iW}	0.930	0.752	0.798	0.903	0.917	0.761	0.824
γ_{0iW}	0.921	0.726	0.798	0.871	0.884	0.733	0.848
γ_{0iA}	0.838	0.866	0.740	0.924	0.823	0.879	0.773
AVGγ_{0i}	0.871	0.795	0.775	0.889	0.860	0.801	0.800

Source: according to the calculation criteria.

Table 13: Grey correlation results of freight development in Guangdong Province (2).

Variable	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{31}	X_{32}
γ_{0iF}	0.850	0.823	0.890	0.870	0.586	0.848	0.816
γ_{0iR}	0.749	0.706	0.779	0.744	0.599	0.712	0.806
γ_{0iT}	0.848	0.819	0.882	0.867	0.590	0.859	0.824
γ_{0iW}	0.885	0.825	0.933	0.899	0.576	0.858	0.770
γ_{0iW}	0.917	0.831	0.912	0.913	0.558	0.877	0.743
γ_{0iA}	0.831	0.764	0.881	0.824	0.579	0.781	0.785
AVGγ_{0i}	0.847	0.795	0.880	0.853	0.582	0.823	0.791

Source: according to the calculation criteria.

import and export, which are most closely related to the development of freight transportation in Guangdong Province, ranking in the top 15 with the correlation degree greater than 0.8. The relevance degree of the added value of the secondary industry in the proportion of GDP is greater than 0.7, and its closeness with the freight development of Guangdong Province cannot be ignored. The relevance degree of science and technology financial

Table 14: Grey correlation results of freight development in Guangdong Province (3).

Variable	X_{41}	X_{42}	X_{43}	X_{44}	X_{45}	X_{46}
γ_{0iF}	0.916	0.947	0.808	0.915	0.908	0.902
γ_{0iR}	0.807	0.785	0.878	0.790	0.790	0.777
γ_{0iT}	0.894	0.923	0.789	0.901	0.903	0.887
γ_{0iW}	0.897	0.919	0.803	0.930	0.894	0.935
γ_{0iW}	0.864	0.876	0.767	0.906	0.829	0.942
γ_{0iA}	0.950	0.922	0.870	0.911	0.885	0.883
AVGγ_{0i}	0.888	0.895	0.819	0.892	0.868	0.888

Source: according to the calculation criteria.

expenditure is only 0.582, and its relevance with the freight development of Guangdong Province can be ignored.

6.3. Analysis of Grey Correlation Results. By using the grey correlation analysis method, we can objectively and reasonably reflect the influence degree of relevant influencing factors on the freight development of Guangdong Province and provide a quantifiable research perspective for the analysis of the relationship between regional economic factors, scientific and technological environment factors, government policy factors, transportation environment factors, and the freight development of Guangdong Province:

(1) The regional economic environment plays an important role in the freight development Guangdong Province.

The average value of the six factors of regional economy (AVG R_{01i}) is 0.827, and the value of four factors is greater than 0.8, which reflects that the economic environment is closely related to the freight development of Guangdong Province. The regional economic environment represents the demand of the freight market in the port market and reflects the impact of the economic environment on the demand of the port freight market; it also reflects the industrial transmission effect of the logistics industry affected by the real economy.

(2) The development of freight transportation in Guangdong Province should pay attention to the environmental factors of science and technology.

The value of science and technology environment factor (AVG R_{02i}) is 0.791. There is a certain degree of correlation between R&D personnel and the level of regional science and technology in Guangdong Province and the development of freight transportation in Guangdong Province, which means that the promotion and development of information technology have greatly promoted the development of freight transportation in Guangdong Province. R&D practitioners reflect the information technology talents, education expenditure reflects the strength of regional talent cultivation, and the amount of Internet broadband access reflects the basic environment of science and technology. The above three factors are greater than 0.8, which shows that the development of freight transportation in Guangdong Province is constrained by the basic conditions of the development of information technology and the cultivation of information talents. However, the government's financial expenditure on science and technology has no direct correlation with the freight development of Guangdong Province, and its correlation coefficient is only 0.582.

(3) Government policy factors guide freight development in Guangdong Province.

The value of government policy factor (AVG R_{03i}) is 0.807. The government's financial capacity and the intensity of transportation expenditure are closely related to the development of freight transportation in Guangdong Province. The amount of government's financial revenue and the intensity of transportation expenditure are helpful to improve the conditions of transportation infrastructure in Guangdong Province.

(4) The transportation environment factors directly affect the development of freight in Guangdong Province.

The value of environmental factors of transportation (AVG R_{04i}) is 0.875. The development of freight transportation in Guangdong Province is closely related to environmental factors of transportation. The correlation degree between the environmental factors of transportation and the freight development of Guangdong Province is greater than 0.8. The operating mileage of railway and oil pipeline reflects the overall infrastructure of freight industry. The number of employees in transportation, warehousing, and postal industry reflects the human resource factors of logistics industry. The number of trucks, ships, and aircraft reflects the basic equipment conditions of logistics. The above indicators are closely related to the freight development of Guangdong Province.

7. Conclusions and Suggestions

This paper makes a quantitative study on the development trend of freight transport in Guangdong from the perspective of freight volume. However, the research needs to be further expanded to fully reflect the overall situation of freight development in Guangdong Province, such as the study of freight market price factors and the change of supply-demand relationship. The empirical study shows the COVID-19 has a certain impact on the freight development of Guangdong, which has the following characteristics: stage characteristics, structural characteristics, and physical transmission characteristics; the total freight volume of Guangdong is affected by −23.001%, and the truck transportation is affected by −29.344%, with −11.296% for water transportation and −3.838% for air transportation. Due to the continuity of transportation and the substitution for other transportation modes, the freight volume of railway and pipeline transportation is affected by 14.343% and 13.057%, respectively.

To further explore the abate measures of COVID-19 impact on Guangdong's freight development, the grey correlation model is introduced to study the correlation factors of freight development of Guangdong Province. This paper selects four factors including regional economic environment, science and technology environment, government policy, and transportation and logistics environment, covering 20 subfactors. Through the research, it is found that the average correlation degree of the four factors is greater than 0.7 and the order of correlation degrees is as follows: transportation and logistics environment factors, government policy factors, and science and technology environment factors.

Through the research on the related factors of freight development of Guangdong Province, this paper puts forward the following measures to promote the freight development in Guangdong Province in the postepidemic era: first, pay attention to improving the freight infrastructure equipment in Guangdong Province, improve the efficiency of infrastructure equipment operation, and improve the connection ability of transportation modes, such as sea rail combined transportation, sea air combined transportation, and sea land combined transportation. Second, the government continues to invest in transportation infrastructure and introduces relevant measures, such as reducing and exempting tolls, operating taxes, and so on. Third, strengthen the linkage between freight development and economy, pay attention to the trend of economic development, and timely adjust the strategy of freight development and operation. Fourth, pay attention to strengthening the construction and application of information technology and intelligent technology in freight operation and improve the intelligent and digital degree of freight operation process.

Data Availability

The data used to support the findings of the study and related data are available from all the authors upon request.

Conflicts of Interest

The authors declare there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was financially supported by the 2021 School Level Scientific Research Project of Guangzhou Huashang College (Grant no.2021HSDS09); 2020 Project of Guangzhou Philosophy and Social Science Development Plan "Research on the Influence Mechanism of Guangzhou Residents' Physical Consumption and Experience Consumption on Happiness" (Grant no .2020GZGJ210); Construction of Guangdong Provincial Characteristic Specialty-Construction of Logistics Management Specialty in 2019 (Grant no. HS2019CXQX04); Humanities and Social Sciences Research Projects of Universities in Guangdong Province "Construction of Key Disciplines in Business Administration" (Grant no. 2015WTSCX126); "Key Discipline Construction of Enterprise Management" of Guangzhou Huashang University (Grant no. 2020HSXK05); School Level Quality Engineering "Supply Chain Management Offline First Class Course" of Guangzhou Huashang University (Grant no. HS2020ZLGC15); and "Key Discipline: International Business Construction and Development" (HS2019CXQX17).

References

[1] M. Li and S. Zhang, "Research on dynamic industrial linkage and impact effect of logistics industry," *World Scientific Research Journal*, vol. 5, no. 7, 2019.

[2] Y. A. Bolumole, D. J. Closs, and F. A. Rodammer, "The economic development role of regional logistics hubs: a cross-country study of interorganizational governance models," *Journal of Business Logistics*, vol. 36, no. 2, 2015.

[3] J. Zhang and L. Chen, "The industrial relations of logistics industry-based on China's 2010 input-output table," *Modern Economy*, vol. 5, no. 12, 2014.

[4] Y. Su, Q. Jin, P. Yang, Q. Jiang, and G.-W. Weber, "A supply chain-logistics super-network equilibrium model for urban logistics facility network optimization," *Mathematical Problems in Engineering*, vol. 2019, Article ID 5375282, 12 pages, 2019.

[5] S. Tang, S. Xu, and J. Gao, "An optimal model based on multifactors for container throughput forecasting," *KSCE Journal of Civil Engineering*, vol. 23, no. 9, 2019.

[6] X. Wang, Y. A. N. Ying, J. Feng, and J. Xiang, "Research on the demand forecasting method of sichuan social logistics based on positive weight combination," *Canadian Social Science*, vol. 14, no. 6, 2018.

[7] T. Liu, S. Li, and S. Wei, "Forecast and opportunity analysis of cold chain logistics demand of fresh agricultural products under the integration of beijing, tianjin and hebei," *Open Journal of Social Sciences*, vol. 5, no. 10, 2017.

[8] D. Knoll, M. Prüglmeier, and G. Reinhart, "Predicting future inbound logistics processes using machine learning," *Procedia CIRP*, vol. 52, pp. 145–150, 2016.

[9] D. Yang, "Logistics demand forecast model for port import and export in coastal area," *Journal of Coastal Research*, vol. 103, no. sp1, 2020.

[10] J. Bruzda, "Multistep quintile forecasts for supply chain and logistics operations: bootstrapping, the GARCH model and quintile regression based approaches," *Central European Journal of Operations Research*, vol. 28, no. 1, 2020.

[11] X. Li and Li Guo, "Research on the driving force of the regional economy to the development of ocean port shipping based on multiple regression analysis," *Journal of Coastal Research*, vol. 111, no. 1, 2020.

[12] H. K. Kim and W. Chang, "Development of a cost forecasting model for air cargo service delay due to low visibility," *Sustainability*, vol. 11, no. 16, 2019.

[13] J. Farhan and G. P. Ong, "Forecasting seasonal container throughput at international ports using SARIMA models," *Maritime Economics & Logistics*, vol. 20, no. 1, 2018.

[14] Y. Xiao, S. Wang, M. Xiao, J. Xiao, and Y. Hu, "The analysis for the cargo volume with hybrid discrete wavelet modeling,"

International Journal of Information Technology & Decision Making, vol. 16, no. 3, 2017.

[15] B.-W. Ko and D.-J. Kim, "Analysis of container shipping market using multivariate time series models," *Journal of Korea Port Economic Association*, vol. 35, no. 3, pp. 61–72, 2019.

[16] B. Hong, R. Song, "Forecast of railway freight ton-kilometers based on the UBGPM-markov model," *Advanced Materials Research*, vol. 3470, pp. 2199–2202, 2014.

[17] W. Li, J. Dai, Y. Xiao, S. Yang, and C. Song, "Estimating waterway freight demand at Three Gorges ship lock on Yangtze River by back propagation neural network modeling," *Maritime Economics & Logistics*, vol. 2020, pp. 1–27, 2020.

[18] M. I. Asborno and S. Hernandez, "Using data from a state travel demand model to develop a multi-criteria framework for transload facility location planning," *Transportation Research Record*, vol. 2672, no. 9, pp. 12–23, 2018.

[19] S. Tadić, M. Krstić, V. Roso, and N. Brnjac, "Planning an intermodal terminal for the sustainable transport networks," *Sustainability*, vol. 11, no. 15, 2019.

[20] Z. Zhang, "Research on the carbon emission diving factors and forecasts of logistics industry in the bohai rim economic zone based on the theory of grey system," *Journal of Management & Decision Sciences*, vol. 3, no. 1, 2020.

[21] H. Zhang, J. Zhu, and L. Zhou, "Research on logistics demand forecasting and transportation structure of beijing based on grey prediction model," *Science Journal of Applied Mathematics and Statistics*, vol. 3, no. 3, 2015.

[22] E. Suryani and S.-Y. Chou, "Chih-Hsien Chen Dynamic simulation model of air cargo demand forecast and terminal capacity planning," *Simulation Modelling Practice and Theory*, vol. 28, 2012.

[23] Y. Huang, "Development prediction of logistics industry in Henan province and its dynamic analysis," *Grey Systems: Theory and Application*, vol. 4, no. 2, 2014.

[24] C.-N. Wang, H.-K. Nguyen, R.-Y. Liao, and E. J. Macías, "Partner selection in supply chain of vietnam's textile and apparel industry: the application of a hybrid DEA and GM (1, 1) approach," *Mathematical Problems in Engineering*, vol. 2017, Article ID 7826840, 16 pages, 2017.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9844357, 8 pages
https://doi.org/10.1155/2021/9844357

Research Article

A Clustering Application Scenario Based on an Improved Self-Organizing Feature Mapping Network System

Qian Cao

College of Information Engineering, Chaohu College, Chaohu 238000, China

Correspondence should be addressed to Qian Cao; 19875069@qq.com

Received 10 May 2021; Revised 14 June 2021; Accepted 9 July 2021; Published 16 July 2021

Academic Editor: Weitao Pan

Categorizing national football teams by level is challenging because there is no standard of reference. Therefore, the self-organizing feature mapping network is used to solve this problem. In this paper, appropriate sample data were collected and an appropriate self-organizing feature mapping network model was built. After training, we obtained the classification results of 4 grades of 16 major Asian football national teams. As for the classification results, it is different to normalize the input data and not to normalize the input data. The classification results accord with our subjective cognition, which indicates the rationality of self-organizing feature mapping network in solving the classification problem of national football teams. In addition, the paper makes a detailed analysis of the classification results of the Chinese team and compares the gap between the Chinese team and the top Asian teams. It also analyses the impact of the normalization of input data on the classification results, taking Saudi Arabia as an example.

1. Introduction

As the biggest sport in the world, football is widely popular in the world, and all the national football teams in the world are fighting for the honor of their country in the international arena, so the level of the national team is widely concerned. However, there has always been disagreement about which level a football team should belong to. In order to objectively and fairly reflect the actual ranking of a football team, the neural network is used to analyse the level of each football team [1–4].

The main contributions of this research are mainly reflected in the following three aspects:

(i) We build a model based on improved self-organizing feature mapping network with the aim to cluster teams more reasonably. To make it clear, we give the specific model parameters and build process, and it includes the collection of data sets, the division of data sets, the normalization and inverse normalization of data, the selection of model parameters, the determination of training functions, the determination of error tolerance, the determination of the number of iterations, the completion of test experiments, and the analysis of results.

(ii) In order to reflect the latest level situation of teams, we trained and tested the proposed model with the latest data set, namely, the last eight major international competitions.

(iii) We quantitatively analysed the performance of the model with a variety of mathematical tools and error analysis methods.

The rest of this paper is organized as follows. Section 1 reviews and summarizes the related work, on this basis, to clarify the significance of this study. In Section 2, the motivation of this research is expounded. Section 3 is preliminaries. In Section 4, the overall scheme of neural network modeling is proposed. In Section 5, an experiment is designed and carried out and the results of the classification of 16 football teams were obtained. Finally, Section 6 concludes this paper.

2. Motivation

The Chinese national football team carries the expectations of hundreds of millions of Chinese fans. However, the Chinese national football team has performed poorly in recent years. So, where does China's national soccer team rank in Asia? Some think it belongs to the Asian second-tier team, others think it belongs to the fourth-tier team. With the help of historical data and the self-organizing feature mapping network, we can make an objective judgment on the level of the Chinese team. We not only hope to accurately reflect the real level of the Chinese team but also hope to verify the rationality of the algorithm.

3. Preliminaries

Self-organizing feature mapping, namely, self-organizing feature mapping network (SOFM or SOM), was proposed by Finnish neural network expert Kohonen in 1981. The biological basis of SOM is as follows. (1) The biological basis of lateral inhibition is like the lateral inhibition between nerve cells which brings out the competition, a degree of excitement from which the strongest nerve cells have obvious inhibitory effect on the peripheral nerve cells, and the excitement in peripheral nerve cells decreases as a result; thus, the neural network is the "winner" of the competition and other nerve cells fail in the competition. (2) When the biological neural network receives specific spatial and temporal information from the outside world, the specific region of the neural network is excited, and similar external information is continuously mapped in the corresponding region. After training, the competing layer neurons of SOM are close to each other with similar functions and far from each other with different functions, which is very similar to the tissue structure of biological neural network.

Each input pattern of self-organizing feature mapping corresponds to a localized region on a two-dimensional grid, and the location and properties of the region vary with the different input patterns. Therefore, there must be a sufficient number of input patterns to ensure that all neurons in the grid are trained and that the self-organizing process converges correctly. An important feature of SOFM is its topological conformal property, that is, the resulting feature map described by the output weight vector can reflect the distribution of the input pattern.

The basic principle of SOFM is that when a certain type of mode is input, a node in the output layer wins by getting the maximum stimulus, and the nodes around the winning node are also stimulated by lateral action. At this time, the network performs a learning operation, and the connection weight vector of the winning node and the surrounding nodes is modified in the direction of the input mode. When the category of input pattern changes, the winning node on the two-dimensional plane also moves from the original node to other nodes. In this way, the network uses a large number of sample data to adjust its connection weight through self-organization, and finally, the network output layer feature graph can reflect the distribution of sample data [3–7].

The SOFM network is a two-layer network consisting of an input layer and an output layer. The output layer establishes the topology of the network to better simulate the phenomenon of lateral inhibition in biology [6]. Figure 1 shows a simple SOFM network in which the output layer is a two-dimensional topology. Of course, the output layer of the SOFM network can also be a higher-dimensional topology. In SOFM networks, input and output neurons are connected by weights, and neighboring output neurons are also connected by weights. The transfer function of the output neuron is usually a linear function, so the output of the network is a linear weighted sum of the input values, as shown in the following formula:

$$Y_j = f\left(\sum_i x_i w_{ij}\right), \tag{1}$$

where w_{ij} represents the weight value, x_i is the input value, and Y_j is the output value [8, 9].

4. Supposed Model

Actually, there are other methods of classification by the level that do exist. In general, classification by machine learning method needs to be defined in advance. However, this method does not apply to the classification of football teams because the classification is supervised learning, that is, the classification of certain football teams must be specified, and then, the other football teams must be evaluated on the basis of those football teams. That is, we have to have standards first, but because of the uncertainty of the football game, it is difficult to find some football teams in the football world as standards. Even the top teams lose sometimes. Even top teams sometimes lose games, and if this top team is used as the standard, the results will be inaccurate.

Therefore, we must consider using the unsupervised clustering method. Self-organizing feature mapping network, as a good unsupervised clustering method, is applied to our research. In this way, as long as the number of categories N that need to be classified is set, the algorithm will convert all samples to N according to the principle of similarity.

This study intends to classify 16 major Asian football national teams. In order to complete this classification, we need to set the number of categories. If the division is too detailed, many teams may be classified into a single category, which is of little significance. If the division is too thick, such as only two categories lose the meaning of classification, so it would be a reasonable choice to divide the 16 teams into four categories.

Since the category number is 4, the competition layer will be set to a 2×2 hexagon structure in the self-organizing feature mapping network.

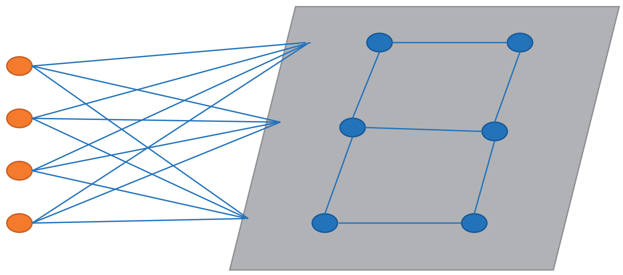


Figure 1: A SOFM network.

4.1. Sample Set Construction. Good classification experiment needs to be supported by good samples. There is no doubt that the results of the World Cup and the Asian cup can fully reflect the real level of a football team. Therefore, we selected the results of the 16 football teams in four international competitions, namely, the 2006 World Cup, the 2010 World Cup, the 2007 Asian cup, and the 2011 Asian cup, as samples. The results of the 16 football teams in the four international competitions are shown in Table 1. In the experiment, the sample data of each team can be represented by an eight-dimensional vector:

$$x = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8]. \quad (2)$$

The components of the vector represent its score in the World Cup 2006, the World Cup 2010, the Asian cup 2007, the Asian cup 2011, the World Cup 2014, the Asian cup 2015, the World Cup 2018, and the Asian cup 2019. Take the Chinese team as an example, the sample data of the Chinese team can be represented by the following 8-dimensional vector:

$$x = [43, 43, 9, 9, 43, 7, 33, 8]. \quad (3)$$

The scores of each team shown in Table 1 are calculated according to the following rules.

Firstly, for the Asian cup 2007 and Asian cup 2011, if a team reaches the final four, its final ranking is its score. If a team reaches the last eight, its score is 5. If a team reaches the last 16, its score is 9. If a team does not reach the final stage of the Asian cup, its score is 17. For the Asian cup 2015 and Asian cup 2019, we use the official final league table as a score for a team.

Secondly, for the World Cup 2006, World Cup 2010, and World Cup 2014, if a team makes it to the finals, its score is its actual ranking in the finals. For a football team that did not make it to the finals, there are two situations. If it enters the top 10 of the qualifiers, then we consider its score to be 33, and if it does not enter the top 10 of the qualifiers, then its score is 43. For the World Cup 2018, if a team makes it to the finals, its score is its actual ranking in the finals. For a football team that did not make it to the finals, there are two situations. If it enters the top 12 of the qualifiers, then we consider its score to be 33, and if it does not enter the top 12 of the qualifiers, then its score is 45.

4.2. Construction of Elman Neural Network. Once we have the data, we can design the experiment. The process of the experiment is first to build the SOFM model, then to train, and finally to test. See Figure 2, for details [10–14].

For Model Creation, the selforgmap function in the matlab neural network toolbox can be directly used to create. The size of the competition layer of the model can be set to 2×2. So, the matlab code to create the model is as follows:

$$\text{Net} = \text{selforgmap}([2 \times 2]). \quad (4)$$

Figure 3 shows the constructed Elman network structure.

5. Experiments

5.1. Experiments without Normalization. As for the test, the training data and the test data are the same, that is, the sample data. Therefore, 16 football teams can be classified by inputting the sample data into the model. Figure 4 is a matlab screenshot of the test results, and it clearly shows the categorization of the 16 teams. Among them, we find that the Chinese team belongs to the Asian third-tier team, indicating that the strength of the Chinese team is not satisfactory.

From this classification, we can see that the top teams in Asia are Japan, Korea, Iran, and Australia. In fact, this is in line with the actual situation, Japan, Korea, and Australia have reached the World Cup finals four times, and Iran has reached the World Cup finals three times.

In general, the selforgmap function tends to divide categories with more elements into finer categories so that categories with fewer elements may therefore merge with other categories so that each category tends to have the same number of elements. Saudi Arabia is the only second-tier team, and it shows that this level is not like other teams.

The Chinese team is classified as the 3rd tier Asian team. In fact, this is in line with the actual situation; after all, the Chinese team's performance in recent years is really very bad, and this is a well-known fact.

In order to see the level of the Chinese team more directly, we drew together the results of the Chinese team and four top Asian teams in the World Cup and compared them. As shown in Figure 5, there are 5 curves in the figure, representing 5 teams. The blue curve at the top represents the Chinese team. We can clearly see that there is a clear gap between the Chinese team and the other 4 teams, while the other 4 curves are intertwined, indicating that the level of the 4 first-class teams is very close. This also fully demonstrates the accuracy of the model we built for this classification.

If the results of the Asian cup are included, the conclusion is still the same. As shown in Figure 6, the five curves in the figure represent the five teams. The blue curve representing the Chinese team is still high, while the other four curves are intertwined. It shows that the gap between the level of Chinese team and the first-class team in Asia is

Table 1: Sample data (the results of the 16 football teams in the four international competitions).

Football teams	World Cup 2006	World Cup 2010	Asian Cup 2007	Asian Cup 2011	World Cup 2014	Asian Cup 2015	World Cup 2018	Asian Cup 2019
China	43	43	9	9	43	7	33	8
Japan	28	9	4	1	29	5	45	2
Korea	17	15	3	3	27	2	19	5
Iran	25	33	5	5	28	6	18	3
Saudi Arabia	28	33	2	9	43	10	26	14
Iraq	43	43	1	5	33	4	45	14
Qatar	43	33	9	5	33	13	33	1
United Arab Emirates	43	33	9	9	43	3	33	4
Uzbekistan	33	33	5	4	33	8	33	10
Thailand	43	43	9	17	43	17	33	13
Vietnam	43	43	5	17	43	17	45	5
Oman	43	43	9	17	33	12	45	16
Bahrain	33	33	9	9	43	11	45	12
North Korea	33	32	17	9	43	14	45	24
Indonesia	43	43	9	17	43	17	45	25
Australia	16	21	4	2	30	1	30	5

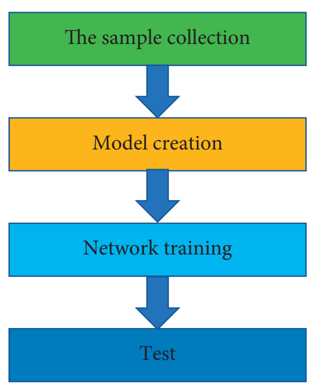


Figure 2: The process of the experiment.

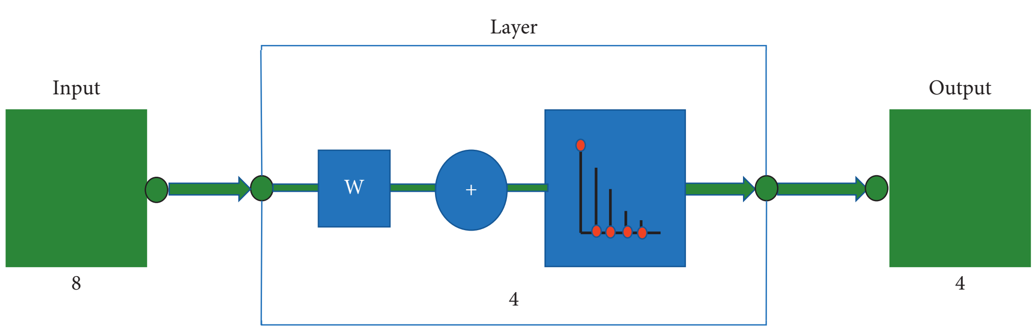


Figure 3: Constructed Elman network structure.

Football teams	Power level
China	3–tier
Japan	1–tier
Korea	1–tier
Iran	1–tier
Saudi Arabia	2–tier
Iran	3–tier
Qatar	3–tier
United Arab Emirates	3–tier
Uzbekistan	3–tier
Thailand	4–tier
Vietnam	4–tier
Oman	4–tier
Bahrain	4–tier
North Korea	4–tier
Indonesia	4–tier
Australia	1–tier

Figure 4: The test results (classification result).

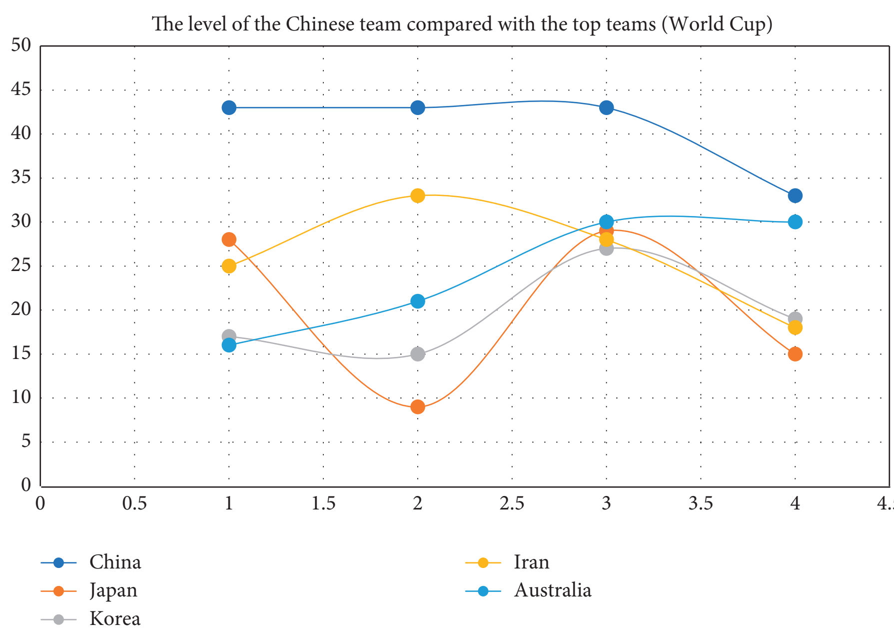


Figure 5: The level of the Chinese team compared with the top teams (World Cup).

relatively large, and it also shows that the classification effect of the model we established is accurate.

The number of fourth-tier teams is the largest, indicating that these teams have a poor track record, as can be seen from Table 1.

5.2. Experiments after Normalization. It needs to be emphasized here that the above results are obtained without normalization of input data. If the input data is normalized by mapminmax function, the results may change. Figure 7 shows the final experimental results obtained after normalization of the input data.

After comparing with the result of the last time, we find that Japan, South Korea, Iran, and Australia are still divided into the category of top Asian teams, which shows the super level of these four teams.

In this category of second-tier teams, Qatar, Iraq, and Uzbekistan have pushed Saudi Arabia, which ranked second in the previous category, into the category of third-tier teams in Asia, which is related to their excellent performances of these three teams in the last two Asian cups. It also shows the level of instability in Saudi Arabia.

For an in-depth analysis of the impact of normalization on the results, the differences between Saudi Arabia and four top teams before and after normalization were compared as examples. Table 2 shows the scores of Saudi Arabia and the four first-class teams before the normalization, and Table 3 shows the scores of Saudi Arabia and the four first-class teams after the normalization.

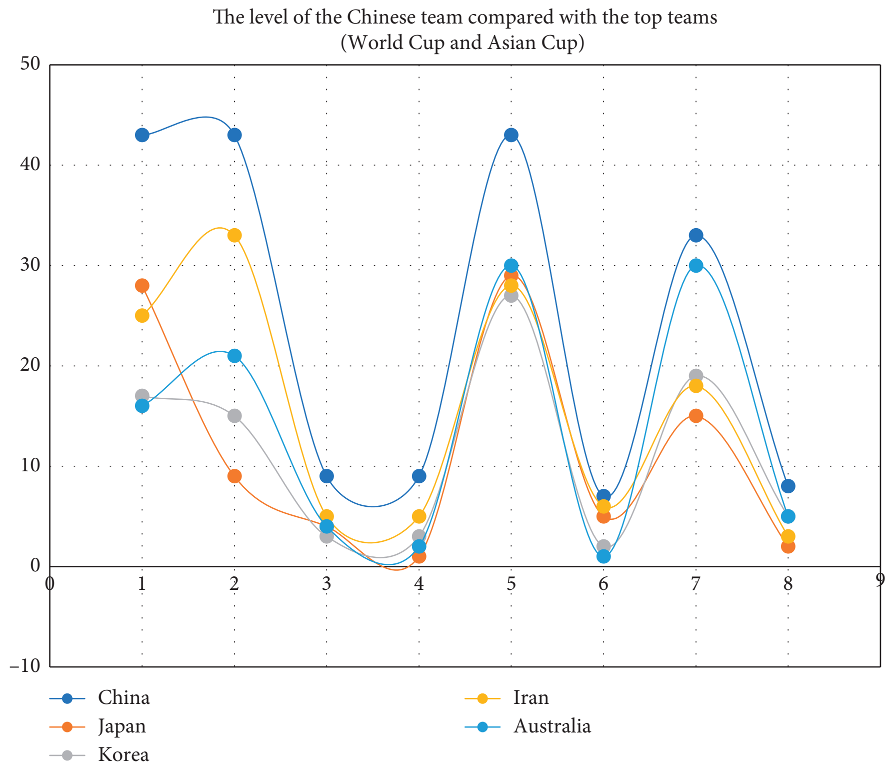


Figure 6: The level of the Chinese team compared with the top teams (World Cup and Asian Cup).

Football teams	Power level
China	3–tier
Japan	1–tier
Korea	1–tier
Iran	1–tier
Saudi Arabia	3–tier
Iran	2–tier
Qatar	2–tier
United Arab Emirates	3–tier
Uzbekistan	2–tier
Thailand	4–tier
Vietnam	4–tier
Oman	4–tier
Bahrain	3–tier
North Korea	3–tier
Indonesia	4–tier
Australia	1–tier

Figure 7: The final experimental results obtained after normalization of the input data.

Table 2: The scores of Saudi Arabia and the four first-class teams before the normalization.

Football teams	World Cup 2006	World Cup 2010	Asian Cup 2007	Asian Cup 2011	World Cup 2014	Asian Cup 2015	World Cup 2018	Asian Cup 2019
Saudi Arabia	28	33	2	9	43	10	26	14
Japan	28	9	4	1	29	5	15	2
Korea	17	15	3	3	27	2	19	5
Iran	25	33	5	5	28	6	18	3
Australia	16	21	4	2	30	1	30	5

Table 3: The scores of Saudi Arabia and the four first-class teams after the normalization.

Football teams	World Cup 2006	World Cup 2010	Asian Cup 2007	Asian Cup 2011	World Cup 2014	Asian Cup 2015	World Cup 2018	Asian Cup 2019
Saudi Arabia	−0.1111	0.4118	−0.875	0	1	0.125	−0.2667	0.0833
Japan	−0.1111	−1	−0.625	−1	−0.75	−0.5	−1	−0.9167
Korea	−0.9259	−0.6471	−0.75	−0.75	−1	−0.875	−0.7333	−0.6667
Iran	−0.3333	0.4118	−0.5	−0.5	−0.875	−0.375	−0.8	−0.8333
Australia	−1	−0.2941	−0.5	−0.875	−0.625	−1	0	−0.6667

To sum up, we can conclude that the relatively stable teams in the clustering of this study are as follows. Firstly, top teams in Asia are Japan, South Korea, Iran, and Australia. Secondly, Asian third-tier teams are China and United Arab Emirates. Thirdly, Asian fourth-tier teams are Thailand, Vietnam, Oman, and Indonesia.

6. Conclusion and Future Work

In order to classify the national football teams, this study took 16 major Asian national teams as samples and eight international competitions as sample features, built a self-organizing feature mapping network model, took matlab as the experimental platform, and finally achieved a reasonable classification result. In this paper, we also focus on the analysis of the situation of the Chinese team. In addition, we further compare and analyse the differences caused by the normalization of the input data. Of course, we know that our model is not perfect and that there may be improper classifications in some classification work; in the future, we will use the same approach to categorize other teams around the world and refine our model based on that. The results of this study are applicable to other scenarios, for example, the results can be used when we want to rank the income levels of a country's residents.

Data Availability

The underlying data supporting the results of this study can be found on the Internet.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the Talent Project of Anhui Province (Grant no. gxgnfx2019034), Anhui Province Key Research and Development Program project (Grant no. 201904a05020091) and its supporting project (Grant no. PT04a05020094), key project of the Natural Science Research of Higher Education Institutions in Anhui Province (Grant no. KJ2020A0680), and 2020 Quality Improvement Project of Chaohu College on Discipline Construction (Grant no. kj20xqyx03).

References

[1] P. Maji and S. Paul, *Neural Network Tree for Identification of Splice Junction and Protein Coding Region in DNA*, Springer International Publishing, New York, NY, USA, 2014.

[2] Y. Zhang, C. Yin, Q. Wu, Q. He, and H. Zhu, "Location-aware deep collaborative filtering for service recommendation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3796–3807, 2021.

[3] R. Qie, X. Guan, X. Yan et al., "Method and its application of natural quality evaluation of arable land based on self-organizing feature map neural network," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 30, no. 23, pp. 298–305, 2014.

[4] Y. Zhang, J. Pan, L. Qi, and Q. He, "Privacy-preserving quality prediction for edge-based IoT services," *Future Generation Computer Systems*, vol. 114, pp. 336–348, 2021.

[5] Q. Rui-Qing, L. Xue, and M. Wang, "Self-organizing feature map (SOFM) neural network method in reservoir quality synthetic evaluation and its application in Jixi basin," *Ence of Surveying & Mapping*, vol. 39, no. 1, pp. 168–174, 2009.

[6] F. Shen and O. Hasegawa, "A fast nearest neighbor classifier based on self-organizing incremental neural network," *Neural Networks*, vol. 21, no. 10, pp. 1537–1547, 2008.

[7] Y. Zhang, K. Wang, Q. He et al., "Covering-based web service quality prediction via neighborhood-aware matrix factorization," *IEEE Transactions on Services Computing*, vol. 1, 2019.

[8] X. Fu, H. Chen, and R. Cai, "Comprehensive evaluation of power quality based on self-organizing feature mapping networks," *Huanan Ligong Daxue Xuebao/Journal of South China University of Technology (Natural Science)*, vol. 42, no. 6, pp. 7–11, 2014.

