Complexity

Applications of Frontiers and Complexity in Optimisation Theory and Algorithms

Lead Guest Editor: Baogui Xin
Guest Editors: Abdelalim A. Elsadany and Lei Xie

Applications of Frontiers and Complexity in Optimisation Theory and Algorithms

Complexity

Applications of Frontiers and Complexity in Optimisation Theory and Algorithms

Lead Guest Editor: Baogui Xin
Guest Editors: Abdelalim A. Elsadany and Lei Xie

Copyright © 2021 Hindawi Limited. All rights reserved.

This is a special issue published in "Complexity." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chief Editor

Hiroki Sayama, USA

Associate Editors

Albert Diaz-Guilera, Spain
Carlos Gershenson, Mexico
Sergio Gómez, Spain
Sing Kiong Nguang, New Zealand
Yongping Pan, Singapore
Dimitrios Stamovlasis, Greece
Christos Volos, Greece
Yong Xu, China
Xinggang Yan, United Kingdom

Academic Editors

Andrew Adamatzky, United Kingdom
Marcus Aguiar, Brazil
Tarek Ahmed-Ali, France
Maia Angelova, Australia
David Arroyo, Spain
Tomaso Aste, United Kingdom
Shonak Bansal, India
George Bassel, United Kingdom
Mohamed Boutayeb, France
Dirk Brockmann, Germany
Seth Bullock, United Kingdom
Diyi Chen, China
Alan Dorin, Australia
Guilherme Ferraz de Arruda, Italy
Harish Garg, India
Sarangapani Jagannathan, USA
Mahdi Jalili, Australia
Jeffrey H. Johnson, United Kingdom
Jurgen Kurths, Germany
C. H. Lai, Singapore
Fredrik Liljeros, Sweden
Naoki Masuda, USA
Jose F. Mendes, Portugal
Christopher P. Monterola, Philippines
Marcin Mrugalski, Poland
Vincenzo Nicosia, United Kingdom
Nicola Perra, United Kingdom
Andrea Rapisarda, Italy
Céline Rozenblat, Switzerland
M. San Miguel, Spain
Enzo Pasquale Scilingo, Italy
Ana Teixeira de Melo, Portugal
Shahadat Uddin, Australia
Jose C. Valverde, Spain
Massimiliano Zanin, Spain

Contents

Hindawi
Complexity
Volume 2021, Article ID 9940866, 12 pages
https://doi.org/10.1155/2021/9940866

Research Article

Developing a Model for the University Course Timetabling Problem: A Case Study

Mozhgan Mokhtari,[1] Majid Vaziri Sarashk,[1] Milad Asadpour,[2,3] Nadia Saeidi,[1] and Omid Boyer[1]

[1]*Department of Industrial Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran*
[2]*Department of Information Systems and Operations Management, Business School, The University of Auckland, Auckland, New Zealand*
[3]*Young Researchers and Elite Club, Najafabad Branch, Islamic Azad University, Najafabad, Iran*

Correspondence should be addressed to Omid Boyer; omidboyer@gmail.com

Received 5 March 2021; Revised 15 November 2021; Accepted 23 November 2021; Published 22 December 2021

Academic Editor: Baogui Xin

Copyright © 2021 Mozhgan Mokhtari et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Over recent years, timetable programming in academic settings has become particularly challenging due to such factors as the growing number of students, the variety of lectures, the inadequacy of educational facilities in some areas, and the incorporation of teachers and students' preferences into the schedule. Many researchers, therefore, have been formulating the problem of timetabling lectures using different methods. In this research, a multiobjective mixed-integer programming model was developed to provide a timetable for the postgraduate courses at the Industrial Engineering Department of Islamic Azad University, Najafabad Branch (IAUN). The proposed model minimized the violation of the lecturers and educational priorities, the student travel time between classes, and the classes' surplus capacity. To convert the multiobjective model into a single one, the ε-constraint method was adopted, and the model's accuracy and feasibility were examined through a real example solved by the CPLEX solver of GAMS software. The results approved the efficiency of this model in preparing a timetable for university lectures.

1. Introduction

The problems associated with timetabling persist in various sectors such as industry, transportation, and sport. However, timetabling continues to be a complicated task in academia. Educational centers widely differ in terms of the available space and the number of courses, teachers, students, and time slots, making the adoption of a universal schedule virtually impossible [1, 2]. Another challenge facing universities in this regard is creating a timetable that meets staff and students' demands and the existing requirements [3]. To deal with this complexity, many educational institutions apply heuristic methods. Mathematical models, such as integer programming (IP), have recently been performed to solve these problems [4]. The IP problems were developed into the mixed-integer programming model by adding realistic assumptions that make the University Course Timetabling Problems (UCTP) more complicated.

These assumptions combine different decisions such as lecturers and students' preferences, free time, class overlapping, students flow, and the maximum capacity of rooms. The scheduling of classes has a significant effect on students' movements, and this means universities with many students attending courses in a single building are highly likely to encounter the problem of congestion in the aisles [5]. In this respect, the objective functions of timetabling problem in the most relevant studies have focused on the maximization of preferences, the minimization of unregistered courses, time interference, free hours between lectures, and the number of students who fail to apply for courses [2].

This research was motivated by the UCTP at the Industrial Engineering Department (IED) of IAUN. The

department has nine classrooms located on the second floor of the Engineering Faculty. There are around 400 postgraduate and 300 undergraduate students majoring in industrial engineering. The department has two chairpersons for post- and undergraduate groups, who provide timetabling programs separately, though they share information about the availability of classrooms and lecturers in each time slot. The postgraduate chair finds it challenging to solve UCTP with three study areas (project management, system optimization, and quality and productivity) because most postgraduate students prefer to take courses on two last days of the week (Wednesday and Thursday Iran). This decision generally results in congestion in the IED.

One possible solution would be to spread courses over the day and week. However, this often leads to free periods in the schedule, thereby causing dissatisfaction among students and teachers. The limited capacity of the classrooms also affects programming despite the chairpersons' attempts to develop the timetable manually by trial and error every semester. Figure 1 illustrates the floor layout of the department.

This study developed a timetabling model to meet the requirements of IED, with three objective functions: minimizing the violation of the teachers and educational priorities, minimizing the travel time for students, and minimizing classrooms' surplus capacity. The model aimed to help the chairperson consider the educational preferences, manage time, and allocate classes more effectively by adding the constraints caused by the department's conditions, including the limited capacity of classrooms and restrictions on the number of groups offered for each lecture.

In summary, the contributions of this study are threefold:

(i) The study applies the mixed-integer programming model to formulate the realistic feature of the timetabling problem, such as the effect of timetabling on the student flow.

(ii) The model has three objectives to satisfy different stakeholders: students, teachers, and the university.

(iii) The feasibility and efficiency of the model have been checked by using the random number and real data.

The rest of the current paper is organized as follows. Section 2 represents a review of the relevant previous studies. In Section 3, the problem will be described, and the proposed model will be formulated. Section 4 will discuss how the model is validated and solved for the random numbers and the case study, followed by an analysis of the results. Finally, the most significant results, along with suggestions for future studies, are discussed in Section 5.

2. Literature Review

Due to the nature of required decisions in scheduling problems such as UCTP, the formulation of this type of mathematical modeling problem often utilizes integer variables. In the following, a comprehensive review of the previous research in the University Course Timetabling Problem is presented.

Formulated objective functions within the literature of UCTP could be summarized in some particular categories, and therefore, we have categorized our reviewed papers based on different objective functions. However, what is challenging among UCTPs is that the problem is completely different from one university to others regarding different constraints and resources. Most of the earlier studies have aimed to provide a university timetable with the minimum cost. For example, Daskalaki et al. [6] developed an IP model of a timetabling problem for the Department of Electrical and Computer Engineering at the University of Patras. The objective function was to minimize the cost to demonstrate the model's capabilities to address the university's timetabling problem, using the MIP solver of CPLEX software. Further, Daskalaki and Birbas [7] proposed a solution approach for the IP model of a university timetabling problem. The model was a cost minimization problem, and the solution was based on a relaxation procedure of certain constraints. These were the constraints that ensured the consecutiveness of the multiperiod sessions assigned to a given course. The applicability of the adopted procedure was investigated by applying real data from a case study. Thepphakorn and Pongcharoen [8] propose a Cuckoo Search-based algorithm for the UCTP of Faculty of Engineering, Naresuan University, with the aim of minimizing the total university operating costs.

Providing a university timetable with particular attention to the preferences of students, teachers, or administrators has been another motivation for the formulation of a UCTP. Ozdemir and Gasimov [3] presented an IP model for the UCTP. The model's objectives were minimizing the average preference level per hour, the average preference level of all instructors, the administration's total preference level, and the total deviation from the upper load limits of the recent instructors. The objectives' weights were calculated using the AHP method, and then the problem was solved using real data from a case study. Gunawan et al. [9] provided a weekly timetable model based on the teachers' preferences. The model's objective function was to maximize the teachers' total preference value of the assignment of courses, days, and time slots. An initial solution was obtained based on the Lagrangian relaxation method; however, this solution was further improved by a Simulated Annealing algorithm. The applicability of the proposed model was examined using real data from a university in Indonesia. Méndez-Díaz et al. [10] considered a timetabling problem arising from a real-world application in a private university in Buenos Aires, *Argentina*. They proposed an integer programming model, which maximizes the global weighted preference (a combination of students' preferences and rank performances) while assigning lessons to the time slots. Afterward, a heuristic algorithm was deployed to solve the model. Jamili et al. [11] proposed a multiobjective integer mathematical model for UCTP. The objectives were to maximize instructors' preferences and the time spent by instructors for teaching, leading to an increase in the available time for research and addressing research affairs of students. The model was coded in GAMS software and solved by the augmented ε-constraint method in the

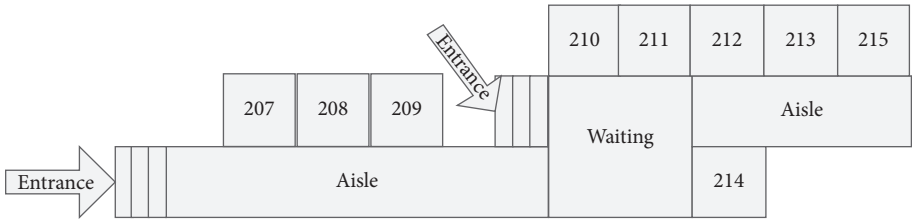


Figure 1: The floor layout of the Industrial Engineering Department of IAUN.

Industrial Engineering and Systems Faculty of a public university in Tehran, Iran. Yasari et al. [12] investigated a UCTP where the registration was implemented in two steps: preregistration and drop/add phases. They developed a two-stage stochastic programming model for this situation, which provided a feasible timetable, maximizing the expected value of satisfaction level among all teachers and students. The performance of the adopted approach was analyzed using random numbers. In the UCTP of Chávez-Bosquez et al. [13], the purpose is to maximize Professors' preferences, and the paper applies a hybrid Tabu Search algorithm to solve the problem. Algethami and Laesanklang [14] propose a multiobjective model for UCTP of Taif University to maximize events assignments and faculty-members preferences satisfaction and minimize student learning days and unassigned events. The model has been solved within IBM ILOG CPLEX software. Colajanni and Daniele [15] formulate an ILP model for UCTP of the University of Catania to maximize teachers' preferences and minimize the daily movements of students among classrooms. The model has been solved within IBM ILOG CPLEX software.

Another group of papers has focused on the minimization of timetable violations. For example, MirHassani [16] formulated an IP model for UCTP to minimize the soft constraint infeasibility and penalize the redundant and nonpreferred times. Azadeh et al. [17] proposed an IP model, intending to minimize the collegians' interferences. The problem was solved by GAMS using random data. The timetabling problem of Vermuyten et al. [5] stated that a timetable should be scheduled based on the teaching preferences of teachers. However, teachers' preferences can be violated by scheduling a lecture at a timeslot when a teacher does not prefer to teach this lecture. The objective functions of the model were the minimization of the violation of teachers' educational preferences and the minimax of the travel time for each series of students. The model was applied to the Faculty of Economics and Business dataset of the KU Leuven Campus Brussels and was solved by using the CPLEX software. Phillips et al. [18] presented an IP model to eliminate violations of an existing timetable while minimizing the disruption to the remainder of the timetable. Real data of the University of Auckland was used to solve the model deploying the Gurobi solver. Bagger et al. [19] formulated an IP model to provide a timetable minimizing a weighted sum of the violations of the soft constraints. This model was validated by solving several real examples from the literature applying the Gurobi solver. In another study, Bagger et al. [20] employed a Dantzig–Wolfe reformulation to solve the model by Column Generation. The objective was to find a feasible timetable while minimizing the soft constraints. In addition, the objective of UCTP studies of Goh et al. [21], Song et al. [22], and Gozali et al. [23] is to minimize violations of soft constraints. In Goh et al. [21], the solution approach is a combined algorithm based on Tabu Search and Simulated Annealing. Song et al. [22] apply a competition-guided multineighborhood local search algorithm to solve their model, while Gozali et al. [23] propose a localized island model genetic algorithm to solve the model. However, in Rezaeipanah et al. [24], the paper maximizes the number of satisfied soft constraints instead of minimizing soft constraints violations. To solve their UCTP, they have presented an Improved Parallel Genetic Algorithm and Local Search.

Meanwhile, in some previous studies, there are capacity limitations on classrooms. In other words, the problem is to schedule courses considering the optimal usage of classrooms. For example, Kaviani et al. [25] presented an IP model for the UCTP with four objectives: minimizing the teacher's idle time, minimizing the surplus capacity of rooms, maximizing the use of the available classrooms, and maximizing teachers' use of the class. The model was solved within LINGO software and using randomly generated data. Phillips et al. [4] presented an IP model for the University of Auckland, which found an efficient partial room assignment and made the best possible use of the available rooms. The objective of the model was the maximization of the number of patterns assigned to the classrooms. Tavakoli et al. [26] formulated a model for the UCTP, where the model considered six objectives: the maximal use of classrooms, lecturers during their attendance at university, the use of a classroom by a lecturer, the assignment of specialized courses to the faculty members, and the quality of the assigned lecturers to courses and minimization of the surplus capacity of rooms. Lemos et al. [27] discussed the problem of room usage optimization by proposing a two-stage IP model. In the first stage, lectures were allocated to the classrooms for the maximum use of classrooms, whereas the second stage minimized the number of transitions from free to occupied (and vice versa) for each classroom. The model was solved using a greedy algorithm and applying real data from Institute Superior Técnico, the Engineering School from Universidade de Lisboa.

Table 1 compares selected papers with the proposed model of this research.

As discussed, the bulk of the recent research has utilized mathematical models with different objective functions for UCPT problems. Nonetheless, most of these models have been formulated as integer programming with binary decision variables. This means the mixed-integer programming model has been paid less attention in the literature, though the continuous variables of MIP models can help consider more realistic assumptions. In this research, a multiobjective mixed-integer programming model was developed for the UCTP, drawing on the model of Vermuyten et al. [5]. It should also be noted that this study considered some particular parameters and constraints in the model to provide a timetable for the IED. These positive points of the model are considering classroom capacity, assignment of teachers based on the expertise, and the number of groups offered for each lecture given the IED's desired circumstances.

3. Materials and Methods

There are several points in the timetabling problem of the IED as follows:

(i) The industrial engineering curriculum entails including a specific number of courses in each semester, which should be scheduled in an organized way.

(ii) The knowledge and experience of a teacher have an effect on the priority of teachers for a lecture.

(iii) Teachers can have a particular number of lectures based on the educational rules in IAUN.

(iv) The congestion and movement of students in aisles disturb other classes. Free hours in the timetable compound this issue. So, the minimization of student flow and also removing free time can alleviate this problem.

(v) Each classroom has different facilities such as a computer and data show. Some classes, therefore, are more suitable for some lectures.

(vi) All model parameters are deterministic, and the problem was scheduled based on the conditions and preferences of the case study.

3.1. Notation. In this section, the required notation for the model formulation is presented in Tables 2–4.

3.2. Model Formulation. A multiobjective mixed-integer programming model with three objective functions is developed to address the IED requirements regarding the mentioned features. The objective functions and the required constraints are explained in this section.

Equations (1) to (3) are objective functions of the proposed MIP model. The first objective function minimizes the violation of the lecturers and educational priorities. The second objective function minimizes the student travel time, and the last objective function minimizes the surplus capacity of classrooms.

$$\text{Min } Z_1 = \sum_{l\varepsilon L}\sum_{t\varepsilon T}\sum_{c\varepsilon C}\sum_{e\varepsilon E} c_{\text{lte}} X_{\text{ltce}}. \tag{1}$$

$$\text{Min } Z_2 = \sum_{c\varepsilon C}\sum_{d\varepsilon D}\sum_{t\varepsilon T} \text{Tarc}_{\text{cdt}}. \tag{2}$$

$$\text{Min } Z_3 = \sum_{c\varepsilon C} cap_c - \sum_{l\varepsilon L}\sum_{t\varepsilon T}\sum_{c\varepsilon C}\sum_{e\varepsilon E} \frac{X_{\text{ltce}}\text{sl}_{\text{sl}} n_s}{\text{cap}_c}. \tag{3}$$

Equation (4) states that only one lecture must be presented in a particular time slot and class by a teacher.

$$\sum_{t\epsilon T}\sum_{c\epsilon C}\sum_{e\epsilon E} X_{\text{ltce}}\text{lr}_{\text{le}} = 1 \quad \forall l. \tag{4}$$

Equation (5) assures each teacher to conduct only one lecture in a certain time slot and a class.

$$\sum_{l\epsilon L}\sum_{c\epsilon C} X_{\text{ltce}}\text{lr}_{\text{le}} \le 1 \quad \forall t, e. \tag{5}$$

Equation (6) indicates that only one lecture can be held in a classroom in a time slot.

$$\sum_{l\epsilon L}\sum_{e\epsilon E} X_{\text{ltce}} \le 1 \quad \forall t, c. \tag{6}$$

Equation (7) states that a group of students can participate solely in one lecture in a particular time slot.

$$\sum_{c\epsilon C}\sum_{e\epsilon E} X_{\text{ltce}}\text{sl}_{\text{sl}} \le 1 \quad \forall t, s, l. \tag{7}$$

Equation (8) states that teachers cannot present more than a specific number of lectures (Δ) in one day.

$$\sum_{l\epsilon L}\sum_{t\epsilon T}\sum_{c\epsilon C} X_{\text{ltce}} \le \Delta_e \quad \forall e. \tag{8}$$

Equation (9) demonstrates the relation between X_{ltcr} and U_{tsp}. In this equation, the percentage of students on a path depends on the flow of students between two classes in two consecutive time slots.

$$U_{\text{tsp}} \ge a_{\text{lm}}\left(\text{pa}_{\text{cp}} X_{\text{ltce}} + \text{pa}_{\text{dp}} X_{m(t+1)\text{de}} - 1\right) \quad \forall (t\epsilon\{1,..,|T-1|\}, p, e, (c, d\epsilon C), \ (l, m\epsilon L, l \ne m). \tag{9}$$

Table 1: Comparison of selected papers with the proposed approach.

Reference number	Type of formulation		Objective function(s)									Solution approach	Case study
			Max			Min							
	Integer programming	Mixed-integer programming	Assigning courses	Classroom usage	Preferences of instructor	Classroom surplus capacity	Travel time	Movements between classrooms	Preferences violation	Timetable disruption	Teaching time		
[4]	*			*								Gurobi IP solver	*
[5]	*	*					*		*			CPLEX software	*
[10]	*				*						*	Augmented ε -constraint method	
[14]	*			*						*		Gurobi IP solver	*
[18]	*		*	*		*						LINGO software	
[19]	*			*				*				Greedy algorithm	*
Present research		*				*	*		*			**Epsilon-constraint method**	*

Table 2: Sets of the model.

Sets	*Description*
L, M	Set of lectures
E	Set of teachers
S	Set of students groups based on areas of study
C, D	Set of all classes
P	Set of all paths in the building
T	Set of available time slots

Table 3: Parameters of the model.

Parameters	*Description*
c_{lte}	Penalty cost when lecture l is scheduled for teacher e in time slot t
n_s	The number of students in group s
$f_{\max}$	Maximum allowed student flow
a_{lm}	Percentage of students moving between lectures l and m
$\mathrm{pa}_{\mathrm{cp}}$	If classroom c is on the path, p equals 1 and 0 otherwise
Δ_e	Teacher e can present a certain number of lectures
$\mathrm{area}_{\mathrm{cd}}$	Area between two classes c and d
$\mathrm{dis}_{\mathrm{cd}}$	Distance between two classes c and d
G_l	The maximum number of lectures l in a semester
$\nu_{\max}$	Maximum walking speed
α	Scale parameter
$\mathrm{lr}_{\mathrm{le}}$	It equals 1 if lecture l is taught by teacher e and 0 otherwise
$\mathrm{sl}_{\mathrm{sl}}$	It equals 1 if lecture l is suggested for group s and 0 otherwise

Table 4: Decision variables of the model.

Decision variables	*Description*
X_{ltce}	It equals 1 if lecture l is scheduled in time slot t by teacher e in classroom c and 0 otherwise
F_{cdt}	Total number of students who move between classes c and d in period t
U_{tsp}	Percentage of group s students has been moved by using path p in period t
$\mathrm{Tarc}_{\mathrm{cdt}}$	The travel time between classrooms c and d in period t

Equation (10) states the total student flow between classes c and d at time t.

$$F_{\mathrm{cdt}} = \sum_{p\epsilon P}\sum_{s\epsilon S} n_s U_{\mathrm{tsp}} \quad \forall t, (c, d\epsilon C). \tag{10}$$

Equation (11) ensures that this flow should not be greater than the maximum allowed flow.

$$F_{\mathrm{cdt}} \le f_{\max} \quad \forall t, (c, d\varepsilon\ C). \tag{11}$$

Equation (12) determines the travel time from class c to class d at time slot t, which varies depending on the total number of students, the distance of two classrooms, the area of the aisle, and the maximum speed of movement.

$$\mathrm{Tarc}_{\mathrm{cdt}} = \left[\left(\frac{\mathrm{dis}_{\mathrm{cd}}}{\alpha}\right)\left(\frac{F_{\mathrm{cdt}}}{\mathrm{area}_{\mathrm{cd}}}\right)\right] + \frac{\mathrm{dis}_{\mathrm{cd}}}{\nu_{\max}} \quad \forall t, (c, d\varepsilon\ C). \tag{12}$$

Equation (13) states that the number of lectures in a semester cannot be more than the predetermined number specified by the department's chair.

$$\sum_{t\varepsilon T}\sum_{c\varepsilon C}\sum_{e\varepsilon E} X_{\mathrm{ltce}} \le G_l \quad \forall l. \tag{13}$$

Finally, the decision variables' domains are defined by equations (14)–(16).

$$X_{\mathrm{ltce}}\varepsilon\{0, 1\}. \tag{14}$$

$$U_{\mathrm{tsp}}\varepsilon[0, 1]. \tag{15}$$

$$F_{\mathrm{cdt}} \ge 0, \mathrm{Tarc}_{\mathrm{cdt}} \ge 0. \tag{16}$$

3.3. Solution Approach. Multiobjective problems include more than one objective function, particularly conflicting objectives. The multiobjective mathematical model area has been widely employed for decades because most real-world problems have various objectives. Many methods and approaches have been developed to tackle multiobjective problems [27].

The ε-constraint method is one of the most efficient methods which can be applied. Accordingly, one of the objective functions is optimized, and other objectives are added to the constraints. The steps of this method are as follows:

(1) The payoff table is calculated for all objectives

(2) One objective is selected to be optimized, and other objectives are considered as constraints

(3) The ε values account for the constrained objectives function

(4) Efficient solutions (Pareto front) to the problem are achieved by the main objective optimization and the epsilon's parametric variation

(5) The Pareto solutions are reported [28, 29]

Figure 2 represents the steps of the ε-constraint method.

In the present study, the ε-constraint method has been deployed to solve the proposed model.