[9] L. Zhao, "Study on development mechanism of regional tourism system based on self-organizing theory: a theoretical analysis framework," *Journal of Beijing International Studies University*, 2010.

[10] S. Yuan and J. Bi, "Comprehensive evaluation approach of power quality based on neural network ensemble and subordinate degree," in *Proceedings of the International Conference on Information Science & Engineering*, IEEE, Harbin, China, January 2011.

[11] X. Wang and L. Wang, "Comprehensive evaluation of students based on SOFM neural networks," *Journal of Hebei Normal University (Natural Science Edition)*, vol. 35, no. 3, 2011.

[12] J. Zhao, H. Peng, J. Zhang et al., "Short-term load forecasting based on self-organizing feature mapping neural network," *Journal of Guizhou University of Technology*, vol. 86, no. 8, pp. 164–167, 2003.

[13] C. Senabre, S. Valero, and J. Aparicio, "Using a self organizing map neural network for short-term load forecasting, analysis

of different input data patterns," in *Proceedings of the Distributed Computing & Artificial Intelligence-International Symposium. DBLP*, pp. 397–400, Avila, Spain, June 2010.

[14] M. López, S. Valero, C. Senabre, J. Aparicio, and A. Gabaldon, "Application of SOM neural networks to short-term load forecasting: the Spanish electricity market case study," *Electric Power Systems Research*, vol. 91, pp. 18–27, 2012.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 2828061, 8 pages
https://doi.org/10.1155/2021/2828061

Research Article

Research of Total Factor Productivity and Agricultural Management Based on Malmquist-DEA Modeling

Binghun Wan [1,2] and Ende Zhou [1,3]

[1] School of Economics and Management, Hubei University of Automotive Technology, Shiyan 442002, Hubei, China
[2] School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China
[3] School of Business, Hubei University, Wuhan 430062, China

Correspondence should be addressed to Ende Zhou; research7102@163.com

Received 16 May 2021; Revised 4 June 2021; Accepted 25 June 2021; Published 2 July 2021

Academic Editor: Wen-Tsao Pan

Based on the Malmquist-DEA Modeling and drawing on the data from 12 cities in Hubei, a central province of China, this paper measures the total factor productivity (TFP) of agricultural management as well as technological change (TC) and technical efficiency change (EC). The Cobb–Douglas (C-D) production function is adopted to empirically estimate the impacts of TFP and its constituent elements on the agricultural management and economic growth comprehensively and further study the effects on different regions in Hubei. The results demonstrate that TC grows at an annual average rate of 6.7% and drives agricultural TFP growth in Hubei. The decline in scale efficiency accounts for the drop of 1.1% of EC. The agricultural TFP growth rates among different regions vary remarkably but overall have a positive and significant effect on agricultural output. The research sheds light on the analysis of agricultural development of Hubei according to the findings based on Malmquist-DEA Modeling and provides practical implications for the future management.

1. Introduction

According to the report delivered at the 19th CPC National Congress held in 2017, efforts should be made for better quality, higher efficiency, and more robust drivers of China's economic growth through reform, and TFP needs to be raised. In neoclassical economics, economic growth is sourced from two parts: the growth of factors of production and the growth of TFP. TFP can be raised significantly if limited input factors are effectively used and allocated. Therefore, with limited resources given, improving TFP is the key to achieve high-quality development. TFP, as previous studies suggested, is a key indicator to measure the quality of economic growth in a country or region and is crucial for economic and social development [1]. If the conventional, unsustainable pattern of high input and excessive waste for high yield were to change to achieve quality agricultural development, the top priority would be to increase TFP and replace old growth drivers with new ones.

It is imperative for China's major agricultural provinces such as Hubei to achieve high-quality development and to grow strong in agriculture by leveraging limited agricultural resources. A key to addressing the issue is to multiply the contribution of TFP to agricultural growth. Thus, this study starts from the measurement and decomposition of agricultural TFP in a new landscape and then introduces TFP and its decomposed elements to the C-D production function to analyze how input factors, such as TC, EC, capital, and labor, contribute to the growth of Hubei's agricultural economy. Meanwhile, this paper offers proposals on how to maintain a stable and high-quality agricultural economy through analysis on regional disparity within Hubei province.

2. Literature Review

In the 1950s, American economist Robert Merton Solow built the aggregate production function and growth model that exhibit constant returns to scale (CRS) and further proposed the concept of TFP, believing that its increase was attributed to TC [2]. This has laid a solid theoretical foundation for research on TFP.

Generally, there are two types of approaches to measuring agricultural TFP: frontier and nonfrontier. A majority of research efforts were based on nonfrontier approaches, which take no account of technical inefficiency but maintain that all changes to TFP are attributed to TC. Moreover, nonfrontier approaches can also be divided into two types: parametric methods, which mainly refer to average function methods, and nonparametric methods, such as methods based on exponential functions [3, 4] and on growth accounting [5, 6]. Scholars started exploring changes in TFP by frontier approaches, which were more rational, as the production frontier model was introduced in the mid-1990s. Unlike nonfrontier approaches, frontier ones prove more advantageous as they take technical inefficiency into account. Among them, there are parametric approaches, including deterministic frontier analysis (DFA) [7] and stochastic frontier analysis (SFA) [8], and there are nonparametric ones, mainly based on the Malmquist productivity index, which can be related to DEA [9–11] or SFA. Superior to other equivalents, the DEA approach can decompose TFP, requires neither priori assumptions of production function nor parametric estimates, and allows inefficiencies [12]. Given that, the particular approach has been widely adopted in related studies of TFP. Most scholars employing the DEA-Malmquist productivity index to study TFP in China's agriculture concluded that agricultural TFP was mainly driven by TC and that the decrease of EC was in sync with that of TC for most parts of China [13, 14]. The paper, therefore, attempts to employ the DEA-Malmquist index-based approach to decompose agricultural TFP of Hubei into TC and EC.

In the past decade or so, China's research on agricultural TFP has not only focused on different measurement approaches of TFP, but also placed greater emphasis on factors affecting agricultural TFP from various perspectives. The previous studies on how agricultural TFP contributes to economic growth, however, are few. Ji et al. demonstrated the positive impact of TFP on total agricultural output by measuring the change of TFP of 13 prefecture-level cities in Jiangsu province and by analyzing its contribution to agricultural output through a production function [15]. Zhang et al. measured the agricultural TFP of 9 cities in Guizhou, a southwestern province of China, from 2010 to 2017 and found that the impact of TC and EC on agricultural production of Guizhou was significant [16]. With the DEA-Malmquist index-based approach, Li et al. calculated agricultural TFP of China from 2004 to 2016 before concluding that agricultural TFP growth accounted for 53.7% of the China's agricultural output [17]. Building on these endeavors, this paper measures the agricultural TFP index of 12 prefecture-level cities in Hubei from 2009 to 2019 with the DEA-Malmquist nonparametric approach and decomposes it into TC and EC, which are then used to construct a C-D production function model. This way, it aims to explore how input factors, such as TC, EC, capital, and labor, contribute to agricultural output of Hubei and, on this basis, to put forward suggestions for reference in formulating agricultural policies in Hubei.

3. Data and Methods

3.1. Malmquist Productivity Index in Agriculture. Enlightened by research on consumption index by Swedish economist Malmquist [18], Caves et al. constructed Malmquist productivity index (Malmquist index, in short) [19], but without further study on how to measure the distance function. It was when Fare et al. had combined DEA with nonparametric linear programming that the index was widely applied [20].

As mentioned above, the Malmquist index takes into account technical inefficiency and decomposes TFP into TC and EC based on the CRS assumption. If returns to scale are variable, EC can be further divided into pure technical efficiency change (PE) and scale efficiency change (SE). Assuming that there are k decision-making units (DMU), where $k = 1, 2, \ldots, K$, the input and output vectors of each period are $x^{k,t} = (x_1^{k,t}, x_2^{k,t}, \ldots, x_N^{k,t}) \in R_+^N$ and $y^{k,t} = (y_1^{k,t}, y_2^{k,t}, \ldots, y_N^{k,t}) \in R_+^M$ respectively, where $t = 1, 2, \ldots, T$. Therefore, the input-oriented Malmquist index can be expressed as (1) under the CRS assumption.

$$\begin{aligned} &M_i^k\left(x^{k,t+1}, y^{k,t+1}, x^{k,t}, y^{k,t}\right) \\ &= \frac{D_i^{k,t+1}\left(x^{k,t+1}, y^{k,t+1}\right)}{D_i^{k,t}\left(x^{k,t}, y^{k,t}\right)} \times \left[\frac{D_i^{k,t}\left(x^{k,t+1}, y^{k,t+1}\right)}{D_i^{k,t+1}\left(x^{k,t+1}, y^{k,t+1}\right)} \times \frac{D_i^{k,t}\left(x^{k,t}, y^{k,t}\right)}{D_i^{k,t+1}\left(x^{k,t}, y^{k,t}\right)}\right]^{1/2} \\ &= \mathrm{EC}_i^k \times \mathrm{TC}_i^k = \mathrm{PE}_i^k \times \mathrm{SE}_i^k \times \mathrm{TC}_i^k. \end{aligned} \tag{1}$$

$D_i^{k,t+1}(x^{k,t+1}, y^{k,t+1})/D_i^{k,t}(x^{k,t}, y^{k,t})$ in (1) measures the EC of DMU k from period t to $t+1$, indicating the impact of EC on TFP for a corresponding period, and EC can be further divided into PE and SE. The section in the square bracket measures TC of DMU k from period t to $t+1$, which indicates the impact of advancement of production technology frontiers on TFP for a corresponding period.

We regard each prefecture-level city in Hubei as an independent DMU and create the optimal frontier of agricultural production in the province for periods under the same technical conditions. It is followed by a comparison of the relationship between the coordinates of agricultural production point of each DMU and the position of the optimal frontier. The technical efficiency of a DMU is at the highest level if the agricultural production point of the DMU is just on the frontier, and if the point is within the frontier, then the DMU is characterized by technical inefficiency. Meanwhile, with the time factor taken into consideration as mentioned earlier, we can compare the agricultural production point of a DMU with the mapping point of the optimal frontier and thus decompose agricultural TFP into TC and EC. Therefore, if $\mathrm{TC} = 1$ for a DMU, this means there is no technical change or innovation for the DMU from t to $t+1$, whereas $\mathrm{TC} > 1$ (or $\mathrm{TC} < 1$) indicates technical progress (or setback). Similarly, $\mathrm{EC} > 1$ ($\mathrm{EC} < 1$) implies there is technical efficiency gain (loss) for the DMU from t to $t+1$. Likewise, $M = 1$ indicates that agricultural TFP in the DMU from t to $t+1$ stays unchanged; $M > 1$ ($M < 1$) denotes an increase (decline) of agricultural TFP.

3.2. Production Function Modeling. Given the above-mentioned measurement formula of the Malmquist index and with the initial year as a base period, the agricultural total factor productivity aggregate rate (TFPA) of a DMU can be calculated through the following equation:

$$\mathrm{TFPA}_i^{k,T} = M_i^{k,t+1} \times M_i^{k,t+2} \times M_i^{k,t+3} \times, \ldots, M_i^{k,T} = \prod_{j=1}^{T-t} M_i^{k,t+j}. \tag{2}$$

Likewise, the agricultural technological change aggregate rate (TCA) and agricultural technical efficiency change aggregate rate (ECA) of a DMU, with the initial year as a base period, can also be calculated by the following equation:

$$\begin{cases} \mathrm{TCA}_i^{k,T} = \prod_{j=1}^{T-t} \mathrm{TC}_i^{k,t+j}, \\ \mathrm{ECA}_i^{k,T} = \prod_{j=1}^{T-t} \mathrm{EC}_i^{k,t+j}. \end{cases} \tag{3}$$

Therefore, inspired by Kumar et al. [21] and Los et al. [22], we decompose the source of economic growth into three parts: TC, EC, and input factors such as capital and labor. We then put them into Cobb–Douglas production function and construct models (4) and (5) as follows. (Production function model (4) is formed by the undecomposed TFPA with input factors such as capital and labor; production function model (5) is comprised of the decomposed TCA and ECA, as well as input factors. Given that TFPA, TCA, and ECA are rates of change, logarithms of these three variables are not taken in the following models.)

$$\mathrm{LnTV}_{it} = \lambda_0 + \lambda_1 \mathrm{TFPA}_{it} + \lambda_2 \mathrm{LnFERT}_{it} + \lambda_3 \mathrm{LnLABOR}_{it} + \lambda_4 \mathrm{LnMCHN}_{it} + \varepsilon_{it}, \tag{4}$$

$$\mathrm{LnTV}_{it} = \delta_0 + \delta_1 \mathrm{TCA}_{it} + \delta_2 \mathrm{ECA}_{it} + \delta_3 \mathrm{LnFERT}_{it} + \delta_4 \mathrm{LnLABOR}_{it} + \delta_5 \mathrm{LnMCHN}_{it} + \eta_{it}. \tag{5}$$

In models (4) and (5), i represents each prefecture-level city in Hubei; t denotes the year; TV_{it} indicates the total output (by 100 million yuan) of the agriculture, forestry, animal husbandry, and fishery of each prefecture-level city over the years; TFPA_{it}, TCA_{it}, and ECA_{it} denote the aggregate rate of agricultural TFP, TC, and EC in each prefecture-level city over the years against the initial year, respectively; FERT_{it} denotes chemical fertilizer consumption (by 1000 tons) in each prefecture-level city over the years; LABOR_{it} indicates the number of workers (by 10,000 people) engaged in the agriculture, forestry, animal husbandry, and fishery of each prefecture-level city over the years; $\mathrm{MCHN}_{\mathrm{it}}$ represents the total power consumption (by 10,000 kW) of agricultural machinery in each prefecture-level city over the years; λ_0 and δ_0 refer to intercepts; and ε_{it} and η_{it} are random terms.

3.3. Data Sources and Descriptive Statistics of Variables. We collected 132-sample data about the agricultural input and output from 12 prefecture-level cities (excluding Enshi Tujia and Miao Autonomous Prefecture) in Hubei province from the year 2009 to 2019. All data are sourced from *Hubei Statistical Yearbook*, *Agricultural Yearbook of China*, and *National Agricultural Costs and Returns Compilation* from 2009 to 2019. The descriptive statistics of variables are shown in Table 1.

Considering the availability of data, we only take labor, land, chemical fertilizer, machinery power, and irrigation as input factors in our models and assume that other input such as agricultural film, seeds, and seedlings exerts little influence on the output. In addition, the chemical fertilizer input is calculated according to the effective net amount applied. Labor input is represented by the year-end number of workers in agriculture, forestry, animal husbandry, and fishery, and machinery input by the total power of agricultural machinery. Land input is represented by the year-end sown area of crops, and irrigation input by effective irrigation area.

4. Empirical Results and Analysis

4.1. Temporal Changes of Agricultural TFP and Its Decomposition in Hubei. We employ the DEAP 2.1 software to compute Malmquist index and its decomposition for Hubei province as a whole from 2009 to 2019 and for each city of Hubei, respectively. The results, as shown in Table 2, illustrate that the Malmquist TFP index of Hubei's agricultural sector grew by 5.6% on average, which was a remarkable increase from 2009 to 2019. In the same period, the annual average growth rate of total agricultural output in Hubei reached about 8.75%, suggesting that 64% of the agricultural output growth was attributed to increased productivity. The agricultural sector of Hubei, undoubtedly, saw considerable fluctuations in the TFP index. For example, the growth rates in 2010 and 2012 reached 20.8% and 15.4%, respectively, while there was a decline of 3.8% and 7.5% in 2017 and 2018, which to some extent reflected the unstable nature of agricultural production.

Table 1: Descriptive statistics for output and input variables.

Variable	Declaration	N	Mean	Standard deviation	Minimum	Maximum
TV	Total output value of agriculture, forestry, animal husbandry, and fishery (RMB 100 million) at 2019 prices	132	413.6	210.7	89.87	836.5
TFPA	TFP aggregate rate (%)	132	1.61	0.59	0.98	3.62
TCA	TC aggregate rate (%)	132	1.67	0.48	1	2.85
ECA	EC aggregate rate (%)	132	0.97	0.19	0.63	1.66
FERT	Chemical fertilizer consumption (1,000 tons)	132	236.7	148.1	43.58	606.1
LABOR	Number of workers in agriculture, forestry, animal husbandry, and fishery (10,000 people)	132	60.79	31.45	16.97	134.3
MCHN	Total power consumption of agricultural machinery (10,000 kW)	132	281.2	165.9	49.73	680.8
LAND	Year-end actual cultivated area (1000 hectares)	132	250.33	137.72	40.4	682.96
IRRI	Effective irrigation area (1000 hectares)	132	172.38	119.21	26	573.3

Note. TFPA, TCA, and ECA are calculated by (2) and (3). TV has been converted to current price based on price index of agricultural production. Source: State Economic Planning Commission and State Statistical Bureau (2009–2019).

Table 2: Temporal changes of the agricultural Malmquist index and its composition in Hubei (2009–2019).

Year	Malmquist index (TFP)	Technical change index (TC)	Technical efficiency change index (EC)	Pure technical efficiency change index (PE)	Scale efficiency change index (SE)
2009–2010	1.208	1.126	1.073	1.067	1.006
2010–2011	1.065	1.060	1.005	0.982	1.023
2011–2012	1.154	1.189	0.970	1.003	0.968
2012–2013	1.067	1.203	0.887	0.943	0.941
2013–2014	1.051	1.042	1.009	1.005	1.004
2014–2015	1.036	1.099	0.943	0.971	0.970
2015–2016	1.086	1.083	1.003	1.007	0.996
2016–2017	0.962	0.980	0.981	1.000	0.981
2017–2018	0.925	0.899	1.029	1.023	1.005
2018–2019	1.030	1.023	1.006	0.997	1.010
2010–2019	1.056	1.067	0.989	0.999	0.990

Note. TFP can be decomposed into TC and EC, whereas EC can be further decomposed into PE and SE. Source: computed by authors based on the data from *Hubei Statistical Yearbook* (2009–2019).

From the perspective of the composition of Malmquist index, it is agricultural TC that drives the growth of agricultural TFP in Hubei. From 2009 to 2019, the agricultural TC in Hubei increased by 6.7% annually, while the agricultural EC decreased by 1.1%. The TC value was greater than 1 throughout the sample period (excluding 2017 and 2018), suggesting that agricultural technology was advancing for most of the time. EC, however, was smaller than 1 in four years of the sample period, indicating a significant loss in technical efficiency. Technological progress, coupled with decreased efficiency, implied that the province came a long way in technological innovation in agriculture for the sample period, despite inefficiency in applying existing agricultural technology. The decomposition of EC showcased the fact that loss in agricultural EC resulted from the poor performance of PE and SE. During the sample period, PE and SE experienced a decline of 0.1% and 1% on average, respectively. This, therefore, explains that loss in agricultural EC for Hubei is mainly caused by decreased SE, a conclusion inconsistent with previous findings in other Chinese provinces [15]. The possible reason behind it is that agricultural production by small household farmers still prevails in Hubei, leading to the lag in promoting and applying cutting-edge technology. With that, greater efforts should be made to promote new agricultural technology and encourage large-scale farming in a way to increase technical efficiency.

4.2. Regional Difference in Agricultural TFP and Its Decomposition in Hubei. According to the official geographical division, Hubei comprises three main regions, namely East Hubei, Central Hubei, and West Hubei. Boasting multiple lakes, the eastern part includes the cities of Wuhan, Huangshi, Ezhou, Xianning, and Huanggang. The central region consists of the cities of Jingmen, Jingzhou, Xiaogan, and Suizhou, and features a large expanse of plains, making it a granary for the province. The mountainous western part, also known as Northwest Hubei, comprises the cities of Shiyan, Yichang, and Xiangyang, as well as Enshi Autonomous Prefecture. (Hubei Province is comprised of 12 prefecture-level cities and Enshi Tujia and Miao Autonomous Prefecture. We only observed the realities of the 12 cities, excluding Enshi Autonomous Prefecture.)

In the view of regional distribution, 12 prefecture-level cities in Hubei saw an increase in agricultural TFP from 2009 to 2019, yet growth rates of the three regions varied remarkably. Central Hubei took the lead with a growth rate in agricultural TFP of 7.6%, followed by West Hubei (5.8%, slightly above the provincial average) and East Hubei (3.9%).

Among the 12 cities, the top three in terms of agricultural TFP growth rate were Jingmen (13.7%), Wuhan (9.2%), and Jingzhou (8.4%). 2 out of the 3 cities are located in Central Hubei. Among the bottom four cities in agricultural TFP, East Hubei accounted for 3 cities, namely Ezhou (3.7%), Huangshi (1.2%), and Huanggang (0.9%).

From the perspective of decomposition, agricultural TFP growth in East Hubei, Central Hubei, and West Hubei from 2009 to 2019 was driven by agricultural TC, which grew at an annual average rate of 6.1%, 8%, and 5.9%, respectively. The three regions saw different degrees of loss in EC. The situation in western Hubei was relatively optimistic, with an average annual efficiency loss of only 0.1%, while the eastern Hubei experienced the maximum efficiency change, with an average annual efficiency loss of 2.1%, which led to the situation that East Hubei ranked at the bottom in agricultural TFP growth despite a relatively developed economy. Moreover, half of the cities in Hubei witnessed technological progress and technical efficiency loss. Among them, half were in East Hubei, which suffered a low TFP growth rate as a result of decreased agricultural EC offsetting the contribution of TC to TFP growth. The increase of TFP in Wuhan, Ezhou, and Shiyan was totally boosted by TC, since EC of the 3 cities stayed unchanged during the study period. Among all the 12 cities, only three cities—Jingmen, Yichang, and Jingzhou—embraced an improvement both in TC and in EC. Through further decomposition of EC, it is not difficult to find that efficiency loss in Central Hubei and West Hubei was attributed to a loss in SE, not in PE, and that in East Hubei was due to a loss in SE and PE, with SE exerting a greater impact. This suggests that large-scale promotion of agricultural technology is expected to be made across Hubei province. Overall, the key to enhancing TFP across the board and to ensuring quality and sustainability in agriculture is the promotion of cutting-edge agricultural technology, the wide and standardized application of new technology, and an increase in agricultural technical efficiency (see Table 3).

4.3. *Contribution of Agricultural TFP and Its Decomposition to the Growth of Hubei's Agricultural Economy.*

To measure the contribution of agricultural TFP and its decomposed factors (agricultural TC and EC) to the growth of the agricultural economy requires an estimate of unknown parameters in models (4) and (5). Prior to that, a coefficient test on variables is conducted to ensure that the models suffer severe multicollinearity (see Tables 4 and 5).

As shown in Tables 4 and 5, the correlation coefficients of independent variables are smaller than 0.8, except for LnMCHN and LnFERT. Thus, the variance inflation factor (VIF) is performed to ensure that the correlation between these two variables does not exert a serious impact on models (4) and (5). The results showed that the VIFs of all independent variables in model (4) are not greater than 7.65 and those in model (5) are not greater than 9. Therefore, the independent variables listed in Tables 4 and 5 can be included in models (4) and (5) at the same time.

According to the result of Hausman test, we chose to specify a two-way fixed-effects model to estimate the parameters, in which city-specific effects and time-specific effects were controlled by the product of time trend and city dummies. In order to ensure the reliability of the regression results, we also reported Pooled OLS estimators as a contrast. Table 6 displays the regression results of models (4) and (5), indicating the impact of the agricultural TFP aggregate rate (TC and EC) and other input factors on Hubei's agricultural output. The regression results below have passed the serial correlation test and the heteroscedasticity test.

The results presented in Table 6 suggest that the goodness of fit using Pooled OLS method was inferior to that controlling for two-way fixed effects. The latter's estimates, therefore, were used to discuss the results of models (4) and (5), respectively.

Model (4) indicates that agricultural TFP played a significant positive role in Hubei's agricultural output. To be specific, agricultural output grew by 0.302% on average with an increase of 1% in the TFP aggregate rate. That means increased agricultural TFP drove the growth of the province's agricultural economy. The estimates in model (5) show that both the decomposed factors of agricultural TFP, i.e., technological progress and efficiency enhancement, promoted the growth of Hubei's agricultural economy in an effective manner. Specifically, an increase of 1% in TCA and ECA led to a rise of 0.279% and 0.322%, respectively, in

Table 3: Regional differences of agricultural Malmquist index and its decomposition in Hubei.

Region	City	TFP	TC	EC	PE	SE
East Hubei	Wuhan	1.092	1.092	1.000	1.000	1.000
	Huangshi	1.012	1.058	0.957	1.000	0.957
	Ezhou	1.037	1.037	1.000	1.000	1.000
	Huanggang	1.009	1.051	0.959	0.997	0.962
	Xianning	1.044	1.068	0.978	0.978	1.000
	Average	**1.039**	**1.061**	**0.979**	**0.995**	**0.984**
Central Hubei	Jingmen	1.137	1.103	1.030	1.034	0.997
	Xiaogan	1.027	1.063	0.966	0.978	0.988
	Jingzhou	1.084	1.081	1.004	1.020	0.984
	Suizhou	1.057	1.073	0.984	0.987	0.997
	Average	**1.076**	**1.080**	**0.996**	**1.005**	**0.992**
West Hubei	Shiyan	1.041	1.041	1.000	1.000	1.000
	Yichang	1.055	1.044	1.010	1.000	1.010
	Xiangyang	1.079	1.093	0.987	1.000	0.987
	Average	**1.058**	**1.059**	**0.999**	**1.000**	**0.999**
Provincial Average		1.056	1.067	0.989	0.999	0.990

Note. The Malmquist index and its decomposition are the annual mean of each city from 2009 to 2019. Source: computed by authors based on the data from *Hubei Statistical Yearbook* (2009–2019). The bold values are the arithmetic means of TFP, TC, EC, PE, and SE for each region of Hubei.

Table 4: Correlation coefficient matrix of variables in model (4).

Variable	LnTV	TFPA	LnFERT	LnLABOR	LnMCHN
LnTV	1				
TFPA	0.361^{***}	1			
LnFERT	0.764^{***}	0.222^{**}	1		
LnLABOR	0.779^{***}	−0.104	0.788^{***}	1	
LnMCHN	0.862^{***}	0.440^{***}	0.834^{***}	0.760^{***}	1

Note. Because of multicollinearity, the input of land and irrigation is excluded from the final model.

Table 5: Correlation coefficient matrix of variables in model (5).

Variable	LnTV	ECA	TCA	LnFERT	LnLABOR	LnMCHN
LnTV	1					
ECA	0.027	1				
TCA	0.425***	−0.088	1			
LnFERT	0.764***	0.229***	0.117	1		
LnLABOR	0.779***	−0.041	−0.076	0.788***	1	
LnMCHN	0.862***	0.175**	0.415***	0.834***	0.760***	1

Note. Because of multicollinearity, the input of land and irrigation is excluded from the final model.

Table 6: Estimates on the impact of agricultural TFP and its decomposition on the growth of agricultural economy in Hubei.

Variable	Model (4)		Model (5)	
	Pooled OLS	Two-way fixed effects	Pooled OLS	Two-way fixed effects
TFPA	0.313*** (3.98)	0.302*** (3.48)		
TCA			0.554*** (6.93)	0.279*** (3.62)
ECA			0.198* (1.71)	0.322** (2.04)
LnFERT	−0.016 (−0.28)	0.094*** (7.73)	0.103* (1.90)	0.208*** (4.15)
LnLABOR	0.654*** (6.13)	0.055*** (6.51)	0.733*** (7.94)	0.405*** (3.49)
LnMCHN	0.234** (2.52)	0.014 (1.55)	0.023 (0.26)	0.290*** (2.64)
Constant	1.591*** (9.14)	0.890*** (6.37)	1.187*** (7.56)	0.588*** (3.24)
N	132.000	132.000	132.000	132.000
Adj-R^2	0.808	0.969	0.854	0.968

Note. The figures in the parentheses are t statistics of estimates. ∗∗∗, ∗∗, and ∗ denote significance at 1%, 5%, and 10%, respectively. The models include 11 dummy variables to control for two-way fixed effects, but the estimated coefficients are not included for brevity.

Hubei's average agricultural output. That means both technical innovation and technical efficiency change in agriculture can considerably enhance agricultural output of Hubei, with technical efficiency change contributing more to the growth of the province's agricultural economy. Given that, Hubei needs to attach greater importance to the R&D of agricultural innovations and the promotion of existing cutting-edge technology.

From the perspective of input factors, estimated results of models (4) and (5) suggest that chemical fertilizer, labor, and machinery input all boosted the growth of Hubei's agricultural economy. Nonetheless, these input factors made a far less contribution to agricultural output than TFP did. This indicates that although the conventional input factors of production drove the growth of the agricultural economy, its contribution was limited. The high-quality and sustainable development of Hubei's agricultural sector hinged more on the innovation of agricultural technology and the improvement of agricultural technical efficiency.

4.4. Contribution of Agricultural TFP and Its Decomposition to Agricultural Economic Growth in East Hubei, Central Hubei, and West Hubei. Table 7 shows how the agricultural TFP aggregate rate (TC and EC) and other input factors in East Hubei, Central Hubei, and West Hubei make an impact on their respective agricultural output. Similarly, we controlled for two-way fixed effects using the product of city dummies and time trend term for each region. As shown in model (4), agricultural TFP in the three regions contributed significantly to their growth of the agricultural economy, with Central Hubei taking the lead and West Hubei ranking at the bottom. To be specific, a rise of 1% in the TFP aggregate rate for Central Hubei and West Hubei brought about an increase of 0.72% and 0.166% in average agricultural output, respectively. The impact of the TFP aggregate rate on agricultural output for West Hubei was even less than that of fertilizer input and machinery input on its agricultural output.

As presented in model (5), TCA and ECA made a greater contribution to agricultural output in Central Hubei than in East Hubei and West Hubei, whereas the impact of TCA on the growth of the agricultural economy was the smallest in West Hubei compared to the other two regions. The estimates on Central Hubei and West Hubei showed that the contribution of ECA to agricultural output was greater than that of TCA, consistent with the above-mentioned regression results about the entire province. As far as regression results about East Hubei were concerned, the impact of TFP growth on agricultural output (0.249%) was overwhelmingly attributed to TCA (0.248%), and ECA contributed little to agricultural output probably due to loss in technical efficiency, which, as mentioned above, was more severe in East

Table 7: Estimates about the impact of agricultural TFP and its decomposition on agricultural economic growth in East Hubei, Central Hubei, and West Hubei.

Variable	Model (4)			Model (5)		
	Eastern Hubei	Central Hubei	Western Hubei	Eastern Hubei	Central Hubei	Western Hubei
TFPA	0.249^{***} (5.59)	0.720^{***} (10.89)	0.166^{**} (2.59)			
TCA				0.248^{***} (5.71)	0.710^{***} (10.39)	0.181^{***} (3.02)
ECA				0.173 (1.16)	1.007^{***} (10.95)	0.828^{***} (3.57)
LnFERT	0.222^{***} (7.75)	−0.293 (−1.23)	0.664^{***} (6.35)	0.212^{***} (7.11)	−0.178 (−0.80)	0.650^{***} (6.46)
LnLABOR	0.066 (1.32)	0.933^{***} (12.20)	−0.249 (−1.10)	0.088^{*} (1.68)	0.930^{***} (12.76)	−0.147 (−0.67)
LnMCHN	0.723^{***} (17.39)	0.045 (0.27)	0.290^{**} (2.76)	0.712^{***} (16.84)	−0.127 (−0.80)	0.266^{**} (2.72)
Constant	0.157 (1.29)	2.210^{***} (5.96)	1.137^{*} (1.82)	0.022 (0.10)	1.531^{***} (4.62)	0.057 (0.07)
N	55.000	44.000	33.000	55.000	44.000	33.000
Adj-R^2	0.995	0.980	0.990	0.995	0.982	0.991

Note. The figures in the parentheses are t statistics of estimates. $***$, $**$, and $*$ denote significance at 1%, 5%, and 10%, respectively. The models include dummy variables to control for two-way fixed effects, but the estimated coefficients are not included for brevity.

Hubei than in the other two regions. In short, model (5) suggests that the influence of ECA on agricultural economic growth was more profound than that of TCA.

5. Conclusions and Implications

Building on statistics from 12 prefecture-level cities in Hubei and employing the DEA-Malmquist productivity index, this paper measures and decomposes agricultural TFP of Hubei. On this basis, the C-D production function is adopted to empirically study the contribution of TFP and its decomposed elements to the agricultural economic growth in Hubei. We can draw the following conclusion based on the analysis above.

First, agricultural TFP of Hubei showed volatile growth, with technical progress and technical efficiency loss coexisting. The TFP index grew by 5.6% on average from 2009 to 2019, and 64% of the increase in agricultural output came from the growth in agricultural TFP, which was driven more by technological progress than by higher technical efficiency. That means Hubei produced fruitful results in innovation in agricultural technology over the decade, yet lacking the application and promotion of existing technologies.

Second, loss in agricultural technical efficiency was mainly attributed to a decline in SE, in addition to decreased PE. Therefore, more efforts need to be made to apply and promote frontier technology, encourage large-scale agricultural production, and develop new technical standards. This will allow for a wide application of new technology and further increase technical efficiency, particularly scale efficiency, for Hubei's agricultural sector.

Third, the TFP growth among cities and regions differed remarkably. From 2009 to 2019, agricultural TFP in East Hubei, Central Hubei, and West Hubei was on the rise but to varying degrees, with Central Hubei, the province's granary, seeing the largest growth, followed by West Hubei and East Hubei. Such a difference depends on the different agricultural resource endowments and the different agricultural output of each region.

Fourth, agricultural TFP growth in East Hubei, Central Hubei, and West Hubei was driven by TC. During the sample period, technical efficiency declined disproportionately across the three regions, with East Hubei seeing the largest drop. An analysis of decomposed factors showed that loss in technical efficiency in Central Hubei and West Hubei resulted from a decrease in SE rather than in PE and that technical efficiency loss in East Hubei was caused by both SE and PE, with the former exerting a greater impact. Thus, to enhance TFP across the province and ensure quality and sustainability in agriculture, the key is to embrace the large-scale and standardized application of technology and increase agricultural technical efficiency.

Fifth, the growth of Hubei's agricultural economy depended on the increase of agricultural TFP (TC and EC). Moreover, an increase of 1% in the TFP aggregate rate led to an uptick of 0.302% in average agricultural output, which was significantly elevated by technological innovation and efficiency. The impact of EC on the growth of the agricultural economy was larger than that of TC. Compared to input factors such as chemical fertilizer, labor, and machinery, TFP growth had a far greater impact on the growth of the agricultural economy. Given that, Hubei should focus more on developing agricultural innovations and promoting existing cutting-edge technology.

Sixth, agricultural TFP growth (especially TC) made a significant, positive contribution to the growth of the agricultural economy in the three regions of Hubei, with Central Hubei being the largest contributor, followed by East Hubei and then by West Hubei. Even on the decomposition of TFP, the contribution of TC and EC in Central Hubei to agricultural output was larger than that in the other two regions. In terms of TC's contribution to agricultural output, West Hubei played a smaller role. In East Hubei, the impact of agricultural TFP growth on agricultural economic growth was overwhelmingly attributed to TC.

Given the aforementioned conclusions, we offer some suggestions as follows. Firstly, the focus should be on how agricultural TFP significantly contributes to the growth of the local agricultural economy, before driving TFP growth as a way to develop a quality and sustainable agricultural economy in Hubei. Secondly, priority should be given to the increase of technical efficiency, particularly scale efficiency, which profoundly affects the growth of the agricultural economy. Governments should focus more on developing new technologies, promoting them on a large scale, setting up new technical standards, and offering relevant training to agricultural technology promoters. Thirdly, Hubei should fully grasp the difference in the regional growth of the agricultural economy before developing tailored and targeted measures and policies on the basis of the distinct realities of each region. Specifically, Central Hubei should maintain its strengths and make up for the shortcomings of low scale efficiency; West Hubei should bolster investment in the R&D of new technology and support for growing industrial chains; East Hubei should step up efforts to promote the large-scale application of agricultural technology. Last but not least, the rational input of production factors should be ensured. To this end, increased efforts should be made to develop and apply agricultural machinery and equipment in East Hubei; the regime of agricultural labor market allocation in Central Hubei should be optimized; the consumption of chemical fertilizer in West Hubei should be effectively controlled, on top of a wide application of agricultural machinery and equipment. In a word, tailored and targeted measures should be adopted to maximize the growth of the agricultural economy throughout Hubei.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by Major Program of National Social Science Foundation of China (Grant no. 20AGL006), Humanities and Social Sciences Project of Ministry of Education of China (Grant no. 19YJC630233), and Educational Commission of Hubei Province of China (Grant no. 19Q129).