4. Experiments

In order to validate the proposed model, a random problem was first generated and solved using the CPLEX solver of GAMS software on a laptop with Intel Core i5, 2.5 GHz, and 4 GB of RAM.

4.1. Validating Model by Random Data. This example considered eight lectures in three days, four classrooms, four lecturers, and six paths between classrooms. There were four time slots per day, which means there were a total

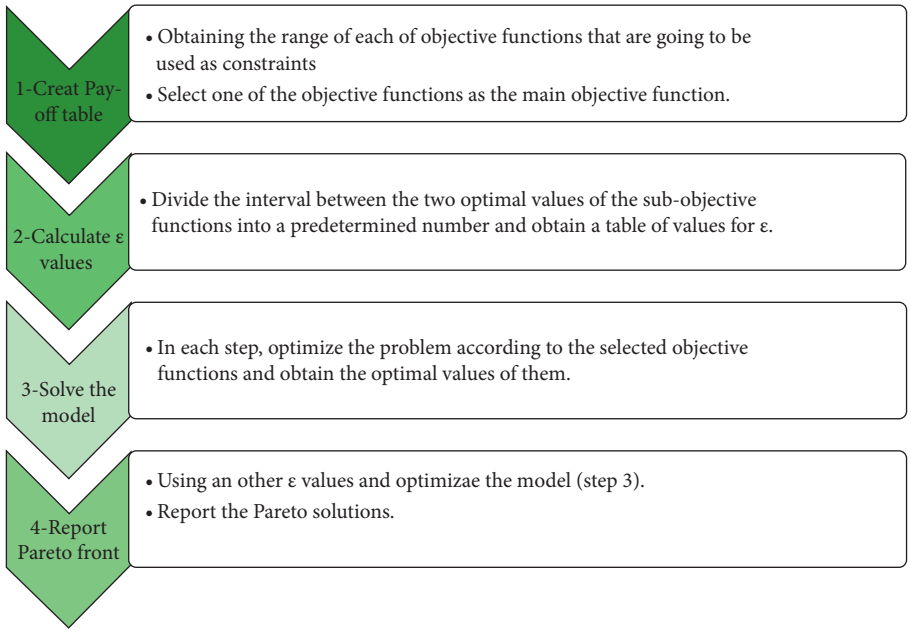


Figure 2: The steps of the ε-constraint method.

of 12 time slots. The floor layout of the classroom's location and paths is shown in Figure 3, and the most important input data of the random problem are presented in Tables 5–8.

Table 5 shows the capacity of each classroom (cap_c). The number of students in each group (n_s) has been presented in Table 6.

Table 7 includes the lectures offered for each group of students (sl_{sl}) and the lectures taught by each teacher (lr_{lr}).

Table 8 shows the paths that connect the classrooms based on Figure 3.

The outputs of the model solved by GAMS software in Table 9 demonstrate the efficiency of the model to create a timetable.

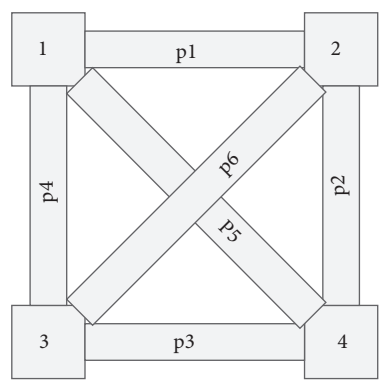


Figure 3: Location of classrooms and paths.

4.2. Applied Model by Real Data. This section will discuss a real example of the IED at IAUN for postgraduate students for the first semester of 2019–2020. It should be mentioned that the IED is located in the Engineering Faculty and offers three different areas of study for postgraduate, including project management, system optimization, and quality and productivity.

This real example considers 14 lectures, eight teachers (denoted by E1 to E8), two days per week, nine classrooms, four time slots on Wednesday, and two time slots on Thursday. The limitation of time slots is due to the availability of the classrooms and preventing overlap with the undergraduate classes. The input data for this instance are presented in Tables 10–16.

Table 10 represents available postgraduate lectures of the IED and teachers' areas of interest.

Table 11 shows the capacity of each classroom (cap_c). For instance, the capacity of classroom number 207 is 30 students, and its code is 1.

Table 5: Capacity of each classroom.

Classroom code	Capacity
1	50
2	50
3	40
4	40

Table 6: The number of students in each group.

Group code	Number of students
1	15
2	20
3	30

The number of students in each group (n_s) is presented in Table 12. For example, the number of students in the project management group (program) with group code S1 is 25.

Table 7: Lectures offered for each group of students that each teacher teaches them.

Students group	Lecture							
	L1	L2	L3	L4	L5	L6	L7	L8
S1	1	0	0	1	1	0	0	1
S2	0	1	0	1	0	1	1	0
S3	1	0	1	0	1	1	0	1
Teachers	Lecture							
E1	1	1	0	0	1	0	1	0
E2	1	1	1	1	0	1	1	1
E3	0	1	0	1	1	1	0	1
E4	0	0	1	1	0	1	1	1

Table 8: All paths that connect classrooms.

Path	Classroom															
	C1 C1	C1 C2	C1 C3	C1 C4	C2 C1	C2 C2	C2 C3	C2 C4	C3 C1	C3 C2	C3 C3	C3 C4	C4 C1	C4 C2	C4 C3	C4 C4
P1	0	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0
P2	0	0	1	0	0	0	1	0	1	1	0	1	0	0	1	0
P3	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0
P4	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0
P5	0	0	1	0	0	0	0	1	1	0	0	1	0	1	1	0
P6	0	1	0	1	1	0	1	0	0	1	0	1	0	0	1	0

Table 9: Timetable of the random example.

Day	Time			
	8 am–10:30 am	10:30 am–13 pm	13 pm–15:30 pm	15:30 pm–16 pm
Tuesday	Lecture 3, classroom 3, and teacher 2	Lecture 7, classroom 4, and teacher 4	—	—
Wednesday	—	Lecture 1, classroom 2, and teacher 1	Lecture 8, classroom 1, and teacher 4	Lecture 4, classroom 2, and teacher 3
Thursday	Lecture 5, classroom 1, and teacher 1	Lecture 6, classroom 1, and teacher 2	Lecture 2, classroom 4, and teacher 2	—

Table 10: Available postgraduate lectures.

Name of lecture	Code of lecture	Code of teachers
Project management information systems	L1	E1 and E2
Project planning and scheduling	L2	E2 and E8
Standards of project management	L3	E2, E5, and E8
Procurement of project	L4	E3 and E8
Design of experiments	L5	E4 and E6
Quality and productivity	L6	E3 and E7
Excellence and quality	L7	E3 and E7
Total quality management	L8	E3 and E7
Integer programming	L9	E5, E6, and E8
Nonlinear programming	L10	E5, E6, and E8
Queuing theory	L11	E4 and E6
Reliability theory	L12	E4
Decision-making methods	L13	E3 and E5
Financial management	L14	E1

Table 13 demonstrates the lectures offered for each group of students (sl_{sl}). As can be seen, if lecture l is offered for a group (program) s, the value is 1 and 0 otherwise.

The area ($area_{cd}$) between the two classrooms is shown in Table 14. For example, the area between classrooms no. 1 and no. 2 is 24 m^2.

Table 11: Capacity of each classroom.

Classroom number	Classroom code	Capacity
207	1	30
208	2	30
209	3	30
210	4	70
211	5	30
212	6	30
213	7	30
214	8	50
215	9	50

Table 12: The number of students in each group.

Group (program) name	Group code	Number of students
Project management	S1	25
System optimization	S2	25
Quality and productivity	S3	25

Table 13: Suitable lectures for each group of students.

Students group	Lectures													
	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14
S1	1	1	1	1									1	1
S2					1	1	1	1				1	1	1
S3				1	1				1	1	1	1	1	1

Table 14: Area between two classrooms.

Area	Classroom								
	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	0	24	96	102	120	126	144	150	168
C2	24	0	24	30	48	54	72	78	96
C3	96	24	0	6	24	30	48	54	72
C4	102	30	6	0	24	30	48	54	72
C5	120	48	24	24	0	6	24	30	48
C6	126	54	30	30	6	0	24	30	48
C7	144	72	48	48	24	24	0	6	24
C8	150	78	54	54	30	30	6	0	24
C9	168	96	72	72	48	48	24	24	0

The distance (dis_{cd}) between two classrooms is illustrated in Table 15. For example, the distance between classrooms no. 1 and no. 2 is 4 m.

By considering the real data of the IED, (G_l) is equal to 1, the number of lessons presented by one professor is four (Δ), the maximum speed is assumed to be 10 ($v_{\max}$), and the scaling parameter is 20.

Then, three objectives are optimized separately to investigate the feasibility and the efficiency of the proposed model for the timetabling problem. Table 16 shows each objective function's values and the allocation of a lecture to a teacher, classroom, and time slot.

As can be seen from Table 16, lecture one is taught by teacher one in time slot six in classroom nine when the first objective has been optimized.

The proposed model is a multiobjective model, so the ε-constraint method is used to convert the model to a single objective model. For this reason, three objective functions are optimized to calculate the payoff table (Table 17).

Then, the first objective was optimized, and the rest were placed in the constraints by six different values of ε to show conflicting objectives function and reach the Pareto solutions. Table 18 shows Pareto solutions based on different six ε values.

It is evident from Figure 4 and Table 18 that the classroom's surplus capacity is increasing to cover the educational priorities as the first objective function, which means the model has been decided to use more classrooms to increase educational priorities.

Table 15: Distance between two classrooms.

Distance	Classroom								
	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	0	4	16	17	20	21	24	25	28
C2	4	0	4	5	8	9	12	13	16
C3	16	4	0	1	4	5	8	9	12
C4	17	5	1	0	4	5	8	9	12
C5	20	8	4	4	0	1	4	5	8
C6	21	9	5	5	1	0	4	5	8
C7	24	12	8	8	4	4	0	1	4
C8	25	13	9	9	5	5	1	0	4
C9	28	16	12	12	8	8	4	4	0

Table 16: Outcomes of the optimized model.

Objective function	Value	Allocating lectures to teacher, class, and time slot					
$\mathbf{z}_1^*$	15	L1.T6.C9.E1	L2.T2.C9.E2	L3.T5.C9.E8	L4.T1.C6.E8	L5.T6.C5.E4	L6.T3.C8.E6
		L7.T6.C6.E7	L8.T3.C6.E3	L9.T5.C8.E6	L10.T2.C6.E5	L11.T6.C2.E6	L12.T5.C6.E4
		L13.T5.C6.E4	L14.T5.C6.E4				
$\mathbf{z}_2^*$	456	L1.T1.C1.E1	L2.T1.C2.E2	L3.T1.C3.E8	L4.T1.C4.E3	L5.T1.C5.E4	L6.T1.C6.E7
		L7.T2.C1.E7	L8.T2.C2.E3	L9.T1.C7.E5	L10.T1.C8.E6	L11.T2.C3.E6	L12.T2.C4.E4
		L13.T2.C5.E5	L14.T2.C6.E1				
$\mathbf{z}_3^*$	315	L1.T6.C7.E1	L2.T3.C7.E2	L3.T4.C5.E1	L4.T2.C1.E3	L5.T5.C1.E4	L6.T5.C7.E3
		L7.T2.C2.E7	L8.T4.C6.E3	L9.T6.C5.E8	L10.T4.C2.E6	L11.T3.C6.E6	L12.T4.C3.E4
		L13.T2.C6.E5	L14.T5.C3.E1				

Table 17: Payoff table for three objective functions.

Objective function	Z1	Z2	Z3
$\mathbf{z}_1^*$	15	63	146
$\mathbf{z}_2^*$	501.6	456	726.15
$\mathbf{z}_3^*$	366.66	336.268	315

Table 18: Pareto solutions of objective functions. Additionally, Figure 4 draws a comparison of the values of objective functions Z1 and Z3 as an example to prove the conflict between the two objectives.