References

[1] W. T. Woo, "Chinese economic growth: sources and prospects," in *The Chinese Economy*, M. Fouquin and F. Lemoine, Eds., Economica, London, UK, 1998.
[2] R. M. Solow, "Technical change and the aggregate production function," *The Review of Economics and Statistics*, vol. 39, no. 3, pp. 312–320, 1957.
[3] J. McMillan, J. Whalley, and L. Zhu, "The impact of China's economic reforms on agricultural productivity growth," *Journal of Political Economy*, vol. 97, no. 4, pp. 781–807, 1989.
[4] C. A. Carter, J. Chen, and B. Chu, "Agricultural productivity growth in China: farm level versus aggregate measurement," *China Economic Review*, vol. 14, no. 1, pp. 53–71, 2003.
[5] S. Fan, "Effects of technological change and institutional reform on production growth in Chinese agriculture," *American Journal of Agricultural Economics*, vol. 73, no. 2, pp. 266–275, 1991.
[6] J. Y. Lin, "Rural reforms and agricultural growth in China," *The American Economic Review*, vol. 82, no. 1, pp. 34–51, 1992.
[7] S. N. Afriat, "Efficiency estimation of production functions," *International Economic Review*, vol. 13, no. 3, pp. 568–598, 1972.
[8] S. Gupta and P. Badal, "Total factor productivity growth of sugarcane in Uttar Pradesh: parametric and non-parametric analysis," *Journal of Pharmacognosy and Phytochemistry*, vol. 10, pp. 134–137, 2021.
[9] C. R. Hulten, *Total Factor Productivity: A Short Biography*, National Bureau of Economic Research, Cambridge, MA, USA, 2000.
[10] T. J. Coelli and D. S. P. Rao, "Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980–2000," *Agricultural Economics*, vol. 32, no. 1, pp. 115–134, 2005.
[11] X. Guo, C. Deng, and D. Wang, "International comparison of the efficiency of agricultural science, technology, and innovation: a case study of G20 countries," *Sustainability*, vol. 13, no. 5, p. 2769, 2021.
[12] S. Malmquist, R. Färe, S. Grosskopf et al., *Index Numbers: Essays in Honour of Sten Malmquist*, Springer Science & Business Media, LLC, Berlin, Germany, 1998.
[13] J. Zhan, F. Zhang, Z. Li et al., "Evaluation of food security based on DEA method: a case study of Heihe river basin," *Annals of Operations Research*, vol. 290, no. 1, pp. 697–706, 2020.
[14] N. Li, Y. Jiang, Z. Yu et al., "Analysis of agriculture total-factor energy efficiency in China based on DEA and Malmquist indices," *Energy Procedia*, vol. 142, pp. 2397–2402, 2017.
[15] X. Ji and S. Zhou, "Total factor productivity and its effect on agricultural economic growth," *Science and Technology Management Research*, vol. 21, no. 36, pp. 104–108, 2016.
[16] Y. Zhang and H. Cui, "Agricultural total factor productivity and its impact on agricultural economic growth in Guizhou," *Chinese Agricultural Science Bulletin*, vol. 26, pp. 153–158, 2020.
[17] J. Li, Y. Chen, and H. Liu, "Growth and coordination of China's agricultural total factor productivity upon urbanization," *World Agriculture*, vol. 02, pp. 35–44, 2021.
[18] S. Malmquist, "Index numbers and indifference surfaces," *Trabajos De Estadistica*, vol. 4, no. 2, pp. 209–242, 1953.
[19] D. W. Caves, L. R. Christensen, and W. E. Diewert, "The economic theory of index numbers and the measurement of input, output, and productivity," *Econometrica*, vol. 50, no. 6, pp. 1414–1939, 1982.
[20] R. Fare, S. Grosskopf, M. Norris et al., "Productivity growth, technical progress, and efficiency change in industrialized countries," *The American Economic Review*, vol. 84, no. 1, pp. 66–83, 2016.
[21] S. Kumar and R. R. Russell, "Technological change, technological catch-up, and capital deepening: relative contributions to growth and convergence," *The American Economic Review*, vol. 92, no. 3, pp. 527–548, 2002.
[22] B. Los and M. P. Timmer, "The 'Appropriate technology' explanation of productivity growth differentials: an empirical approach," *Journal of Development Economics*, vol. 77, no. 2, pp. 517–531, 2005.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 2133255, 7 pages
https://doi.org/10.1155/2021/2133255

Research Article

Is the Long Memory Factor Important for Extending the Fama and French Five-Factor Model: Evidence from China

Yicun Li,[1] Yuanyang Teng,[1] Wei Shi,[1] and Lin Sun[2]

[1]*School of Management, Zhejiang University, Hangzhou, China*
[2]*School of Applied Mathematics, Guangdong University of Technology, Guangzhou, China*

Correspondence should be addressed to Lin Sun; yssl12@gdut.edu.cn

Received 10 May 2021; Accepted 10 June 2021; Published 24 June 2021

Academic Editor: Wen-Tsao Pan

This paper proposes a new factor model, which is built upon the marriage of the Fama and French five-factor model and a long memory factor based on the monthly data of the A-share market in the Chinese stock market from January 2010 to July 2020. We first examine the explanatory power of the Fama and French five-factor model. We find strong market factor return of market (RM), size factor small minus big (SMB), and value factor high minus low (HML) but weak factor robust minus weak (RMW) and investment factor conservative minus aggressive (CMA). Then, both the Hurst exponent and the momentum factors (MOM) are added to the model to test the improvement of the explanatory power of these two new factors. We find that both the momentum factor and the Hurst exponent factor can effectively improve the explanatory power of the model. The momentum factor captures the short-term trend, but it cannot completely replace the Hurst exponent, which reflects the long memory effect.

1. Introduction

In the field of quantitative investment, factor models have always attracted much attention. In 1993, Fama and French proposed a celebrated three-factor model including a size factor (SMB) and value factor (HML) in addition to the market beta, which captures the cross-sectional variation in average stock returns. Moreover, Fama and French [1] found that the three-factor model can explain many regularities that are anomalous under the capital asset pricing model, including firm size, book-to-market (BM), past sales growth, long-run reversals, cash-flow-to-price, and earnings-to-price. However, Fama and French [2] claimed that their five-factor model, which adds the profitability factor (RMW) and investment factor (CMA) to the three-factor model, is superior to their original three-factor model for US firms with new and longer data from July 1963 to December 2013. However, the search for factors that explain the cross section of expected stock returns has produced hundreds of potential candidates. A fundamental task facing the asset pricing field today is to bring more discipline to the proliferation of factors. In particular, a question that remains open is how to judge whether a new factor adds explanatory power for asset pricing, relative to the hundreds of factors the literature has so far produced?

In recent years, two more possible factors have been discovered, including momentum factor (MOM) and long memory factor, which is denoted by the Hurst parameter. Initially, Jegadeesh and Titman [3] proposed the momentum effect. In 1997, Carhart [4] observed the momentum effect of different maturities and extracted the momentum factor (MOM), which is the difference in the equal-weighted average return of the top 30% stocks and the last 30% stocks with a one-month lag in the past 11 months, to incorporate into the asset pricing model. The model explains the inertia of most fund performance. Ouyang and Fei [5] studied the applicability of the four-factor pricing model in China's stock market. They tested the four-factor asset pricing model with a six-month lagging momentum factor by region and industry and found that it has higher explanatory power than the three-factor and CAPM model.

In addition to the momentum effect, researchers have also conducted a lot of discussions on whether the time series of stock returns has the property of the long memory.

Indeed, the Hurst exponent is often used to describe the long memory of a time series. The commonly used method for estimating the Hurst exponent is the R/S analysis method (Rescaled Range Analysis) proposed by Hurst [6]. Mandelbrot [7] first applied the R/S analysis method to securities market research. However, some scholars [8–10] have shown that when there is short-term memory in the time series, the results obtained by R/S analysis are biased. Lo [9] proposed a revised R/S analysis method, but Teverovasky et al. [11] believed that the revised R/S still has a big flaw because the method must be selected for parameters, and improper selection of parameters often results in large deviations. As far as we know, the commonly used nonparametric estimation method is log-periodogram regression. Its advantage is that the algorithm is relatively simple, but the accuracy and stability are poor. To overcome this obstacle, Robinson [12] proposed another semiparametric estimation method: local Whittle estimation method (LW). He proved that LW estimation is better than the log-periodogram regression method despite the need for numerical optimization. The detrended fluctuation analysis (DFA) method is a scale index method proposed by Peng et al. [13] based on DNA mechanism, which is used to analyze the long-range correlation of time series. This method is mainly to remove the local trend of the data on different time scales, but for a time series, if there is no trend and the specific form of the trend, there will be certain limitations. In addition, the DFA method and the R/S analysis method have a common defect-insufficient accuracy when the time series length is too short. Later, some algorithms dedicated to improving the estimation accuracy appeared gradually to be more effectively applied to the analysis of financial time series, including Quasi Maximum Likelihood (QML) analysis, Generalized Hurst Exponent (GHE), wavelet analysis, Centered Moving Average (CMA), multifractal detrended fluctuation analysis (MFDFA), a nonlinear tool similar with the Lyapunov exponent, geometric method-based procedures (GM), and fractal dimension algorithms (FD). The disadvantage of the maximum likelihood estimation method is weak consistency. The wavelet transform in the wavelet analysis method involves the selection of the fundamental wavelet function. If the selection is improper, the analysis result will be greatly biased. The CMA method has better stability when n is small. Vitanov et al. [14] introduced the estimation method of Hurst exponent by MFDFA and used methods such as Lyapunov exponent and PCA to estimate the chaos of the system and compress the dimensions.

Researchers not only study the differences of Hurst exponent estimation methods but also incorporate the Hurst exponent's long memory interpretation of time series into the factor model and compare it with the momentum factor. For example, semiparametric estimation approaches involve the celebrated the R/S statistic introduced by Hurst [6]; the parameter estimation method includes the exact maximum likelihood estimation proposed by Beran [15]; Whittle maximum likelihood estimation provided by Fox and Taqqu [16] and Dahlhaus [17]; the quadratic variations approach proposed by Guyon and Leon [18] and Istas and Lang [19]; the modified R/S statistic provided by Lo [9]; the Higuchi's method (see, for example, Higuchi [20]); the detrended fluctuation analysis provided by Peng et al. [13]; the log-periodogram regression method proposed by Geweke and Porter-Hudak [21] and Robinson [22]; and the local Whittle method developed by Robinson [12]; Velasco [23]; Phillips and Shimotsu [24]; Shimotsu and Phillips [25]; Bardet and Kammoun [26]; and Shimotsu [27]. Nonparametric estimation includes the increment ratio method proposed in Surgailis et al. [28] and extended in Bardet and Surgailis [29]; the wavelet based method provided by Bardet and Kammoun [26]. López-García et al. [30] first analyzed the explanatory power of five-factor model on U.S. stock returns, and they introduced the fractal dimension algorithm (FD method), compared the Hurst exponent calculated by the FD method with the momentum factor, and pointed out the superiority of the Hurst exponent over the momentum factor in model interpretation. However, this paper based on Fama–French three-factor model only uses the FD method to calculate the Hurst exponent and has no robustness test of the Hurst estimation method. In this paper, we will analyze the explanatory power of the Hurst exponent factor based on the Fama and French five-factor model and estimate Hurst exponent by two methods to test the results for robustness.

For this purpose, we first use five factors to analyze the explanatory power of China's A-share stocks and establish a factor model that includes momentum factor and the Hurst exponent estimated by two methods. Then, we will fully compare the performance of Hurst exponent and momentum factor on model improvement and test the momentum effect and long memory of the time series in the Chinese capital market. In order to further explore the robustness of the results, we will use two popular methods to estimate the Hurst exponent, which are based on least squares method by Berzin et al. [31].

The paper is organized as follows. Section 2 introduces the five-factor model as proposed by Fama and French [2] and explains two-parameter methods for the Hurst parameter and momentum factor. Section 3 provides several empirical applications of the procedure and explores the robustness of the results. Section 4 gives the concluding remarks.

2. Five-Factor Model and Hurst Exponent

CAPM model is a classical model which describes stock returns as risk-free rate plus market premium risk return as follows:

$$R_{it} - R_{ft} = a_i + b_i\left(\mathrm{Mkt}_t - R_{ft}\right) + e_{it}, \qquad (1)$$

where R_{it} is the return on security i for period t, R_{ft} is the risk-free return, and Mkt_t is the difference between R_{it}.

From the empirical evidence on U.S. stocks and the applications of CAPM, Fama and French [1, 32] proposed an extension of (1) by introducing two new factors and capturing patterns associated with the size and value versus growth stocks. The three-factor empirical asset pricing model is defined then as follows:

$$R_{it} - R_{ft} = a_i + b_i \mathrm{Mkt}_t + s_i \mathrm{SMB}_t + h_i \mathrm{HML}_t + e_{it}, \quad (2)$$

where SMB_t is the returns on a diversified portfolio of small stocks minus the returns on a diversified portfolio of big stocks and HML_t is the difference between the returns on diversified portfolios of high book-to-market and low book-to-market stocks.

Fama and French [2] introduced a five-factor asset pricing model that adds the profitability and investment factors to the three-factor model of Fama and French [32] as follows:

$$\begin{aligned} R_{it} - R_{ft} = a_i &+ b_i \mathrm{Mkt}_t + s_i \mathrm{SMB}_t + h_i \mathrm{HML}_t \\ &+ r_i \mathrm{RMW}_t + c_i \mathrm{CMA}_t + e_{it}, \end{aligned} \quad (3)$$

where RMW_t is the difference between the returns on diversified portfolios of stocks with robust and weak profitability correspondingly, CMA_t is the difference between the returns on diversified portfolios of the stocks of companies with low and with high investment practices, and e_{it} is a zero-mean residual.

Long memory in economics and finance has attached a great attention since a ground-breaking work of Mandelbrot and Van Ness [33]. In fact, the long memory property of time series means a significant dependence between very distant observations and a pole in the neighborhood of the zero frequency of their spectrum. When stock market returns have the property of long memory, the Efficient Market Hypothesis is not confirmed. In this case, the distribution of the stock return has fat tails and is persistent. Thus, stock returns are highly correlated, and there is black noise and a trend in the market. Some early studies in long memory process in finance include the studies by Hurst [6]; Mandelbrot and Wallis [34]; and Lo [9]. To consider the long memory fact, we should estimate the Hurst parameter. In fact, there exists a vast literature that describes different methods for estimating the Hurst parameter of the fBm.

In this paper, we use two strongly consistent and asymptotically normal estimators of Berzin et al. [31].

Let $X_t = \log(\sigma_t)$, $n \in \mathbb{N}^+ - 1$, $k \in \mathbb{R}^+$, and $M_k(n) = (1/n-1)\sum_{i=0}^{n-k} (X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta})^k$.

For $i = 1, \ldots, \ell$, Berzin et al. [31] introduced the following least squares estimator H_k of H as

$$\widehat{H}_k = -\frac{1}{k} \sum_{i=1}^{\ell} z_i \log(M_k(n_i)), \quad (4)$$

where $n_i = r_i n$, $r_i \in \mathbb{N}^*$, $i = 1, \ldots, \ell$ and $z_i = (y_i / \sum_{i=1}^{l} y_i^2)$ and $y_i = \log(r_i) - (1/\ell)(\sum_{i=1}^{\ell} \log(r_i))$. Moreover, let $M_{\log}(n) = (1/(n-1))\sum_{i=0}^{n-2} \log(|X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta}|)$. Berzin et al. [31] introduced another least squares estimator $\widetilde{H}_{\log}$ of H as follows:

$$\widetilde{H}_{\log} = -\sum_{i=1}^{\ell} z_i M_{\log}(n_i). \quad (5)$$

From Remark 3.12 and Remark 3.15 of Berzin et al. [31], we can state the following asymptotic theory.

Corollary 1. *The estimator $\widehat{H}_k$ is an asymptotically unbiased strongly consistent estimator of H, and the estimator $\widetilde{H}_{log}$ is unbiased weakly consistent estimator of H. Furthermore, for $k = 2r_i = 2^{i-1}$ and $i = 1, \ldots, \ell$, we have*

$$\begin{aligned} &\sqrt{n}\left(\widehat{H}_k - H\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\widehat{H}_k}\right), \\ &\sqrt{n}\left(\widehat{H}_{\log} - H\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\widetilde{H}_{\log}}\right), \end{aligned} \quad (6)$$

where

$$\begin{aligned}
\sigma^2_{\widehat{H}_k} &= \left(\frac{6}{\log(2)}\right)^2 \frac{1}{\ell^2\left(\ell^2-1\right)^2} 2\left(\sum_{i<j;\,i,j=1}^{\ell} 2^{-j}(2i-(\ell+1))(2j-(\ell+1)) \times \sum_{r=-\infty}^{+\infty} \rho^2_{1,2j-i}(r) + \sum_{i=1}^{\ell} 2^{-i}(2i-(\ell+1))^2 \sum_{r=-\infty}^{+\infty} \rho_H^2(r)\right), \\
\sigma^2_{\widetilde{H}_{\log}} &= \left(\frac{3}{\log(2)}\right)^2 \frac{1}{\ell^2\left(\ell^2-1\right)^2} \left(2\sum_{i<j;\,i,j=1}^{\ell} 2^{-j+1}(2i-(\ell+1))(2j-(\ell+1))\right) \times \sum_{p=1}^{+\infty} (2p)! \left(\frac{1}{p(2p-1)!!}\right)^2 \sum_{r=-\infty}^{+\infty} \rho^{2p}_{1,2j-i}(r) \\
&\quad + \sum_{i=1}^{\ell} 2^{-i+1}\left(2i-(\ell+1)^2\right) \times \sum_{p=1}^{+\infty} (2p)! \left(\frac{1}{p(2p-1)!!}\right)^2 \sum_{r=-\infty}^{+\infty} \rho_H^{2p}(r), \\
\rho_{b,c}(x) &= \frac{1}{2\left(4-2^{2H}\right)} (bc)^{-H} \left[-|x|^{2H} + 2|x-b|^{2H} - |x-2b|^{2H} + 2|x+c|^{2H} - 4|x+c-b|^{2H}\right. \\
&\quad \left. + 2|x+c-2b|^{2H} - |x+2c|^{2H} + 2|x+2c-b|^{2H} - |x+2c-2b|^{2H}\right], \\
\rho_H(x) &= \frac{-6|x|^{2H} + 4|x+1|^{2H} - |x+2|^{2H} - |x-2|^{2H} + 4|x-1|^{2H}}{2\left(4-2^{2H}\right)}.
\end{aligned} \quad (7)$$

Because of the price limits policy in Chinese A-Share Market, in this paper, the momentum factor is compiled as follows:

$$\mathrm{MOM} = \frac{X_{i+k}}{X_i}, \quad i = 1, 2, 3, \ldots, n-k, \tag{8}$$

where x is the stock price and K is the momentum factor parameter, usually 12 months which will be adjusted for robust test later.

3. Empirical Application

3.1. Five-Factor Model in Chinese Stock Market. We are going to examine the performance of the Fama–French five-factor model in Chinese stock market and analyze the performance of the Hurst factor and MOM in the factor model. The stock data in this study are daily data from January 2010 to July 2020. By sampling from each month, the monthly data of the time period are obtained. The financial data needed mainly come from the quarterly financial reports of listed companies. Since listed companies announce their financial reports at different times, there are always differences in the financial data collected at the end of each month in the A-share market. Hurst exponent and momentum factor are calculated based on monthly stock price data.

Similar to Fama and French [2], the four factors of SMB, HML, RMW, and CMA in the five-factor model are calculated based on the grouping of financial data on the monthly return rate of stocks. Then, we combine the monthly return rate of the market index with the return rate of the four factors to obtain the final value of the five factors in the month, cycle the calculation of the five factors every month, and finally get the five-factor data from February 2010 to June 2020.

To study which factor in the five-factor model is more significant to explain stock returns, we use a combination of five factors as explanatory variables to construct a regression model. Five factors can form $C_5^1 + C_5^2 + C_5^3 + C_5^4 + C_5^5$ kinds of combinations; that is, 31 kinds of combinations can be formed. The Akaike [35] test (AIC) is used to select the optimal model; that is, when the AIC is the smallest, the model is regarded as the optimal model.

For all A shares, ignoring the fact that data errors cannot be regressed, the mean of r square of the regression is 0.4002. Table 1 reports the proportion of the five factors in the optimal model of each stock. The MKT is the factor with the highest proportion, which means that MKT has a universal explanatory power for A-share returns. RMA and SMB are also very important, accounting for close to 50%, which means the company's profit fundamentals have a greater impact on stock returns of A shares and small market value effect is common. CMA performances are the weakest, which may be related to the large amount of data in the A-share financial statements.

Table 1: The proportion of the five factors in the optimal model of each stock.

Factor	MKT	SMB	HML	RMW	CMA
The proportion	0.9091	0.4736	0.3824	0.5110	0.0418

3.2. Seven-Factor Model. The calculation of Hurst exponent and momentum factor both needs to determine a time series length. In this paper, the length of 12 month is selected for a preliminary study, and then the parameter will be changed for a systematic study.

The momentum factor and the Hurst exponent, which are calculated by the $\widehat{H}_k$ algorithm and the $\widehat{H}_{\log}$ algorithm, are added into the five-factor model to form a seven-factor model.

Seven factors can form $C_7^1 + C_7^2 + C_7^3 + C_7^4 + C_7^5 + C_7^6 + C_7^7$ kinds of combinations. The AIC criterion is still used for model selection. For all A shares, using the Hurst exponent calculated by the $\widehat{H}_k$ algorithm and the $\widehat{H}_{\log}$ algorithm, the mean of r square is 0.4731 and 0.4729, respectively.

The two different algorithms have no difference in the improvement of r square, which is maintained at about 47% and has an increase of 7% relative to the five-factor model.

The proportion of the seven factors in the optimal model is shown in Table 2. Using the $\widehat{H}_k$ and $\widehat{H}_{\log}$ algorithm to calculate Hurst exponent, the market factor MKT is the most significant, while MOM's proportion is second only to the market factor. RMW and SMB take a proportion about 50%, and the result is consistent with the five-factor model, which shows that the newly added MOM and Hurst exponent have a certain degree of substitution to the five-factor model. The Hurst exponent also has a certain effect, and it often appears in the model at the same time as MOM.

From Table 2, we can find different Hurst exponent performance about the same, accounting for roughly 23%. The proportions of the other 6 factors have not changed much, indicating that different algorithms of Hurst exponent have no substitute influence on other factors.

From Table 3, we check the cross effect of H and MOM. Firstly, we can conclude that both H and MOM are very important factors because there is a very low percent (roughly 9%) of models which does not contain neither H nor MOM. Secondly, this shows that although these two factors portray the trend performance, they are complementary to each other, rather than substituting previously guessed because the percent of "H and MOM" is as high as about 20%.

3.3. Portfolio Factor Analysis. In order to see if our result above is stable, in this section, we will use random portfolio which consists of 10 and 30 stocks, respectively, to run the seven-factor model and check the result for consistency with that in Section 3.2.

The mean of r square of 10,000 random portfolios of 10 stocks is 0.781194 for $\widehat{H}_k$ and 0.777463 for $\widehat{H}_{\log}$. This result is higher than the r square of the single stock model. And Table 4 shows a higher percent of presence of factors. This is because the diversity of 10 stock portfolios lowers the unsystematic risk and improves the explanatory power of seven factors.

Table 2: The proportion of the seven factors in the optimal model.

Factor	MKT	SMB	HML	RMW	CMA	MOM	Hurst
$\widehat{H}_k$	0.940152	0.485249	0.329306	0.505198	0.045518	0.876651	0.220849
$\widehat{H}_{\log}$	0.938465	0.482719	0.330149	0.507165	0.04608	0.87918	0.239112

Table 3: Percent of presence in the models of factors H and MOM.

Factor	No one	H but no MOM	MOM but no H	H and MOM
$\widehat{H}_k$	0.090475	0.030346	0.670413	0.208767
$\widehat{H}_{\log}$	0.094409	0.028941	0.684743	0.191908

Table 4: Percent of presence in the models of each factor for 10,000 random portfolios of 10 stocks.

Factor	MKT	SMB	HML	RMW	CMA	MOM	Hurst
$\widehat{H}_k$	0.972744	0.552933	0.330231	0.525843	0.051853	0.347017	0.280372
$\widehat{H}_{\log}$	0.972114	0.554504	0.319419	0.524421	0.052222	0.355248	0.272773

Table 5: Percent of presence in the models of each factor for 10,000 random portfolios of 30 stocks.

Factor	MKT	SMB	HML	RMW	CMA	MOM	Hurst
$\widehat{H}_k$	0.990253	0.50288	0.393886	0.682322	0.100576	0.473638	0.40895
$\widehat{H}_{\log}$	0.99359	0.516026	0.409341	0.689103	0.091117	0.462454	0.391026

Table 6: Mean of r square under different parameters and Hurst exponent algorithms.

Algorithm/parameter	12 months	24 months	36 months
$\widehat{H}_k$	0.4731	0.472543	0.490673
$\widehat{H}_{\log}$	0.4729	0.471874	0.508291

Table 7: The proportion of each factor under different Hurst exponent algorithms.

Panel A: Time parameter equals 12 months							
Factor	MKT	SMB	HML	RMW	CMA	MOM	Hurst
$\widehat{H}_k$	0.938466	0.48272	0.330149	0.507165	0.04608	0.87918	0.239112
$\widehat{H}_{\log}$	0.940152	0.485249	0.329306	0.505198	0.045518	0.876651	0.220849
Panel B: Time parameter equals 24 months							
Factor	MKT	SMB	HML	RMW	CMA	MOM	Hurst
$\widehat{H}_k$	0.934807	0.518141	0.374717	0.47619	0.061791	0.617063	0.268141
$\widehat{H}_{\log}$	0.933107	0.516156	0.376701	0.474206	0.060374	0.61763	0.25085
Panel C: Time parameter equals 36 months							
Algorithm/factor	MKT	SMB	HML	RMW	CMA	MOM	Hurst
$\widehat{H}_k$	0.940908	0.536095	0.401157	0.490405	0.088943	0.524825	0.271703
$\widehat{H}_{\log}$	0.921729	0.527268	0.406329	0.494207	0.088726	0.526137	0.291325

Table 5 shows the result of 10,000 random portfolios of 30 stocks. The mean of r square of 10,000 random portfolios of 30 stocks is 0.899578 for $\widehat{H}_k$ and 0.902907 for $\widehat{H}_{\log}$. Moreover, percent of presence of factor is even higher than that in Table 4. This result enhances the conclusion of the effect of diversity.

3.4. Robustness Test. Next, we will change the parameters of the momentum factor and the Hurst exponent to make a systematic comparison and observe whether the parameter changes have a significant impact on the results. The parameters are 12 months, 24 months, and 36 months.

Table 6 shows the mean value of r square of the seven-factor model of the 2 Hurst exponent under different parameters. The mean value of r square obtained by different time parameters under the same algorithm has little change. Overall, the mean value of r square fluctuates not much when parameters vary, swinging between 47%–50%.

Table 7 summarizes the proportion of each factor in the seven-factor model under different Hurst exponent algorithms. Among all the factors, the market factor accounts for the highest proportion, maintaining at around 93%, which means that the market factor is undoubtedly the most explanatory factor in the Chinese A-share market. The MOM also accounts for a relatively high proportion, but the proportion gradually decreases as the time parameter increases, which means that the momentum effect usually has a large explanatory power in the short term, and the explanatory power gradually declines with the increase in time. SMB, HML, and RMW also have good explanatory power, and the proportions are relatively stable. The Hurst exponent gradually stabilizes with the increase in parameters, maintaining at about 27%, which indicates Hurst exponent reflects a long-term trend. MOM captures more short-term trends, while Hurst exponent captures long-term trends. The two trends often appear in the same model instead of replacing each other.

4. Conclusions

In this paper, we first examine the performance of the Fama–French five-factor model in the Chinese A-share market. Choosing the AIC criterion as the criterion for model selection, we get the average value of r square in the five-factor model, which is equal to 0.4002. The most efficient factor is the market factor, and CMA performances are the weakest.

Then, we compile two kinds of Hurst exponents and add Hurst exponent and MOM to construct a seven-factor model. When the time parameter is 12 months, the mean value of r square is about 47%, which is 7% higher than that of the five-factor model. In terms of the explanatory power of each factor, the market factor is still the strongest (about 93%), and the newly added MOM also has a strong explanatory power (about 87%), and the SMB, HML, and RMW also have a certain degree efficiency, and the weakest is still CMA. The Hurst exponent has a strong explanatory power (about 23%) and is complementary to MOM to a certain extent.

Finally, we study the sensitivity of time parameters. Set the time parameter to 12 months, 24 months, and 36 months, calculate MOM and Hurst exponent, and screen the seven-factor model. We find that the explanatory power of MOM gradually decreases with the increase in the parameters, and the Hurst exponent stabilizes at about 27% as the parameters increase, which explains that the Hurst exponent and MOM have complementary effects. MOM explains the short-term trend, while the Hurst exponent explains the long-term trend. The proportions of other factors are consistent with the previous model, and the difference between the two Hurst algorithms is not obvious.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Humanities and Social Sciences Research and Planning Fund of the Ministry of Education of China (Grant no. 20YJA630053).

References

[1] E. F. Fama and K. R. French, "The CAPM is wanted, dead or alive," *Journal of Finance*, vol. 51, no. 5, pp. 1947–1958, 1996.
[2] E. F. Fama and K. R. French, "A five-factor asset pricing model," *Journal of Financial Economics*, vol. 116, no. 1, pp. 1–22, 2015.
[3] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: implications for stock market efficiency," *The Journal of Finance*, vol. 48, no. 1, pp. 65–91, 1993.
[4] M. M. Carhart, "On persistence in mutual fund performance," *Journal of Finance*, vol. 52, no. 1, pp. 57–82, 1997.
[5] Z. Ouyang and L. Fei, "The applicability of the four-factor asset pricing model in Chinese stock market," *Journal of Financial Economics*, vol. 31, no. 2, pp. 84–96, 2016, in Chinese.
[6] H. E. Hurst, "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, vol. 116, pp. 770–808, 1951.
[7] B. Mandelbrot, "The variation of certain speculative prices," *Journal of Business*, vol. 36, no. 4, pp. 394–419, 1963.
[8] M. Couillard and M. Davison, "A comment on measuring the Hurst exponent of financial time series," *Physica A: Statistical Mechanics and Its Applications*, vol. 348, pp. 404–418, 2005.
[9] A. W. Lo, "Long-term memory in stock market prices," *Econometrica*, vol. 59, no. 5, pp. 1279–1313, 1991.
[10] M. A. Sánchez Granero, J. E. Trinidad Segovia, and J. García Pérez, "Some comments on Hurst exponent and the long memory processes on capital markets," *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 22, pp. 5543–5551, 2008.
[11] V. Teverovsky, M. S. Taqqu, and W. Willinger, "A critical look at Lo's modified R/S statistic," *Journal of Statistical Planning and Inference*, vol. 80, no. 1-2, pp. 211–227, 1999.
[12] P. M. Robinson, "Gaussian semiparametric estimation of long range dependence," *Annals of Stats*, vol. 23, no. 5, pp. 1630–1661, 1995.
[13] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Physical Review. E, Statal Physics, Plasmas, Fluids, and Related Interdiplinary Topics*, vol. 49, no. 2, pp. 1685–1689, 1994.
[14] N. K. Vitanov, N. P. Hoffmann, and B. Wernitz, "Nonlinear time series analysis of vibration data from a friction brake: SSA, PCA, and MFDFA. Chaos," *Solitons & Fractals*, vol. 69, pp. 90–99, 2014.

[15] J. Beran, *Statistics for Long-Memory Processes*, CRC Press, New York, NY, USA, 1994.
[16] R. Fox and M. S. Taqqu, "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *Annals of Statistics*, vol. 14, no. 2, pp. 517–532, 1986.
[17] R. Dahlhaus, "Efficient parameter estimation for self-similar processes," *Annals of Statistics*, vol. 17, no. 4, pp. 1749–1766, 1989.
[18] X. Guyon and J. León, "Convergence en loi des H-variations d'un processus gaussien stationnaire sur R," *Annales de l'IHP Probabilités et Statistiques*, vol. 25, no. 3, pp. 265–282, 1989.
[19] J. Istas and G. Lang, "Quadratic variations and estimation of the local holder index of a Gaussian process," *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, vol. 33, pp. 407–436, 1997.
[20] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.
[21] J. Geweke and S. Porter-Hudak, "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, vol. 4, no. 4, pp. 221–238, 1983.
[22] P. M. Robinson, "Log-periodogram regression of time series with long range dependence," *Annals of Statistics*, vol. 23, no. 3, pp. 1048–1072, 1995.
[23] C. Velasco, "Gaussian semiparametric estimation of non-stationary time series," *Journal of Time Series Analysis*, vol. 20, no. 1, pp. 87–127, 1999.
[24] P. C. Phillips and K. Shimotsu, "Local whittle estimation in nonstationary and unit root cases," *Annals of Statistics*, vol. 32, no. 2, pp. 656–692, 2004.
[25] K. Shimotsu and P. C. Phillips, "Exact local whittle estimation of fractional integration," *Annals of Statistics*, vol. 33, no. 4, pp. 1890–1933, 2005.
[26] J.-M. Bardet and I. Kammoun, "Asymptotic properties of the detrended fluctuation analysis of long-range-dependent processes," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2041–2052, 2008.
[27] K. Shimotsu, "Exact local whittle estimation of fractional integration with unknown mean and time trend," *Econometric Theory*, vol. 26, no. 2, pp. 501–540, 2010.
[28] D. Surgailis, G. Teyssiere, and M. Vaiciulis, "The increment ratio statistic," *Journal of Multivariate Analysis*, vol. 99, no. 3, pp. 510–541, 2008.
[29] J.-M. Bardet and D. Surgailis, "Measuring the roughness of random paths by increment ratios," *Bernoulli*, vol. 17, no. 2, pp. 749–780, 2011.
[30] M. N. López-García, J. E. Trinidad-Segovia, and M. Sánchez-Granero, "Extending the Fama and French model with a long term memory factor," *European Journal of Operational Research*, vol. 291, no. 2, pp. 421–426, 2021.
[31] C. Berzin, A. Latour, and J. R. León, *Inference on the Hurst Parameter and the Variance of Diffusions Driven by Fractional Brownian Motion*, Springer, Cham, Switzerland, 2014.
[32] E. F. Fama and K. R. French, "Common risk factors in returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56, 1993.
[33] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.
[34] B. B. Mandelbrot and J. R. Wallis, "Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence," *Water Resources Research*, vol. 5, no. 5, pp. 967–988, 1969.
[35] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 1475781, 18 pages
https://doi.org/10.1155/2021/1475781

Research Article

Research on the Operating Efficiency of Chinese Listed Pharmaceutical Companies Based on Two-Stage Network DEA and Malmquist

Tsung-Xian Lin,[1] **Zhong-huan Wu**,[1] **Xiao-xia Ji**,[2] **and Jia-jia Yang**[3]

[1]*Department of Business Administration, Guangzhou Huashang College, Guangzhou 511300, China*
[2]*School of Business, Guangdong University of Foreign Studies, Guangzhou 510000, China*
[3]*Department of Public Policy, King's College London, London WC2R 2LS, UK*

Correspondence should be addressed to Xiao-xia Ji; 976224056@qq.com

Received 20 May 2021; Accepted 16 June 2021; Published 23 June 2021

Academic Editor: Yiwen Zhang

The development of pharmaceutical companies, which is an important part of the national economy and industry, is closely related to people's livelihood issues. With the era of big data, this paper uses the two-stage DEA and Malmquist method to evaluate the efficiency of listed Chinese pharmaceutical companies. From a static and dynamic perspective, it analyses the total factor productivity index, pure efficiency change index, scale efficiency index, and so on. The results show that government subsidies have not had a positive impact on most Chinese pharmaceutical companies, and for films, diseconomies of scale caused by rapid expansion should be avoided.

1. Introduction

Since the outbreak of the COVID-19, medical health has become the primary topic of concern for the government and the public. The operation of pharmaceutical companies has become the focus of discussion. At present, China's pharmaceutical manufacturing industry has long been in a state of low industry concentration, high product homogeneity, and weak technological innovation capabilities. It has a large gap with the international advanced level. In recent years, China's various medical reform policies and regulations have been promulgated one after another. It has brought many uncertain factors to the development of the pharmaceutical industry. How to adapt to changes in the external environment and further improve operating efficiency has become an important issue facing the government and enterprises. Based on the above background, this article focuses on the research on the operating efficiency of pharmaceutical companies, innovatively combining the DEA method with Malmquist, taking 164 listed companies in China as a sample, and studying their operating conditions during the five-year period from 2015 to 2019.

The structure of the following parts of this article is as follows. Section 2 collects current scholars' research on the efficiency evaluation of pharmaceutical companies; Section 3 introduces the related content of the two-stage network DEA and Malmquist index method in this article; Section 4 outlines the research of this article process, sample selection, and two-stage DEA data analysis. Section 5 uses Malmquist to analyse the dynamic effects of the data and finally summarizes the research results of this article.

2. Literature Review

At present, many scholars use the DEA method to measure the actual situation of the operating efficiency of pharmaceutical companies and provide suggestions for improvement of the company's operations. For example, Wang [1] selected Chinese biopharmaceutical companies from 2017 to 2019 as a research sample and established a static DEA model to measure their financing efficiency. The results show that although the overall financing efficiency of Chinese biotech companies is not high, the level of it is increasing year by year. And Zhang Zicheng [2] innovatively combined the AHP method with the DEA method, finding that, compared with the scale factor, the deficiency of technology hinders the operating performance of companies such as Lunan Pharmaceutical firm. Meanwhile, Li et al. [3] used a two-step method of factor analysis and SE-DEA model to calculate the financial data of 58 listed pharmaceutical companies in China from 2009 to 2013 and concluded that the overall inefficiency of the pharmaceutical industry is also due to insufficient investment on scale and technology.