Objective function	ε0	ε1	ε2	ε3	ε4	ε5
Z1	15	35	45	65	85	95
Z2	546	516	531	531.3	501	456.15
Z3	336.2	330	327.7	322.5	317.5	315

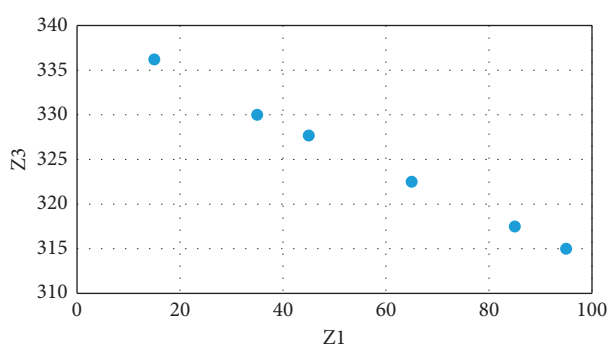


Figure 4: Pareto solutions of Z_1 and Z_3.

Ultimately, Table 19 compares the conventional university lecture timetable systems and the outcomes of the mathematical model for the Department of Industrial Engineering of IAUN. The results indicate the amount of educational priorities objective in the conventional method stands at 17, whereas by means of the mathematical model, this figure arrives at 15, which indicates an improvement of approximately 13% in the quality. Another significant

Table 19: The proposed timetable for the case study.

Lecture	The conventional method Day/time slots/classroom/teacher	The mathematical method Day/time slots/classroom/teacher
Total penalty	17	15
Project planning and scheduling	Wednesday/8 am–10:30 am/209/E8	Wednesday/10:30 am–13 pm/215/E2
Standards of project management	Wednesday/10:30 am–13 pm/208/E2	Thursday/8 am–10:30 am/215/E8
Procurement of project	Wednesday/13 pm–15:30 pm/214/E3	Wednesday/8 am–10:30 am/212/E8
Design of experiments	Thursday/8 am–10:30 am/213/E4	Thursday/10:30 am–13 pm/211/E4
Quality and productivity	Thursday/10:30 am–13 pm/214/E3	Thursday/10:30 am–13 pm/210/E3
Excellence and quality	Thursday/10:30 am–13 pm/211/E7	Thursday10:30 am–13 pm/212/E7
Total quality management	Thursday/8 am–10:30 am/211/E7	Wednesday/13 pm–15:30 pm/212/E3
Integer programming	Wednesday/8 am–10:30 am/207/E6	Thursday/8 am–10:30 am/214/E6
Nonlinear programming	Thursday/8 am–10:30 am/210/E5	Wednesday/10:30 am–13 pm/212/E5
Queuing theory	Thursday/8 am–10:30 am/207/E6	Thursday/10:30 am–13 pm/208/E6
Reliability theory	Thursday/10:30 am–13 pm/213/E4	Thursday/8 am–10:30 am/212/E4
Decision-making methods	Wednesday/15:30 pm–18 pm/214/E3	Thursday/8 am–10:30 am/208/E3
Financial management	Wednesday/10:30 am–13 pm/215/E1	Thursday/8 am–10:30 am/207/E1

change resulting from the mathematical model is the run time. The previous method was based on the programmer's experience, and it demands a great deal of time to modify the timetable.

5. Conclusion

This study provided multiobjective mixed-integer programming to prepare a timetable for postgraduate students at the Industrial Engineering Department of IAUN. The model has three objective functions: minimizing the violation of the lecturers and educational priorities, minimizing students' travel time, and minimizing classrooms' surplus capacity. The study also investigated the related constraints created by classrooms, days, time slots, and the expertise of teachers.

The model was applied to schedule all postgraduate lectures according to the data from the first semester of 2019–2020. The results approved the proposed model's applicability and practicality in providing a case study timetable considering the limitations and preferences of the case study. The multiobjective model was transformed into a single objective model using the ε-constraint method and solved by GAMS software, CPLEX solver. The parametric variation of ε resulted in the Pareto optimal solutions, which indicated a trade-off between objective functions.

Researchers in future studies can apply the proposed model in this study for scheduling lectures in other universities and educational centers. However, creating software capable of scheduling lectures is one of the other exciting directions for future research. Further research should also focus on developing heuristic or metaheuristic techniques that can solve large-scale lecture timetabling problems.

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.

References

[1] A. Schaerf, "A survey of automated timetabling," *Artificial Intelligence Review*, vol. 13, no. 2, pp. 87–127, 1999.

[2] S. A. MirHassani and F. Habibi, "Solution approaches to the course timetabling problem," *Artificial Intelligence Review*, vol. 39, no. 2, pp. 133–149, 2013.

[3] M. S. Ozdemir and R. N. Gasimov, "The analytic hierarchy process and multiobjective 0-1 faculty course assignment," *European Journal of Operational Research*, vol. 157, no. 2, pp. 398–408, 2004.

[4] A. E. Phillips, H. Waterer, M. Ehrgott, and D. M. Ryan, "Integer programming methods for large-scale practical classroom assignment problems," *Computers & Operations Research*, vol. 53, pp. 42–53, 2015.

[5] H. Vermuyten, S. Lemmens, I. Marques, and J. Beliën, "Developing compact course timetables with optimized student flows," *European Journal of Operational Research*, vol. 251, no. 2, pp. 651–661, 2016.

[6] S. Daskalaki, T. Birbas, and E. Housos, "An integer programming formulation for a case study in university timetabling," *European Journal of Operational Research*, vol. 153, no. 1, pp. 117–135, 2004.

[7] S. Daskalaki and T. Birbas, "Efficient solutions for a university timetabling problem through integer programming," *European Journal of Operational Research*, vol. 160, no. 1, pp. 106–120, 2005.

[8] T. Thepphakorn and P. Pongcharoen, "Performance improvement strategies on Cuckoo Search algorithms for solving the university course timetabling problem," *Expert Systems with Applications*, vol. 161, Article ID 113732, 2020.

[9] A. Gunawan, K. M. Ng, and K. L. Poh, "A hybridized Lagrangian relaxation and simulated annealing method for the course timetabling problem," *Computers & Operations Research*, vol. 39, no. 12, pp. 3074–3088, 2012.

[10] I. Méndez-Díaz, P. Zabala, and J. J. Miranda-Bront, "An ILP based heuristic for a generalization of the post-enrollment course timetabling problem," *Computers & Operations Research*, vol. 76, pp. 195–207, 2016.

[11] A. Jamili, M. Hamid, H. Gharoun, and R. Khoshnoudi, "Developing a comprehensive and multi-objective mathematical model for university course timetabling problem: a real case study," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Paris, France, July 2018.

[12] P. Yasari, M. Ranjbar, N. Jamili, and M.-H. Shaelaie, "A two-stage stochastic programming approach for a multi-objective course timetabling problem with courses cancelation risk," *Computers & Industrial Engineering*, vol. 130, pp. 650–660, 2019.

[13] O. Chávez-Bosquez, J. Hernández-Torruco, B. Hernández-Ocaña, and J. Canul-Reich, "Modeling and solving a Latin American university course timetabling problem instance," *Mathematics*, vol. 8, no. 10, p. 1833, 2020.

[14] H. Algethami and W. Laesanklang, "A mathematical model for course timetabling problem with faculty-course assignment constraints," *IEEE Access*, vol. 9, pp. 111666–111682, 2021.

[15] G. Colajanni and P. Daniele, "A new model for curriculum-based university course timetabling," *Optimization Letters*, vol. 15, no. 5, pp. 1601–1616, 2021.

[16] S. A. MirHassani, "A computational approach to enhancing course timetabling with integer programming," *Applied Mathematics and Computation*, vol. 175, no. 1, pp. 814–822, 2006.

[17] A. Azadeh, H. Gholizadeh, and M. Jeihoonian, "A multi-objective optimisation model for university course timetabling problem using a mixed integer dynamic non-linear programming," *International Journal of Services and Operations Management*, vol. 15, no. 4, pp. 467–481, 2013.

[18] A. E. Phillips, C. G. Walker, M. Ehrgott, and D. M. Ryan, "Integer programming for minimal perturbation problems in university course timetabling," *Annals of Operations Research*, vol. 252, no. 2, pp. 283–304, 2017.

[19] N.-C. F. Bagger, G. Desaulniers, and J. Desrosiers, "Daily course pattern formulation and valid inequalities for the curriculum-based course timetabling problem," *Journal of Scheduling*, vol. 22, no. 2, pp. 155–172, 2019.

[20] N.-C. F. Bagger, M. Sørensen, and T. R. Stidsen, "Dantzig-Wolfe decomposition of the daily course pattern formulation for curriculum-based course timetabling," *European Journal of Operational Research*, vol. 272, no. 2, pp. 430–446, 2019.

[21] S. L. Goh, G. Kendall, N. R. Sabar, and S. Abdullah, "An effective hybrid local search approach for the post enrolment course timetabling problem," *Opsearch*, vol. 57, no. 4, pp. 1131–1163, 2020.

[22] T. Song, M. Chen, Y. Xu, D. Wang, X. Song, and X. Tang, "Competition-guided multi-neighborhood local search algorithm for the university course timetabling problem," *Applied Soft Computing*, vol. 110, Article ID 107624, 2021.

[23] A. A. Gozali, B. Kurniawan, W. Weng, and S. Fujimura, "Solving university course timetabling problem using localized island model genetic algorithm with dual dynamic migration policy," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 15, no. 3, pp. 389–400, 2020.

[24] A. Rezaeipanah, S. S. Matoori, and G. Ahmadi, "A hybrid algorithm for the university course timetabling problem using the improved parallel genetic algorithm and local search," *Applied Intelligence*, vol. 51, no. 1, pp. 467–492, 2021.

[25] M. Kaviani, H. Shirouyehzad, and S. M. Sajadi, "A mathematical model for university course timetabling problems by considering multi functions," *International Journal of Modelling in Operations Management*, vol. 3, no. 3/4, pp. 282–295, 2013.

[26] M. M. Tavakoli, H. Shirouyehzad, F. H. Lotfi, and S. E. Najafi, "Proposing a new multi objective mathematical model for university course timetabling problem regarding optimisation of the quality of lecturers," *International Journal of Modelling in Operations Management*, vol. 7, no. 1, pp. 75–94, 2018.

[27] A. Lemos, F. S. Melo, P. T. Monteiro, and I. Lynce, "Room usage optimization in timetabling: a case study at Universidade de Lisboa," *Operations Research Perspectives*, vol. 6, Article ID 100092, 2019.

[28] K. Deb and K. Deb, "Multi-objective optimization," in *Search Methodologies*, pp. 403–449, Springer, Boston, MA, USA, 2014.

[29] M. Ehrgott and X. Gandibleux, "Multiobjective combinatorial optimization - theory, methodology, and applications," in *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*, pp. 369–444, Springer, New York, NY, USA, 2003.

Hindawi
Complexity
Volume 2021, Article ID 9975071, 17 pages
https://doi.org/10.1155/2021/9975071

Research Article

A Robust Mathematical Model for Sustainable and Resilient Supply Chain Network Design: Preparing a Supply Chain to Deal with Disruptions

Zahra Sadeghi, Omid Boyer, Shila Sharifzadeh, and Nadia Saeidi

Department of Industrial Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran

Correspondence should be addressed to Omid Boyer; omidboyer@gmail.com

Received 5 March 2021; Revised 24 May 2021; Accepted 20 September 2021; Published 27 October 2021

Academic Editor: Lei Xie

Copyright © 2021 Zahra Sadeghi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supply chains suffer from serious vulnerabilities and disruptions with increasing global crises, including pandemics and natural disasters. Dynamic and complex supply chain environments have constantly led companies to modern management approaches such as resilience to address disruptions. Besides, the sustainability approach enhances the strength of the supply chain in disruptions by considering economic, social, and environmental aspects. This paper develops a mathematical model for designing a supply chain network considering resilience and sustainability. In this model, suppliers were exposed to disruption with different probabilities. The model has three objectives: minimizing total costs and maximizing suppliers' social and environmental scores. A robust scenario-based stochastic programming approach has been used for potential disruption scenarios. The multiobjective model is solved by the ε-constraint method in GAMS software. The numerical results show the performance of the model in a different situation. Also, the robust scenario-based stochastic programming approach allows the average performance of the supply chain in each objective to improve.