Among them, some scholars even divide medical companies into groups to study their operational status in different regions [4, 5]. The research found that there are indeed differences in the operating efficiency as well as in terms of technological innovation of pharmaceutical companies in different regions [6, 7]. For example, Xiong [8] used the panel data of technological innovation of medical companies listed in Guangdong, Shandong, Zhejiang, and Jiangsu as samples to study the allocation of technological innovation resources of pharmaceutical manufacturers in four provinces in China from 2015 to 2017 and finally came to the following conclusions: in the province, Jiangsu pharmaceutical companies have the advantage of pure technical efficiency in innovation activities, while Guangdong pharmaceutical companies performed better in scale efficiency with regard to technological innovation. However, the traditional DEA model cannot study the influencing factors in the process. Therefore, the research studies on the two-stage network DEA model and the Malmquist index method have received widespread attention from many scholars.

Regarding the two-stage network DEA, such method has been applied to plenty of fields. For example, Lewis et al. applied the undirected network DEA method to the efficiency evaluation of Major League Baseball [9]. Additionally, Liang et al. used a two-stage network DEA model to analyse the input-output efficiency of 50 universities in China [10], while Li et al. applied the DEA model under the two-stage expansion structure to the research on the efficiency of R&D in China's provincial regions [11]. At the same time, some researchers also combine two-stage network DEA with other methods. For instance, Chen et al. [12] and Kao [13] combined it with the two-stage additive efficiency decomposition DEA method to study the relative efficiency of 24 non-life insurance companies in Taiwan. Lee and Johnson combined Malmquist under the network DEA structure to study the performance of the semiconductor manufacturing industry [14]. It can be seen that the two-stage network DEA has been widely used in insurance companies, universities, and other industry research. However, research in the pharmaceutical industry is still relatively rare.

Regarding the Malmquist index method, scholars have also made great achievements. In the field of sustainable e-agriculture, Pan and others used the 31 provinces as the research objects and explored the sustainable development efficiency of agriculture in mainland China through DEA and Malmquist productivity index models [15]. In the medical industry, Hashimoto and Haneda [16] used the conventional DEA method and the Malmquist productivity index method to measure the R&D efficiency of the Japanese pharmaceutical industry from the enterprise level and the industry level, respectively. Empirical evidence shows that the total factor productivity of Japan's pharmaceutical industry is declining, and the main reason for the decline is the sharp decrease in technological changes. What is more, Pannu et al. [17] used the output-oriented VRS model and the Malmquist productivity index method to measure the increase in efficiency and productivity of the Indian pharmaceutical industry over a 10-year period, finding that the increase was mainly due to the growths in technical efficiency. Furthermore, Zhiyue and Qiu [18] also used the Malmquist index method to conduct an empirical analysis of the operating efficiency of China's biopharmaceutical industry from both horizontal and vertical aspects. The results show that the overall operating efficiency of the biopharmaceutical industry is not ideal, and there is a large difference in efficiency between provinces and cities.

In summary, it can be seen that scholars have used many different methods to study the operating efficiency of pharmaceutical manufacturing enterprises, but the research still has the following shortcomings. Firstly, most research studies on the efficiency of pharmaceutical manufacturing enterprises use nonparametric methods. When measuring enterprise efficiency, some scholars only consider a certain aspect of static or dynamic and thus cannot comprehensively analyse the efficiency level and development trend of pharmaceutical manufacturing enterprises. Secondly, there are few literatures on the research of listed pharmaceutical companies using the two-stage network DEA and Malmquist index method, most of which focus on the traditional DEA method. Finally, in the literature on efficiency influencing factors, the selection of variables is relatively limited, and there are few literatures that consider the R&D capabilities of enterprises. For pharmaceutical manufacturing companies, environmental variables are very important and have a very large impact on the efficiency of the company. Therefore, the external environment of the company should be considered when studying the efficiency of the company. Based on the above deficiencies, this paper uses the two-stage network DEA and Malmquist index method to study the operating efficiency of enterprises from both static and dynamic perspectives. When studying the factors affecting the operating efficiency of enterprises, environmental variables have been added and considered from multiple angles in the article, striving for a more comprehensive selection of influencing factors.

3. Research Method

3.1. Two-Stage Network DEA. In the traditional DEA model, we only know the final efficiency values of the entire process, but the specific situation in the whole process is unknown. The information provided by the traditional DEA model is not enough, and the guidance to managers is limited. The two-stage network DEA model can open the "black box" of the production system, which can effectively measure the complex production network. Therefore, this paper also adopts the two-stage network DEA model for performance evaluation and pays more attention to the progressive relationship between the two stages based on the research results of other scholars. Its internal structure is shown in Figure 1.

Among them, $X_{ij}^1 (i = 1, 2, \ldots, I)$ represents the i-th input of DMU_j in the first stage; Z_{dj} $(d = 1, 2, \ldots, D)$ represents the intermediate variable, namely, it is not only the d-th output of DMU_j in the first stage, but also the d-th input of DMU_j in the second stage; $X_{kj}^2 (k = 1, 2, \ldots, K)$ represents the k-th input of the newly added DMU_j in the second stage; and $y_{rj} (r = 1, 2, \ldots, R)$ represents the r-th output of DMU_j in the second stage. First of all, calculate the efficiency of the first stage and then calculate the efficiency of the second stage on this basis, which means keeping the efficiency of the first stage unchanged. Finally, the product of the efficiency of the two stages is regarded as the total efficiency of the system. At this point, the model can be established as follows:

The first-stage model (model 1) is given by

$$\mathrm{Max} \quad \frac{\sum_{d=1}^{D} W_d Z_{d0}}{\sum_{i=1}^{I} V_i X_{i0}^1}, \tag{1}$$

$$\mathrm{s.t.} \frac{\sum_{d=1}^{D} W_d Z_{dj}}{\sum_{i=1}^{I} V_i X_{ij}^1} \le 1, \quad j = 1, 2, \ldots, n, \tag{2}$$

$$\frac{\sum_{r=1}^{R} U_r Y_{rj}}{\sum_{d=1}^{D} z_{dj} + \sum_{k=1}^{k} t_k x_{kj}^2} \le 1, \quad j = 1, 2, \ldots, n, \tag{3}$$

$$U_r, V_i, W_d, t_k \ge 0, \quad \forall r, i, d, k. \tag{4}$$

Model 1 adds a constraint on the basis of the traditional CCR model, that is, the last constraint. Its purpose is to ensure that the optimal solution of the first stage makes the efficiency value of the second stage not more than 1, so as to ensure that the second stage model must have a feasible solution; otherwise, there may be no feasible solution. Therefore, this constraint is necessary, which was not considered in the previous two-stage DEA model. In model 1, U_r, V_i, W_d, t_k, respectively, represent the weights of the corresponding variables, after considering the study conducted by Cheng and Zheng [19]. The efficiency of each DMU can be obtained by model 1. Record the efficiency of the DMU_0 as $\theta_1^{0\,*}$.

Then second-stage model (model 2) is given by

$$\mathrm{Max} \quad \frac{\sum_{r=1}^{R} U_r Y_{r0}}{\sum_{d=1}^{D} W_d Z_{d0} + \sum_{k=1}^{k} t_k x_{k0}^2}, \tag{5}$$

$$\mathrm{s.t} \frac{\sum_{d=1}^{D} W_d Z_{dj}}{\sum_{i=1}^{I} V_i X_{ij}^1} \le 1, \ j = 1, 2, \ldots, n, \tag{6}$$

$$\frac{\sum_{r=1}^{R} U_r Y_{rj}}{\sum_{d=1}^{D} W_d Z_{dj} + \sum_{k=1}^{k} t_k x_{kj}^2} \le 1, \quad j = 1, 2, \ldots, n, \tag{7}$$

$$\frac{\sum_{d=1}^{D} W_d Z_{d0}}{\sum_{i=1}^{I} V_i X_{i0}^1} = \theta_1^{0\,*}, \tag{8}$$

$$U_r, V_i, W_d, t_k \ge 0, \quad \forall r, i, d, k. \tag{9}$$

Solving (5)–(9) can get the efficiency of the second-stage DMU, noting $\theta_2^{0\,*}$ as the efficiency of the second stage. So far, it can be concluded that the total efficiency of the two-stage system DMU_0 is $\theta^{0*} = \theta_1^{0\,*} \theta_2^{0\,*}$ (10).

3.2. Malmquist Index. The two-stage network DEA model just horizontally compared the efficiency of listed pharmaceutical enterprises. So we will build the Malmquist index model to make a longitudinal analysis of efficiency and dynamically analyse the change of efficiency.

TFP is total factor productivity index:

$$\mathrm{TFP} = \frac{D_t^u (x_t, y_t)}{D_{t-1}^u (x_{t-1}, y_{t-1})} \times \left[\frac{\left(D_t (x_t, y_t) / D_{t-1}^C (x_{t-1}, y_{t-1}) \right)}{\left(D_t^u (x_t, y_t) / D_{t-1}^u (x_{t-1}, y_{t-1}) \right)} \right] \times \left[\frac{D_{t-1}^C (x_{t-1}, y_{t-1})}{D_t^C (x_{t-1}, y_{t-1})} \times \frac{D_{t-1}^C (x_t, y_t)}{D_t^C (x_t, y_t)} \right]^{(1/2)}. \tag{10}$$

PEC is pure efficiency change index:

$$\mathrm{PEC} = \frac{D_t^u (x_t, y_t)}{D_{t-1}^u (x_{t-1}, y_{t-1})}, \tag{11}$$

SE is scale efficiency index:

$$\mathrm{SE} = \frac{\left(D_t^C (x_t, y_t) / D_{t-1}^C (x_{t-1}, y_{t-1}) \right)}{\left(D_t^4 (x_t, y_t) / D_{t-1}^u (x_t - 1, yt - 1) \right)}, \tag{12}$$

TC is technical change index:

$$\mathrm{TC} = \left[\frac{D_{t-1}^C (x_{t-1}, y_{t-1})}{D_t^C (x_{t-1}, y_{t-1})} \times \frac{D_{t-1}^C (x_t, y_t)}{D_t^C (x_t, y_t)} \right]^{(1/2)}, \tag{13}$$

The formula of total factor productivity is

$$\mathrm{TFP} = \mathrm{PEC} \times \mathrm{SE} \times \mathrm{TC}. \tag{14}$$

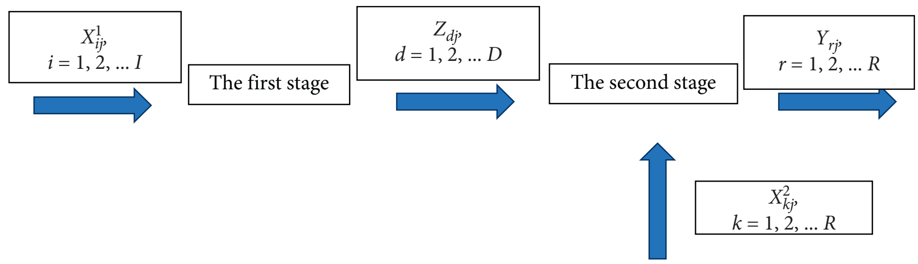


Figure 1: The two-stage network DEA model.

Table 1: Variable selection and measurement.

	Variable type	Variable name	Measure
The first stage	Output	Gross costs	Gross costs in corporate annual reports
	Input	Total number of employees	Total number of in-service staff
		Net value of fixed assets	Net value of fixed assets in corporate annual reports
		Gross revenue	Gross revenue in corporate annual reports
The second stage	Output	Net profit	Total profit − income tax
	Input	Gross costs	—
		Government subsidies	Government subsidies that are included in the current profit and loss

4. Empirical Analysis

4.1. Sample Selection. With reference to the definition of pharmaceutical companies, combined with the description of the main business in the annual report of the A-share listed pharmaceutical company and the proportion of the main business in operating income, the study sample is determined. At the same time, to ensure the validity of the analysis results, ST, PT, and ST* companies were excluded, and 164 listed pharmaceutical companies were finally identified as the research samples. The data of inputs and outputs comes from Cathay Pacific database and the annual public report of enterprise.

4.2. Variable Selection

4.2.1. The First Stage of Input and Output Variables. The selection of variables is based on character of listed pharmaceutical companies, and we fully inspect the business characteristics and operation of listed pharmaceutical companies. The inputs selected are gross costs (X_1), total number of employees (X_2), and net value of fixed assets (X_3), and the output is gross revenue (Y_1).

4.2.2. The Second Phase of Input and Output Variables. This paper comprehensively examines the business characteristics and operation of listed pharmaceutical companies. Government subsidies (X_5) as a new input are added to the second stage, and the other input is the gross revenue (X_4) that is the output of first stage, and the final output is net profit (Y_2). More details are shown in Table 1.

4.3. DEA Efficiency Analysis in the First Stage. In first stage (shown in Table 2), no listed pharmaceutical company's technical efficiency reached 1 during 2015–2019. There are 119 companies with technical efficiency between 0.6 and 1.0, accounting for 73%, indicating that the technical efficiency of these companies is good. The efficiency values of the eight securities companies (Chongqing Taiji Industry, Baiyunshan, Kelun, Huabei, Hisun, Harbin, China Medicine, and Renfu) are below 0.4, indicating that the technical efficiency level of these companies is relatively low, and they need to increase investment.

Five securities companies, Adisseo, Changchun High-Tech, Hualan Biological Engineering, Hengrui, and Zhifei, have achieved technology effectiveness in some years. Adisseo's technical efficiency is effective in 2016 and has dropped significantly after 2016, and the technical efficiency can be improved by referring to the operation method of 2016 when the technology is effective (see Table 3 for specific annual data).

4.4. DEA Efficiency Analysis in the Second Stage. In the second stage (shown in Table 2), there is no company that has reached technical efficiency of 1 during 2015–2019, indicating that all the companies were not effective. There are 97 companies with a technical efficiency value of 0.6–1.0, accounting for 59.5%, which is significantly lower than the first stage. Only 3 companies had technical efficiency below 0.4 (Sinopharm, Xinbang, and Baiyunshan). The technical efficiency in second stage is generally low, but companies with lower efficiency have been promoted, which may be correlated with government subsidies.

Adisseo, Dezhan Health, Jiao, Hengrui, Livzon Group, Shanghai RAAS, and Zhifei Biotechnology have achieved technical efficiency of 1 in some years. And Adisseo is the same as the first stage, the technical efficiency is effective in 2016 and has dropped dramatically after 2016. The technical efficiency of Dezhan Health and Jiao pharmaceutical companies in the second stage has increased; we believe that the first stage is relevant to the second stage in these companies. However, the average technical efficiency of Hualan

Table 2: Summary of means of technical efficiency.

Company name	The first stage	The second stage	Total
Adisseo	0.731807447	0.70427645	0.718041949
Anke Bio	0.701258545	0.65681619	0.679037367
Osaikon	0.77517144	0.677756486	0.726463963
Baiyunshan	0.383760611	0.329568568	0.35666459
Bdyy	0.621063253	0.668172657	0.644617955
Beilu	0.755456729	0.662521895	0.708989312
Porton	0.626443198	0.653894142	0.64016867
None	0.708205489	0.644792119	0.676498804
Changchun High-Tech	0.754140117	0.760920762	0.757530439
Changjiang Runfa Medicine	0.622444299	0.581956694	0.602200497
Changshan Pharma	0.705379957	0.663215003	0.68429748
DAJY	0.611416622	0.570090491	0.590753556
Dezhan	0.800138709	0.776221753	0.788180231
Jiao	0.68813071	0.761262056	0.724696383
DBBT	0.715738024	0.641593159	0.678665591
VC	0.463121349	0.46703195	0.465076649
Dongcheng	0.67384118	0.680702372	0.677271776
Nhwa Pharm	0.655734649	0.667349718	0.661542183
Ekzy	0.679041427	0.578708106	0.628874767
Fangsheng	0.700194471	0.656633368	0.678413919
Fengyuan	0.58132929	0.49193138	0.536630335
Fczy	0.678039919	0.651980945	0.665010432
Fayy	0.64776847	0.593446924	0.620607697
Fosun Pharm	0.469682442	0.744647296	0.607164869
Fuxiang	0.699215567	0.600348021	0.649781794
Guangji	0.686416693	0.640838606	0.66362765
Kwong Sang Hong	0.739991067	0.650835677	0.695413372
Guang Yuyuan	0.697803931	0.643573172	0.670688552
Lark	0.630381464	0.657746723	0.644064093
Glsj	0.677120987	0.682796241	0.679958614
Sinopharm Hyundai	0.496221436	0.397513751	0.446867593
Harbin Pharm	0.34428928	0.475791143	0.410040211
Haili Bio	0.727883095	0.652682123	0.690282609
Hnhy	0.609983775	0.528580996	0.569282386
Hepalink	0.64389566	0.594403859	0.619149759
Haishun New Pharma	0.750928424	0.661675715	0.706302069
Haisco	0.720451696	0.487743987	0.604097841
Hisoar	0.645544623	0.65460421	0.650074417
Haixin	0.627599087	0.672097746	0.649848417
Hisun	0.353460494	0.409015554	0.381238024
Han Sen Pharm	0.668777752	0.663899646	0.666338699
Hybio	0.636033086	0.580810237	0.608421662
Hengrui	0.837490527	0.961302739	0.899396633
Chase Sun	0.547733919	0.410016	0.478874959
NCPC	0.36259657	0.447801928	0.405199249
Huahai	0.543995409	0.553533665	0.548764537
Hualan Bio	0.807931213	0.595846914	0.701889064
Huaren	0.595328472	0.659644838	0.627486655
HRSJ	0.541774861	0.574425738	0.558100299
China Resources Double-Crane	0.568038828	0.54622275	0.557130789
Huashen Technology	0.719199418	0.674035569	0.696617493
Walter Dyne	0.664596536	0.687196219	0.675896377
Yanbian FC	0.778142258	0.63590963	0.707025944
Kyrgyzstan	0.601616926	0.496027976	0.548822451
Jichuan	0.683807204	0.487854907	0.585831056
Jimin	0.681185309	0.657322504	0.669253906
JYPC	0.685562921	0.649641211	0.667602066
Joincare pharm	0.610082457	0.454142296	0.532112376
Jiangzhong	0.642567847	0.736567664	0.689567756
Jincheng	0.639785389	0.569178368	0.604481878

Table 2: Continued.

Company name	The first stage	The second stage	Total
Jinhe Bio	0.623354577	0.649028661	0.636191619
Jinling	0.546235719	0.635576844	0.590906281
Jinshiya	0.715796667	0.644757882	0.680277274
Jingxin	0.68571291	0.570388142	0.628050526
Jinghua	0.636944978	0.641649287	0.639297132
Jingfeng	0.606488959	0.543530495	0.575009727
Jiuqiang	0.786956228	0.700046207	0.743501218
Jiuzhitang	0.645201026	0.516603024	0.580902025
Jiuzhou	0.628567115	0.584155422	0.606361268
CONBA	0.543861375	0.481004348	0.512432862
Kanghong	0.665245011	0.527460836	0.596352924
Kangyuan	0.61414204	0.577926992	0.596034516
Kangzhi	0.684816087	0.656415871	0.670615979
KHB	0.727976507	0.591841298	0.659908903
Kelun	0.378799441	0.493231698	0.43601557
Sunflower	0.556170097	0.511400244	0.533785171
Kunming Pharm	0.534324416	0.551964191	0.543144303
Lummy	0.625056228	0.628112755	0.626584491
LAYN	0.721614484	0.661566479	0.691590481
Lisheng Pharma	0.657353855	0.679177139	0.668265497
Livzon Pharm	0.71738134	0.628871094	0.673126217
LEADMAN	0.714944289	0.624498345	0.669721317
Lianhuan pharm	0.698888411	0.670367118	0.684627765
Lingkang	0.725743874	0.584694518	0.655219196
Lingrui Pharm	0.666997747	0.677070779	0.672034263
Long jin Pharm	0.753570272	0.647133166	0.700351719
Lukang Pharm	0.509027638	0.460965245	0.484996442
Mike Bio	0.742422513	0.650407694	0.696415103
M.k.	0.654220437	0.562356283	0.60828836
Palin Bio	0.703269836	0.65109186	0.677180848
PIEN TZE HUANG	0.763824558	0.726005071	0.744914814
Julie Plec	0.709624166	0.469789228	0.589706697
plyy	0.489613809	0.53054052	0.510077164
CHEEZHENGTTM	0.714689814	0.613582646	0.66413623
Qidi	0.691053939	0.654977334	0.673015636
Qianhong Biopharma	0.70228214	0.679203457	0.690742799
Qianjin Pharm	0.612647181	0.607590824	0.610119003
Qianyuan	0.664973414	0.556845497	0.610909455
Renfu	0.322406425	0.42016555	0.371285988
Renhe Pharmacy	0.596351472	0.672679619	0.634515546
rpsw	0.651166925	0.5912946	0.621230762
Saisheng	0.761947778	0.701836844	0.731892311
SAM	0.671926336	0.659700513	0.665813424
Shanhe Pharmacy	0.734791349	0.650221186	0.692506268
Shkb	0.698069521	0.680387876	0.689228698
Shanghai RAAS Blood Products	0.676407378	0.605561928	0.640984653
Shenqi	0.646589105	0.581969742	0.614279423
Biological Stock	0.760202981	0.756042924	0.758122952
Salvage Pharm	0.463144872	0.502413614	0.482779243
Sts	0.739781567	0.689149244	0.714465406
Scyy	0.699966811	0.603380282	0.651673546
Beijing SL Pharm	0.773760654	0.68883442	0.731297537
Stellite	0.667760149	0.642702048	0.655231099
Shsw	0.73692057	0.657203864	0.697062217
Tat	0.60265023	0.544697259	0.573673745
Taiji Group	0.399279671	0.498518625	0.448899148
Taloph Pharm	0.61924255	0.644434623	0.631838587
Teyi	0.698667383	0.66716684	0.682917112
Tasly	0.41666853	0.630275487	0.523472009
Tiantan Biological	0.733010668	0.657600328	0.695305498

Table 2: Continued.

Company name	The first stage	The second stage	Total
Tianyao Pharm	0.646240088	0.613143357	0.629691722
Thdb	0.77883326	0.706539764	0.742686512
Thjm	0.567809163	0.564386742	0.566097953
TRT	0.499628521	0.691621404	0.595624963
Wanbangde	0.427406852	0.537992679	0.482699766
Wondfo	0.715009555	0.631798868	0.673404211
WEDGE INDUSTRIAL	0.736184308	0.643355439	0.689769873
Weiming	0.711826016	0.650321829	0.681073922
Wowu	0.770567021	0.675840362	0.723203692
Wohua	0.703903392	0.661188132	0.682545762
Wosen	0.671912477	0.552904143	0.61240831
AMD	0.750945781	0.61052309	0.680734436
Xianju Pharm	0.609803878	0.607517262	0.60866057
Xiangxue Pharm	0.56837251	0.537740093	0.553056301
Sunflower	0.55841059	0.566683692	0.562547141
NHU	0.632654979	0.4122508	0.52245289
Xinhua	0.488947377	0.556767052	0.522857214
Xinbang	0.440203247	0.395079286	0.417641267
SALUBRIS	0.715240078	0.63026292	0.672751499
BROTHER	0.697594871	0.621553604	0.659574237
Yabao	0.558060463	0.567098443	0.562579453
Yatai	0.59468608	0.568673089	0.581679584
Yananbikang	0.585080092	0.503340182	0.544210137
Yiling Pharm	0.591814993	0.635212897	0.613513945
Yifan	0.659880029	0.609050879	0.634465454
Yibai	0.564093323	0.437330699	0.500712011
Yisheng	0.661646927	0.603737031	0.632691979
Yiduoli	0.648099384	0.597623845	0.622861614
Chinataurine	0.692881108	0.647370414	0.670125761
Gloria Pharm	0.479739963	0.433218753	0.456479358
Baiyao	0.50362558	0.819177469	0.661401524
Zhejiang Medicine	0.500873109	0.515343568	0.508108338
Zhenbao Island	0.66468048	0.494783002	0.579731741
zdzy	0.568909321	0.561712443	0.565310882
Zhifei	0.821043844	0.783585599	0.802314721
Zhongguancun	0.603226518	0.546917339	0.575071929
China Medicine	0.326022175	0.619902235	0.472962205
Zhongheng Group	0.699336493	0.716431982	0.707884238
Zhongmu	0.534061049	0.625219552	0.579640301
Zhongxin	0.523656226	0.64705675	0.585356488
Zsyy	0.728308394	0.611906374	0.670107384
JLZX	0.70032643	0.554290101	0.627308266
Zuoli	0.658487779	0.492143686	0.575315732

Biological is below 0.6 in the two stages, for the resources cannot be well utilized (see Table 4 for specific annual data).

4.5. Overall Efficiency Analysis. In the overall efficiency analysis (shown in Table 2), there are 109 companies with efficiency between 0.6 and 1.0, accounting for 66.9%, and it shows that the technical efficiency of the second stage is less than that of the first stage.

Comparing Hengrui (the highest efficiency) and Baiyunshan (the lowest efficiency), we found that Hengrui did not receive government subsidies in the 2018 and 2019, but the technical efficiency reached 1, and Baiyunshan has received government subsidies, but the technical efficiency rises first and then decreases.

For Zhifei Bio with the second highest efficiency, its efficiency in 2018 and 2019 was significantly higher than in 2016 and 2017, and the government subsidies received by Zhifei Bio in 2018 and 2019 were significantly lower than before. The second-to-last-ranked company, Medicare, reached a low point in 2018, followed by a significant rebound next year, when it was not subsidized by the government in 2019.

It can be concluded that government subsidies have no obvious effect for most companies, but it has a positive impact on enterprises with low efficiency in a short term. The government may need to reconsider the way of subsidies to pharmaceutical companies, such as the capital subsidies to equipment upgrades and talent introduction (see Table 5 for specific annual data).

Table 3: Analysis of DEA efficiency in the first stage of Chinese pharmaceutical companies.

Company name	Year			
	2015-2016	2016-2017	2017-2018	2018-2019
Adisseo	1	0.538826377	0.748769622	0.639633789
Anke Bio	0.69893892	0.731620495	0.671113197	0.703361568
Osaikon	0.713077361	0.75673785	0.813339088	0.817531461
Baiyunshan	0.328542111	0.356320523	0.546507045	0.303672764
Bdyy	0.589761232	0.62883886	0.592320492	0.673332429
Beilu	0.701152973	0.769964623	0.718534493	0.832174827
Porton	0.617509415	0.622015319	0.606350945	0.659897114
None	0.709527587	0.749839674	0.680641517	0.692813179
Changchun High-Tech	0.644238901	0.693875856	0.678445711	1
Changjiang Runfa Medicine	0.661073921	0.633858532	0.577628241	0.617216502
Changshan Pharma	0.702574435	0.737444456	0.654896914	0.726604025
DAJY	0.608034181	0.612562673	0.565649446	0.659420187
Dezhan	0.778248014	0.8859713	0.801172953	0.735162569
Jiao	0.695876014	0.752555196	0.784627533	0.519464099
DBBT	0.698427402	0.732910456	0.683799871	0.747814366
VC	0.356746951	0.369605175	0.596620054	0.529513216
Dongcheng	0.685298529	0.713356357	0.65224054	0.644469293
Nhwa Pharm	0.566050744	0.595888879	0.559629612	0.90136936
Ekzy	0.695893135	0.642142688	0.56111062	0.817019263
Fangsheng	0.690430635	0.7072366	0.668234842	0.734875807
Fengyuan	0.507954055	0.513382048	0.500312529	0.803668528
Fczy	0.673318722	0.706559734	0.637287549	0.694993671
Fayy	0.640199162	0.622882581	0.47487985	0.853112289
Fosun Pharm	0.440829161	0.453177832	0.491647136	0.49307564
Fuxiang	0.69509172	0.713319861	0.655317677	0.733133012
Guangji	0.665875845	0.685465131	0.666509377	0.727816419
Kwong Sang Hong	0.731211949	0.758118876	0.703598927	0.767034517
Guang Yuyuan	0.705727447	0.728802756	0.672683165	0.684002357
Lark	0.592069879	0.607690808	0.548144601	0.77362057
Glsj	0.646076312	0.70291605	0.642642634	0.716848951
Sinopharm Hyundai	0.550500314	0.31927442	0.586702033	0.528408978
Harbin Pharm	0.232904004	0.239750504	0.452243584	0.452259029
Haili Bio	0.720958825	0.755724092	0.683383826	0.751465636
Hnhy	0.593723249	0.615663681	0.533902785	0.696645384
Hepalink	0.613624689	0.594913319	0.565038631	0.802005999
Haishun New Pharma	0.730858479	0.770599806	0.71368573	0.788569679
Haisco	0.694950719	0.689640795	0.627724818	0.86949045
Hisoar	0.536160514	0.598345145	0.58053481	0.867138023
Haixin	0.598921852	0.651676624	0.60413451	0.655663364
Hisun	0.279462381	0.291252982	0.396970805	0.446155807
Han Sen Pharm	0.622641145	0.670338249	0.638264052	0.743867561
Hybio	0.706990021	0.762095736	0.569815841	0.505230747
Hengrui	0.659908876	0.690053233	1	1
Chase Sun	0.590616729	0.546391865	0.485056058	0.568871023
NCPC	0.261478601	0.273294764	0.476531644	0.439081271
Huahai	0.510939772	0.532978902	0.422450272	0.70961269
Hualan Bio	0.724489053	0.75903868	0.74819712	1
Huaren	0.547811923	0.60655341	0.573018429	0.653930126
HRSJ	0.421806748	0.424780128	0.642346121	0.678166446
China Resources Double-Crane	0.396338204	0.412017952	0.746295359	0.717503796
Huashen Technology	0.692282897	0.748644224	0.673121394	0.762749156
Walter Dyne	0.64905781	0.687320156	0.604331968	0.717676209
Yanbian FC	0.728434735	0.796358227	0.622973596	0.964802474
Kyrgyzstan	0.690648563	0.723721435	0.676558269	0.315539437
Jichuan	0.654593403	0.687752464	0.658917181	0.733965769
Jimin	0.667207312	0.715203145	0.639062034	0.703268743
JYPC	0.676163429	0.660551878	0.674564338	0.730972038
Joincare Pharm	0.42609937	0.779556709	0.42013106	0.814542688
Jiangzhong	0.577983092	0.660234497	0.621771461	0.710282338

Table 3: Continued.

Company name	Year 2015-2016	2016-2017	2017-2018	2018-2019
Jincheng	0.604349975	0.635628024	0.540328079	0.778835478
Jinhe Bio	0.620488457	0.622315827	0.593163862	0.657450163
Jinling	0.457693647	0.464485434	0.485300003	0.77746379
Jinshiya	0.739579461	0.794884842	0.642073972	0.686648392
Jingxin	0.615013047	0.645371162	0.600849303	0.881618128
Jinghua	0.654292046	0.695723085	0.644925134	0.552839647
Jingfeng	0.655470879	0.625155781	0.56168705	0.583642128
Jiuqiang	0.775537026	0.80576769	0.741353597	0.825166597
Jiuzhitang	0.638432251	0.632345693	0.526816311	0.783209851
Jiuzhou	0.531129929	0.563082683	0.553227164	0.866828683
CONBA	0.424959572	0.448091924	0.76318362	0.539210385
Kanghong	0.649678902	0.676913482	0.617801313	0.716586348
Kangyuan	0.547173659	0.538412405	0.521974436	0.84900766
Kangzhi	0.681459875	0.718856205	0.658583409	0.68036486
KHB	0.679909201	0.70601526	0.646219781	0.879761787
Kelun	0.309440068	0.306632161	0.481370093	0.417755441
Sunflower	0.503468625	0.542712714	0.541164765	0.637334286
Kunming Pharm	0.487752766	0.494592469	0.456784207	0.698168222
Lummy	0.59477378	0.650148325	0.620139476	0.63516333
LAYN	0.706666281	0.764250345	0.658484063	0.757057246
Lisheng Pharma	0.652358213	0.658225585	0.633802648	0.685028972
Livzon Pharm	0.483127888	0.90147395	0.765854345	0.719069177
LEADMAN	0.685277868	0.739693609	0.685479669	0.749326011
Lianhuan Pharm	0.672599294	0.711757269	0.670159541	0.741037539
Lingkang	0.698902905	0.743402231	0.696586279	0.764084079
Lingrui Pharm	0.674401636	0.672389024	0.614906134	0.706294196
Long Jin Pharm	0.731845112	0.770452444	0.718767022	0.793216512
Lukang Pharm	0.427468486	0.46570682	0.460332769	0.682602477
Mike Bio	0.708027483	0.72319852	0.645149742	0.893314305
M.k.	0.708395924	0.726482926	0.56873759	0.613265307
Palin Bio	0.671391314	0.718181986	0.666425242	0.757080803
PIEN TZE HUANG	0.684930611	0.750405659	0.661039196	0.958922766
Julie Plec	0.707704329	0.723714045	0.673462726	0.733615566
Plyy	0.406959748	0.414853778	0.427735839	0.708905869
CHEEZHENGTTM	0.686682951	0.725135357	0.678237865	0.768703082
Qidi	0.666806547	0.68236924	0.64498398	0.770055989
Qianhong Biopharma	0.71938682	0.719385025	0.656560098	0.713796618
Qianjin Pharm	0.530656791	0.548629069	0.522713754	0.848589111
Qianyuan	0.638506467	0.671543368	0.62791026	0.721933561
Renfu	0.334218803	0.413272744	0.149231487	0.392902669
Renhe Pharmacy	0.533268479	0.506914813	0.519999497	0.8252231
Rpsw	0.638039028	0.654974398	0.617845497	0.693808777
Saisheng	0.754473392	0.796748489	0.72770648	0.76886275
SAM	0.680909431	0.689598358	0.594020614	0.72317694
Shanhe Pharmacy	0.713262981	0.747755843	0.701291618	0.776854953
Shkb	0.690613654	0.721026937	0.655361928	0.725275566
Shanghai RAAS Blood Products	0.815937216	0.711632966	0.29512233	0.882937
Shenqi	0.646515892	0.665122901	0.604905717	0.669811909
Biological Stock	0.776300364	0.857731961	0.733310511	0.673469086
Salvage Pharm	0.405423981	0.471147411	0.570580408	0.405427689
Sts	0.745342868	0.789809576	0.695943443	0.728030381
Scyy	0.621324992	0.736624764	0.67014276	0.771774728
Beijing SL Pharm	0.754959569	0.822479251	0.739066637	0.778537157
Stellite	0.64699861	0.679791002	0.648708242	0.695542742
Shsw	0.706490688	0.749047972	0.701812101	0.790331518
Tat	0.583302035	0.579802833	0.530612746	0.716883304
Taiji Group	0.368697953	0.288373447	0.505506669	0.434540617
Taloph Pharm	0.60505421	0.644707882	0.569453781	0.657754327
Teyi	0.685639388	0.710137044	0.667309716	0.731583384

Table 3: Continued.

Company name	Year			
	2015-2016	2016-2017	2017-2018	2018-2019
Tasly	0.344736323	0.364398246	0.547507961	0.410031592
Tiantan Biological	0.56685493	0.735139685	0.705597237	0.924450822
Tianyao Pharm	0.567784075	0.640100799	0.568366328	0.808709151
Thdb	0.690420152	0.76073724	0.692058634	0.972117015
Thjm	0.700759308	0.682066065	0.62099078	0.2674205
TRT	0.449941637	0.452561856	0.578565313	0.517445278
Wanbangde	0.350398153	0.410898312	0.464547495	0.483783447
Wondfo	0.706033933	0.749163259	0.666377487	0.738463539
WEDGE INDUSTRIAL	0.718041181	0.768397698	0.690260111	0.76803824
Weiming	0.748177009	0.769128321	0.619853486	0.710145248
Wowu	0.739298451	0.784243922	0.7386125	0.820113211
Wohua	0.674322938	0.710157045	0.659012071	0.772121516
Wosen	0.619303412	0.571491875	0.797393821	0.699460802
AMD	0.689966569	0.789941263	0.71329527	0.810580022
Xianju Pharm	0.514464857	0.566490622	0.540109586	0.818150446
Xiangxue Pharm	0.537783667	0.540068623	0.481275233	0.714362519
Sunflower	0.540029523	0.598998632	0.388532054	0.706082152
NHU	0.517692055	0.603883517	0.66053792	0.748506424
Xinhua	0.402109863	0.430156322	0.437174929	0.686348395
Xinbang	0.43011624	0.428808192	0.361237255	0.540651302
SALUBRIS	0.744881615	0.779294136	0.697651488	0.639133074
BROTHER	0.663383853	0.719528977	0.583598178	0.823868476
Yabao	0.453418706	0.507914392	0.515088977	0.755819777
Yatai	0.687624616	0.723497946	0.658661044	0.308960713
Yananbikang	0.615088274	0.598667978	0.63509745	0.491466664
Yiling Pharm	0.535399897	0.54629301	0.521004694	0.764562368
Yifan	0.62137248	0.706145232	0.543427556	0.768574846
Yibai	0.601819983	0.583245952	0.344474902	0.726832453
Yisheng	0.640708057	0.679203014	0.623539381	0.703137255
Yiduoli	0.61878647	0.601097854	0.566968291	0.80554492
Chinataurine	0.65392248	0.715865276	0.674513572	0.727223103
Gloria Pharm	0.619802462	0.583161465	0.534749776	0.181246147
Baiyao	0.407703292	0.415325832	0.598683278	0.592789916
Zhejiang Medicine	0.39694149	0.393998731	0.627437898	0.585114319
Zhenbao Island	0.615637011	0.630467145	0.565464187	0.847153576
Zdzy	0.53930568	0.541502435	0.447281659	0.74754751
Zhifei	0.659011263	0.779383924	0.84578019	1
Zhongguancun	0.607984942	0.597661057	0.559189938	0.648070137
China Medicine	0.298293719	0.284321785	0.389396559	0.332076638
Zhongheng Group	0.6583369	0.730823738	0.660884453	0.747300882
Zhongmu	0.46697542	0.505701728	0.480753041	0.682814008
Zhongxin	0.425869997	0.478604566	0.475336345	0.714813998
Zsyy	0.672284742	0.721839471	0.64125814	0.877851223
JLZX	0.658045275	0.724715509	0.596666189	0.821878746
Zuoli	0.647050689	0.667106722	0.616372459	0.703421246

Table 4: Analysis of DEA efficiency in the second stage of Chinese pharmaceutical companies.