1. Introduction

In the new age, the expansion of information and communication technologies and a decrease in geographic borders' effect caused the development of complicated supply chains. Supply chains, especially in large industries, are installed in different geographical places. They are driven to complicated strategies and processes such as global outsourcing, just-in-time, and lean activities to provide higher quality end products at the proper time and the lowest possible price to customers who are scattered everywhere. Hence, it is needed to create a smooth and uninterrupted flow of materials and information throughout the supply chain. Although these measures lead to lower operating costs, higher product quality, and upgraded commerce agility for the supply chains, companies face various risks by making and implementing these decisions [1]. In general, the globalization of the supply chain and the resulting complexity of the chain structure make them vulnerable to events and disorders such as natural disasters, political instability, strikes, unexpected legal issues, and terrorist acts. Such events, which lead to the disruption of information flow and materials, even if such events are in a remote area, can cause widespread disruptions throughout the supply chain and have many negative effects, called the bullwhip effect [2].

Statistics show a sharp increase in the number of unexpected events and catastrophes that firms have experienced in the recent past. Recent disasters such as tsunami (2004 and 2011), Hurricane Katrina (2005), earthquakes in 1999, 2009, and 2010 in Taiwan, earthquake in Turkey (2012), flood in Thailand (2011), terrorist attacks in New York (2001), Madrid (2004), London (2005), Jakarta (2009), and Mumbai (2008), diseases, and recession show dire consequences, including production downtime, disruption of productivity, and consumed capacity. These issues have unexpected effects on supply chains [1,3]. These disruptions have substantial negative consequences on sales returns,

profit returns, stock returns, brand image, companies' hiring, buyers' safety, and overall supply chain performance [4–6]. COVID-19 is one of the newest events with which the world has been faced. According to the world meter website, most countries will be infected by March 5, 2021. The number of people diagnosed up to that date was 116 million, and the number of reported deaths was 2,570,000 [7].

The pandemic's impact on public health also affects the supply chains' performance, sustainable economic growth, and the environmental performance of supply chains [8]. For example, all 1,000 large companies have been severely affected because all have multiple facilities in quarantined areas [9]. Even before the outbreak was declared a pandemic, production and supply of materials for 938 companies out of 1,000 Fortune companies were severely affected because their 1st and 2nd tier suppliers were in Wuhan, China, where the origin is generally believed to be of this disease (Fortune, 2020). Moreover, the severe spread of the virus into Europe and the United States has blocked the movement of products and materials worldwide [10]. Similarly, because of the cessation of operations in some parts of the supply chain, the continuation of supply chain operations has stopped [11, 12]. While almost all manufacturing firms in various industries have been affected by COVID-19 [9], the effect varies depending on the nature of the products, for example, the impact on high-demand items or low-demand items. In this regard, it can be pointed to the significant increase in demand for some products such as toilet paper, hand detergents and disinfectants, food, and medicine, while demand for some other products such as garments and sports items decreased sharply [13, 14]. During the pandemic, the impacts on these high-demand items are more immediate and visible, given that these products are essential for daily life and in some cases for survival [12, 15, 16]. Moreover, while firms experience an immediate and sharp increase in demand for these products, they also face a substantial shortage of raw materials during this pandemic [9, 12, 17]. This increase in worldwide demand is a clear example of the consequences of coronavirus disruption.

Supply chains to deal with disturbances, unexpected issues, and reduce risk should be designed to have the ability to face different events and provide an effective response to disruptions. They should have the ability to return to the initial status, or even better condition than before disturbance that is the resilience definition in the supply chains. The resilience of companies assesses their abilities in a fast return to the functional levels before the crisis in the cases such as the production level, the level of services, and the amount of storage [3]. Resilience is the adaptive capability of a supply chain to reduce the probability of sudden disruptions, resist the spread of disturbances, and maintain control over structures and functions. Also, it recovers and responds through immediate and effective reactive plans to transcend the disturbance and restore the supply chain to a robust state of operations. In fact, two concepts of flexibility and redundancy are involved in resilience. For example, having a flexible transport system and flexible production facilities is included in the flexibility, and having multiple suppliers, safety stock, and backup suppliers is a redundancy concept [18].

Various studies have shown that resilience and sustainability are two inspiring approaches in responding to concerns related to living standards in the face of unexpected events. In general, sustainability, on the one hand, according to economic, social, and environmental considerations for both current and future generations, is focused on raising the standard of living. On the other hand, resilience focuses on the response of systems, including economic, social, and environmental systems, to widespread disruptions, persistent external pressure, and systems' ability to return to predisruptive status [19, 20]. Given the conditions mentioned in today's world, the need to design resilient and sustainable supply chains to deal with natural disasters and contagious diseases is fully felt.

According to the numerous reviewed literature on related research to the sustainable and resilient supply chain, no specific research proposes a robust mathematical model for designing a sustainable and resilient supply chain in four echelons of suppliers, manufacturers, distribution centers, and consumers based on redundancy methods. Hence, this study's main contributions to fill the gap are as follows: (1) development of a robust scenario-based stochastic programming model to design resilient supply portfolios by considering different objectives to achieve sustainable and resilient strategies. The model objectives include minimizing the total cost, minimizing the environmental and social performance scores of suppliers to design a sustainable supply chain, and (2) definition of four strategies for a resilient supply chain based on redundancy practices, which is given as follows.

The first and second strategies have been proposed for suppliers, as the upstream layer of a supply chain plays a vital role in a product's value chain. Pandemics and disasters cause some disruptions in supplying products. When the supplies are disrupted, the whole supply chain becomes vulnerable. Hence, one of the proper strategies to prevent this situation is to have backup suppliers. The second one is to force suppliers to carry the excess capacity of raw materials at a higher cost in the case of any disruption to other suppliers. This strategy helps supply raw materials through suppliers who are less affected by the disruption.

In the third strategy, a certain amount of raw materials is stored as safety stock at the factory before any disruption. The limited capacity of the factory and purchasing price that is less than at disruption time are two main factors for this decision. This strategy enhances the strength of the supply chain to meet demand in any disruption.

In the fourth strategy, in the distribution center, a certain amount of product is stored as a safety stock based on the capacity of the store and its storage cost. As mentioned, some disruptions increase consumers' demand; therefore, this strategy helps answer the growth of needs in the market, such as panic buying.

This article is formatted as follows: the literature review is indicated in Section 2, developing a mathematical model and robust scenario-based programming is indicated in Section 3, the numerical results and sensitivity analysis are explained in Section 4, and the article is concluded in Section 5.

2. Literature Review

In the last decade, one of the main fields in operations research and mathematical modelling is supply chain design (SCD). SCD helps a company to have a competitive advantage in the market [21]. The companies seek to increase the SC performance that prevents vulnerability [22]. The importance of accessing and sharing supply chain disruption information is essential for the proper deployment of disruption reduction strategies [23]. Because of the importance of this issue and reduction of systems' disruption effects, the authors of [24–26] examined the disruption recovery problems.

Supply chains today are more uncertain than ever. In the face of unforeseen disruptions, sustainability for supply chains is rewarding in terms of competitive advantage. However, the literature is still far from having a sustainable supply chain configuration [27]. Recently, considerable research has been done to simultaneously examine disruptions and the role of sustainability in supply chains that researchers such as Al-Saidi et al. [28], Babbitt et al. [29], Karmaker et al. [30], and Yadav et al. [31] studied these issues when encountered with COVID-19. Gholizadeh et al. [32] developed a sustainable logistic model with four objectives, which include minimizing the total cost, carbon emission coverage for vehicle selection, and the fraud function which gains from the big sharing of supply chain data as well as maximizing demand coverage for vehicle selection simultaneously. Robust fuzzy stochastic programming was used to deal with some uncertainty parameters in this research. They used the augmented ε-constraint approach to solve the multiobjective problem.

Infrastructures with reliability and safety are critical to the sustainability of advanced societies. To deal with the increase in destructive events such as man-made and natural disasters that attack infrastructure, resilience must be used as an integrated perspective in the process of system planning [33]. Various resilience strategies have been applied by researchers. Elluru et al. [34] proposed a location-routing problem with time window by considering reactive and proactive scenarios for disaster resilient supply chain. A reactive and proactive approach helps to solve disasters caused by disruption and design of the distribution system, which leads to create a resilient supply chain. Taleizadeh et al. [35] utilized surplus inventory as one of most important resilience strategies in distribution centers in supply chain competition. They proposed a mixed-integer programming model to deal with disruption. Ivanov and Dolgui [36] examined the supply chain resilience for COVID-19 outbreaks. They stated that survival at the level of intertwined supply network requires resilience in each part of the supply chain.

Previous research in this area shows that the theories of sustainability and resilience theory are often studied separately. However, considering the issue of sustainability alongside resilience helps to solve such problems more effectively. One of the first research studies investigating the sustainability-resilience relationship for SCD was conducted by Fahimnia and Jabbarzadeh [19]. In this study, a multiobjective optimization model quantifies the social and environmental performance of the supply chain by using a performance scoring method and a fuzzy goal programming approach was utilized to find trade-off solutions to solve a sustainable-resilient supply chain model. Some researchers studied the design of resilient supply chain network such as Zahiri et al. [37] and developed a multiobjective location-allocation model to design a sustainable-resilient supply chain network under uncertainty. To address the uncertainty, a new fuzzy stochastic-probabilistic programming model was developed. Also, the new measures of sustainability and resilience such as minimizing total cost, maximizing social satisfaction, and minimizing predefined environmental resilience measures were proposed. Jabbarzadeh et al. [1] proposed a hybrid approach to design a sustainable supply chain network. This approach remained sustainable while facing stochastic disruptions in a resilient way. In this article, the C-means fuzzy clustering method was applied to quantify and evaluate suppliers' sustainable performance while facing disruption. Ivanov [38] examined the intersections between sustainability and environmental features in supply chains. This simulation-based research focuses mainly on the analysis of sustainability factors and their role in reducing or increasing the ripple effect in supply chains. This model analyzed the dissemination of disruption in a supply chain by considering sustainability features (economic, social, and environmental aspects) to design a resilient supply chain structure. Souza et al. [39] stated designing a sustainable supply chain that can maximize profit and minimize the environmental effects, but how such policies can influence resilience is not clear. Hence, the ecosystem network analysis approach was proposed to evaluate the resilience when designing the sugar beet supply chain. In this study, a mixed-integer linear programming model was developed to minimize greenhouse gas emissions and maximize profit. Pavlov et al. [40] investigated a gap in research for designing resilient and sustainable supply chain. They presented a model to optimize network redundancy and proactive contingency plans as well. Disruption scenarios led to mitigating supply chain levels. In other words, in each scenario supply chain disruption restructured. In this study, an optimization-simulation method was developed. Table 1 summarizes the studies related to the present study.

Regarding the literature review published in supply chain management, it is perceived that the sustainability and resilience of the supply chain have recently been a heated debate among researchers. According to the literature on sustainable and resilient supply chains, there is no specific research that can propose a robust mathematical model for designing a sustainable and resilient three-echelon supply chain (supplier, manufacturer, and distribution centers) based on redundancy concepts. Accordingly, to cover the existing gap, in this study, an RSSP model is developed for designing resilient portfolios to achieve a sustainable and resilient strategy by considering conflicting goals. In the proposed model, the total cost, suppliers' environmental and social performance scores, and lead time are optimized. The concepts of redundancy-based flexibility and stability are examined in particular.

Table 1: Literature review of sustainability and resilience in supply chain network design.

Research	Sustainability	Resilience	MODM	Solving methods						Network redundancy
				*RFSP	**RP	***ε C	Hybrid approach	Simulation	Scoring method	
Fahimnia and jabbarzadeh [19]	✓	✓	✓						✓	
Zahiri et al. [37]	✓	✓	✓	✓						
Jabbarzadeh et al. [1]	✓	✓			✓		✓	✓		
Ivanov [38]	✓	✓						✓		
Pavlov et al. [40]	✓	✓								✓
Souza et al. [39]	✓	✓	✓							
Elluru et al. [34]		✓			✓					
Gholizadeh et al. [32]	✓		✓	✓		✓				
Taleizadeh et al. [35]		✓								
Ivanov and dolgui [41]		✓								
The research proposal	✓	✓	✓	✓					✓	✓

*RFSP = robust fuzzy stochastic programming; **RP = reactive/proactive approach; ***ε $C = e$-constraint approach.

The most remarkable innovations in the present paper compared to other research studies done previously are as follows:

(i) Considering a resilient supply portfolio along with designing a sustainable supply chain with economic, social, and environmental considerations

(ii) Proposing a new optimization model to design a sustainable and resilient supply chain considering sustainable objectives in both normal and disrupted conditions

(iii) Proposing four strategies for disruption risk management in a supply chain

(iv) Robust scenario-based programming method for facing disruption scenarios and reducing the decision risks resulted from the model

Accordingly, a multiobjective optimization model is developed for the problem. Different strategies are examined for supply chain resilience and facing disruption risks while providing so that the supply chain performs more effectively under disruption scenarios. Therefore, an RSSP method is applied so as to make the best of these strategies. In the end, for the trade-off between objectives, the LP metric method is used.

3. Mathematical Model

In this study, a resilient and sustainable supply chain is designed to deal with pandemics and natural disasters. The supply chain in this paper includes four levels. The first level is suppliers of raw materials, including current and backup suppliers. Although suppliers are independent of the supply chain in this paper, they are evaluated about their environmental and economic performance. The second and third levels are manufacturers and distribution centers that can keep the safety stock of raw material and the final product. The fourth level is customers that trigger the flow of the product in the supply chain by their demand. The model can decide to select suppliers, allocation order, location of capacitated centers, and flow of goods through the supply chain echelons. The information about customers' demand and the place is supposed to be available. Figure 1 indicates a schematic flow of the supply chain.

In this paper, minimizing the total cost is one of the strategic objectives to have a sustainable supply chain, and the model has been completed by adding the maximization of environmental and social scores of the supply chain. In the end, the final solution should be in a way that the supply chain would be resilient under any disruption scenario, and it can keep its performance. Four strategies are defined for a resilient supply chain under disruption scenarios that managers can use a combination of any of these strategies.

The first strategy: have a contract with some of the evaluated and selected suppliers as a backup to provide raw material during disruption.

The second strategy: have a contract with suppliers to use their overcapacity of raw material in any disruption. The price of purchasing the raw material is more than the usual conditions.

The third strategy: storage of raw material as safety stock at manufacturer centers before disruptions. The safety stock has holding costs, and the manufacturer has to make the right decision about the amount of safety stock.

: storage of finished product as safety stock at distribution centers to meet market demand under disruptions. The capacity of the distribution center and holding cost are two effective factors in this decision.

The model has the following assumptions:

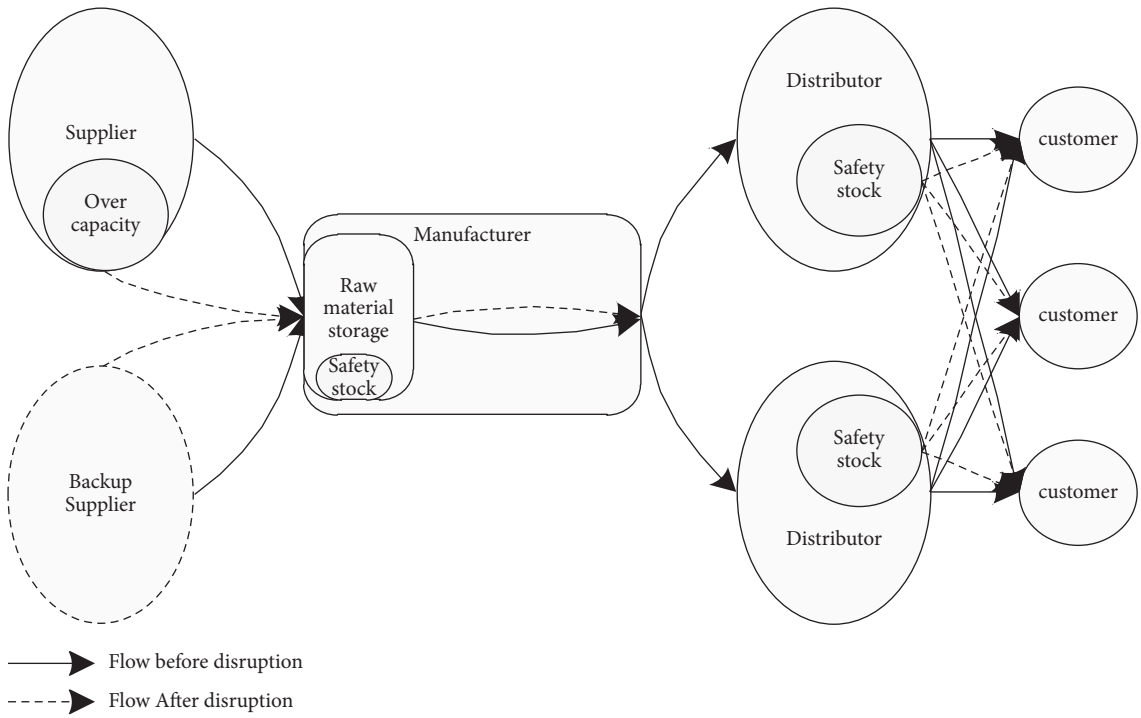


Figure 1: The supply chain network.

(i) The location of the supplier, their capacity, fixed cost of selection, and the purchase cost of raw materials are assumed to be determined

(ii) The backup supplier has been evaluated and selected

(iii) Social and environmental performance scores are already calculated (scores are from 0.1 to 1, 1 is for the best performance)

(iv) Potential locations for location of manufacturing centers have been determined

(v) Every factory can produce one or multiple productions with different capacities

(vi) Potential locations for installing distribution centers have been determined

(vii) Raw materials are preservable and not perishable

(viii) Produced goods are preservable and not perishable

(ix) The location and transportation are capacitated

Indices, parameters, and variables of the model are illustrated in Table 2.

Sets and indices:

Set of raw material suppliers: $\mathbf{S} = 1, 2, \ldots, \mathbf{s}$

Set of raw material backup suppliers: $\mathbf{S\prime} = 1, 2, \ldots, \mathbf{s\prime}$

Set of raw materials: $\mathbf{R} = 1, 2, \ldots, \mathbf{r}$

Set of products: $\mathbf{P} = 1, 2, \ldots, \mathbf{p}$

Set of manufacturer: $\mathbf{M} = 1, 2, \ldots, \mathbf{m}$

Set of distribution centers: $\mathbf{D} = 1, 2, \ldots, \mathbf{d}$

Set of demand centers, customers, or markets: $\mathbf{C} = 1, 2, \ldots, \mathbf{c}$

Set of capacity levels for manufacturers: $\mathbf{A} = 1, 2, \ldots, \mathbf{a}$

Set of capacity levels for distributing centers: $\mathbf{B} = 1, 2, \ldots, \mathbf{b}$

Set of probabilistic scenarios of supply disruption: $\mathbf{G} = 1, 2, \ldots, \mathbf{g}$

Set of disrupted supplier that is subset of suppliers: $\mathbf{ds} \subseteq \mathbf{S}$

Parameters:

Unit cost of purchasing raw material from supplier s: $\mathbf{pc_{sr}}$

Unit cost of purchasing raw material r from overcapacity of supplier s under scenario g: $\mathbf{pcg_{srg}}$

Fixed cost of making contract supplier s for material r: $\mathbf{fc_{sr}}$

Fixed cost of making contract backup supplier ss for material r: $\mathbf{fc\prime_{s\prime r}}$

Capacity of supplier s for raw material r: $\mathbf{cas_{sr}}$

Social performance score of supplier s for raw material r: $\mathbf{scs_{sr}}$

Social performance score of backup supplier s' for raw material r: $\mathbf{scs\prime_{s\prime r}}$

Social performance score of manufacturing center m for product p: $\mathbf{scm_{mp}}$

Social performance score of distribution center d for product p: $\mathbf{scd_{dp}}$

Environmental performance score of supplier s for raw material r: $\mathbf{ens_{sr}}$

Environmental performance score of supplier s' for raw material r: $\mathbf{ens\prime_{s\prime r}}$

Table 2: Indices, parameters, and decision variables.

Sets and indices	
Set of raw material suppliers	$\mathbf{S} = 1, 2, \ldots, \mathbf{s}$
Set of raw material backup suppliers	$\mathbf{S\prime} = 1, 2, \ldots, \mathbf{s}'$
Set of raw materials	$\mathbf{R} = 1, 2, \ldots, \mathbf{r}$
Set of products	$\mathbf{P} = 1, 2, \ldots, \mathbf{p}$
Set of manufacturer	$\mathbf{M} = 1, 2, \ldots, \mathbf{m}$
Set of distribution centers	$\mathbf{D} = 1, 2, \ldots, \mathbf{d}$
Set of demand centers, customers, or markets	$\mathbf{C} = 1, 2, \ldots, \mathbf{c}$
Set of capacity levels for manufacturers	$\mathbf{A} = 1, 2, \ldots, \mathbf{a}$
Set of capacity levels for distributing centers	$\mathbf{B} = 1, 2, \ldots, \mathbf{b}$
Set of probabilistic scenarios of supply disruption	$\mathbf{G} = 1, 2, \ldots, \mathbf{g}$
Set of disrupted supplier that is subset of suppliers	$\mathbf{ds} \subseteq \mathbf{S}$
Parameters	
Unit cost of purchasing raw material from supplier s	$\mathbf{pc_{sr}}$
Unit cost of purchasing raw material r from overcapacity of supplier s under scenario g	$\mathbf{pcg_{srg}}$
Fixed cost of making contract supplier s for material r	$\mathbf{fc_{sr}}$
Fixed cost of making contract backup supplier ss for material r	$\mathbf{fc_{sr}}$
Capacity of supplier s for raw material r	$\mathbf{cas_{sr}}$
Social performance score of supplier s for raw material r	$\mathbf{scs_{sr}}$
Social performance score of backup supplier $s\prime$ for raw material r	$\mathbf{scs'_{s'r}}$
Social performance score of manufacturing center m for product p	$\mathbf{scm_{mp}}$
Social performance score of distribution center d for product p	$\mathbf{scd_{dp}}$
Environmental performance score of supplier s for raw material r	$\mathbf{ens_{sr}}$
Environmental performance score of supplier $s\prime$ for raw material r	$\mathbf{ens'_{s'r}}$
Environmental performance score of manufacturing center m for product p	$\mathbf{enm_{mp}}$
Environmental performance score of distribution center d for product p	$\mathbf{end_{dp}}$
Disruption rate of raw material r for supplier s under scenario g	$\alpha_{\mathbf{srg}}$
If supplier s can provide raw material r, equal 1, otherwise 0	$\mathbf{av_{sr}}$
Amount of raw material r for producing a unit of product p	$\rho_{\mathbf{rp}}$
Volume of a unit of raw material r	$\mathbf{vl_r}$
Fixed cost of installing manufacturing center m to produce product p with capacity a	$\mathbf{fcm_{mpa}}$
Unit cost of producing product p at manufacturing center m	$\mathbf{pcm_{mp}}$
Production capacity of manufacturing center m for product p with capacity level a	$\mathbf{cap_{mpa}}$
Capacity of manufacturing center m for storage of raw materials r	$\mathbf{car_{mr}}$
Unit holding cost of raw material r (safety stock)	$\mathbf{hcm_r}$
Fixed cost of installing distribution center d with capacity level b	$\mathbf{fcd_{db}}$
Capacity of distribution d for holding products with capacity level b	$\mathbf{cad_{db}}$
Unit holding cost of product p in distributing center d (safety stock)	$\mathbf{hcd_{dp}}$
Unit volume of product p	$\mathbf{vlp_p}$
Demand of product p in market c	$\mathbf{dem_{cp}}$
Unit cost of transporting raw material r from supplier s to manufacturer m	$\mathbf{trs_{smr}}$
Unit cost of transporting raw material r from backup supplier $s\prime$ to manufacturer m	$\mathbf{trs_{s'mr}}$
Unit cost of transporting product p from manufacturer m to distributing center d	$\mathbf{trm_{m\,dp}}$
Unit cost of transporting product p from distributing center d to customer c	$\mathbf{trd_{dc\,p}}$
Probability of scenario g	$\mathbf{p_g}$
Flexibility rate of supplier s for raw material r under scenario g	$\mathbf{fl_{srg}}$
Rate of safety stock of raw material r that should be kept hold in manufacturing center m	$\mathbf{rsm_{mr}}$
Rate of safety stock of product d that should be kept hold in distributing center d	$\mathbf{rsg_{dp}}$
Decision variables	
If a supplier s is selected for raw material r, equal 1, otherwise 0	$\mathbf{w_{sr}}$
If a backup supplier $s\prime$ is selected for raw material r, equal 1, otherwise 0	$\mathbf{w'_{s'r}}$
If a manufacturing center m with capacity a for product p is installed, equal 1, otherwise 0	$\mathbf{x_{mpa}}$
If a distributing center d with capacity b is installed, equal 1, otherwise 0	$\mathbf{y_{db}}$
Amount of transported raw material r from supplier s to manufacturer m ordered before disruption	$\mathbf{qsm_{sm}r}$
Amount of transported product p from manufacturer m to distribution center d before disruption	$\mathbf{qmd_{m\,dp}}$
Amount of transported product p from distribution center d to market c before disruption	$\mathbf{q\ dc_{dc\,p}}$
Amount of producing product p at manufacturer m before disruption	$\mathbf{pm_{mp}}$
Amount of safety stock of raw material r at manufacturer m	$\mathbf{ssm_{mr}}$
Amount of safety stock of product p in distribution center d	$\mathbf{ssd_{p\,d}}$
Amount of transported raw material r from supplier s to manufacturer m under disruption scenario g	$\mathbf{qsmg_{smrg}}$
Amount of transported raw material r from backup supplier ss to manufacturer m under disruption scenario g	$\mathbf{qsmg'_{s'mrg}}$
Amount of transported product p from manufacturer m to distribution center d under disruption scenario g	$\mathbf{qm\ dg_{m\,dp\,g}}$
Amount of transported product p transported from distribution center d to market c under disruption scenario g	$\mathbf{q\ dcg_{dc\,pg}}$
Amount of produced product p at manufacturer m under disruption scenario g	$\mathbf{pmg_{mpg}}$
Amount of raw material r allocated to supplier s	$\mathbf{or_{sr}}$
Amount of raw material r allocated to backup suppliers$'$ under disruption scenario g	$\mathbf{or'_{s'rg}}$