Company name	Year			
	2015-2016	2016-2017	2017-2018	2018-2019
Adisseo	1	0.649837545	0.625737431	0.541530825
Anke Bio	0.765162026	0.476189062	0.772255149	0.613658522
Osaikon	0.843778066	0.441445985	0.850780731	0.575021161
Baiyunshan	0.128029731	0.338472809	0.44681541	0.404956323
Bdyy	0.890834395	0.445222754	0.727676509	0.608956969
Beilu	0.808544991	0.447781474	0.735202266	0.658558851
Porton	0.790606856	0.452864356	0.741151893	0.630953461
None	0.899553352	0.412224769	0.692164738	0.575225616
Changchun High-Tech	0.697331634	0.534086802	0.974210731	0.83805388
Changjiang Runfa Medicine	0.758278202	0.457208597	0.757178228	0.355161748
Changshan Pharma	0.839430913	0.452268392	0.722949426	0.63821128
DAJY	0.498749812	0.453046771	0.736062102	0.592503277
Dezhan	1	0.540144516	0.907325608	0.65741689
Jiao	1	0.695657355	1	0.349390869
DBBT	0.79160979	0.443887279	0.725436161	0.605439407
VC	0.52176685	0.452560475	0.4923262	0.401474275
Dongcheng	0.875632334	0.456954862	0.768638857	0.621583435
Nhwa Pharm	0.902685534	0.481101741	0.812506347	0.47310525
Ekzy	0.663965694	0.492139653	0.741589175	0.417137904
Fangsheng	0.83728348	0.445912668	0.729448377	0.613888946
Fengyuan	0.381700772	0.448248178	0.729255255	0.408521314
Fczy	0.810829455	0.450323096	0.733612414	0.613158813
Fayy	0.853851088	0.462652138	0.626676033	0.430608438
Fosun Pharm	0.405879589	0.887608474	0.863039604	0.822061518
Fuxiang	0.686147937	0.40302621	0.65968793	0.652530008
Guangji	0.781783744	0.438862422	0.728105584	0.614602675
Kwong Sang Hong	0.833600898	0.445235421	0.721873605	0.602632782
Guang Yuyuan	0.873544633	0.467173716	0.78987118	0.443703159
Lark	0.856359723	0.507355275	0.834541643	0.432730249
Glsj	0.762864498	0.498932775	0.802401553	0.666986138
Sinopharm Hyundai	0.304326441	0.467872765	0.522609903	0.295245894
Harbin Pharm	0.471907832	0.498927974	0.51892983	0.413398936
Haili Bio	0.845981594	0.452963586	0.715761969	0.596021344
Hnhy	0.546132286	0.451209071	0.73929906	0.377683569
Hepalink	0.570692829	0.455773586	0.833748227	0.517400795
Haishun New Pharma	0.875191121	0.441283929	0.718583719	0.61164409
Haisco	0.265247759	0.462094574	0.770404638	0.453228976
Hisoar	0.806777693	0.483733043	0.841284576	0.486621529
Haixin	0.869599312	0.45122025	0.748838002	0.618733421
Hisun	0.295330677	0.469935051	0.434672128	0.436124359
Han Sen Pharm	0.842792274	0.449054927	0.739269782	0.624481603
Hybio	0.741190863	0.482159427	0.64940653	0.450484128
Hengrui	0.994006619	0.851204335	1	1
Chase Sun	0.510934768	0.252963632	0.388183145	0.487982456
NCPC	0.446958138	0.442893623	0.485019512	0.416336438
Huahai	0.483631419	0.517819876	0.744562547	0.468120819
Hualan Bio	0.488018688	0.463275133	0.825155867	0.606937968
Huaren	0.858322442	0.445957989	0.72614299	0.608155932
HRSJ	0.39055996	0.606296731	0.659827456	0.641018804
China Resources Double-Crane	0.532023886	0.544202986	0.588679129	0.519985
Huashen Technology	0.899423723	0.461862909	0.723906959	0.610948685
Walter Dyne	0.842502696	0.497242721	0.767143201	0.641896259
Yanbian FC	0.442623386	0.657329408	0.887815963	0.555869763
Kyrgyzstan	0.455180857	0.466836149	0.763056261	0.299038639
Jichuan	0.692456078	0.269019483	0.480891191	0.509052877
Jimin	0.843175982	0.447230193	0.72593401	0.612949828
JYPC	0.902264504	0.405752965	0.710764503	0.579782874
Joincare Pharm	0.320679344	0.446019187	0.442365325	0.607505328
Jiangzhong	0.968578156	0.488910163	0.806995681	0.681786657

Table 4: Continued.

Company name	Year			
	2015-2016	2016-2017	2017-2018	2018-2019
Jincheng	0.680587626	0.446883725	0.726463433	0.422778688
Jinhe Bio	0.767555971	0.454800853	0.751896726	0.621861095
Jinling	0.882151804	0.462632532	0.777149818	0.420373223
Jinshiya	0.818824975	0.430337738	0.708029409	0.621839405
Jingxin	0.754271786	0.3997261	0.669011225	0.458543459
Jinghua	0.818896414	0.46479897	0.771305173	0.511596589
Jingfeng	0.659628837	0.462586912	0.756245434	0.295660795
Jiuqiang	0.888800059	0.475114126	0.779541159	0.656729486
Jiuzhitang	0.682939594	0.388385296	0.576221175	0.418866032
Jiuzhou	0.70068303	0.459376494	0.75017596	0.426386203
CONBA	0.450840367	0.532007534	0.572159801	0.36900969
Kanghong	0.630861917	0.383874254	0.632910424	0.462196751
Kangyuan	0.569889462	0.483812562	0.800075831	0.457930115
Kangzhi	0.860448512	0.447077729	0.722807424	0.595329819
KHB	0.75874045	0.446262253	0.731865553	0.430496937
Kelun	0.392192554	0.491878068	0.573939315	0.514916857
Sunflower	0.397069408	0.402221867	0.677747189	0.568562512
Kunming Pharm	0.484712539	0.482852557	0.787863272	0.452428396
Lummy	0.757591911	0.446024294	0.735699424	0.57313539
LAYN	0.82616477	0.46673708	0.735088104	0.618275961
Lisheng Pharma	0.892289898	0.44823865	0.743782606	0.632397402
Livzon Pharm	0.330476519	1	0.621269109	0.563738749
LEADMAN	0.744491132	0.439674197	0.708587278	0.605240773
Lianhuan Pharm	0.883381664	0.449882686	0.733590338	0.614613785
Lingkang	0.487342648	0.461147315	0.755609423	0.634678684
Lingrui Pharm	0.820570294	0.469108287	0.768403276	0.650201258
Long Jin Pharm	0.825269505	0.445474296	0.7213483	0.596440566
Lukang Pharm	0.709307544	0.271967764	0.448405964	0.41417971
Mike Bio	0.829672491	0.491963544	0.816516452	0.463478288
M.k.	0.702540171	0.461798568	0.750504723	0.334581671
Palin Bio	0.800366688	0.44299407	0.733746558	0.627260125
PIEN TZE HUANG	0.9045343	0.538290845	0.947536174	0.513658965
Julie Plec	0.549521085	0.305111815	0.499939327	0.524584684
Plyy	0.533402894	0.421972728	0.708018213	0.458768245
CHEEZHENGTTM	0.565148659	0.465429448	0.761471867	0.662280612
Qidi	0.86117892	0.441527153	0.709508207	0.607695056
Qianhong Biopharma	0.874606048	0.456315537	0.750563573	0.635328671
Qianjin Pharm	0.739246108	0.471814504	0.779924867	0.439377818
Qianyuan	0.831338985	0.404575277	0.655154668	0.336313058
Renfu	0.334414205	0.618045999	0.176055483	0.552146513
Renhe Pharmacy	0.888111873	0.496037076	0.837040815	0.469528713
Rpsw	0.594394568	0.433880869	0.709930487	0.626972475
Saisheng	0.929682934	0.476157778	0.77599003	0.625516632
SAM	0.905264682	0.451150287	0.680555861	0.60183122
Shanhe Pharmacy	0.847924844	0.432928146	0.710008765	0.610022991
Shkb	0.848089711	0.471499157	0.758909287	0.643053348
Shanghai RAAS Blood Products	1	0.54470828	0.409520354	0.468019077
Shenqi	0.714762096	0.381344561	0.617613041	0.61415927
Biological Stock	0.965359357	0.549223179	0.870620483	0.638968678
Salvage Pharm	0.689364433	0.415308201	0.706749477	0.198232344
Sts	0.935757517	0.472141333	0.74357052	0.605127606
Scyy	0.739153533	0.426485993	0.676414771	0.571466831
Beijing SL Pharm	0.733749939	0.506947781	0.832842848	0.681797114
Stellite	0.803378896	0.437370993	0.715251262	0.614807042
Shsw	0.890410616	0.434761769	0.700177597	0.603465476
Tat	0.842586294	0.360807923	0.584823244	0.390571577
Taiji Group	0.67919953	0.452980777	0.474001541	0.387892653
Taloph Pharm	0.842636824	0.438224781	0.689161525	0.607715363
Teyi	0.836358694	0.453682077	0.749014352	0.629612238

Table 4: Continued.

Company name	Year			
	2015-2016	2016-2017	2017-2018	2018-2019
Tasly	0.717299163	0.615735065	0.675224149	0.512843572
Tiantan Biological	0.678613559	0.591061254	0.860246774	0.500479724
Tianyao Pharm	0.857677116	0.448393948	0.737240947	0.409261416
Thdb	0.900688131	0.545722415	0.888741777	0.491006733
Thjm	0.744976147	0.47311369	0.785334315	0.254122817
TRT	0.836177972	0.656151609	0.703560723	0.570595311
Wanbangde	0.792851861	0.453600664	0.482306406	0.423211786
Wondfo	0.718234154	0.423654128	0.712394655	0.672912535
WEDGE INDUSTRIAL	0.78623338	0.450705667	0.727173666	0.609309043
Weiming	0.801846721	0.489019559	0.697530662	0.612890373
Wowu	0.915178281	0.444228865	0.732530156	0.611424148
Wohua	0.849951067	0.451158165	0.726657243	0.616986053
Wosen	0.293538834	0.36896297	0.927581915	0.621532855
AMD	0.565789697	0.46605309	0.755941006	0.654308568
Xianju Pharm	0.722129826	0.46975224	0.788610822	0.449576159
Xiangxue Pharm	0.56594413	0.450246609	0.728601595	0.406168036
Sunflower	0.775727247	0.435119495	0.472627581	0.583260445
NHU	0.744446192	0.084792409	0.174517515	0.645247085
Xinhua	0.686150026	0.418580131	0.691475644	0.430862405
Xinbang	0.37146641	0.482379256	0.29706599	0.42940549
SALUBRIS	0.828950689	0.45514634	0.744067755	0.492886897
BROTHER	0.867361567	0.491290966	0.722969193	0.404592688
Yabao	0.62538223	0.465981424	0.776164146	0.400865973
Yatai	0.791869911	0.458579905	0.749488501	0.274754037
Yananbikang	0.609592291	0.525754839	0.495889903	0.382123694
Yiling Pharm	0.732199742	0.505748671	0.83520904	0.467694136
Yifan	0.895380366	0.428163195	0.614491215	0.498168738
Yibai	0.622820274	0.32775112	0.386250987	0.412500415
Yisheng	0.615818561	0.450233859	0.733856106	0.615039599
Yiduoli	0.768370215	0.454151698	0.749092834	0.418880633
Chinataurine	0.790572254	0.445496958	0.734787151	0.618625291
Gloria Pharm	0.425375138	0.473021826	0.735198757	0.099279291
Baiyao	0.676420897	0.831225519	0.898534261	0.870529199
Zhejiang Medicine	0.648977621	0.470928238	0.508877262	0.43259115
Zhenbao Island	0.366714214	0.44796952	0.719005835	0.445442437
Zdzy	0.666462777	0.477272075	0.688004343	0.415110578
Zhifei	0.794312561	0.48858374	1	0.851446092
Zhongguancun	0.638431487	0.361048475	0.600829723	0.58735967
China Medicine	0.600385661	0.632935767	0.705012728	0.541274783
Zhongheng Group	0.779488337	0.516431725	0.842980772	0.726827094
Zhongmu	0.891791426	0.461595881	0.758629587	0.388861314
Zhongxin	0.784299166	0.500035996	0.833734288	0.470157549
Zsyy	0.712236556	0.493776655	0.805881833	0.435730452
JLZX	0.56839467	0.487351822	0.7538412	0.407572713
Zuoli	0.506830635	0.327842837	0.530045226	0.603856046

Table 5: Analysis of the overall efficiency of Chinese pharmaceutical companies.

Company name	Year			
	2015-2016	2016-2017	2017-2018	2018-2019
Adisseo	1	0.594331961	0.687253527	0.590582307
Anke Bio	0.732050473	0.603904778	0.721684173	0.658510045
Osaikon	0.778427713	0.599091917	0.832059909	0.696276311
Baiyunshan	0.228285921	0.347396666	0.496661228	0.354314544
Bdyy	0.740297813	0.537030807	0.6599985	0.641144699
Beilu	0.754848982	0.608873049	0.726868379	0.745366839
Porton	0.704058135	0.537439838	0.673751419	0.645425288
None	0.804540469	0.581032222	0.686403127	0.634019398
Changchun High-Tech	0.670785268	0.613981329	0.826328221	0.91902694
Changjiang Runfa Medicine	0.709676062	0.545533564	0.667403235	0.486189125
Changshan Pharma	0.771002674	0.594856424	0.68892317	0.682407653
DAJY	0.553391996	0.532804722	0.650855774	0.625961732
Dezhan	0.889124007	0.713057908	0.85424928	0.696289729
Jiao	0.847938007	0.724106276	0.892313767	0.434427484
DBBT	0.745018596	0.588398867	0.704618016	0.676626886
VC	0.4392569	0.411082825	0.544473127	0.465493745
Dongcheng	0.780465431	0.58515561	0.710439698	0.633026364
Nhwa Pharm	0.734368139	0.53849531	0.686067979	0.687237305
Ekzy	0.679929415	0.567141171	0.651349898	0.617078583
Fangsheng	0.763857057	0.576574634	0.69884161	0.674382377
Fengyuan	0.444827414	0.480815113	0.614783892	0.606094921
Fczy	0.742074088	0.578441415	0.685449982	0.654076242
Fayy	0.747025125	0.542767359	0.550777941	0.641860364
Fosun Pharm	0.423354375	0.670393153	0.67734337	0.657568579
Fuxiang	0.690619828	0.558173035	0.657502803	0.69283151
Guangji	0.723829795	0.562163776	0.697307481	0.671209547
Kwong Sang Hong	0.782406423	0.601677149	0.712736266	0.684833649
Guang Yuyuan	0.78963604	0.597988236	0.731277172	0.563852758
Lark	0.724214801	0.557523041	0.691343122	0.603175409
Glsj	0.704470405	0.600924412	0.722522094	0.691917544
Sinopharm Hyundai	0.427413377	0.393573592	0.554655968	0.411827436
Harbin Pharm	0.352405918	0.369339239	0.485586707	0.432828982
Haili Bio	0.783470209	0.604343839	0.699572898	0.67374349
Hnhy	0.569927767	0.533436376	0.636600922	0.537164476
Hepalink	0.592158759	0.525343452	0.699393429	0.659703397
Haishun New Pharma	0.8030248	0.605941868	0.716134725	0.700106885
Haisco	0.480099239	0.575867685	0.699064728	0.661359713
Hisoar	0.671469103	0.541039094	0.710909693	0.676879776
Haixin	0.734260582	0.551448437	0.676486256	0.637198392
Hisun	0.287396529	0.380594016	0.415821467	0.441140083
Han Sen Pharm	0.732716709	0.559696588	0.688766917	0.684174582
Hybio	0.724090442	0.622127582	0.609611185	0.477857438
Hengrui	0.826957748	0.770628784	1	1
Chase Sun	0.550775748	0.399677749	0.436619602	0.528426739
NCPC	0.354218369	0.358094194	0.480775578	0.427708854
Huahai	0.497285595	0.525399389	0.58350641	0.588866754
Hualan Bio	0.606253871	0.611156906	0.786676493	0.803468984
Huaren	0.703067183	0.5262557	0.649580709	0.631043029
HRSJ	0.406183354	0.51553843	0.651086789	0.659592625
China Resources Double-Crane	0.464181045	0.478110469	0.667487244	0.618744398
Huashen Technology	0.79585331	0.605253566	0.698514176	0.686848921
Walter Dyne	0.745780253	0.592281438	0.685737585	0.679786234
Yanbian FC	0.58552906	0.726843818	0.75539478	0.760336119
Kyrgyzstan	0.57291471	0.595278792	0.719807265	0.307289038
Jichuan	0.673524741	0.478385974	0.569904186	0.621509323
Jimin	0.755191647	0.581216669	0.682498022	0.658109286
JYPC	0.789213967	0.533152421	0.692664421	0.655377456
Joincare Pharm	0.373389357	0.612787948	0.431248192	0.711024008
Jiangzhong	0.773280624	0.57457233	0.714383571	0.696034497

Table 5: Continued.

Company name	Year			
	2015-2016	2016-2017	2017-2018	2018-2019
Jincheng	0.6424688	0.541255874	0.633395756	0.600807083
Jinhe Bio	0.694022214	0.53855834	0.672530294	0.639655629
Jinling	0.669922726	0.463558983	0.63122491	0.598918507
Jinshiya	0.779202218	0.61261129	0.67505169	0.654243898
Jingxin	0.684642416	0.522548631	0.634930264	0.670080794
Jinghua	0.73659423	0.580261028	0.708115153	0.532218118
Jingfeng	0.657549858	0.543871346	0.658966242	0.439651462
Jiuqiang	0.832168543	0.640440908	0.760447378	0.740948042
Jiuzhitang	0.660685922	0.510365495	0.551518743	0.601037942
Jiuzhou	0.615906479	0.511229588	0.651701562	0.646607443
CONBA	0.43789997	0.490049729	0.667671711	0.454110037
Kanghong	0.640270409	0.530393868	0.625355868	0.589391549
Kangyuan	0.558531561	0.511112483	0.661025134	0.653468887
Kangzhi	0.770954193	0.582966967	0.690695416	0.63784734
KHB	0.719324825	0.576138757	0.689042667	0.655129362
Kelun	0.350816311	0.399255114	0.527654704	0.466336149
Sunflower	0.450269017	0.47246729	0.609455977	0.602948399
Kunming Pharm	0.486232652	0.488722513	0.622323739	0.575298309
Lummy	0.676182846	0.548086309	0.67791945	0.60414936
LAYN	0.766415526	0.615493713	0.696786083	0.687666604
Lisheng Pharma	0.772324056	0.553232118	0.688792627	0.658713187
Livzon Pharm	0.406802204	0.950736975	0.693561727	0.641403963
LEADMAN	0.7148845	0.589683903	0.697033474	0.677283392
Lianhuan Pharm	0.777990479	0.580819977	0.70187494	0.677825662
Lingkang	0.593122777	0.602274773	0.726097851	0.699381382
Lingrui Pharm	0.747485965	0.570748655	0.691654705	0.678247727
Long Jin Pharm	0.778557308	0.60796337	0.720057661	0.694828539
Lukang Pharm	0.568388015	0.368837292	0.454369366	0.548391093
Mike Bio	0.768849987	0.607581032	0.730833097	0.678396297
M.k.	0.705468047	0.594140747	0.659621157	0.473923489
Palin Bio	0.735879001	0.580588028	0.7000859	0.692170464
PIEN TZE HUANG	0.794732455	0.644348252	0.804287685	0.736290866
Julie Plec	0.628612707	0.51441293	0.586701026	0.629100125
Plyy	0.470181321	0.418413253	0.567877026	0.583837057
CHEEZHENGTTM	0.625915805	0.595282402	0.719854866	0.715491847
Qidi	0.763992733	0.561948196	0.677246094	0.688875522
Qianhong Biopharma	0.796996434	0.587850281	0.703561836	0.674562645
Qianjin Pharm	0.634951449	0.510221787	0.651319311	0.643983464
Qianyuan	0.734922726	0.538059322	0.641532464	0.529123309
Renfu	0.334316504	0.515659371	0.162643485	0.472524591
Renhe Pharmacy	0.710690176	0.501475944	0.678520156	0.647375907
Rpsw	0.616216798	0.544427634	0.663887992	0.660390626
Saisheng	0.842078163	0.636453134	0.751848255	0.697189691
SAM	0.793087056	0.570374323	0.637288238	0.66250408
Shanhe Pharmacy	0.780593912	0.590341994	0.705650192	0.693438972
Shkb	0.769351682	0.596263047	0.707135607	0.684164457
Shanghai RAAS Blood Products	0.907968608	0.628170623	0.352321342	0.675478039
Shenqi	0.680638994	0.523233731	0.611259379	0.641985589
Biological Stock	0.870829861	0.70347757	0.801965497	0.656218882
Salvage Pharm	0.547394207	0.443227806	0.638664943	0.301830016
Sts	0.840550192	0.630975455	0.719756982	0.666578994
Scyy	0.680239263	0.581555378	0.673278766	0.671620779
Beijing SL Pharm	0.744354754	0.664713516	0.785954743	0.730167136
Stellite	0.725188753	0.558580997	0.681979752	0.655174892
Shsw	0.798450652	0.591904871	0.700994849	0.696898497
Tat	0.712944165	0.470305378	0.557717995	0.553727441
Taiji Group	0.523948741	0.370677112	0.489754105	0.411216635
Taloph Pharm	0.723845517	0.541466331	0.629307653	0.632734845
Teyi	0.760999041	0.581909561	0.708162034	0.680597811

Table 5: Continued.

Company name	Year			
	2015-2016	2016-2017	2017-2018	2018-2019
Tasly	0.531017743	0.490066656	0.611366055	0.461437582
Tiantan Biological	0.622734244	0.66310047	0.782922006	0.712465273
Tianyao Pharm	0.712730596	0.544247373	0.652803637	0.608985283
Thdb	0.795554142	0.653229828	0.790400206	0.731561874
Thjm	0.722867728	0.577589877	0.703162547	0.260771659
TRT	0.643059805	0.554356733	0.641063018	0.544020294
Wanbangde	0.571625007	0.432249488	0.473426951	0.453497617
Wondfo	0.712134043	0.586408694	0.689386071	0.705688037
WEDGE INDUSTRIAL	0.75213728	0.609551683	0.708716889	0.688673642
Weiming	0.775011865	0.62907394	0.658692074	0.661517811
Wowu	0.827238366	0.614236394	0.735571328	0.715768679
Wohua	0.762137003	0.580657605	0.692834657	0.694553785
Wosen	0.456421123	0.470227422	0.862487868	0.660496829
AMD	0.627878133	0.627997177	0.734618138	0.732444295
Xianju Pharm	0.618297341	0.518121431	0.664360204	0.633863302
Xiangxue Pharm	0.551863899	0.495157616	0.604938414	0.560265277
Sunflower	0.657878385	0.517059063	0.430579818	0.644671298
NHU	0.631069124	0.344337963	0.417527717	0.696876755
Xinhua	0.544129945	0.424368226	0.564325286	0.5586054
Xinbang	0.400791325	0.455593724	0.329151623	0.485028396
SALUBRIS	0.786916152	0.617220238	0.720859622	0.566009985
BROTHER	0.76537271	0.605409971	0.653283686	0.614230582
Yabao	0.539400468	0.486947908	0.645626561	0.578342875
Yatai	0.739747264	0.591038926	0.704074772	0.291857375
Yananbikang	0.612340283	0.562211409	0.565493676	0.436795179
Yiling Pharm	0.633799819	0.526020841	0.678106867	0.616128252
Yifan	0.758376423	0.567154214	0.578959385	0.633371792
Yibai	0.612320129	0.455498536	0.365362945	0.569666434
Yisheng	0.628263309	0.564718437	0.678697744	0.659088427
Yiduoli	0.693578342	0.527624776	0.658030562	0.612212776
Chinataurine	0.722247367	0.580681117	0.704650362	0.672924197
Gloria Pharm	0.5225888	0.528091646	0.634974266	0.140262719
Baiyao	0.542062095	0.623275676	0.74860877	0.731659557
Zhejiang Medicine	0.522959555	0.432463484	0.56815758	0.508852734
Zhenbao Island	0.491175612	0.539218332	0.642235011	0.646298006
Zdzy	0.602884228	0.509387255	0.567643001	0.581329044
Zhifei	0.726661912	0.633983832	0.922890095	0.925723046
Zhongguancun	0.623208215	0.479354766	0.58000983	0.617714904
China Medicine	0.44933969	0.458628776	0.547204643	0.43667571
Zhongheng Group	0.718912618	0.623627731	0.751932612	0.737063988
Zhongmu	0.679383423	0.483648804	0.619691314	0.535837661
Zhongxin	0.605084581	0.489320281	0.654535316	0.592485773
Zsyy	0.692260649	0.607808063	0.723569987	0.656790838
JLZX	0.613219972	0.606033666	0.675253694	0.61472573
Zuoli	0.576940662	0.497474779	0.573208842	0.653638646

Table 6: Annual Malmquist index and its decomposition indexes.

	EC	SC	TC	PC	TFP
2015-2016	1.1841	2.7910	1.0109	0.4242	1.1970
2016-2017	0.5162	1.4472	1.1211	0.3567	0.5787
2017-2018	0.9815	2.4676	1.0018	0.3978	0.9833
2018-2019	0.5188	1.2176	1.0034	0.4260	0.5205
Mean	0.8001	1.9808	1.0343	0.4012	0.8219

5. Dynamic Effect Analysis

The efficiency of the two-stage network DEA model varies from year to year, and the efficiency value of different years is not comparable, so time series analysis cannot be carried out. To make up for the shortcomings of the traditional two-stage network DEA, this paper adds the Malmquist index to study the total factor productivity of listed pharmaceutical companies in 2015–2019 and quantify its decomposition limit. The results are shown in Table 6.

The average total factor productivity (TFP) is 0.8219 that has fallen by an average 17.81%. Viewed from the decomposition index, the mean of EC is 0.8001; that is, EC has decreased by an average of 19.99%. The mean of PC is 0.4012, with the rate of decline in each averaging over 59.88% a year. SC is 1.9808, with an average annual growth rate of 98.08%. It shows that the operating efficiency of listed pharmaceutical enterprises depends on the scale expansion and makes up for low management efficiency. The average technology change (TC) is 1.0343, and it has risen by nearly 3.43% per year. The technology change has been improved between 2015 and 2019.

The listed pharmaceutical companies rely on product development and can be combined with innovative technologies. For the pharmaceutical industry, the level of research and development of products indirectly affects the level of industry development. The drugs or pharmaceutical equipment is very important; if the level of medical technology research and development is not advanced enough, the progress of medical level will be affected. Therefore, the listed pharmaceutical enterprises should rely on the existing advanced technology achievements, improving their own technology, to improve operating efficiency.

6. Conclusion

This paper firstly divides the two subsystems by using the two-stage network DEA and analyses the operating efficiency of 1,63 listed pharmaceutical companies in China from 2015 to 2019. Secondly, Malmquist index is used for dynamic analysis; the total factor productivity and decomposition limit were obtained. Finally, we make some suggestions based on the results of the study.

From the results, the technical efficiency of the second phase is less than that of the first stage; government subsidies have no positive impact on most companies. It is possible that enterprises move government subsidies elsewhere rather than pharmaceutical companies. It is also possible that government subsidies have increased, enterprises are more willing to invest in product development and enterprises expansion, and it is difficult to see the improvement of operational efficiency in the short term. However, the government subsidies have a positive impact on enterprises with low efficiency in a short term. To ensure the efficiency of investment and avoid waste of resources, government needs to choose the object of subsidies carefully and reformulate policies that encourage pharmaceutical listed companies. And according to the Malmquist index results, enterprises should pay attention to risk prevention and avoid rapid expansion bringing in diseconomies of scale. All in all, enterprises should improve management level and technological capabilities and shift scale growth to total factor productivity.

However, there are also some limitations. Regarding the data resources, the data we chose cannot exactly predict the operational situations among current Chinese medical firms, since there is more uncertainty in the market, especially during the COVID-19 period, which is likely to be a potential direction that other scholars can study further in the future. Concerning variables, this article analyses the operational efficiency of 164 firms; researchers can only choose several companies to make an in-depth analysis instead of the whole industrial analysis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was financially supported by Guangdong Planning Office of Philosophy and Social Sciences Project (Youth): Research on cross-border social responsibility of private foreign trade enterprises in Guangdong and the reconstruction of legitimacy—Perspective of the organization to piece together. (Project number: GD20YGL09). And it also supported by Department of Education of Guangdong Province "Innovative projects with characteristics of ordinary universities" project: Research on Sustainable Development of Foreign Trade in Guangdong Province Based on Energy Footprint (Project no. 2019WTSCX158). Moreover, it is also supported by Key Discipline-International Business Construction and Development Project (Project no. HS2019CXQX17).

References

[1] Y. Wang, "Research on the financing efficiency of biomedicine industry based on DEA model," *Hebei Enterprise*, no. 01, pp. 104-105, 2021.

[2] Z. Zhang, *Research on Performance Evaluation of Pharmaceutical Enterprises Based on AHP-DEA Method*, Tianjin University, Tianjin, China, 2018.

[3] Z. Li, R. Liu, H. He, and J. Qin, "An empirical study on the efficiency of the pharmaceutical industry based on the SE-DEA model," *Accounting Communications*, no. 14, pp. 26–28, 2016.

[4] X. Zhou and C. Luo, "Research on the relative effectiveness of China's pharmaceutical industry technological innovation based on DEA model," *Science and Technology Management Research*, no. 9, pp. 252–254, 2009.

[5] P. Shi and L. Tao, "Evaluation of innovation performance of pharmaceutical industry based on DEA model," *Economic Perspective*, no. 3, pp. 28-29, 2011.

[6] H. Chen, *Research on Technology Innovation Efficiency of China's Pharmaceutical Industry—Analysis Based on Panel Data*, Jinan University, Guangzhou, China, 2011.

[7] J. Liu, "Study on performance evaluation of science and technology investment in China's pharmaceutical high-tech industry," *Science Management Research*, vol. 27, no. 4, pp. 35–38, 2009.

[8] A. Xiong, *Research on Technological Innovation Efficiency of Pharmaceutical Companies Based on DEA Method*, Guangdong Pharmaceutical University, Guangzhou, China, 2020.

[9] H. F. Lewis, S. Mallikarjun, and T. R. Sexton, "Unoriented two-stage DEA: the case of the oscillating intermediate products," *European Journal of Operational Research*, vol. 229, no. 2, pp. 529–539, 2013.

[10] L. Liang, Z.-Q. Li, W. D. Cook et al., "Data envelopment analysis efficiency in two-stage networks with feedback," *IIE Transactions*, vol. 43, no. 5, pp. 309–322, 2010.

[11] L. Liang, D. Wade, and J. Z. Cook, "DEA models for two-stage processes: game approach and efficiency decomposition," *Naval Research Logistics*, vol. 55, no. 7, pp. 643–653, 2008.

[12] Y. Chen, J. Du, H. David Sherman et al., "DEA model with shared resources and efficiency decomposition," *European Journal of Operational Research*, vol. 207, no. 1, pp. 339–349, 2010.

[13] C. Kao, "Efficiency decomposition in network data envelopment analysis with slacks-based measures," *European Journal of Operational Research*, vol. 192, no. 3, pp. 949–962, 2009.

[14] C.-Y. Lee, L. Andrew, and Z. Johnson, "A decomposition of productivity change in the semiconductor manufacturing industry," *International Journal of Production Research*, vol. 49, no. 16, pp. 4761–4785, 2011.

[15] W. T. Pan, M. E. Zhuang, Y. Y. Zhou et al., "Research on sustainable development and efficiency of China's E-agriculture based on a data envelopment analysis-malmquist model," *Technological Forecasting and Social Change*, vol. 162, Article ID 120298, 2021.

[16] A. Hashimoto and S. Haneda, "Measuring the change in R&D efficiency of the Japanese pharmaceutical industry," *Research Policy*, vol. 37, no. 10, pp. 1829–1836, 2008.

[17] H. Pannu, U. Kumar, and J. Farooqui, *Impact of Innovationon the Performance of Indian Pharmaceutical Industry Using Data Envelopment Analysis*, IIMB working paper, Karnataka, India, 2010.

[18] X. Zhiyue and J. Qiu, "Evaluation and analysis of the operating efficiency of my country's biopharmaceutical industry," *Shanghai Medicine*, vol. 36, no. 13, pp. 50–53, 2015.

[19] P. Cheng and C. Zheng, "Research on my country's science and technology input efficiency based on decision preference two-stage network DEA," *Science and Technology Progress and Policy*, vol. 31, no. 08, pp. 125–129, 2014.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9974585, 11 pages
https://doi.org/10.1155/2021/9974585

Research Article

Research on the Relationship between Data Empowerment and Service Innovation Capability of Logistics Platform Enterprise

Yuhua Zhang [1] and Mengdie Hu [2]

[1] *Department of Business Administration, Guangzhou Huashang College, Guangzhou 510000, China*
[2] *School of Business, Guangdong University of Foreign Studies, Guangzhou 510000, China*

Correspondence should be addressed to Yuhua Zhang; zhyh58@163.com

Received 18 March 2021; Revised 24 April 2021; Accepted 15 May 2021; Published 28 May 2021

Academic Editor: Shianghau Wu

Based on the application of big data, this paper constructed a theoretical model focusing on the mechanism of data empowerment on the service innovation capability of logistics platform enterprises, with value cocreation as the mediating variable and environmental dynamism as the moderating variable. The research hypothesis was empirically tested based on the results obtained from the questionnaire survey. The results demonstrate that data empowerment can promote the value co-creation between logistics platform enterprises and users, and value co-creation is an important factor to promote the service innovation capability of logistics platform enterprises. Meanwhile, the moderating variable of environmental dynamism is found to inhibit the interaction between cooperation and service innovation capability. The findings expand the theoretical research on data empowerment and raise important inspiration for practical activities of logistics platform enterprises.

1. Introduction

The rapid development of the Internet and big data has made the service forms and applications of logistics enterprises more and more extensive, and various types of logistics platform enterprises are constantly emerging, i.e., intra-city freight logistics platform, vehicle-cargo matching bulk freight logistics platform, and highway-port logistics platform. Yunmanman, China's first freight dispatch platform enterprise that operates based on cloud computing, big data, mobile Internet, and artificial intelligence technologies, had developed into China's largest vehicle capacity scheduling platform, smart logistics information platform, and vehicle-free carrier in just 4 years. Platform enterprises extensively connect to bilateral (multilateral) users of the platform and establish an open innovation ecosystem relied on common infrastructure (e.g., platform architecture) [1]. A basic feature of platform enterprises is to reduce of transaction cost that prevents the value exchange among parties gathering together [2]. This improves the operating efficiency of participants and benefits both parties. Therefore, platform enterprises not only enable themselves to develop rapidly but also play a huge role in creating employment opportunities, innovating technological patterns, and benefiting the public.

The emergence of logistics platform enterprises has fundamentally solved the problems of fragmented traditional logistics resources, blocked transaction links, and low information concentration. In other words, it has completely changed the situation of "small, chaotic, and scattered" in traditional logistics industry in China, improving utilization rate of social resources and logistics operation efficiency, while reducing social logistics costs. However, it can also be seen that after several years of development, the intense competition among logistics platform enterprises generated by homogenization, weak overall innovation capabilities, and poor profitability has seriously hampered the healthy development of platform enterprises. Therefore, major logistics platform enterprises have been concerning on issues such as improving the service innovation capabilities of logistics platforms, obtaining more customer resources, and increasing market share.

Big data changes the original production and operation patterns of enterprises, integrating and efficiently utilizing social resources, improving the efficiency of economic, and achieving sound and rapid economic development. Profoundly based on the progress of the Internet and big data, the developments of logistics platform enterprises provide timely and effective information for both parties in transaction, change the original logistics transaction patterns, and create higher value for logistics enterprises and customers. The application of big data provides efficient means for the development of logistics platform enterprises. Platform enterprises are committed to maximizing the sharing of information and resources, which is conducive to attracting more users and acquiring more data resources, thereby promoting the innovation and sustainable development of the platform-based logistics enterprises. The nature of the Internet platform is empowerment. In the future development of the logistics platform, data empowerment will become the main approach. Meanwhile, service innovation has added increasing significance and value to the development of logistics enterprises, and it is the main way for enterprises to achieve competitive advantages. Based on the existing problems of logistics platform enterprises and considering the importance of data empowerment and service innovation in logistics platforms, this research will clarify the mechanism and context of data empowerment to promote the improvement of service innovation capabilities for logistics platform enterprises, and it helps to provide practical reference for the sustainable and healthy development of logistics platform enterprises.

The research on data empowerment mainly focuses on business practice and the empowerment of social public affairs. For the empowerment of social public affairs, the research focuses more on the empowerment by digital technology for medical treatment and nursing of patients [3, 4], women's rights [5, 6], minors' health [7], social citizenship [8, 9], and the data enabling between teachers and students in the realization of classroom and distance education [10, 11]. In recent years, due to the disruptive impact of data and data technology on business practices, data empowerment has attracted widespread attention in terms of patterns and means of empowerment. Moreover, research on capacity building and mechanism of data empowerment process, such as enterprise manufacturing transformation and upgrading, business model innovation, and value cocreation, has increased [12–14]. The research on service innovation capability focuses on the connotation and evaluation, the function mechanism, and the influence factors. For the connotation and influence factors, the research perspectives are based on supply chain integration [15–17], relationship network [18, 19], knowledge learning [20–22], and resource management [23]. For the evaluation of service innovation, the research measures are concerned from the aspect of capability characteristics [24] and comprehensive evaluation from the perspective of process [25, 26]. For the function mechanism, the research focuses are regarding the mechanism of service innovation capability on enterprise performance and competitive advantage [27] and the mechanism of interaction orientation on innovation capability from three perspectives: value co-creation [28], absorptive capacity [29], and initiative improvement [30–32].

As of now, although there is enormous research focused on data empowerment as well as service innovation capability, very few studies explore the intersection and interaction of the two. Meanwhile, the practical activities which data empowers platform enterprise to promote innovation capability are very active, but there is scarce research on relationship between data empowerment and service innovation capability for platform enterprise. Therefore, this study aims to theoretically explain the relationship between data empowerment and service innovation capability, and using logistics platform enterprise as the research object, to study how data empowerment can promote the improvement of service innovation capability, thus provides theoretical guidance for the enterprise practice and provides a new starting point for the academic research on the data empowerment and innovation capability of platform enterprises.

In practice, the big data capability of platform enterprise based on Internet big data technology has an important impact on the improvement of service innovation capability. In addition, the realization of information communication between the buyer and the seller requires consumer to feed back demand information to enterprise actively, so as to realize the value of cocreation, and hence improve the service innovation capability of platform enterprise. However, platform enterprises, born in the Internet era, are bound to live in a highly variable environment. Therefore, environmental dynamism must be considered as an important factor to measure the improvement of service innovation capability of platform enterprises. This study will be carried out following the context of data empowerment, service innovation capability, value co-creation, and environmental dynamism.

2. Theoretical Basis and Research Hypothesis

2.1. Data Empowerment. In recent years, empowerment has received attention and emphasis in multiple professions and trades. In particular, data empowerment is applied more frequently in Internet industry, and the opinions, such as digital empowerment and Internet empowerment, have emerged, which indicate the rapid development of data applications. Empowerment is a purposeful and continuous process, emphasizing the concepts of group-centered and mutual respect, and concerns the participation of group members. By involving in these processes, members who lack equal and valuable resources could obtain and control these resources [33]. Empowerment is mainly classified into structural empowerment, psychological empowerment, and resource empowerment, and some scholars have claimed that structural empowerment and psychological empowerment do not truly reflect the connotation of empowerment. Therefore, the focus on the research of resource empowerment has been raised by more and more researchers. In the era of big data, the core function of resource empowerment is gradually reflected in data empowerment. Data empowerment is a process of realizing the value in which a specific system innovates the application scenarios of data and applies skills and methods to gain or improve the overall capability of the system. The realization of data empowerment depends on the enterprises' capabilities of information collection, processing, transmission, and storage. By improving these capabilities, enterprises can realize data potential to a greater extent, integrate corporate resources efficiently to better cater customer needs, promote the value of data, and improve corporate performance. Lenka et al. [14] classified data capabilities into intelligence capabilities, connection capabilities, and analysis capabilities. Intelligence capabilities represent hardware configuration capabilities and intelligence collection capabilities. Besides, connection capabilities represent information transmission, processing capabilities, and connectivity capabilities among other smart products. Analysis capabilities are the ability to transform the existing data into valuable insights and feasible instructions for enterprises.

2.2. Data Empowerment and Service Innovation Capability. With the continuous development and application of big data technology, digital technology continues to empower multiple industries to meet customer needs and improve service innovation capabilities through efficient integration of corporate resources with intelligence, connectivity, and analysis capabilities. Digital capabilities have promoted the generation of new forms of knowledge and provided the necessary channels for the realization of complex innovations [34]. In the era of digital information, the information platform is the base of information sharing. The information sharing of the platform can update information in a timely manner and stimulate more originalities and ideas, thereby promoting the achievement of service innovation projects among communicators and realizing the improvement of service innovation capability for all parties. With the continuous deepening of the application of digital technology for platform enterprises, data empowerment has penetrated deeply into the products and services of platform enterprises, fundamentally changing the product categories and service methods. Logistics platform enterprises, relying on digital information technology, can improve their information processing capabilities through enhancing the application of digital information technology and hence better catering customer needs, improving existing services, simulating product innovation, and realizing the improvement of the service innovation capabilities of logistics platform enterprises [34]. Therefore, the data empowerment of the logistics service platform has a positive effect on the realization of service innovation capabilities. Thus, the following hypothesis is proposed:

H1: data empowerment has a positive effect on the service innovation capabilities of logistics platform enterprises.

2.3. Data Empowerment and Value Co-creation. Value co-creation is the process of creating value and experience through interaction between enterprises and consumers [35]. Gummesson and Mele have divided the value co-creation process into two aspects: interaction and resource integration [36]. The platform organization integrates into the multidirectional relationship of stakeholders and promotes value creation and value transfer through interaction and resource integration, achieving value co-creation among organizations. The integration of resources in the process of value co-creation can improve the value efficiency of oneself and the counterparty at the same time. Once the supplementary, redundant, or mixed resources in the value co-creation network are matched with each other among participants, the utility of resources will be maximized [37]. Therefore, the interactive cooperation and resource integration between users and enterprises are conducive to the realization of value co-creation.

Empowerment can realize value co-creation through co-creation activities such as interaction, cooperation, and resource integration, among different behavioral subjects (organizations, enterprises, consumers, etc.).

In indoor decoration platform enterprises, data empowerment makes customers change from passively accepting the design scheme to actively participate in the design process. At the same time, the design process balances the information asymmetry between customers and enterprises, which is conducive to promoting the value co-creation between customers and enterprises. In data-based tourism platform enterprises, represented by Didi, data empowerment is manifested in connection capabilities (people to people and people to objects), intelligence capabilities (user behavior perception), and analysis capabilities (information exchange), and the realization of data empowerment can promote the mutual cooperation and resource integration between both parties and realize the value co-creation of data-based travel platform enterprises. It can be seen that the data empowerment of platform enterprises can enhance the interaction between the enterprises and the customers and promote the integration of resources between the two parties, achieving mutual value. Since logistics platform enterprises have similar operating mechanisms and processes with platform enterprises, the data empowerment of logistics service platforms has a positive effect on the realization of value co-creation. Thus, the following hypotheses are proposed:

H2: data empowerment has a positive effect on the interaction and cooperation of logistics platform enterprises.

H3: data empowerment has a positive effect on the resource integration of logistics platform enterprises.

2.4. Value Co-creation and Service Innovation Capability. Under the influence of value co-creation, enterprises enhance the openness of the service innovation platform, enlarge the scope of participation, and improve the voluntariness of participation and effort of the behavioral subjects in the process of service innovation, so as to enhance the interaction between the users and enterprises, which has a positive impact on the improvement of the service innovation capability of enterprises [38]. The participants corresponding to the platform enterprises with a higher degree of openness tend to be more active in participating in innovation activities, and the interaction and cooperation between enterprises and users are more frequent. The high degree of interactive cooperation is conducive to promoting platform enterprises to improve their service innovation capabilities. Platform openness offered vast opportunities for creating value that goes beyond enterprises' core competencies. Platform enterprises can make full use of the advantages of the platform architecture, integrate complementary assets in the ecosystem, and establish an ecological governance mechanism and a mutually beneficial co-creation mechanism, fully activating the potential of platform users and achieving collaborative innovation effects [39]. If enterprises can effectively integrate its internal resources, external resources, or both internal and external resources at the same time, it can gain competitive advantage to achieve value creation. Therefore, interactive cooperation and resource integration have a positive effect on the improvement of service innovation capabilities of platform enterprises. Thus, the following hypotheses are proposed:

H4: interactive cooperation has a positive effect on the service innovation capability of logistics platform enterprises.

H5: resource integration has a positive effect on the service innovation capability of logistics platform enterprises.

2.5. The Mediating Effect of Value Co-creation. The Didi Chuxing platform, representing for the Internet big data platform, interacts and cooperates with other social networking platforms such as WeChat and Ganji.com through the application of data technology capabilities, to achieve a win-win situation for both parties. It has a strategic merger with Kuaidi and realizes the resource integration of mobile traveling platform, promoting the service innovation of platform enterprises. Platform enterprises can effectively use the data capabilities of the Internet to screen and process a wide range of information resources, break down the barriers of communication between enterprises and users through information resource integration, reduce communication costs, improve communication and decision-making efficiency, and realize the improvement of enterprises services innovation capabilities. Based on the effective use of big data, the logistics service platform enhances the interaction and cooperation with users, selects and integrates scattered information resources, reduces the communication cost of enterprises, enhances the management and decision-making capabilities of enterprises, and provides users with more suitable products and services. The service innovation capabilities of logistics platform enterprises have been improved. Therefore, interactive cooperation and resource integration play an intermediary role in the data empowerment and service innovation capabilities of logistics platform enterprises. Thus, the following hypotheses are proposed:

H6: interactive cooperation plays an intermediary role between data empowerment and logistics platform enterprise service innovation capabilities.

H7: resource integration plays an intermediary role between data empowerment and logistics platform enterprise service innovation capabilities.

2.6. The Moderating Effect of Environmental Dynamism. Environmental dynamism is manifested as changes in the market environment and market conditions, which can specifically be regarded as fluctuations in consumer demand in the market, adaptations in the marketing strategies of competitors, and the emergence of new technologies. When facing a complex market environment and uncertain changes, enterprises will adapt to the market environment in which they are located, and by examining the causes of environmental changes, they will improve the organization's change and innovation capabilities [40]. Environmental dynamism is manifested not only in the variability of the external market environment, but also in the variability of different stages of the enterprises' developing cycles. In different life stages of an enterprise in the development process, the impact of environmental dynamism on its service innovation capabilities will have different regulatory effects [41].

Under high environmental dynamism, the interactive cooperation between enterprises and consumers are more frequent. When the enterprise's own knowledge reserves fail to meet the requirements of service innovation capabilities, it can achieve value creation by learning external knowledge through interactive cooperation with consumers and suppliers and also enhance service innovation capabilities through the interactive cooperation between the two parties. Therefore, environmental dynamism has a positive moderating effect on knowledge sharing and service innovation capabilities [42]. In addition to positive adjustment, the moderating effect of environmental dynamism may also be U-shaped and inverted U-shaped [43]. When the environment becomes more turbulent, the impact on the enterprise is, in turn, uncertain or difficult to estimate. Since a considerable number of platform enterprises are still in the initial stage of corporate development, and limited by their own capabilities, high environmental dynamism may negatively regulate their service innovation capabilities.

With the development of platform enterprises, environmental volatility (market turbulence, competition turbulence, and technological turbulence) has become a norm in platform service industries. With the continuous improvement of environmental dynamism, when facing the intensified competition among enterprises and changes of user requirements, an enterprise can only better enhance its service innovation ability by satisfying the needs of users in a timely manner. Therefore, high environmental dynamism plays a moderating effect between interactive cooperation, resource integration, and platform enterprise service innovation capabilities. Thus, the following hypotheses are proposed:

H8: high environmental dynamism plays a positive regulatory effect between interactive cooperation and logistics platform enterprise service innovation capabilities.

H9: high environmental dynamism plays a positive regulatory effect between resource integration and logistics platform enterprise service innovation capabilities.

In summary, the theoretical model constructed in this study is shown in Figure 1.

3. Research Design

3.1. Data Source. According to principles of typical representativeness, data collection, and research matching for logistics platform enterprises, this study selected logistics platform enterprises, Yunmanman, Lalamove, and Truck Alliance for research, and the respondents of the survey were managers, employees, drivers, and consumers of the platform enterprises. The items of the questionnaire adopt mature scale items that domestic and foreign scholars have and were measured using 7-point Likert scale. Based on the small-scale preinvestigation and testing, the scale was appropriately revised and adjusted in combination with the management practices of China's logistics platform enterprises.

160 questionnaires were distributed in the preinvestigation, and 82 of returned 143 were valid. The questionnaire response rate was 89.4%, and the questionnaire validity rate was 57.3%. According to the presurvey test results, the questionnaire items are revised and adjusted to form a formal questionnaire. The formal survey issued 838 questionnaires, of which 615 were returned and 360 were valid. The questionnaire response rate was 73.4% and the questionnaire validity rate was 58.5%, which met the questionnaire collection indicators.

3.2. Variable Measurement. The measurement of data empowerment starts from the three aspects, intelligence ability, connection ability, and analysis ability, and it proposes 6 items for data empowerment. For the measurement of value co-creation, this research is conducted from two aspects: interactive cooperation and resource integration. The measurement of interactive cooperation is considered from three aspects. As logistics platform enterprises are dependent on the Internet, their operation patterns are mainly realized by connecting the two parties of the transaction. The existing fixed and basic resources are limited, and the development of the platform mainly depends on pioneering resources. The measurement of resource integration is considered from three items. Since the realization of service innovation capability is based on the enterprises' resource acquisition and the ability to capture and utilize external opportunities, the measurement of logistics platform enterprise service innovation capability is not much different from that of the enterprise service innovation capability. Therefore, this research measures the service innovation capability of logistics platform enterprises from four aspects. The existing logistics platform enterprises rely on the Internet, experiencing high volatile industry policies and markets, and inadequate consumer loyalty. Meanwhile, entry barriers of enterprises are relatively low, and competitors are prone to appear. The industry is highly competitive and various ways of competition are emerging. In addition, the Internet data platform where logistics platform enterprises are located presents fast and frequent technological development and updates. Therefore, this research measures environmental dynamism from three aspects: market volatility, competition intensity, and technological volatility. The results of measurement item are shown in Table 1.

4. Data Analysis and Hypothesis Testing

4.1. Reliability and Validity Analysis. Using Cronbach'α value to test the reliability of each variable, the data were analyzed with the help of Amos21.0 and Spss23.0, and the reliability and validity indicators of related measurement variables were obtained as shown in Table 2. It shows that both the factor loading and Cronbach'α coefficient of each measured variable exceed 0.7, indicating that the reliability of the scale is solid, while the KMO value of each variable is between 0.6 and 0.8, which can be used for factor analysis.

With the help of Amos21.0 to conduct confirmatory factor analysis, the results show that the five-factor model has ideal fitting indexes, with high goodness of fit for the research model. Details are shown in Table 3 (index value).

4.2. Direct Effect Test. This study assumes that H1 believes data empowerment has a positive effect on service innovation capabilities. Therefore, based on the direct effect model M1, the relevant fitting index of the direct effect model is calculated by using the Amos21.0. It can be seen from Table 3 that the various fitting indicators of the direct effect model are relatively fine, indicating that the direct effect has a higher goodness of fit. The direct effect model M1 only considers the direct effect of data empowerment on service innovation capabilities, and further analysis and testing using Spss23.0 showed that the correlation coefficient between them was 0.846, p value was less than 0.001, and the t value was 30.075. Considering the criterion that the t value is greater than 1.96, R^2chge is close to 1, and the Bootstrap interval test result does not contain 0, it is shown that data empowerment has a positive effect on the service innovation ability, and hence the hypothesis H1 is verified. The test results are shown in Table 4.

4.3. Mediating Effect Test. For this research, the mediating effect of the model should be tested first, followed by the moderating effect of the model. For the test of the mediation model, this study separately verified the interactive cooperation and resource integration. On the basis of the mediation models M2 and M3, the relevant fitting index of mediating effect was analyzed by using Amos21.0. Detailed values are shown in Table 3. It can be seen from Table 3 that mediating effect has a high goodness of fit. It can be seen from Table 4 that the mediating effect model M2 adds interactive cooperation to the direct effect model M1. It is found that the correlation coefficient (β) of data empowerment to interactive cooperation is 0.828^{***}, while the t value is 27.908. The correlation coefficient of interactive

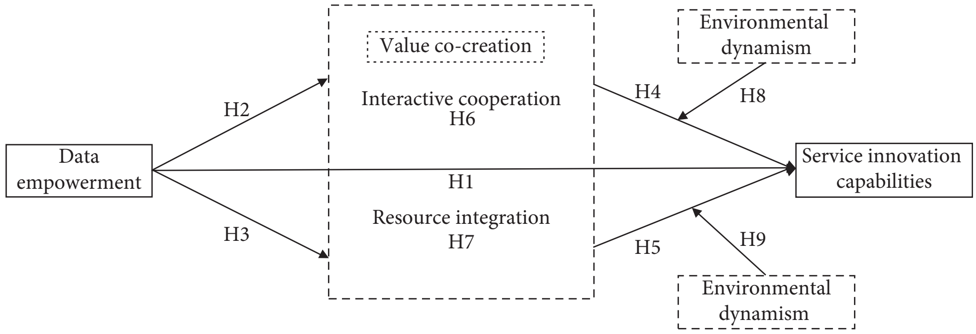


Figure 1: Theoretical model.

Table 1: Variable measurement item.

Variable		Measurement item
Data empowerment	Intelligence capability 1	The platform has strong software development and vulnerability repair capabilities
	Intelligence capability 2	The platform can flexibly classify products and provide corresponding products for different customers
	Connectivity capability 1	The platform can provide other related product or service information for customers
	Connectivity capability 2	The platform interface or system can be accessed from other ports
	Analysis capability 1	The platform can analyze business conditions and exploit emerging markets based on the results
	Analysis capability 2	The platform can combine the location of the user and match the order to the most suitable driver through analysis
Cooperative interaction	Cooperative interaction 1	During the operation of the platform, users should raise requests to the platform, engage in dialogue at any time, and actively participate in interaction
	Cooperative interaction 2	During the operation of the platform, the platform will conduct survey visits to users and pass on relevant information
	Cooperative interaction 3	During the operation, the platform allows registered drivers to interact with consumers
Resources integration	Resources integration 1	The platform can combine and utilize new resources
	Resources integration 2	The platform can use innovative methods to combine resources to expand business
	Resources integration 3	The platform can realize the creative combination of new resources and existing resources
Services innovation capability	Services innovation capability 1	The platform setup service innovation process
	Services innovation capability 2	The platform has the ability to respond to changing market conditions
	Services innovation capability 3	The platform values the new service innovation projects
	Services innovation capability 4	The platform uses new information technology to promote information sharing
Environmental dynamism	Environmental dynamism 1	The platform develops targeted marketing strategies to cater to different consumers
	Environmental dynamism 2	The industry that the platform is located in is highly competitive, with price wars or similar products appearing from time to time
	Environmental dynamism 3	The technology development and innovation of the industry that the platform is located in is fast

cooperation on service innovation capability is 0.847^{***}, and the t value is 30.179. According to the criteria that the t value >1.96 and $p < 0.001$, the hypotheses H2 and H4 are verified.

When interactive cooperation is involved in the relationship between data empowerment and service innovation capabilities, if the significant relationship between data empowerment and service innovation capabilities disappears, it can be assumed that interactive cooperation plays a completely mediating role between data empowerment and service innovation capabilities. If the relationship between data empowerment and service innovation capability is significant, it can be assumed that interactive cooperation plays a part of the mediating role between data empowerment and service innovation capability. The research results (from Figure 2 and Table 4) show that the correlation coefficient and t value between data empowerment and service innovation ability are reduced from $\beta = 0.846^{***}$ and $t = 30.075$ to $\beta = 0.461^{***}$ and $t = 10.533$, respectively, and the

Table 2: Reliability and validity analysis of measured variable.

Variable	Item	Factor loading	Cronbach'α coefficient	KMO value
Data empowerment	Intelligence capability 1	0.811	0.873	0.868
	Intelligence capability 2	0.786		
	Connectivity capability 1	0.791		
	Connectivity capability 2	0.697		
	Analysis capability 1	0.830		
	Analysis capability 2	0.776		
Cooperative interaction	Cooperative interaction 1	0.877	0.813	0.671
	Cooperative interaction 2	0.802		
	Cooperative interaction 3	0.880		
Resources integration	Resources integration 1	0.830	0.763	0.670
	Resources integration 2	0.767		
	Resources integration 3	0.874		
Services innovation capability	Services innovation capability 1	0.824	0.831	0.776
	Services innovation capability 2	0.772		
	Services innovation capability 3	0.831		
	Services innovation capability 4	0.831		
Environmental dynamism	Environmental dynamism 1	0.889	0.801	0.673
	Environmental dynamism 2	0.769		
	Environmental dynamism 3	0.878		

Table 3: Model fitting coefficient.

Fitting coefficient	χ^2	df	χ^2/df	RMR	RMSEA	GFI	AGFI	NFI	CFI
Index value	426.744	142	3.005	0.043	0.075	0.881	0.840	0.913	0.940
Direct effect model	79.15	28	2.827	0.035	0.071	0.957	0.915	0.961	0.974
Mediating effect model	315.1	99	3.18	0.042	0.078	0.897	0.859	0.920	0.943
Criteria			2~5	<0.08	<0.08	≥0.90	≥0.80	≥0.90	≥0.90

Table 4: Test result.

Hypothesis	Coeff.	t value	R^2chge	F chge	Bootstrap95% (1000 times)		Result
					Lower limit	Upper limit	
H1	0.846***	30.075	0.716	904.485***	0.789	0.957	Valid
H2	0.828***	27.908	0.685	778.881***	0.813	0.984	Valid
H4	0.847***	30.179	0.718	910.769***	0.828	0.949	Valid
H6	0.466***	10.643	0.785	650.720***	0.283	0.588	Valid
H3	0.828***	27.921	0.685	779.577***	0.745	0.923	Valid
H5	0.869***	33.241	0.755	1104.982***	0.827	0.937	Valid
H7	0.535***	12.894	0.807	744.177***	0.382	0.686	Valid

*** indicates $p < 0.001$; ** indicates $p < 0.01$; * indicates $p < 0.05$.

F value is reduced from 904.485*** under the direct effect M1 to 650.720***. For the mediation effect of interactive cooperation, $\beta = 0.466^{***}$, $t = 10.643$, and R^2chge is 0.785. The results indicate that although the relationship between data empowerment and service innovation ability is weakened, the correlation between these two is still significant, and the t value is reduced but still greater than 1.96. It indicates that interactive cooperation plays a part of the mediating role between data empowerment and service innovation capabilities. Hypothesis H6 is verified. In summary, after the direct effect M1 is added into interactive cooperation, the mediating effect of M2 is still significant. The above hypotheses of H2, H4, and H6 are all established.

After resource integration is added into the mediating effect model M3 on the basis of the direct effect M1 and tested by Spss23.0, the correlation coefficient of data empowerment to resource integration is 0.828*** and the t value is 27.921. The correlation coefficient of resource integration on service innovation capability is 0.869***, and the t value is 33.241. According to the standard of t value greater than 1.96 and $p < 0.001$, hypotheses H3 and H5 have been verified.

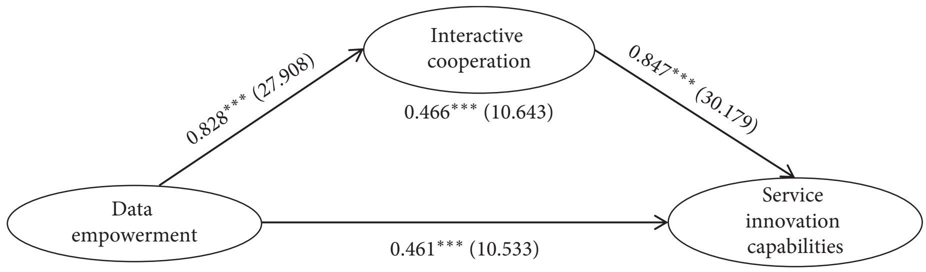


Figure 2: Mediating effect model M2.

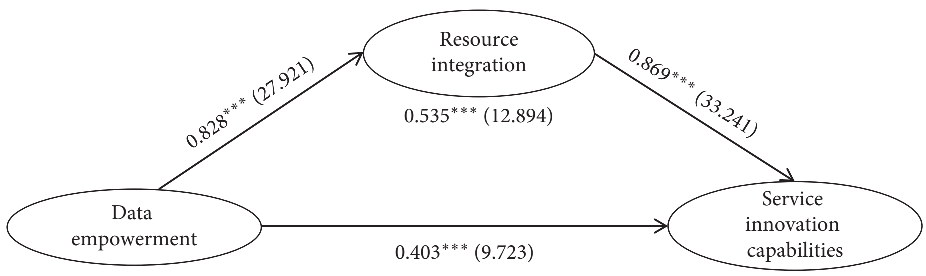


Figure 3: Mediating effect model M3.

Table 5: Test result of the moderating effect of environmental dynamism.

Hypothesis	Path	R^2chge	Coeff	t value	Result
H8	Cooperative interaction × environmental dynamism⟶ services innovation capability	0.003	−0.341	2.046*	Not valid
H9	Resources integration × environmental dynamism⟶ services innovation capability	0.001	−0.260	1.630	Not valid

When resource integration is involved in the relationship between data empowerment and service innovation capabilities, if the significant relationship between data empowerment and service innovation capabilities disappears, it can be considered that resource integration plays a completely intermediary role between data empowerment and service innovation capabilities. If the relationship between data empowerment and service innovation capability is significant, it can be considered that resource integration plays a part of the intermediary role between data empowerment and service innovation capability. According to the results (from Figure 3 and Table 4), the correlation coefficient and t value between data empowerment and service innovation ability are reduced from $\beta = 0.846^{***}$ and $t = 30.075$ when the direct effect is M1 to $\beta = 0.403^{***}$ and $t = 9.723$, respectively, and the F value is reduced from 904.485^{***} to 744.177^{***}. The mediating effect of interactive cooperation is $\beta = 0.535^{***}$, $t = 12.894$, and R^2chge is 0.807. The results show that although the relationship between data empowerment and service innovation capability is weakened, the correlation between the two is still significant. Although the t value is reduced, it is still greater than 1.96, indicating that resource integration plays a part of the intermediary between data empowerment and service innovation capability effect, and hence Hypothesis H7 is verified. In summary, after the direct effect model M1 is added to resource integration, the test result of the mediation effect M3 is still significant. Hypotheses H3, H5, and H7 are all established. The hypothesis test results are shown in Table 4.

4.4. Moderating Effect Test. The moderating effect of this study is mediated moderation, which is to test the moderating role of environmental dynamism in interactive cooperation and service innovation capabilities, resource integration, and service innovation capabilities. The path coefficients of the moderating effects are obtained by Spss23.0 analysis, as shown in Table 5. From the results of the path coefficient, the interaction coefficient between the interaction of interactive cooperation and environmental dynamism and the service innovation capability is −0.341, and the t value is 2.046. The interaction coefficient between the interaction of resource integration and environmental dynamism and the service innovation capability is −0.260, and the t value is 1.630. Both interaction coefficients are negative, indicating that environmental dynamism failed to positively mediate the relationship between interactive cooperation and service innovation capabilities, resource integration, and service innovation capabilities and even inhibits the relationship between value cocreation and service innovation capabilities. Therefore, the hypotheses H8 and H9 have not been verified.

5. Conclusions and Implications

5.1. Conclusions. This research constructs a theoretical model that takes the intermediary variables of value co-creation and environmental dynamism as the moderating variables and divides the intermediary variables of value co-creation into interactive cooperation and resource integration. It discusses the mechanism of data empowerment on the service innovation capability of logistics platform enterprises and analyzes the moderating effect of environmental dynamism (market turbulence, competition turbulence, and technological turbulence) on value co-creation and service innovation capabilities in detail. Empirical analysis shows that data empowerment not only directly affects the service innovation capability of logistics platform enterprises but also indirectly affects service innovation capability through the intermediary effect of value co-creation (interactive cooperation and resource integration). In addition, environmental dynamism does not have a positive moderating effect on interaction and cooperation, resource integration, and service innovation capabilities but exhibits a certain inhibitory effect instead.

5.2. Implications for Practice. The research conclusion expands the theory of data empowerment and has important implications for the practice of logistics platform enterprises.

(1) Logistics platform enterprises should give full play to their intelligence, connectivity, and analysis capabilities to form their own data empowerment. Logistics platform enterprises should strengthen and make good use of their intelligence, connectivity, and analysis capabilities to keep abreast of data development trends and quickly find information conducive to their own development, so as to enhance their service innovation capabilities to gain a competitive advantage.

(2) It is empirically verified that data empowerment can affect service innovation capabilities by influencing value co-creation. This influence mechanism has an exemplary significance for improving the service innovation capabilities of logistics platform enterprises. Logistics platform enterprises can improve service innovation capabilities through effective interactive cooperation and resource integration with the help of data empowerment.

(3) When making decisions, logistics platform enterprises should consider not only their own capabilities and conditions, but also their actual tolerance to the external environment. With the rapid application and development of the Internet, the dynamism of the environment becomes more and more significant and becomes a variable factor that cannot be ignored in corporate decision-making. Under the high environmental dynamism, enterprises should consider their own capabilities, appropriately carry out or maintain interaction and cooperation with users, and focus their energy and resources on improving their own capabilities. Enterprises should create and develop their own logistics resources, meet user demands in a timely and efficient manner, and enhance the service innovation capabilities of platform enterprises.

5.3. Limitations and Future Research. As with most researches, the design of this research is subject to limitations, which opens up opportunities for future research. Firstly, we collect data from specific logistics platform enterprises; there may be difference for the applicability of the model to platform enterprises in other industries, and further research should extend our model to other industries. Secondly, the informant bias could be a concern, as only managers, employees, drivers, and consumers completed the survey. Future research could attempt to avoid such concerns by recruiting multiple informants, e.g., senior managers. Finally, our research is cross-sectional, which limits the test of the causal inferences for platform enterprise data empowerment and service innovation capabilities. Due to the fact that practicing platform enterprise data empowerment and developing enterprise service innovation capabilities require enterprises to continuously create bundles of new resources and knowledge, future research should examine the co-evolution between data empowerment and service innovation capabilities with a panel data research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was financially supported by the National Natural Science Foundation of China (no. 71974039).

References

[1] A. Tiwana, "Evolutionary competition in platform ecosystems," *Information Systems Research*, vol. 26, no. 2, pp. 266–281, 2015.

[2] D. S. Evans, "Attention platforms, the value of content, and public policy," *Review of Industrial Organization*, vol. 54, no. 4, pp. 775–792, 2019.

[3] A. T. Johnsen, N. B. Eskildsen, T. G. Thomsen, M. Grønvold, L. Ross, and C. R. Jørgensen, "Conceptualizing patient empowerment in cancer follow-up by combining theory and qualitative data," *Acta Oncologica*, vol. 56, no. 2, pp. 232–238, 2017.

[4] C. K. H. Wong, C. L. K. Lam, E. Y. F. Wan et al., "Evaluation of patient-reported outcomes data in structured diabetes education intervention: 2-year follow-up data of patient empowerment programme," *Endocrine*, vol. 54, no. 2, pp. 422–432, 2016.

[5] J. Heckert and M. S. Fabic, "Improving data concerning women's empowerment in sub-saharan africa," *Studies in Family Planning*, vol. 44, no. 3, pp. 319–344, 2013.

[6] P. Pierre and S. J. Floyd, "Refining the conceptualization and measurement of women's empowerment in sub-saharan africa using data from the 2013 nigerian demographic and

health survey," *Social Indicators Research*, vol. 140, no. 2, pp. 777–793, 2018.
[7] M. W. Alsem, M. Verhoef, J. Braakman et al., "Parental empowerment in paediatric rehabilitation: exploring the role of a digital tool to help parents prepare for consultation with a physician," *Child Care Health*, vol. 45, no. 12, pp. 623–636, 2019.
[8] M. N. Hossain, M. S. Talukder, M. R. Hoque, and Y. Bao, "The use of open government data to citizen empowerment: an empirical validation of a proposed model," *Foresight*, vol. 20, no. 6, pp. 665–680, 2018.
[9] Y. Tim, S. L. Pan, S. Bahri, and A. Fauzi, "Digitally enabled affordances for community-driven environmental movement in rural Malaysia," *Information Systems Journal*, vol. 28, no. 1, pp. 48–75, 2017.
[10] B. Akkoyunlu and A. Yilmaz, "Prospective teachers' digital empowerment and their information literacy self-efficacy," *Eurasian Journal of Educational Research*, vol. 44, no. 11, pp. 33–50, 2011.
[11] C. Costa, M. Murphy, and A. L. Pereira, "Higher education students' experiences of digital learning and (dis) empowerment," *Australasian Journal of Educational Technology*, vol. 34, no. 3, pp. 140–152, 2018.
[12] D. J. Teece, G. Pisano, and A. Shuen, "Dynamic capabilities and strategic management," *Strategic Management Journal*, vol. 18, no. 7, pp. 509–533, 1997.
[13] J. B. Barney, D. J. Ketchen Jr., and M. Wright, "The future of resource-based theory: revitalization or decline?" *Journal of Management*, vol. 37, no. 5, pp. 1299–1315, 2011.
[14] S. Lenka, V. Parida, and J. Wincent, "Digitalization capabilities as enablers of value co-creation in servitizing firms," *Psychology & Marketing*, vol. 34, no. 1, pp. 92–100, 2017.
[15] L. Z. Song, M. Song, and C. A. Di Benedetto, "A staged service innovation model," *Decision Sciences*, vol. 40, no. 3, pp. 571–599, 2009.
[16] B. B. Flynn, B. Huo, and X. Zhao, "The impact of supply chain integration on performance: a contingency and configuration approach," *Journal of Operations Management*, vol. 28, no. 1, pp. 58–71, 2010.
[17] L. A. Bettencourt, S. W. Brown, and N. J. Sirianni, "The secret to true service innovation," *Business Horizons*, vol. 56, no. 1, pp. 13–22, 2013.
[18] P. den Hertog, W. van der Aa, and M. W. de Jong, "Capabilities for managing service innovation: towards a conceptual framework," *Journal of Service Management*, vol. 21, no. 4, pp. 490–514, 2010.
[19] D. Kindström, C. Kowalkowski, and E. Sandberg, "Enabling service innovation: a dynamic capabilities approach," *Journal of Business Research*, vol. 66, no. 8, pp. 1063–1073, 2013.
[20] N. Evers, S. Andersson, and M. Hannibal, "Stakeholders and marketing capabilities in international new ventures: evidence from Ireland, Sweden, and Denmark," *Journal of International Marketing*, vol. 20, no. 4, pp. 46–71, 2012.
[21] A. Ganguly, A. Talukdar, and D. Chatterjee, "Evaluating the role of social capital, tacit knowledge sharing, knowledge quality and reciprocity in determining innovation capability of an organization," *Journal of Knowledge Management*, vol. 23, no. 6, pp. 1105–1135, 2019.
[22] C. A. Lengnick-Hall and R. J. Griffith, "Evidence-based versus tinkerable knowledge as strategic assets: a new perspective on the interplay between innovation and application," *Journal of Engineering and Technology Management*, vol. 28, no. 3, pp. 147–167, 2011.
[23] L. J. Menor and A. V. Roth, "New service development competence in retail banking: construct development and measurement validation," *Journal of Operations Management*, vol. 25, no. 8, pp. 825–846, 2007.
[24] T. Bauer and K. Borodako, "Trade show innovations - organizers implementation of the new service development process," *Journal of Hospitality and Tourism Management*, vol. 41, no. 12, pp. 197–207, 2019.
[25] G. Akman and C. Yilmaz, "Innovative capability, innovation strategy and market orientation: an empirical analysis in Turkish software industry," *International Journal of Innovation Management*, vol. 12, no. 1, pp. 69–111, 2008.
[26] N. Lindberg and F. Nordin, "From products to services and back again: towards a new service procurement logic," *Industrial Marketing Management*, vol. 37, no. 3, pp. 292–300, 2008.
[27] D. Kafetzopoulos, K. Gotzamani, and V. Gkana, "Relationship between quality management, innovation and competitiveness. evidence from Greek companies," *Journal of Manufacturing Technology Management*, vol. 26, no. 8, pp. 1177–1200, 2015.
[28] H. Perks, T. Gruber, and B. Edvardsson, "Co-creation in radical service innovation: a systematic analysis of microlevel processes," *Journal of Product Innovation Management*, vol. 29, no. 6, pp. 935–951, 2012.
[29] C. Jaw, J.-Y. Lo, and Y.-H. Lin, "The determinants of new service development: service characteristics, market orientation, and actualizing innovation effort," *Technovation*, vol. 30, no. 4, pp. 265–277, 2010.
[30] C. M. Wallenburg, "Innovation in logistics outsourcing relationships: proactive improvement by logistics service providers as a driver of customer loyalty," *Journal of Supply Chain Management*, vol. 45, no. 2, pp. 75–93, 2009.
[31] Y. C. Chen, P. C. Li, and K. R. Evans, "Effects of interaction and entrepreneurial orientation on organizational performance: insights into market driven and market driving," *Industrial Marketing Management*, vol. 41, no. 6, pp. 1019–1034, 2012.
[32] S. J. Hogan, G. N. Soutar, J. R. McColl-Kennedy, and J. C. Sweeney, "Reconceptualizing professional service firm innovation capability: scale development," *Industrial Marketing Management*, vol. 40, no. 8, pp. 1264–1273, 2011.
[33] Cornell Empowerment Group, "Empowerment and family support," *Networking Bulletin*, vol. 1, no. 2, pp. 1–23, 1989.
[34] D. Dougherty and D. D. Dunne, "Digital science and knowledge boundaries in complex innovation," *Organization Science*, vol. 23, no. 5, pp. 1467–1484, 2012.
[35] C. K. Prahalad and V. Ramaswamy, "Co-creation experiences: the next practice in value creation," *Journal of Interactive Marketing*, vol. 18, no. 3, pp. 5–14, 2004.
[36] E. Gummesson and C. Mele, "Marketing as value co-creation through network interaction and resource integration," *Journal of Business Market Management*, vol. 4, no. 4, pp. 181–198, 2010.
[37] C. Neghina, M. C. J. Caniëls, J. M. M. Bloemer, and M. J. H. van Birgelen, "Value cocreation in service interactions," *Marketing Theory*, vol. 15, no. 2, pp. 221–242, 2015.
[38] N. O. Ommen, M. Blut, C. Backhaus, and D. M. Woisetschläger, "Toward a better understanding of stakeholder participation in the service innovation process: more than one path to success," *Journal of Business Research*, vol. 69, no. 7, pp. 2409–2416, 2016.
[39] W. Fu, Q. Wang, and X. Zhao, "Platform-based service innovation and system design: a literature review," *Industrial*

Management & Data Systems, vol. 118, no. 5, pp. 946–974, 2018.