Environmental performance score of manufacturing center m for product p: $\mathbf{enm_{mp}}$

Environmental performance score of distribution center d for product p: $\mathbf{end_{dp}}$

Disruption rate of raw material r for supplier s under scenario g: $\alpha_{\mathbf{srg}}$

If supplier s can provide raw material r, equal 1, otherwise 0: $\mathbf{av_{sr}}$

Amount of raw material r for producing a unit of product p: $\rho_{\mathbf{rp}}$

Volume of a unit of raw material r: $\mathbf{vl_r}$

Fixed cost of installing manufacturing center m to produce product p with capacity a: $\mathbf{fcm_{mpa}}$

Unit cost of producing product p at manufacturing center m: $\mathbf{pcm_{mp}}$

Production capacity of manufacturing center m for product p with capacity level a: $\mathbf{cap_{mpa}}$

Capacity of manufacturing center m for storage of raw materials r: $\mathbf{car_{mr}}$

Unit holding cost of raw material r (safety stock): $\mathbf{hcm_r}$

Fixed cost of installing distribution center d with capacity level b: $\mathbf{fcd_{db}}$

Capacity of distribution d for holding products with capacity level b: $\mathbf{cad_{db}}$

Unit holding cost of product p in distributing center d (safety stock): $\mathbf{hcd_{dp}}$

Unit volume of product p: $\mathbf{vlp_p}$

Demand of product p in market c: $\mathbf{dem_{cp}}$

Unit cost of transporting raw material r from supplier s to manufacturer m: $\mathbf{trs_{smr}}$

Unit cost of transporting raw material r from backup supplier s' to manufacturer m: $\mathbf{trs_{s'mr}}$

Unit cost of transporting product p from manufacturer m to distributing center d: $\mathbf{trm_{mdp}}$

Unit cost of transporting product p from distributing center d to customer c: $\mathbf{trd_{dcp}}$

Probability of scenario g: $\mathbf{p_g}$

Flexibility rate of supplier s for raw material r under scenario g: $\mathbf{fl_{srg}}$

Rate of safety stock of raw material r that should be kept hold in manufacturing center m: $\mathbf{rsm_{mr}}$

Rate of safety stock of product d that should be kept hold in distributing center d: $\mathbf{rsg_{dp}}$

Decision variables:

If a supplier s is selected for raw material r, equal 1, otherwise 0: $\mathbf{w_{sr}}$

If a backup supplier s' is selected for raw material r, equal 1, otherwise 0: $\mathbf{w'_{s'r}}$

If a manufacturing center m with capacity a for product p is installed, equal 1, otherwise 0: $\mathbf{x_{mpa}}$

If a distributing center d with capacity b is installed, equal 1, otherwise 0: $\mathbf{y_{db}}$

Amount of transported raw material r from supplier s to manufacturer m ordered before disruption: $\mathbf{qsm_{sm}r}$

Amount of transported product p from manufacturer m to distribution center d before disruption: $\mathbf{qmd_{mdp}}$

Amount of transported product p from distribution center d to market c before disruption: $\mathbf{qdc_{dcp}}$

Amount of producing product p at manufacturer m before disruption: $\mathbf{pm_{mp}}$

Amount of safety stock of raw material r at manufacturer m: $\mathbf{ssm_{mr}}$

Amount of safety stock of product p in distribution center d: $\mathbf{ssd_{pd}}$

Amount of transported raw material r from supplier s to manufacturer m under disruption scenario g: $\mathbf{qsmg_{smrg}}$

Amount of transported raw material r from backup supplier ss to manufacturer m under disruption scenario g: $\mathbf{qsmg'_{s'mrg}}$

Amount of transported product p from manufacturer m to distribution center d under disruption scenario g: $\mathbf{qmdg_{mdpg}}$

Amount of transported product p transported from distribution center d to market c under disruption scenario g: $\mathbf{qdcg_{dcpg}}$

Amount of produced product p at manufacturer m under disruption scenario g: $\mathbf{pmg_{mpg}}$

Amount of raw material r allocated to supplier s: $\mathbf{or_{sr}}$

Amount of raw material r allocated to backup suppliers$'$ under disruption scenario g: $\mathbf{or'_{s'r}g}$

The total cost objective (economic objective) in the normal conditions:

$$\begin{aligned}\min Z1 = &\sum_s\sum_r \mathrm{fc}_{\mathrm{sr}}.w_{sr} + \sum_{s'}\sum_r \mathrm{fc}'_{s'r}.w'_{s'r} + \sum_m\sum_p\sum_a \mathrm{fcm}_{\mathrm{mpa}}.x_{\mathrm{mpa}} + \sum_d\sum_b \mathrm{fcd}_{\mathrm{db}}.y_{\mathrm{db}}\\ &+ \sum_s\sum_m\sum_r \mathrm{trs}_{\mathrm{smr}}.\mathrm{qsm}_{\mathrm{smr}} + \sum_m\sum_d\sum_p \mathrm{trm}_{\mathrm{mdp}}.\mathrm{qmd}_{\mathrm{mdp}} + \sum_d\sum_c\sum_p \mathrm{trd}_{\mathrm{dcp}}.\mathrm{qdc}_{\mathrm{dcp}} + \sum_s\sum_r \mathrm{pc}_{\mathrm{sr}}.\mathrm{or}_{\mathrm{sr}}\\ &+ \sum_m\sum_p \mathrm{pcm}_{\mathrm{mp}}.\mathrm{pm}_{\mathrm{mp}} + \sum_m\sum_r \mathrm{hcm}_r.\mathrm{ssm}_{\mathrm{mr}} + \sum_d\sum_p \mathrm{hcd}_{\mathrm{dp}}.\mathrm{ssd}_{\mathrm{dp}}.\end{aligned} \tag{1}$$

Costs under disruption scenario g:

$$Z1_g = \sum_s \sum_r \mathrm{pc}_{\mathrm{sr}}.\mathrm{os}_{\mathrm{sr}} + \sum_s \sum_r \mathrm{pcg}_{\mathrm{srg}} \cdot \left(\sum_m \mathrm{qsmg}_{\mathrm{smrg}} - \mathrm{or}_{\mathrm{sr}} \right) + \sum_{s'} \sum_r \mathrm{pcg}_{s'\mathrm{rg}}'.\mathrm{or}_{s'\mathrm{rg}}' + \sum_s \sum_m \sum_r \mathrm{trs}_{\mathrm{smr}}.\mathrm{qsmg}_{\mathrm{smrg}}$$
$$+ \sum_{s'} \sum_m \sum_r \mathrm{trs}_{s'\mathrm{rm}}.\mathrm{qsmg}_{s'\mathrm{rmg}}' + \sum_m \sum_d \sum_p \mathrm{trm}_{\mathrm{mdp}}.\mathrm{qmdg}_{\mathrm{mdpg}} + \sum \sum \sum \mathrm{trd}_{\mathrm{dcp}}.\mathrm{qdcg}_{\mathrm{dcpg}} + \sum_m \sum_p \mathrm{pcm}_{\mathrm{mp}}.\mathrm{pmg}_{\mathrm{mpg}};\ \forall g \in G. \tag{2}$$

Social objective:

$$\max Z2 = \sum_s \sum_m \sum_r \mathrm{scs}_{\mathrm{sr}}.\ \mathrm{qsm}_{\mathrm{smr}} + \sum_m \sum_d \sum_p \mathrm{scm}_{\mathrm{mp}}.\mathrm{qmd}_{\mathrm{mdp}} + \sum_d \sum_c \sum_p \mathrm{scd}_{\mathrm{dp}}.\mathrm{qdc}_{\mathrm{dcp}}. \tag{3}$$

Social objective under disruption scenario s:

$$Z2_g = \sum_s \sum_m \sum_r \mathrm{scs}_{\mathrm{sr}}.\ \mathrm{qsmg}_{\mathrm{smrg}} + \sum_{s'} \sum_m \sum_r \mathrm{scs}_{\mathrm{rs'}m}'.\mathrm{qsmg}_{\mathrm{rs'mg}}' + \sum_m \sum_d \sum_p \mathrm{scm}_{\mathrm{mp}}.\mathrm{qmdg}_{\mathrm{mdpg}} + \sum_d \sum_c \sum_p \mathrm{scd}_{\mathrm{dp}}.\mathrm{qdcg}_{\mathrm{dcpg}}, \quad \forall g \in G. \tag{4}$$

Environmental objective:

$$\max Z3 = \sum_r \sum_s \sum_m \mathrm{ens}_{\mathrm{sr}}.\mathrm{qsm}_{\mathrm{rsm}} + \sum_p \sum_m \sum_d \mathrm{enm}_{\mathrm{mp}}.\mathrm{qmd}_{\mathrm{pmd}} + \sum_p \sum_d \sum_c \mathrm{end}_{\mathrm{dc}}.\mathrm{qdc}_{\mathrm{pdc}}. \tag{5}$$

Environmental objective under disruption scenario s:

$$Z3_g = \sum_r \sum_s \sum_m \mathrm{ens}_{\mathrm{sr}}.\mathrm{qsmg}_{\mathrm{rsmg}} \sum_r \sum_{s'} \sum_m \mathrm{ens}_{s'r}'.\mathrm{qsmg}_{\mathrm{rs'mg}}' + \sum_p \sum_m \sum_d \mathrm{enm}_{\mathrm{mp}}.\mathrm{qmdg}_{\mathrm{pmdg}} + \sum_p \sum_d \sum_c \mathrm{end}_{\mathrm{dp}}.\mathrm{qdcg}_{\mathrm{pdcg}}, \quad \forall g \in G. \tag{6}$$

In equation (1), total cost minimization as the economic objective of the resilient-sustainable supply chain design problem is considered before the disruption. The total costs consist of five main terms. First is the cost of facility establishment, including factory and distribution center, as well as supplier and backup supplier selection cost. The second term is the cost of transportation throughout supply chain networks. The third term relates to the cost of purchasing raw material. Fourth is the cost of production, and finally, the fifth term is the holding cost of safety stock of raw materials and final products in factories and distribution centers, respectively. In equation (2), the cost of every disruption scenario is formulated, including the raw material purchasing cost from supplier and backup supplier, production cost, and total transportation cost in the supply chain under scenario g. This equation is considered in scenario-based stochastic programming. In equation (3), the maximization of the social score of supply chain is defined. In equation (4), this score is calculated for every disruption scenario. In equation (5), the maximization of the environmental score of supply chain is presented. Moreover, the primary raw materials are green (environmental-friendly). Equation (6) calculates the environmental score for every disruption scenario.