[40] I. P. Mccarthy, T. B. Lawrence, B. Wixted, and B. R. Gordon, “A multidimensional conceptualization of environmental velocity,” *Academy of Management Review*, vol. 35, no. 4, pp. 604–626, 2010.

[41] C. Dibrell, J. Craig, and E. Hansen, “Natural environment, market orientation, and firm innovativeness: an organizational life cycle perspective,” *Journal of Small Business Management*, vol. 49, no. 3, pp. 467–489, 2011.

[42] Z. Yang, V. T. Nguyen, and P. B. Le, “Knowledge sharing serves as a mediator between collaborative culture and innovation capability: an empirical research,” *Journal of Business & Industrial Marketing*, vol. 33, no. 7, pp. 958–969, 2018.

[43] G. Li, W. Yaru, and S. Feng, “A study on the relationship among strategic orientation, absorption capacity and service innovation performance: the regulatory role of environmental uncertainty,” *Journal of North China University of Water Resources and Electric Power (Social Science Edition)*, vol. 35, no. 5, pp. 24–32, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5582067, 10 pages
https://doi.org/10.1155/2021/5582067

Research Article

Exploiting the Chain Convenience Store Supplier Selection Based on ANP-MOP Model

Yaqin Ou [1] and Bo Liu [2]

[1]*School of Business Administration, Chaohu College, Chaohu 238000, China*
[2]*School of Information Engineering, Chaohu College, Chaohu 238000, China*

Correspondence should be addressed to Bo Liu; 623321161@qq.com

Received 29 January 2021; Revised 26 February 2021; Accepted 6 May 2021; Published 19 May 2021

Academic Editor: Zong-Yao Sun

The purchase cost of chain convenience stores accounts for a large proportion of the total cost. Enterprises are facing the problem of how to reasonably reduce the purchase cost while ensuring the quality of service. This paper first considers the uniqueness of chain convenience stores, draws on existing research results, and establishes an evaluation index system for chain convenience store suppliers through field research. Then, we use the principal component analysis method of selection to determine the weight of each index and make a preliminary ranking of the importance of chain convenience store suppliers. Secondly, according to the relevant weights determined by the network hierarchy method, the collected data are substituted into the multiobjective mathematical programming model to analyze the distribution of supplier procurement and calculate the procurement cost. The level of procurement expenditures accounted for by suppliers confirms the importance of suppliers. The results show that through the evaluation and ranking of suppliers and the effective management of suppliers according to the ranking results, not only the goal of cost minimization is achieved but also the reasonable service level is guaranteed, which is scientific.

1. Introduction

The beginning of the entire supply chain management is the selection of suppliers. Among all types of enterprises, the purchase cost of chain retail accounts for the largest total cost. The current research on supplier evaluation and selection in chain retail is not thorough enough. Especially in chain retail, the research on supplier selection based on the two major factors of procurement cost and service level, procurement allocation, and follow-up management is very much lacking [1–3].

Guo used AHP to conduct preliminary supplier selection and comprehensive supplier selection and determined its primary and alternative suppliers [4]. Liao and Zhang adopted the AHP-TOPSIS composite method to determine the weight of each index, thus constructing a standardized model for evaluating and selecting green suppliers for papermaking enterprises. And through the application of empirical examples, the feasibility of the way is verified, and an effective method is provided for papermaking enterprises to choose green suppliers [5]. Zhou and Liao used the analytic hierarchy process (AHP) to calculate the weight of a total of 14 indicators in 5 categories, including green environmental indicators, and built an objective and operable supplier evaluation process [6]. Song used the AHP and took candy products as an example to conduct an empirical analysis of the supplier evaluation index system [7]. From the current research on the evaluation of supplier selection by scholars, most scholars still use AHP to construct the supplier index system; still, AHP cannot fully consider the interrelationship between the various index systems. More importantly, most of the scholars' evaluation and selection of suppliers is aimed at manufacturing companies, ignoring the high proportion of chain retail purchase costs in total costs [8–12].

From the existing supply chain management research, according to the current supply chain management research, there is not much research on the selection management of chain retail suppliers whose procurement costs account for the total costs. Therefore, it is necessary to further explore the evaluation and selection of chain retail suppliers. This paper takes chain convenience stores in chain retail as an example to examine the evaluation choices of chain convenience store suppliers. The content of this paper is arranged as follows. In Section 2, ANP is described in detail and is used to construct, select, and determine the weight of the supplier's index system. In Section 3, firstly, MOP model is constructed based on ANP model; secondly, taking Rosen, the leader in chain convenience stores, as an example, the collected relevant data which are substituted into the MOP to determine the weight of its supplier evaluation index system and the allocation of procurement volume. The empirical results analysis is presented in Section 4, and a summary of this paper is given in Section 5.

2. ANP Model Construction

2.1. ANP Model Construction and Index Optimization. The elements $p_1, p_2, \ldots, p_n$are in the ANP control layer, and the elements $c_1, c_2, \ldots, c_n$ are in network layer, where c_1 has elements $e_{i1}, e_{i2}, \ldots, e_{in}$, where $i = 1, 2, \ldots, N$. The criterion is the control layer $p_s\,(s = 1, 2, \ldots, m)$, and the secondary criterion is the element $e_{jl}\,(l = 1, 2, \ldots, n_j)$. The elements in the element group c_i are compared according to their influence on e_{jl}, and the judgment matrix [13, 14] is constructed under the criterion ps, as shown in Table 1.

$$W_{ij} = \begin{bmatrix} w_{i1}^{(j1)} & w_{i1}^{(j2)} & \ldots & w_{i1}^{(jni)} \\ w_{i2}^{(j1)} & w_{i2}^{(j2)} & \ldots & w_{i2}^{(jni)} \\ & & \vdots & \\ w_{in_j}^{(j1)} & w_{in_j}^{(j2)} & \ldots & w_{in_j}^{(jni)} \end{bmatrix}, \quad (1)$$

$$\begin{array}{ccc} & & \begin{array}{ccccccccc} C_1 & & & C_2 & & & C_N & & \\ e_{11} & \ldots & e_{1n_1} & e_{21} & \ldots & e_{2n_2} & e_{N1} & \ldots & e_{Nn_n} \end{array} \\ & \begin{array}{c} e_{11} \\ \ldots \\ e_{1n_1} \\ e_{21} \\ \ldots \\ e_{2n_2} \\ e_{N1} \\ \ldots \\ e_{Nn_n} \end{array} & \\ W = \begin{array}{c} C_1 \\ \\ \\ C_2 \\ \\ \\ C_N \end{array} & & \begin{bmatrix} W_{11} & W_{12} & \ldots & W_{1N} \\ W_{21} & W_{22} & \ldots & W_{2N} \\ \vdots & \vdots & & \vdots \\ W_{N1} & W_{N2} & \ldots & W_{NN} \end{bmatrix}. \end{array} \quad (2)$$

The weight vector is obtained by the characteristic-root method. Repeat the above steps to obtain the matrix record as shown in formula (1):

The column vector of here is the sorting vector of the influence degree of the element in the element. If the middle element is not affected by the middle element, then = 0. Repeat the above steps for i = 1, 2, ..., N, j = 1, 2, ..., N. Finally, the supermatrix under the criterion can be obtained as equation (2). There are m of these supermatrices, and they are all nonnegative matrices.

Table 1: Judgment matrix between each element.

e_{ij}	$e_{i1}, e_{i2}, \ldots, e_{in_j}$	Normalized feature vector
e_{i1}		$w_{i1}^{(ji)}$
e_{i2}		$w_{i2}^{(ji)}$
$\vdots$		$\vdots$
e_{in_i}		$w_{in_i}^{(ji)}$

2.2. Construction and Optimization of Evaluation Index System. On the basis of summarizing the research of scholars, a questionnaire was designed in combination with the characteristics of chain convenience stores; after distribution, recovery, and calculation, the evaluation index system of chain convenience store procurement suppliers was initially constructed [9–13] (as shown in Table 2).

In response to the initially constructed indicator system, a questionnaire on the importance of indicators was designed and distributed to multiple experts to score them. Then, the questionnaire was returned to the minimum and maximum statistics of all indicators, and mean and standard deviation were calculated. By comparing and analyzing the results, the criterion layer and the index layer are selected to obtain a new and more realistic index system [14–17], and then the data on the importance of the criterion layer and the index layer are sorted out.

Table 3 shows the mean and standard deviation of the importance scores of the five categories of indicators by respondents. It can be seen from the table that the mean (8.13) of the price factor is the largest, which is almost close to the maximum value of 9.00, which also shows that price is the most crucial consideration. Secondly, the larger mean is cooperation ability (7.31), which shows that the cooperation ability is also vital. The mean of technical ability (7.11) and the mean of cooperation ability (7.31) are very close in size, so they are even more important. However, the internal competitiveness (3.11) and external competitiveness (4.07) have a smaller mean.

The standard deviation of price factor, cooperation ability, and internal competitiveness is relatively small. It can be seen that the respondents have basically no objection to the importance of these three indicators, while the standard deviation of external competitiveness and technical ability is relatively large, indicating that interviewers have a specific difference in their opinions on the importance of these two indicators.

In the same way, the subdivision secondary indicators of other standard levels are shown in Table 4.

2.3. Establishment of the Index Weights for Supplier Evaluation Index System. The following only lists the evaluation data of one group of five groups of experts on the pairwise comparison matrix of the criterion-level indicators. The evaluation data of all indicators are directly input into the software for calculation [18–21]. The calculation results are shown in Table 5. The indicators of the standard level in the software are technical capabilities B_1, cooperation capabilities B_2, external competitiveness B_3, internal competitiveness B_4, and price factors B_5.

Table 2: Evaluation index system for purchasing suppliers of chain convenience store.

Target layer	Criterion layer	Index layer	Indicator type
Supplier selection evaluation	Technical skills B1	Logistics facilities and equipment C1	Qualitative
		Information team processing ability C2	Qualitative
		Proportion of supply chain management professionals C3	Qualitative
		Scheme design ability C4	Qualitative
		Process integration capability C5	Qualitative
	Teamwork ability B2	Supply capacity C6	Qualitative
		Reliability C7	Qualitative
		After-sales service C8	Qualitative
		Delivery accuracy C9	Quantitative
		Quick response ability C10	Qualitative
	Internal competitiveness B3	Operating resources C11	Qualitative
		R & D capabilities C12	Qualitative
		Production management ability C13	Qualitative
		Organizational effectiveness analysis C14	Qualitative
		Operational capability C15	Qualitative
	External competitiveness B4	Market environment and competition C16	Qualitative
		Customer needs and feedback C17	Qualitative
		Economic environment C18	Qualitative
		Social environment C19	Qualitative
	Price factor B5	Transportation cost C20	Quantitative
		Product cost C21	Quantitative
		Storage cost C22	Quantitative
		Quality cost C23	Quantitative
		Out-of-stock cost C24	Quantitative

Table 3: Descriptive statistical analysis of criterion-level indicators.

	Technical skills	Teamwork ability	Internal competitiveness	External competitiveness	Price factor
Number	100	100	100	100	100
Minimum	4	5	2	3	7
Maximum	9	9	5	8	9
Mean (mean)	7.11	7.31	3.11	4.07	8.13
Standard deviation	0.877	0.803	0.810	1.143	0.802

Table 4: The optimized index system.

Target	Criterion layer	Index layer	Indicator type	Data sources
Supplier selection evaluation	Technical skills	Logistics facilities and equipment	Qualitative	Scoring by experts
		Proportion of supply chain management	Quantitative	Formula calculation
		Process integration capability	Qualitative	Scoring by experts
	Teamwork ability	Supply capacity	Qualitative	Scoring by experts
		Reliability	Qualitative	Scoring by experts
		Delivery accuracy	Quantitative	Formula calculation
	Internal competitiveness	Operating resources	Qualitative	Scoring by experts
		Production management ability	Qualitative	Scoring by experts
	External competitiveness	Organizational effectiveness analysis	Qualitative	Scoring by experts
		Customer needs and feedback	Qualitative	Scoring by experts
	Price factor	Transportation cost	Quantitative	Formula calculation
		Storage cost	Quantitative	Formula calculation
		Quality cost	Quantitative	Formula calculation
		Out-of-stock cost	Quantitative	Formula calculation

TABLE 5: Pairwise comparison matrix of expert group 1's evaluation of criterion-level indicators.

B_i	B_1	B_2	B_3	B_4	B_5	W_i
B_1	1	3	2	2	1/2	0.252
B_2	1/3	1	1	3	1/2	0.173
B_3	1/2	1	1	2	1/3	0.143
B_4	1/2	1/3	1/2	1	1/4	0.076
B_5	2	2	3	4	1	0.356

Note: the eigenvalue of the matrix is 5.200; the matrix consistency is 0.045 which is less than 0.1. Expert Group 1: = (0.252,0.173,0.143,0.076,0.356). Expert Group 2: = (0.220,0.228,0.099,0.065,0.388). Expert Group 3: = (0.187,0.348,0.063,0.107,0.295). Expert Group 4: = (0.241,0.199,0.118,0.072,0.370). Expert Group 5: = (0.241,0.115,0.076,0.199,0.369).

$$\begin{aligned}
&\text{Expert Group 1: } wi = (0.252, 0.173, 0.143, 0.076, 0.356)\\
&\text{Expert Group 2: } wi = (0.220, 0.228, 0.099, 0.065, 0.388)\\
&\text{Expert Group 3: } wi = (0.187, 0.348, 0.063, 0.107, 0.295)\\
&\text{Expert Group 4: } wi = (0.241, 0.199, 0.118, 0.072, 0.370)\\
&\text{Expert Group 5: } wi = (0.241, 0.115, 0.076, 0.199, 0.369)
\end{aligned}$$

Using the geometric mean, the final weight of the B-level indicators can be determined as

$$u_j = \left[\prod_{i=1}^{m} w_i\right]\frac{1}{m}, \qquad u_B = (0.231, 0.210, 0.105, 0.100, 0.354). \tag{3}$$

The same method is used to calculate the weighted hypermatrix of all indicators. From the limit matrix, all indicator weights between indicators can be obtained as shown in Table 6 and Figure 1.

The top seven indicators are product cost, out-of-stock cost, transportation cost, quality cost, delivery accuracy, supply capacity, and reliability.

TABLE 6: Index weights of the index layer.

Name	Weighted value	Limit value
Logistics facilities and equipment	0.30647	0.04045
Proportion of supply chain management	0.28752	0.04076
Process integration capability	0.40601	0.04385
Supply capacity	0.35325	0.05102
Reliability	0.22543	0.05023
Delivery accuracy	0.42132	0.06102
Operating resources	0.54124	0.00731
Production management ability	0.45876	0.03621
Organizational effectiveness analysis	0.44676	0.00821
Customer needs and feedback	0.55324	0.03205
Transportation cost	0.22432	0.13385
Storage cost	0.33492	0.24137
Quality cost	0.21975	0.11265
Out-of-stock cost	0.22101	0.14102

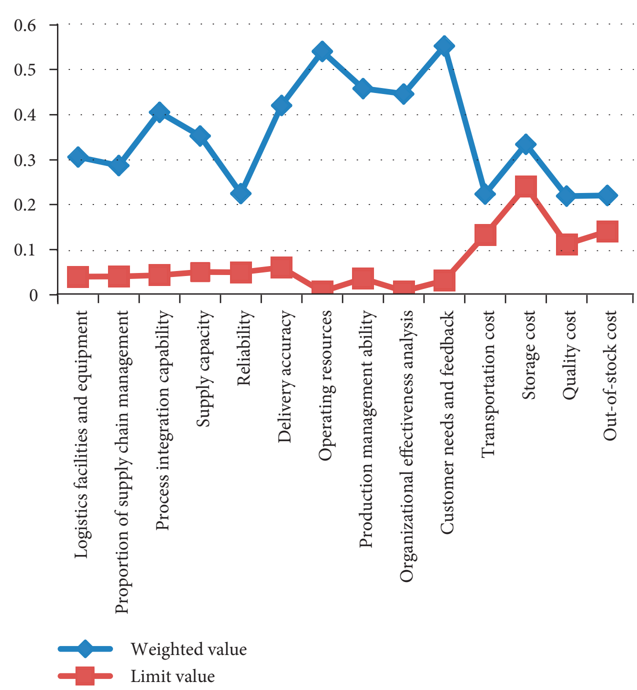


FIGURE 1: Index weight distribution chart of the index layer.

3. MOP Model Construction

We establish a specific model for chain convenience store purchasing supplier selection and purchasing volume allocation.

3.1. Meaning of Symbols.

3.1. Meaning of Symbols. The mathematical symbols of the following model are defined as follows:

(1) Two types of decision variables:

One is the allocation ratio $x_{ij}(0 < x_{ij} < 1)$ of the purchased quantity when purchasing goods from each candidate supplier g_i; the other is the variable $y_{ij} = 1$ that indicates whether the company purchases product j from a supplier g_i, if the company purchases product j from that supplier, then $y_{ij} = 1$, otherwise $y_{ij} = 0$, where i is the serial number of the supplier, i = 1,2, ..., m, j is the serial number of the purchased product, $j = 1, 2, \ldots, n$.

(2) Parameter description:

c_j: the total amount of the company's purchase of product j during the planning period; u_{ij}: the minimum purchase cost required for the product j purchased by the supplier g_i; k_{ij}: the maximum supply of product j provided by the supplier g_i; n_i: the company can choose the number of suppliers; n: the number of products that the company intends to purchase; q_{ij}: the importance factor of the products provided by the supplier g_i; j_{ij} the product cost of buying j products from the supplier g_i; j_i: the company expects to limit the maximum cost of the supplier's product j. t_{ij}: the out-of-stock cost when the g_i supplier provides the j product; t_j: the maximum out-of-stock cost when the company limits the g_i supplier to provide the j product; d_{ij}: the transportation of the j product purchased from the g_i supplier cost; d_j: the highest transportation cost of the company's purchase of j product limited by the g_i supplier; l_{ij}: the quality cost of the j product purchased by the g_i supplier; l_j: the highest quality cost of the company's limited g_i supplier j product; r_{ij}: g_i supplier j product delivery accuracy rate; r_j: represents the lowest delivery accuracy rate that the company expects supplier j's product; e_{ij}: g_i supplier j's product supply capacity; e_j: the company expects the minimum supply capacity of g_i supplier; n_{ij}: the reliability of the product provided by the supplier g_i; n_j: indicates the lowest limit of the reliability of the product expected by the supplier g_i to provide j; p_{ij}: the proportion of the purchase cost of the g_i supplier's purchase of j products.

3.2. Model Establishment. According to the concept and symbolic assumptions of cooperation with essential suppliers, the following MOP models for supplier selection can be established:

(1) Objective function:

$$\max Z_1 = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}x_{ij}, \tag{4}$$

$$\max Z_2 = \sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij}x_{ij}. \tag{5}$$

(2) Constraints:

St:

$$u_{ij} \le c_j x_{ij} \le k_{ij}, \tag{6}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}x_{ij} \le f_j, \tag{7}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} t_{ij}x_{ij} \le t_j, \tag{8}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij}x_{ij} \le d_j, \tag{9}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} l_{ij}x_{ij} \le l_j, \tag{10}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij}x_{ij} \le r_j, \tag{11}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} e_{ij}x_{ij} \le e_j, \tag{12}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij}x_{ij} \le r_j, \tag{13}$$

$$\sum_{i=1}^{m} x_{ij} = 1, \tag{14}$$

$$x_{ij}\left(y_{ij} - 1\right) = 0, \tag{15}$$

$$x_{ij} \in (0, 1), \tag{16}$$

$$y_{ij} \in (0, 1). \tag{17}$$

(3) Formula meaning:

Formula (4): represents the lowest purchasing cost.

Formula (5): represents the amount of priority to purchase from g_i suppliers when purchasing product j.

Formula (6): the upper and lower limits of the ability of supplier g_i to provide products.

Formula (7): the product cost constraint when supplier g_i purchases product j.

Formula (8): the constraint of the out-of-stock cost when supplier g_i purchases product j.

Formula (9): the transportation cost constraint when supplier g_i purchases product j.

Equation (10): constraint of quality cost when g_i supplier purchases product j.

Equation (11): constraint on the delivery accuracy of product j of g_i supplier.

Formula (12): requirements for the supply capacity of g_i supplier j's products.

Formula (13): g_i supplier provides product j reliability requirements.

Equation (14): when purchasing product j from the g_i supplier, y_{ij} must be 1.

4. Lawson Convenience Store Supplier Evaluation Selection

4.1. The Weight of Index System Determined Based on ANP Model. According to the indicator system constructed in the previous paper and the determination of related weights, the relevant data are substituted for the evaluation and selection of Lawson convenience store suppliers [22–26], and now 5 suppliers = $(g_1, g_2, g_3, g_4, g_5)$ are screened and one is selected as a partner of the company; a tripartite group $(p_1, p_2, p_3, p_4, p_5)$ is formed by a team of senior purchasers, a team of logistics managers, and a team of experts in related industries. The expert group decided to select the top seven indicators to evaluate the five suppliers. The weights of the tripartite experts in the decision-making are: senior purchaser team $\varpi = 0.5$, logistics manager team $\varpi = 0.3$, and related industry expert team$\varpi = 0.2$. In the evaluation index, the evaluation data of five quantitative indicators come from 5 procurement suppliers, where product cost, out-of-stock cost, quality cost, and transportation cost are the ratios of the average cost of the four products to the related expenses in the same industry, which are shown in Table 7 and Figure 2.

The tripartite experts score the supplier's performance under the qualitative indicators and the data of all qualitative indicators. The scoring adopts a nine-point system. The supplier's performance is represented by 9-1 from high to bottom. 9 points means very good, and 1 point represents extremely poor performance. The larger the score, the better the performance.

The three-party experts' specific scores on the 5 candidates for selection and the final scores after calculation are shown in Table 8 and Figure 3.

The first step is to determine the weights. In the previous paper, the results of ANP have been obtained and the weights of the seven evaluation indicators are

$$W = (0.24137, 0.14102, 0.13385, 0.11265, 0.06102, 0.05102, 0.05023). \tag{18}$$

The second step is to construct an evaluation matrix of five procurement suppliers under seven evaluation indicators in the software.

The third step is to define the attributes of the index evaluation. For each evaluation index, define its evaluation attribute and construct the evaluation attribute matrix for pairwise comparison.

The fourth step is to sort the suppliers, input the final score value of each supplier, and further comprehensively calculate the corresponding ranking result: $g_1 > g_2 > g_5 > g_3 > g_4$, and you can get 5 purchases. The supplier's importance coefficient is q = (0.23310, 0.23193, 0.18241,0.17214, 0.21423).

4.2. Purchase Volume Allocated Based on MOP

4.2.1. Data Collection. There are too many types of chain convenience stores. Here, 4 of them are selected as an example [27]. Substitute the 5 buyers qualified in the preliminary selection and sorted into the model and calculate and evaluate further. The parameter values and evaluation information of each supplier's supply capacity and level and the parameter values required by retail enterprises for the supplier's supply capacity and level are shown in Tables 9 and 10 .

4.2.2. Data Processing. According to the symbols established in formulas (4) to (17) and the supplier's importance coefficient established in the previous article, the MOP model that can be established is as follows:

$$\begin{aligned}
&\max z_1 = \left(0.12x_{11} + 0.11x_{21} + 0.15x_{32} + 0.14x_{52} + 0.13x_{23} + 0.12x_{33} + 0.11x_{14} + 0.12x_{44}\right)\\
&\max z_2 = 0.233x_{11} + 0.232x_{21} + 0.18x_{32} + 0.214x_{52} + 0.232x_{23} + 0.182x_{33} + 0.233x_{14} + 0.138x_{44}\\
&\text{s.t.}\quad \begin{cases}
9 \le 20x_{11} \le 15, 7 \le 20x_{21} \le 13, x_{32} \le 10, 8 \le x_{52} \le 11\\
5 \le 15x_{23} \le 12, 5 \le 20x_{21} \le 13, 6 \le 10x_{14} \le 8, 4 \le 10x_{44} \le 5\\
6x_{11} + 4x_{21} \le 7, 5x_{32} + 4x_{52} \le 5, 3x_{23} + 3x_{33} \le 6, 3x_{14} + 2x_{44} \le 5\\
x_{11} + 0.5x_{21} \le 1.2, x_{32} + 0.5x_{52} \le 0.8, 0.5x_{32} + 0.5x_{33} \le 1, 0.5x_{14} + 0.5x_{44} \le 0.7\\
0.2x_{11} + 0.1x_{21} \le 0.5, 0.2x_{32} + 0.1x_{52} \le 0.4, 0.1x_{32} + 0.1x_{33} \le 0.4, 0.1x_{14} + 0.1x_{44} \le 0.3\\
0.9x_{11} + 0.95x_{21} \ge 0.95, 7x_{32} + 7x_{52} \ge 9, 8x_{32} + 6x_{33} \ge 9.7x_{14} + 6x_{44} \ge 8\\
7x_{11} + 8x_{21} \ge 8, 7x_{32} + 6x_{52} \ge 9, 7x_{32} + 6x_{33} \ge 9.6x_{14} + 7x_{44} \ge 8\\
x_{11} + x_{21} = 1, x_{32} + x_{52} = 1, x_{23} + x_{33} = 1, x_{14} + x_{44} = 1\\
x_{11} \times \left(y_{11} - 1\right) = 0, x_{21} \times \left(y_{21} - 1\right) = 0, x_{32} \times \left(y_{32} - 1\right) = 0, x_{52} \times \left(y_{52} - 1\right) = 0\\
x_{23} \times \left(y_{23} - 1\right) = 0, x_{33} \times \left(y_{33} - 1\right) = 0, x_{14} \times \left(y_{14} - 1\right) = 0, x_{44} \times \left(y_{4} - 1\right) = 0\\
0 \le x_{ij} \le 1, 0 \le y_{ij} \le 1\\
i = 1.2, \ldots, 5, j = 1.2, \ldots, 4.
\end{cases}
\end{aligned} \tag{19}$$

Table 7: Five quantitative indicators of five suppliers.

	g_1	g_2	g_3	g_4	g_5
Storage cost	65%	60%	70%	58%	76%
Out-of-stock cost	60%	64%	65%	70%	66%
Transportation cost	60%	58%	60%	68%	59%
Quality cost	59%	67%	72%	55%	75%
Delivery accuracy	97%	96%	92%	92%	91%

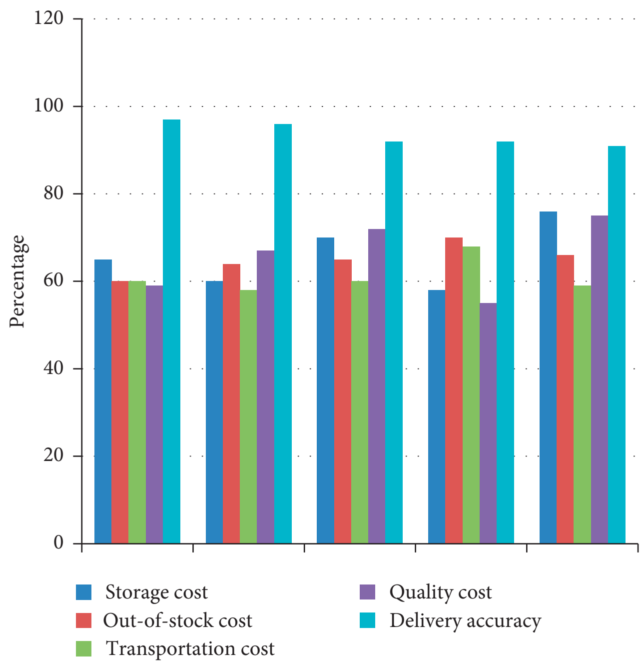


Figure 2: Distribution map of five quantitative indicators of five suppliers.

Table 8: Specific scoring results of 5 candidate suppliers.

	g_1	g_2	g_3	g_4	g_5
Storage cost	8	7.7	6.3	3.8	6.5
Out-of-stock cost	7.7	6.8	5	4.5	7.1
Transportation cost	7.3	8	6.4	6.5	7
Quality cost	6.7	5.9	6	6.7	6.7
Delivery accuracy	6.6	7.3	5.7	6.2	6.5
Supply capacity	8.5	7.2	5.5	6.4	6.5
Reliability	7.1	6.8	6.4	5.7	7.1

4.2.3. Solution. According to the algorithm introduced in the previous paper, we can solve

$$\begin{aligned} z_{1^*} &= 49.1\%, \\ x_{11} &= 0.427, \\ x_{21} &= 0.573, \\ x_{32} &= 0.516, \\ x_{52} &= 0.484 \\ x_{23} &= 0.785, \\ x_{33} &= 0.215, \\ x_{14} &= 0.548, \\ x_{44} &= 0.452 \\ y_{11} &= 1, \\ y_{21} &= 1, \\ y_{32} &= 1, \\ y_{52} &= 1, \\ y_{23} &= 1, \\ y_{33} &= 1, \\ y_{14} &= 1, \\ y_{44} &= 1, \end{aligned} \tag{20}$$

and adopt $\lambda = 0.05$. Then, the above model can be further transformed into the following form:

$$\begin{aligned} \max z_2 = {} & 0.233x_{11} + 0.232x_{21} + 0.182x_{32} + 0.214x_{52} \\ & + 0.232x_{23} + 0.182x_{33} + 0.233x_{14} + 0.138x_{44}. \end{aligned} \tag{21}$$

The constraint condition is added on the original basis, namely:

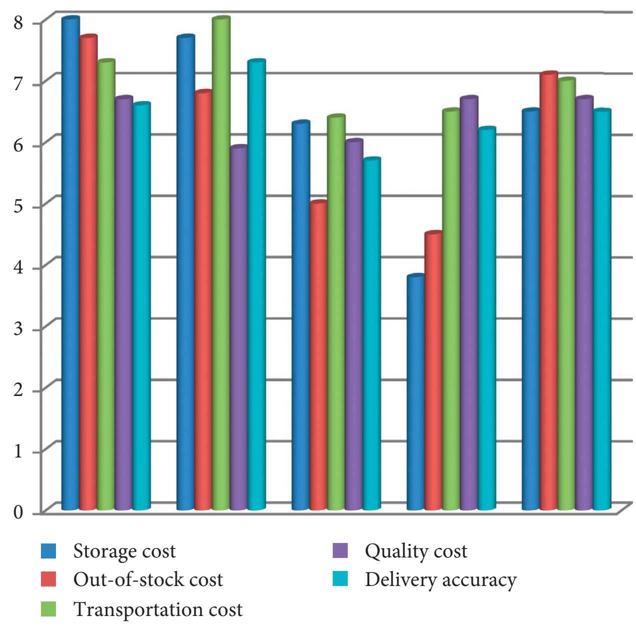


FIGURE 3: Distribution of specific scoring results of 5 candidate suppliers.

TABLE 9: Parameter values and evaluation information for marking the supply capacity and level of each supplier.

Product serial number	Supplier	p_{ij} (%)	q_{ij} (%)	k_{ij} (ten thousand yuan)	u_{ij} (ten thousand yuan)	f_{ij} (ten thousand yuan)	t_{ij} (ten thousand yuan)	d_{ij} (ten thousand yuan)	l_{ij} (ten thousand yuan)	r_{ij} (%)	e_{ij}	n_{ij}
1	1	12	23.3	15	9	6	1	0.2	0.1	90	6	7
	2	11	23.2	13	7	4	0.5	0.1	0.1	95	8	8
2	3	15	22.3	10	8	5	1	0.2	0.12	90	7	7
	5	14	22.5	11	8	4	0.5	0.1	0.15	90	7	6
3	2	13	18.9	12	5	3	0.5	0.1	0.1	95	8	7
	3	12	19.2	8	5	3	0.5	0.1	0.12	90	6	6
4	1	11	21.5	8	6	3	0.5	0.1	0.05	90	7	6
	4	12	20.2	5	4	2	0.5	0.1	0.05	95	6	5

TABLE 10: The requirements of a retail company on the supply capacity of its suppliers.

Purchase products (ten thousand yuan)	c_j	f_j	t_j	d_j	r_j	e_j	l_j	n_j
1	20	7	1.2	0.5	95	8	0.5	8
2	12	5	0.8	0.4	92	9	0.3	9
3	15	6	1	0.4	95	9	0.2	9
4	10	5	0.7	0.3	95	8	0.3	8

$$0.12x_{11} + 0.11x_{21} + 0.15x_{32} + 0.14x_{52} + 0.13x_{23} + 0.12x_{33} + 0.11x_{14} + 0.12x_{44} \leq (1 - \lambda) \times 49.1\%. \tag{22}$$

The solution of the single objective mathematical programming model is

$$\begin{aligned} x_{11} &= 0.712, \\ x_{21} &= 0.288, \\ x_{32} &= 0.195, \\ x_{52} &= 0.805, \\ x_{23} &= 0.751, \\ x_{33} &= 0.249, \\ x_{14} &= 0.378, \\ x_{44} &= 0.622, \\ y_{11} &= 1, \\ y_{21} &= 1, \\ y_{32} &= 1, \\ y_{52} &= 1, \\ y_{23} &= 1, \\ y_{33} &= 1, \\ y_{14} &= 1, \\ y_{44} &= 1. \end{aligned} \tag{23}$$

5. The Data Analysis

From the above calculation results, it can be seen that the best procurement plan for the company when purchasing product 1 is to choose g_1 and g_2 suppliers, and the optimal procurement volume is 71.2% for g_1 and 28.8% for g_2; when the company purchases product 2, the best procurement plan is to choose g_3 and g_5 suppliers. The best procurement volume is 19.5% for g_3 and 80.5% for g_5. The best solution for the company to purchase product 3 is to choose g_2 and g_3 as suppliers, and the optimal purchase quantity is 75.1% for g_2 and 24.9% for g_3.The best solution for the company to purchase product 4 is to choose g_1 and g_4 as suppliers, and the optimal purchase quantity is 37.8% for g_1 and 62.2% for g_4.

Based on the above analysis, the total procurement costs of these five suppliers are calculated as follows:

$$\begin{aligned} &\mathrm{M}\ (g_1) = 0.712 * 20 + 0.378*10 = 19.02, \\ &\mathrm{M}\ (g_2) = 0.288*20 + 0.751*15 = 17.025, \\ &\mathrm{M}\ (g_3) = 0.195*12 + 0.249*15 = 6.075, \\ &\mathrm{M}\ (g_4) = 0.622*10 = 6.22, \\ &\mathrm{M}\ (g_5) = 0.805*12 = 9.66. \end{aligned}$$

Through the calculation of the procurement costs of the above suppliers, when the purchase cost is constant, the higher the company's purchase cost for a certain supplier, the more transactions the supplier has with the company, and the more important it is to the company. [28–31], so it can be concluded that the order of the 5 suppliers to be selected is $g_1 > g_2 > g_5 > g_3 > g_4$. According to the situation of these 5 suppliers, a multiobjective mathematical programming model is established to calculate the distribution of the calculated purchase amount. From the above calculation results, it can be seen that the procurement expenditures of the 5 suppliers are basically in line with the importance of the suppliers in the previous section, and the amount is g, which accounts for 190,200 RMB, 170,250RMB, 60,750 RMB, 62,200RMB, and 96,600 RMB.

6. Conclusion

Reasonably reducing the purchase cost while ensuring the level of service quality is always a problem to be solved by chain convenience stores. This paper constructs ANP to select the chain convenience store supplier indicators, takes the Lawson convenience store as an example to collect relevant data, and combines the MOP to choose and rank Lawson chain convenience store suppliers. Then, the reasonable selection of Rosen's suppliers is carried out to determine the procurement allocation of each supplier, to provide a theoretical basis for the later maintenance of suppliers, which is helpful for the supervision and classification management of suppliers, ensuring the supply quality of suppliers, deepening the cooperation with suppliers, controlling the procurement cost, and effectively improving the service level and competitiveness of enterprises. Existing research has neglected the importance of suppliers and later management. This article considers the service level quality while considering the purchase cost, which has certain guiding significance for the development of modern enterprises. At the same time, it also verified that the established model is scientific and practical. With the continuous changes in consumer demand, multiple goals can be incorporated in the future to select and manage the suppliers of chain retail stores.

Data Availability

All of the underlying data supporting the results of this study are publicly available online and can be easily accessed.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by Anhui Province Key Research and Development Program Project (grant no. 201904a05020091) and its supporting projects (grant no. PT04a05020094), Natural Science Research Project of Anhui Universities in 2020 (grant no. KJ2020A0680), Chaohu University Discipline Construction Quality Improvement Project in 2020 (grant no. kj20xqyx03), Chaohu University Natural Science Research (grant no. XWY-201801), a key project of Chaohu University Natural Science Research (grant no. XLZ-201905), and a scientific research project commissioned by Hefei Hongzu Education Technology Co., Ltd (grant no. hxkt20210001).

References

[1] N. Merino, M. Marín, M. Fuentes et al., "Information and digital technologies of Industry 4.0 and Lean supply chain management: a systematic literature review," *International Journal of Production Research*, vol. 58, no. 16, 2020.

[2] Y. Li, "Research on supply chain CSR management based on differential game," *Journal of Cleaner Production*, vol. 268, 2020.

[3] B. Sarkar, M. Sarkar, B. Ganguly et al., "Combined effects of carbon emission and production quality improvement for fixed lifetime products in a sustainable supply chain management," *International Journal of Production Economics*, vol. 231, 2021.

[4] N. Guo, "research on optimization of supplier selection for A retail enterprise based on analytic hierarchy process," *China Logistics and Procurement*, vol. 7, pp. 71–73, 2020.

[5] J. Liao and R. Zhang, "Research on evaluation and selection of green suppliers of papermaking enterprises based on AHP-TOPSIS method," *Logistics Engineering and Management*, vol. 42, no. 02, pp. 91–93, 2020.

[6] S. Zhou and J. Liao, "Research on evaluation and selection of green suppliers of pharmaceutical enterprises based on AHP method," *Logistics Technology*, vol. 42, no. 10, pp. 28–31, 2020.

[7] B. Song, "Research on evaluation and selection of suppliers in L supermarket," *Henan University of Technology*, vol. 7, 2019.

[8] M. Modak, S. Sinha, A. Raj et al., "Corporate social responsibility and supply chain management: framing and pushing forward the debate," *Journal of Cleaner Production*, vol. 141, p. 273, 2020.

[9] C. Machado, M. Vivaldini, and O. Oliveira, "Production and supply-chain as the basis for SMEs' environmental management development: a systematic literature review," *Journal of Cleaner Production*, vol. 1, p. 273, 2020.

[10] M. Beheshti, J. Clelland, and K. Harrington, "Competitive advantage with vendor managed inventory," *Journal of Promotion Management*, vol. 26, no. 6, 2020.

[11] J. Lee and G. Johnson, "Contracting for vendor-managed inventory with a time-dependent stockout penalty," *International Transactions in Operational Research*, vol. 27, no. 3, 2020.

[12] D. Maio and D. Laganà, "The effectiveness of vendor managed inventory in the last-mile delivery: an industrial application," *Procedia Manufacturing*, vol. 42, 2020.

[13] L. Lin, "Application of analytic hierarchy process in supplier selection-taking M food manufacturing enterprise as an example," *Modern Commerce*, vol. 7, pp. 32-33, 2020.

[14] P. Liu, C. Wei, J. Liu et al., "Using Analytic Hierarchy Process to select suppliers for project procurement," *China Management Information*, vol. 23, no. 9, pp. 111–113, 2020.

[15] D. Li and G. Chen, "Construction of evaluation index system for selection of agricultural products cold chain logistics suppliers," *Market Research*, vol. 6, pp. 64-65, 2019.

[16] Y. Zhang, K. Wang, Q. He et al., "Covering-based web service quality prediction via neighborhood-aware matrix factorization," *IEEE Transactions on Services Computing*, vol. 1, 2019.

[17] L. Pan, "Analysis on the selection and evaluation of suppliers by using activity-based costing," *Accounting Learning*, vol. 17, pp. 167–169, 2020.

[18] W. Tong, "Research on supplier evaluation and selection of MC enterprises in the environment of internet of things," *Donghua University*, vol. 8, 2019.

[19] X. Han, "Research on supplier selection under multiple constraints based on genetic algorithm," *Value Engineering*, vol. 39, no. 14, pp. 123-124, 2020.

[20] Y. Xu, "Research on E-commerce supplier selection based on big data analysis Technology," *Modern Electronic Technology*, vol. 43, no. 15, pp. 152–158, 2020.

[21] G. Xiao and G. Zhang, "Supplier selection and coordination of B2C enterprises based on green supply chain theory," *Business Economic Research*, vol. 3, pp. 127–129, 2020.

[22] W. Luong, "A vendor managed inventory policy with emergency orders," *Journal of Industrial and Production Engineering*, vol. 37, pp. 2-3, 2020.

[23] Q. Wei, J. Zhang, G. Zhu et al., "Vendor managed inventory with considering stochastic learning effect," *Journal of the Operational Research Society*, vol. 71, no. 4, 2020.

[24] S. Amiri, A. Zahedi, M. Kazemi et al., "Determination of the optimal sales level of perishable goods in a two-echelon supply chain network," *Computers & Industrial Engineering*, vol. 139, 2020.

[25] H. Golpîra, "Optimal integration of the facility location problem into the multi-project multi-supplier multi-resource Construction Supply Chain network design under the vendor managed inventory strategy," *Expert Systems With Applications*, vol. 139, 2020.

[26] Q. Chen and Z. Zeng, "Supplier selection and evaluation based on partnership," *Car Practical Technology*, vol. 14, pp. 219–221, 2020.

[27] Y. Zhang, J. Pan, L. Qi, and Q. He, "Privacy-preserving quality prediction for edge-based IoT services," *Future Generation Computer Systems*, vol. 114, pp. 336–348, 2021.

[28] S. Khoshfetrat, M. Rahiminezhad Galankashi, and M. Almasi, "Sustainable supplier selection and order allocation: a fuzzy approach," *Engineering Optimization*, vol. 52, no. 9, pp. 1494–1507, 2020.

[29] D. He, *Research on Improvement of Supplier Evaluation of A Company*, Jilin University, Changchun, China, 2020.

[30] M. Zhao, "Application of triangular intuitionistic fuzzy number theory in supplier selection," *Wireless Internet Technology*, vol. 17, no. 05, pp. 166–168, 2020.

[31] O. Olanrewaju, Z. Dong, and S. Hu, "Supplier selection decision making in disaster response," *Computers & Industrial Engineering*, vol. 143, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5534583, 7 pages
https://doi.org/10.1155/2021/5534583

Research Article

A Study of Efficiency Evaluation of National Quality Online Courses during the Epidemic: Based on Fuzzy Logic Calculation and Bootstrap-DEA

Zhen-Hui Li,[1] **Jia-Jia Yang**,[2] **Hai-Qing Qin**,[1] **Yi-Wei Xia**,[3] **and Mei-er Zhuang**[3]

[1]*State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China*
[2]*Department of Public Policy, King's College London, London WC2R 2LS, UK*
[3]*School of Business, Guangdong University of Foreign Studies, Guangzhou 510000, China*

Correspondence should be addressed to Jia-Jia Yang; 1298256371@qq.com

Received 24 February 2021; Revised 29 March 2021; Accepted 1 April 2021; Published 12 April 2021

Academic Editor: Weilin Xiao

The study combines fuzzy logic calculation method and bootstrap-DEA to explore the efficiency of National Quality Online Courses (NQOOCs) during the new coronavirus epidemic. We choose Project 985 universities in China as our sample considering their academic influence. The results show that the efficiency value of NQOOCs calculated by bootstrap-DEA differs from the efficiency value measured by traditional DEA. The difference of universities` ranking is more pronounced in bootstrap-DEA. The amount of input resources (including human resources, science and technology resources, and financial resources) may not be directly reflected on the efficiency of NQOOCs. There is still much room for improvement in the efficiency of NQOOCs in most universities.

1. Introduction

Under the influence of the new coronavirus epidemic, most of universities have actively responded to the Ministry of Education's initiative of "no suspension of classes, no suspension of classes" and launched online courses. During this period, "National Quality Online Open Courses (NQOOCs)" played a leading role in demonstration. Many universities have opened high-quality online courses and virtual simulation experimental teaching resources free of charge to students across the country through online education websites such as "Chinese Universities MOOC." However, due to the lack of online teaching experience and the different input levels and methods of online courses, the effectiveness of online education construction in various universities is uneven. In order to improve the quality of online courses education in universities, it is particularly important to explore the relationship between the investment and effectiveness of online courses in universities.

Although there have been studies on the advantages and disadvantages of online courses, the development model, and the problems faced by online teaching during the epidemic, most of them are qualitative analysis, and there is a lack of quantitative evaluation of the construction methods and effects of online education in different universities. As the pioneer of online education in China, universities sponsored by Project 985 have accumulated relative experience. The National Quality Online Open Courses represent the best quality in online courses. Therefore, this article takes universities sponsored by Project 985 as the representative and uses the DEA-SBM model to analyze the relationship between the input elements and output elements of the national quality online open courses during the epidemic in 2020 and obtains the optimal benchmarking model by comparison. It is hoped that this research can provide reference suggestions for improving the quality of online education courses in universities.

2. Literature Review

E-learning, also known as distance education and online learning, is a learning method that can spread content across time and space through the use of information technology and Internet technology, bringing great convenience to our lives, learning, and work [1, 2]. Its forms include micro-classes, MOOCs, SPOC, and so on. Among them, MOOCs are the main form of China's online education designated quality online courses. In 2015, the "Opinions of the Ministry of Education on Strengthening the Application and Management of Online Open Curriculum Construction in Colleges and Universities" pointed out that China's universities should strengthen the construction of an online open curriculum system and platform with Chinese characteristics, and actively promote the update of college education concepts and the optimization of teaching models [3]. Major universities in China actively respond to the call of the Ministry of Education, and many famous universities led by Peking University and Tsinghua University actively participate in the construction of online open courses. In 2019, the Ministry of Education identified 801 courses as 2018 national quality online open courses, including 690 undergraduate education courses and 111 vocational higher education courses.

Online education in colleges and universities has emerged as a teaching model in recent years, and many scholars have analyzed its advantages, disadvantages, and development models. Online education can increase the interactivity of courses through the information database management technology and two-way interactive function of the computer network, and the asynchronous or synchronous learning network model reduces the time and space constraints [4, 5], and the learners of the course have no background and experience restrictions [6], which provides participants with convenient learning opportunities. However, the existing online courses have the problems of insufficient wisdom, low teachers' new media literacy, and lack of innovation in instructional design [7]. Zhang and Wang found that there is still a lot of room for improvement in the quantity, technology, and quality of the MOOC and MOOC platform through the observation of relevant data of the national quality online open courses [8]. In response to the above problems, a lot of research has begun to focus on the factors that affect the efficiency of online education and the ways and means to improve online education [9]. The study mainly explores the influence of learner factors, teacher factors, and technical factors on the satisfaction of online education courses from the TAM model [10, 11]. The exploration of the efficiency of online courses in colleges and universities under the background of new coronavirus is also very important. For example, Zou et al. combined with the situation in the Guangdong-Hong Kong-Macao Greater Bay Area to build an online teaching model for universities and implemented it in emergency to solve the urgent need for online teaching organization and management during the outbreak [12].

However, the relevant research on online education, especially MOOCs, mainly focuses on the influencing factors of course efficiency, such as course design factors. However, such factors mainly affect learners' satisfaction and the willingness to learn. Besides, the measurement of the outcome of MOOCs is mostly based on singular variable, and the influence of the learner factor is exaggerated [12]. For the national quality online courses, the input of universities has a more significant impact on their quantity and quality, so it is of great practical significance to explore the relationship between university inputs and the output of quality online courses. In addition, the existing literature has proposed many strategies for online education in colleges and universities, but it has not obtained the optimal model by comparing the efficiency of online education construction in many colleges.

Therefore, this article takes universities sponsored by Project 985 as the representative and combines the fuzzy logic calculation method and bootstrap-DEA model to analyze the relationship between the input elements and output elements of the national excellent courses from January to May 2020 to find the optimal benchmarking model. This article will also provide reference suggestions for universities to improve the quality of online education courses.

3. Data Sources and Variables

Data for the input of universities for education are derived from panoramic data platform of universities, the platform comprehensively covers more than 150 high-interest development related data indicators, and the collected data cover double first-class construction, faculty, scientific research, courses construction, education and teaching, and so on. The data are accurate and provide a good data source for researchers.

The data of Project 985 national high-quality online courses are mainly from the Chinese University MOOC website. This website is a domestic high-quality Chinese MOOC learning platform, which was created by the Love Course Network and Netease Cloud Classroom. The platform has more than a thousand courses including Project 985 universities, among which are the first batch of certified national quality online open courses data, including the number of courses, course ratings, and student participation. It can provide data support for our analysis of the efficiency of national quality online courses (https://www.icourse163.org/).

Combining the data from the two data platforms and the list of 985 national universities, a total of 39 universities were identified. After that, we excluded three schools that did not open online courses on the MOOC platform of Chinese universities during the epidemic, namely, Tsinghua University, Northwest A&F University, Lanzhou University, South China University of Technology, and National University of Defense Technology that lacked input data. Finally, 34 Chinese Project 985 universities were identified in our research sample.

The input variables we choose are human resources, science, and technology resources and financial resources universities invested in 2019, which are essential elements for universities` courses design and construction. We use high-end talent (HET), tech talent (TT), and course professionals (CP) to measure human resources. And to measure science and technology resources, we use scientific research projects (SRP), research scale (RS), and discipline construction (DC) as variables. Financial resources are measured by project investment (PI), social donation (SD), and fund investment (FI). Besides, to measure the output of NQOOC, we use total courses (TC), teaching quality (TP), and total number of participants (TNP) during the epidemic situation (from Jan 2020 to May 2020). Definitions and statistic descriptions of the above variables are shown in Table 1.

HET represents scholars who have made outstanding contributions to the development of related fields. TT are scholars who have a high world influence. CP are members of the College Teaching Steering Committee of the Ministry of Education. These experts have strong teaching ability and high academic attainments and can provide consultation for the discipline construction of universities.

SRP are projects supported by the National Natural Science Foundation of China usually have high research value and significance. RS refers to number of papers included in the Scopus database, which is used as an important data source in various domestic and foreign university rankings. The selection of “double first-class” DC is a great affirmation of the effectiveness of college discipline construction.

Funding for scientific research projects is an important material basis for the smooth development of scientific research in colleges and universities. So, we choose PI as the representor of scientific and technological research funds. SD is one of the important sources of funding for colleges and universities, and it can play a role in helping colleges and universities. Through FI, we can see the overall scale and overall scientific research strength of basic research and scientific research workers in each school.

4. Mathematical Model and Data Analysis

4.1. Fuzzy Logic Calculation. Based on multi-valued logic, fuzzy logic is a science using fuzzy sets to study uncertainty concept and ambiguity phenomena as well as their laws. Fuzzy logic can imitate the way of thinking of human brain, express qualitative knowledge and experience with unclear boundaries, and judge and reason about uncertain concepts. It uses membership function to distinguish fuzzy sets and handle fuzzy relationships. For description systems where the model is unknown or uncertain, fuzzy sets and fuzzy rules are used for reasoning and expressing transitional boundaries or qualitative knowledge experience. Meanwhile, fuzzy comprehensive judgment is carried out and solves regular fuzzy information problems that conventional methods are difficult to deal with.

In this paper, by using fuzzy logic, the second-level indicators of input and output can be reasoned and judged comprehensively, thus scientifically and effectively merged into the first-level indicators of input and output, which lays the foundation for the efficiency evaluation below. Specific steps are as follows.

First, each column of sample data of each secondary index is subjected to min-max normalization processing. The min-max normalization method is to linearly transform the original data. The formula is new data = (original data-minimum value)/(maximum value-minimum value).

Second, define the variables in the fuzzy logic and determine the degree of membership. Convert the input value of fuzzy logic into the degree of membership of each set.

Third, conduct the process of fuzzification. Determine the relationship between the input value and membership, so that the input value can find the membership of the corresponding set at any point. Define Gaussian, trapezoidal, and triangular membership function parameters, and add input and output variable membership functions to the fuzzy inference system.

Fourth, design the fuzzy logic judgment operation. Each membership degree decomposed, imitating human judgment using the concept of fuzzy, the fuzzy rule base and fuzzy inference of the previous step are used to obtain fuzzy logic control result signal. In this study, we use the minimum membership method (MIN implication) for processing.

Fifth, add fuzzy logic decision rules. After the input is fuzzified, rules need to be set and recombined with the operation of fuzzy logic.

Finally, conduct the process of de-fuzzification. This step is to convert the fuzzy value of the inference result into a clear control signal value. After the fuzzy logic converts the input value into the membership of each set through fuzzy, then several outputs can be obtained through rules and operations (see Table 2), and the membership function graphs are drawn (see Figure 1).

As can be seen from Figure 1 and Table 2, after fuzzy logic processing, the number of variables is effectively reduced to three inputs and one output. Since the set data output interval is between 0 and 20, the index results after fuzzy logic control are all within 20, which provides good data conditions for further empirical analysis.

4.2. NQOOCs Efficiency Analysis of Universities Based on the Bootstrap-DEA Method. Bootstrap-DEA is an extension of the traditional DEA method. Although the DEA method for some technical parameters estimation method has several advantages, the estimation results are susceptible to random interference factors, with sample sensitivity [13], and the bootstrap-DEA is very essential to overcome the efficiency value of the inner dependency. It makes statistical inferences from the raw data, without making any assumptions about the unknown population. And it generates pseudorandom numbers by taking back samples from existing samples, thus inferring the characteristics of the population [14]. Efficiency value calculation with bootstrap-DEA includes the following steps:

Firstly, each DMU (X_k, Y_k), $k = 1, \ldots, n$, uses the traditional DEA method to calculate the efficiency value θ_k of the sample data.

Secondly, based on the efficiency θ_k, $k = 1, \ldots, n$ the bootstrap method produces n efficiency value $\theta^*_{1b}, \theta^*_{2b}, \ldots, \theta^*_{nb}$, where b represents the b-th iteration using the bootstrap method

Table 1: The definitions and summary statistics of the variables.

Variables	Definition	Mean	Std. dev.	Min	Max
HET	Total number of academicians of Chinese Academy of Sciences and Chinese Academy Of Engineering	1.235	1.478	0.000	7.000
TT	Total number of highly cited scholars of Elsevier China	30.735	25.826	1.000	107.000
CP	Total number of members of the College Teaching Steering Committee of the Ministry of Education	49.353	23.604	9.000	111.000
SRP	Total number of National Natural Science Foundation Projects	416.853	279.227	15.000	1261.000
RS	Total number of published research papers included in the Scopus database	2.389	1.242	0.090	5.445
DC	Total number of double first-class construction disciplines in colleges and universities	8.059	7.651	1.000	41.000
PI	The average amount of scientific and technological research funds in universities (ten thousand yuan)	35.974	16.422	12.358	79.926
SD	School foundation annual social donation income (ten thousand yuan)	11.148	12.560	0.000	57.050
FI	Total amount of National Natural Science Foundation (100 million yuan)	3.053	2.252	0.183	11.039
TC	The number of available “National Quality Online Open Courses” during the epidemic	13.029	10.050	1.000	42.000
TQ	The viewer rating of the course on “China University MOOC” (0–5).	4.689	0.084	4.400	4.857
TNP	The number of participants in the course during the epidemic, that is, the total number of participants in the course divided by the times of the specific courses opened	2263.364	1736.402	558.992	8408.004

Table 2: Results of fuzzy logic calculation.

University	Human resource	Technology resource	Financial resource	NQOOCs
Peking University	15.000	15.000	15.000	12.021
Beihang University	5.250	8.066	5.002	11.379
Beijing Institute of Technology	5.000	5.000	10.000	12.518
Beijing Normal University	7.983	5.000	5.000	11.423
Dalian University of Technology	7.104	8.734	5.000	11.928
University of Electronic Science and Technology of China	5.000	5.000	5.000	8.724
Northeastern University	5.000	5.000	9.181	11.128
Southeast University	8.605	9.012	5.000	11.438
Fudan University	11.182	10.279	10.249	10.798
Harbin Institute of Technology	9.062	10.646	9.996	10.056
Hunan University	5.000	5.000	5.000	10.927
East China Normal University	5.000	5.000	5.000	10.183
Huazhong University of Science and Technology	9.860	10.471	8.813	10.890
Jilin University	9.857	11.350	5.000	14.759
Nanjing University	10.101	9.824	7.516	15.000
Nankai University	5.000	5.000	5.000	10.003
Xiamen University	7.723	5.000	10.007	10.892
Shandong University	9.951	12.828	5.000	8.486
Shanghai Jiao Tong University	14.161	10.295	10.849	9.922
Sichuan University	9.995	11.027	6.883	12.037
Tianjin University	9.705	9.972	5.000	10.032
Tongji University	10.103	10.438	5.000	9.020
Wuhan University	9.972	9.912	5.000	13.032
Xi’an Jiaotong University	9.640	10.386	5.000	10.662
Northwestern Polytechnical University	5.000	5.000	5.000	10.778
Zhejiang University	14.315	10.474	13.653	14.086
Ocean University of China	5.000	5.000	5.000	10.173
University of Science and Technology Of China	9.428	9.362	8.001	10.032
China Agricultural University	5.000	5.000	5.000	10.142
Renmin University of China	7.983	6.376	5.000	5.659
Central South University	6.309	11.853	5.000	10.419
Sun Yat-sen University	11.507	10.068	9.710	10.251
Minzu University of China	5.000	5.000	5.000	9.139
Chongqing University	6.733	6.610	5.000	10.079

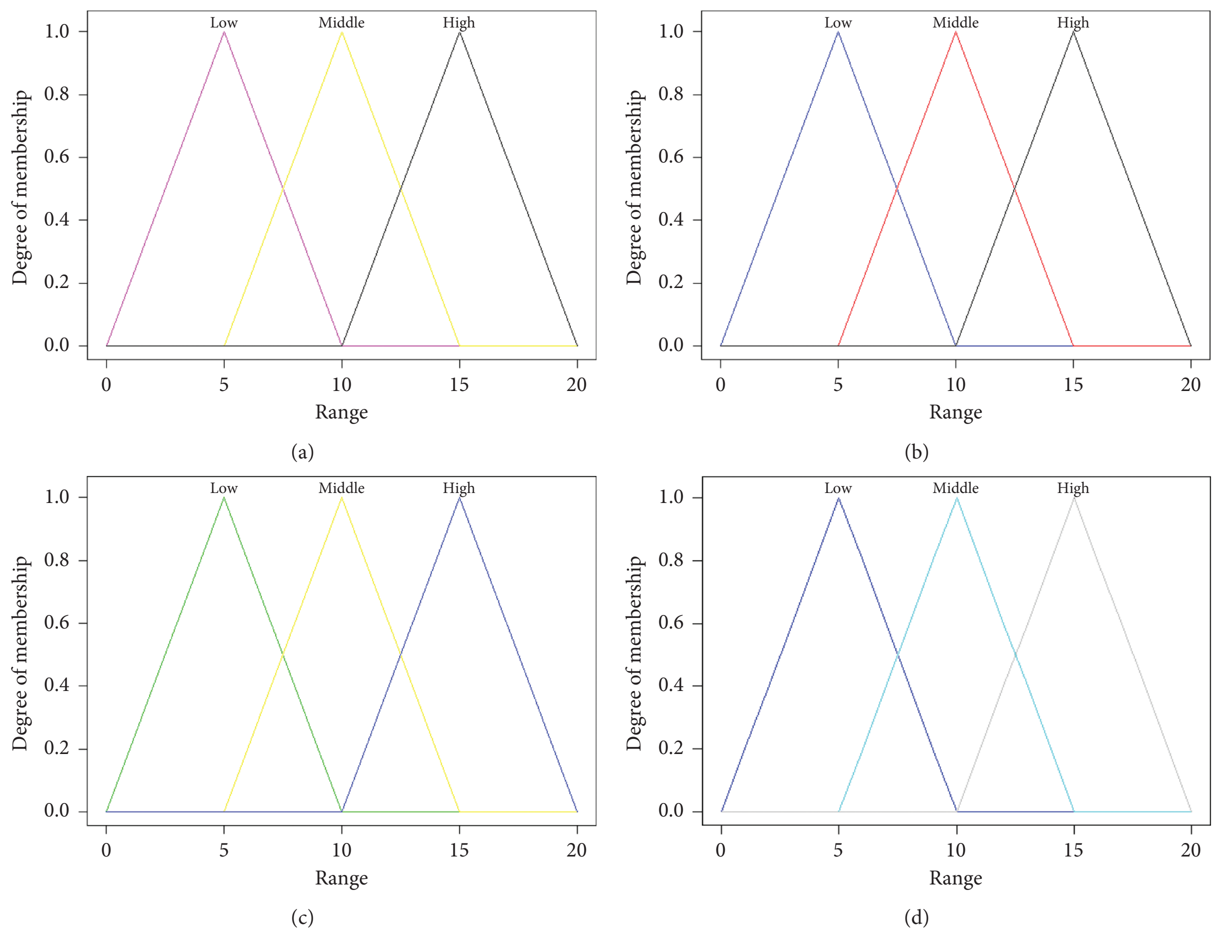


FIGURE 1: Membership functions of fuzzy logic.

Thirdly, calculate bootstrap simulation sample (X^*_{kb}, Y_k), $K = 1, \ldots, n$, where $X^*_{ab} = \theta_k \times (x_k/x^*_{ab})$, $(k = 1, \ldots, n)$

Fourthly, the traditional DEA method is used to simulate each bootstrap method sample, and the efficiency value θ_{kb}, $k = 1, \ldots, n$, is calculated again

Finally, repeat steps 2–4 B times to generate the efficiency value θ_{kb}, $b = 1, \ldots n$

Generally speaking, the more iterations of bootstrap-DEA, the more accurate the calculation results of efficiency value will be. The greater the confidence, the greater the distance between the upper and lower limits of the confidence interval [15]. According to existing studies, without loss of generality, this paper sets the number of bootstrap-DEA analysis iterations as 2000 and the confidence as 95%.

In the previous paper, the fuzzy logic calculation method is used to obtain the fuzzy values of input variables and output variables. On this basis, bootstrap-DEA was used to analyze the operational efficiency of online education in 34 Project 985 universities. The results are shown in Table 3.

From the perspective of ranking, the ranking of universities before and after rectifying deviation has changed greatly. From a comprehensive perspective, except for Hunan University and Northwestern Polytechnical University, the ranking of other universities has decreased by different degrees compared with the traditional DEA results. Except for Hunan University, the former top universities all dropped in different degrees. Beihang University, Beijing Normal University, and Jilin University ranked third, fourth, and fifth, respectively. The ranking of Beijing Institute of Technology dropped to 12, and the efficiency value after rectifying deviation was 0.845, which was the university with the largest decline. In addition, Northwestern Polytechnical University ranked second both before and after the correction.

The NQOOCs of Hunan University was 0.932 after rectifying deviation, and the lowest value was 0.342 after rectifying deviation of Peking University. There is still much room for improvement in the efficiency of NQOOCs in most universities.

Among the top 10 ranking universities, there are three polytechnic universities, namely, Northwestern Polytechnical University, Beihang University, and Dalian University of Technology, which indicates that the inputs of these universities are well utilized to courses construction. Three comprehensive universities, namely, Hunan University, Jilin University, and Nankai university, also perform

Table 3: Efficiency evaluation results of Project 985 universities NQOOCs, with traditional DEA and bootstrap-DEA.

DMU	Efficiency value	Original ranking	Efficiency value after rectifying deviation	Ranking after rectifying deviation	Deviation	Lower limit	Upper limit
Peking University	0.367	30.000	0.342	34.000	0.025	0.327	0.376
Beihang University	1.000	1.000	0.917	3.000	0.083	0.873	0.991
Beijing Institute of Technology	1.000	1.000	0.845	12.000	0.155	0.726	0.978
Beijing Normal University	1.000	1.000	0.902	4.000	0.098	0.840	0.983
Dalian University of Technology	0.941	3.000	0.869	6.000	0.072	0.831	0.920
University of Electronic Science and Technology of China	0.798	16.000	0.744	20.000	0.054	0.711	0.819
Northeastern University	0.908	9.000	0.784	17.000	0.124	0.689	0.898
Southeast University	0.851	12.000	0.785	15.000	0.066	0.746	0.852
Fudan University	0.478	26.000	0.446	30.000	0.032	0.429	0.485
Harbin Institute of Technology	0.500	25.000	0.465	29.000	0.035	0.445	0.505
Hunan University	1.000	1.000	0.932	1.000	0.068	0.891	1.026
East China Normal University	0.932	4.000	0.868	7.000	0.063	0.830	0.956
Huazhong University of Science and Technology	0.541	24.000	0.506	28.000	0.035	0.489	0.542
Jilin University	1.000	1.000	0.879	5.000	0.121	0.796	1.000
Nanjing University	0.835	15.000	0.784	16.000	0.051	0.759	0.825
Nankai University	0.915	8.000	0.853	10.000	0.062	0.816	0.939
Xiamen University	0.870	10.000	0.764	19.000	0.106	0.685	0.862
Shandong University	0.575	21.000	0.510	27.000	0.065	0.463	0.583
Shanghai Jiao Tong University	0.426	29.000	0.390	33.000	0.035	0.370	0.423
Sichuan University	0.688	19.000	0.635	23.000	0.052	0.607	0.673
Tianjin University	0.715	18.000	0.653	22.000	0.062	0.613	0.730
Tongji University	0.632	20.000	0.573	24.000	0.059	0.533	0.646
Wuhan University	0.931	6.000	0.852	11.000	0.079	0.801	0.951
Xi'an Jiaotong University	0.748	17.000	0.676	21.000	0.072	0.629	0.761
Northwestern Polytechnical University	0.986	2.000	0.919	2.000	0.067	0.879	1.012
Zhejiang University	0.574	22.000	0.521	25.000	0.053	0.488	0.563
Ocean University of China	0.931	5.000	0.868	8.000	0.063	0.829	0.955
University Of Science and Technology of China	0.546	23.000	0.513	26.000	0.033	0.497	0.547
China Agricultural University	0.928	7.000	0.865	9.000	0.063	0.827	0.952
Renmin University of China	0.469	28.000	0.434	32.000	0.035	0.414	0.458
Central South University	0.857	11.000	0.792	13.000	0.066	0.757	0.836
Sun Yat-Sen University	0.473	27.000	0.441	31.000	0.032	0.425	0.475
Minzu University of China	0.836	14.000	0.779	18.000	0.057	0.745	0.858
Chongqing University	0.841	13.000	0.790	14.000	0.051	0.764	0.831

well. Besides, the only two normal universities of Project 985 universities, namely, Beijing Normal University and East China Normal University, are both on the top 10 list. The possible reason is that normal universities have more teaching techniques and experience to improve the courses effect, and can better promote the conversion of invested resources in teaching performance. Moreover, Ocean University of China and China Agricultural University achieved the 8th and 9th rankings, which suggests that course characteristics and professionalism are also important factors in ensuring course quality.

5. Conclusion

Based on the fuzzy logic calculation analysis and bootstrap-DEA analysis, this study evaluated the efficiency of national quality online open courses of Chinese Project 985 universities. The results have some implications.

The combination of fuzzy logic calculation analysis and bootstrap-DEA analysis may provide some reference for the process of dealing with variables that are of similarities. And the use of bootstrap-DEA can overcome several disadvantages of traditional DEA, especially that there are significant difference in the ranking results of bootstrap-DEA, which is essential for us to probe into the possible causes.

There are some practical implications of this paper. Firstly, universities with abundant resources should pay more attention to the efficiency of curriculum construction, because there are a large number of participants that rely on the NQOOCs to gain knowledge. Secondly, ensuring the distinguishing feature and professional level of NQOOCs is a vital way to promote teaching efficiency. What is more, the focus of this research is online education, which can provide reference value for scholars in this field in the future, enriching literature materials.

Additionally, this study still has some limitations that may offer future research direction. Although we have evaluated the efficiency of NQOOCs during the new coronavirus epidemic, the previous invested resources may be for a whole year of output. Future research can investigate the efficiency of NQOOCs measured by the whole year outputs. Besides, the results of bootstrap-DEA only offer the efficiency of NQOOCs, but it is not clear how different input elements produce a marked effect. Using other data-mining methods is necessary to test the mechanism of input elements and outcomes.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was financially supported by Asian Media Research Center Project, "Research on the Improvement PATH of China's International Communication Ability in the New Media Era" (Grant no. AMRC2018-3).

References

[1] W. Bhuasiri, O. Xaymoungkhoun, H. Zo, J. J. Rho, and A. P. Ciganek, "Critical success factors for e-learning in developing countries: a comparative analysis between ICT experts and faculty," *Computers & Education*, vol. 58, no. 2, pp. 843–855, 2012.

[2] P.-C. Sun, R. J. Tsai, G. Finger, Y.-Y. Chen, and D. Yeh, "What drives a successful e-learning? an empirical investigation of the critical factors influencing learner satisfaction," *Computers & Education*, vol. 50, no. 4, pp. 1183–1202, 2008.

[3] L. Zhao, C. Yun, and Bo Yu, "Research on online teaching mode of ordinary colleges and universities from the perspective of smart education," *Computer Knowledge and Technology*, vol. 15, no. 30, pp. 158-159, 2019, in Chinese.

[4] G. Piccoli, R. Ahmad, and B. Ives, "Web-based virtual learning environments: a research framework and a preliminary assessment of effectiveness in basic IT skills training," *MIS Quarterly*, vol. 25, no. 4, pp. 401–426, 2001.

[5] Y. J. Katz, "Attitudes affecting college students' preferences for distance learning," *Journal of Computer Assisted Learning*, vol. 18, no. 1, pp. 2–9, 2002.

[6] A. McAuley, "Massive open online courses: digital ways of knowing and learning," *The MOOC Model for Digital Practice*, vol. 5, pp. 3–6, 2010.

[7] A. Margaryan, M. Bianco, and A. Littlejohn, "Instructional quality of massive open online courses (MOOCs)," *Computers & Education*, vol. 80, pp. 77–83, 2015.

[8] L. Zhang and J. Wang, "Data analysis of the national excellent online open courses of the Mu class platform," *Digital Education*, vol. 5, no. 5, pp. 36–41, 2019.

[9] A. H. Mousa, "Advance acceptance status model for E-learning based on university academics and students," *IOP Conference Series: Materials Science and Engineering*, vol. 671, no. 1, 2020.

[10] L. L. Martins and F. W. Kellermanns, "A model of business school students' acceptance of a web-based course management system," *Academy of Management Learning & Education*, vol. 3, no. 1, 2011.

[11] R. Edmunds, M. Thorpe, and G. Conole, "Student attitudes towards and use of ICT in course study, work and social activity: a technology acceptance model approach," *British Journal of Educational Technology*, vol. 43, no. 1, pp. 77–84, 2012.

[12] Y. Zou, C. Li, and Y. Xie, "Construction and implementation of online teaching "Bay Area Model" in colleges and universities during the epidemic situation," *China Audio-Visual Education*, vol. 4, pp. 22–28, 2020, in Chinese.

[13] M.-L. Song, L.-L. Zhang, W. Liu, and R. Fisher, "Bootstrap-DEA analysis of BRICS' energy efficiency based on small sample data," *Applied Energy*, vol. 112, pp. 1049–1055, 2013.

[14] Qi. Zhang, "Efficiency evaluation of Chinese commercial banks based on DEA and bootstrap method," *Economic Research Guide*, vol. 32, pp. 123–125, 2009.

[15] W. Yin and W. Yuan, "Evaluation of scientific research efficiency of universities directly under the ministry of education of China based on Bootstrap-DEA method," *Statistics and Information Forum*, vol. 28, no. 6, pp. 61–69, 2013.