Constraint:

$$\sum_{m} \mathrm{qsm}_{\mathrm{smr}} \leq \mathrm{or}_{\mathrm{sr}}, \quad \forall, r \in R, s \in S. \tag{7}$$

Constraint (7) states that the transported raw material to the manufacturer cannot exceed the allocated order to the supplier.

$$\mathrm{or}_{\mathrm{sr}} \leq \mathrm{cas}_{\mathrm{sr}}.\mathrm{av}_{\mathrm{sr}}.w_{\mathrm{sr}}, \quad \forall, r \in R, s \in S. \tag{8}$$

Constraint (8) ensures that when a supplier is selected and a contract is made, a factory can order the raw materials, and the total amount of order of raw material to an established supplier cannot exceed the supplier's capacity.

$$\mathrm{pm}_{\mathrm{mp}} \leq \sum_{a} \mathrm{cap}_{\mathrm{mpa}}.x_{\mathrm{mpa}} \sum_{a}, \quad \forall, m \in M, p \in P. \tag{9}$$

Constraint (9) shows that the production level of the installed factory does not exceed its capacity.

$$\sum_{a} \mathbf{x}_{\mathbf{mpa}} \leq 1, \quad \forall m \in M, p \in P. \tag{10}$$

Constraint (10) represents that there is, at most, a specific capacity for each factory and product.

$$\sum_{p} \rho_{\mathrm{rp}}.\mathrm{pm}_{\mathrm{mp}} + \mathrm{ssm}_{\mathrm{mr}} = \sum_{s} \mathrm{qsm}_{\mathrm{smr}}, \quad \forall. r \in R, m \in M. \tag{11}$$

Constraint (11) indicates the need for raw materials for production and safety stock provided by suppliers.

$$\mathrm{ssm}_{\mathrm{mr}} \geq \mathrm{rsm}_{\mathrm{mr}}. \sum_{p} \rho_{\mathrm{rp}}.\mathrm{pm}_{\mathrm{mp}}, \quad \forall, r \in R, m \in M. \tag{12}$$

Constraint (12) shows the minimal amount of raw material as safety stock in the manufacturing center.

$$\sum_{s'} W_{s'r}{}' \geq 1, \quad \forall r \in R. \tag{13}$$

Constraint (13) guarantees that the company contracts with at least a backup supplier before the disruption.

$$\mathrm{vl}_{r}.\mathrm{ssm}_{\mathrm{mr}} \leq \mathrm{car}_{\mathrm{mr}}.x_{\mathrm{mpa}}, \quad \forall, m \in M, r \in R, p \in P, a \in A. \tag{14}$$

Constraint (14) controls the capacity limitation of factories for the safety stock of raw materials.

$$\mathrm{pm}_{\mathrm{mp}} = \sum_{d} \mathrm{qmd}_{\mathrm{mdp}}, \quad \forall m \in M, p \in P. \tag{15}$$

Constraint (15) guarantees the balance between the production amount and the output flow in every factory.

$$\sum_{m} \sum_{p} \mathrm{vlp}_{p}.\mathrm{qmd}_{\mathrm{pmd}} \leq \sum_{b} \mathrm{cad}_{\mathrm{db}}. y_{\mathrm{db}}, \quad \forall, d \in D. \tag{16}$$

Constraint (16) ensures that transporting products from a factory to a distribution center is possible when locating a distribution center. Moreover, the maximum amount of products that can be transferred cannot be exceeding than the capacity of distributing centers.

$$\sum_{b} y_{db} \leq 1, \quad \forall, d \in D. \tag{17}$$

Constraint (17) indicates that only one distributing center can be established with a certain capacity in a potential location.

$$\sum_{m} \mathrm{qmd}_{\mathrm{mdp}} - \mathrm{ssd}_{\mathrm{dp}} = \sum_{c} \mathrm{qdc}_{\mathrm{dcp}}, \quad \forall, d \in D, \ p \in P. \tag{18}$$

Constraint (18) guarantees the balance between input and output flows in a distributing center.

$$\mathrm{ssd}_{\mathrm{dp}} \geq \mathrm{rsg}_{\mathrm{dp}}. \sum_{m} \mathrm{qmd}_{\mathrm{mdp}}, \quad \forall, d \in D, p \in P. \tag{19}$$

Constraint (19) shows that the at least amount of product as safety stock in the distributing center.

$$\sum_{d} \mathrm{qdc}_{\mathrm{dcp}} = \mathrm{dem}_{\mathrm{cp}}, \quad \forall, c \in C, p \in P. \tag{20}$$

Constraint (20) is to satisfy the market's demand for every product before any disruption.

The abovementioned constraints are for the normal condition without any disruption. Equations related to disruption scenarios are defined as follows to make the supply chain resilient.

$$\sum_{m} \mathrm{qsmg}_{\mathrm{smrg}} \leq or_{sr}.\left(1 - \alpha_{\mathrm{srg}}\right), \quad \forall, r \in R, s \in \mathrm{sd}, \ g \in G. \tag{21}$$

Constraint (21) expresses that orders are reduced based on disruption rate under every disruption scenario for some suppliers.

$$\sum_{m} \mathrm{qsmg}_{\mathrm{smrg}} \leq \mathrm{or}_{\mathrm{sr}}.\left(1 + \mathrm{fl}_{\mathrm{srg}}\right), \quad \forall, r \in R, s \in \frac{S}{\mathrm{sd}}, \ g \in G. \tag{22}$$

Constraint (22) shows undisrupted supplier's flexibility to provide the raw material from overcapacity.

$$\mathrm{or}_{\mathrm{sr}}.\left(1 + \mathrm{fl}_{\mathrm{srg}}\right) \leq \mathrm{cas}_{\mathrm{sr}}.\mathrm{av}_{\mathrm{sr}}.w_{\mathrm{sr}}, \quad \forall, r \in R, s \in \frac{S}{\mathrm{sd}}. \tag{23}$$

Constraints (23) expresses that it is possible to use overcapacity of suppliers that are not disrupted under any disruption scenario, and the total amount of order of raw material cannot exceed the supplier's capacity.

$$\sum_{m} \mathrm{qsmg}_{s'\mathrm{mrg}}{}' \leq \mathrm{or}_{s'r}{}', \quad \forall, r \in R, s' \in S', g \in G, \tag{24}$$

$$\mathrm{or}_{s'\mathrm{rg}}{}' \leq \mathrm{cas}_{s'r}{}'.w_{s'r}{}', \quad \forall, r \in R, s \in S. \tag{25}$$

Constraints (24) and (25) state that the transported raw material to the manufacturer cannot exceed the allocated order to the backup supplier.

$$\sum_{p} \rho_{\mathrm{mp}}.\mathrm{pmg}_{\mathrm{mpg}} = \sum_{s} \mathrm{qsmg}_{\mathrm{smrg}} + \sum_{s'} \mathrm{qsmg}_{s'\mathrm{mrg}}' + \mathrm{ssm}_{mr}, \quad \forall, r \in R, m \in M, g \in G. \tag{26}$$

Constraint (26) explains that raw materials for producing products in a factory are equal to the total raw materials supplied under that disruption scenario by supplier and backup supplier plus safety stock before the disruption scenario.

$$\mathrm{pmg}_{\mathrm{mpg}} \leq \sum_{a} \mathrm{cap}_{\mathrm{mpa}}.x_{\mathrm{mpa}}, \quad \forall, m \in M, p \in P. \tag{27}$$

Constraint (27) guarantees that, under any disruption scenario, production level in every factory (if established) does not exceed its capacity.

$$\mathrm{pmg}_{\mathrm{mpg}} = \sum_{d} \mathrm{qmdg}_{\mathrm{pmdg}}, \quad \forall, m \in M, p \in P, \ g \in G. \tag{28}$$

Constraint (28) determines the balance between the production amount and output flow in every factory under disruption scenarios.

$$\sum_{m} \sum_{p} \mathrm{vlp}_{p}.\mathrm{qmdg}_{\mathrm{pmdg}} \leq \sum_{b} \mathrm{cad}_{\mathrm{db}}.\ y_{\mathrm{db}}, \quad \forall, d \in D, \ g \in G. \tag{29}$$

Constraint (29) ensures that transporting products from factories to distributing centers requires locating a distributing center under a disruption scenario, and the accomplishment of this process depends on the capacity of the distributing center.

$$\sum_{m} \mathrm{qmdg}_{\mathrm{pmdg}} + \mathrm{ssd}_{\mathrm{pd}} = \sum_{c} \mathrm{qdcg}_{\mathrm{pdcg}}, \quad \forall, d \in D, \ p \in P, \ g \in G. \tag{30}$$

Constraint (30) shows the balance between inputs and outputs in distribution centers under disruption scenarios.

$$\sum_{d} \mathrm{qdcg}_{\mathrm{pdcg}} = \mathrm{dem}_{\mathrm{cp}}, \quad \forall, c \in C, \ p \in P, g \in G. \tag{31}$$

Constraint (32) explains the necessity of fulfilling market demand under disruption scenarios.

$$\begin{aligned} &\mathrm{qsm}_{\mathrm{rsm}}, \mathrm{qmd}_{\mathrm{pmd}}, \mathrm{qdc}_{\mathrm{pdc}}, \mathrm{pm}_{\mathrm{mp}}, \mathrm{ssm}_{\mathrm{mr}}, \mathrm{ssd}_{\mathrm{pd}}, \mathrm{qsmg}_{\mathrm{rsmg}}, \\ &\quad \mathrm{qmdg}_{\mathrm{pmdg}}, \mathrm{qdcg}_{\mathrm{pdcg}}, \mathrm{pmg}_{\mathrm{mpg}}, \mathrm{or}_{\mathrm{sr}} \geq 0 \\ &\quad w_{sr}, \ x_{\mathrm{mpa}}, y_{\mathrm{db}} \in \{0, 1\}. \end{aligned} \tag{32}$$

Moreover, finally, (31) determines the decision variables' range.

3.1. Robust Scenario-Based Stochastic Programming. Scenario-based stochastic programming (SSP) is one of the approaches that can be used in stochastic optimization. In this type of problem, there are two kinds of variables. One group of variables is independent of scenarios, and another group is impacted by scenarios [42]. The objective function in SSP is the mean value of the system under entire scenarios [43]. Generally, this approach can be illustrated as follows:

$$\begin{cases} \mathrm{Min}\, E(Z) = \sum_{s \in S} p_g.z_g, \\ z_g = c_g^{\mathrm{T}}.x_g + d_g^{\mathrm{T}} y \ \forall, g \in G, \\ A_g x_g + K_g y = b_g, \quad \forall, g \in G, \\ Ry = q, \\ y \in Y, \quad x_g \geq 0. \end{cases} \tag{33}$$

Here, z_g is the amount of objective function under scenario $g \in G$ and x_g is a dependent decision variable and y is an independent decision variable to the scenario. Also, c_g, d_g, A_g, K_g, and b_g are dependent and R and q are independent parameters to the scenario and p_g is the probability of occurring scenario $g \in G$.

The proposed method in [43] can consider optimality robustness called robust scenario-based stochastic programming (RSSP) as a quadratic model, so Yu and Li [44] linearized the model as follows:

$$\begin{cases} \mathrm{Min} \sum_{g \in G} pr_g.z_g + \lambda \sum_{g \in G} p_g.\left(z_g - \sum_{g' \in G} p_{g'}.z_{g'} + 2\theta_g \right) + \omega \sum_{g \in G} \pi_g\left(\xi_g^{+} + \xi_g^{-}\right), \\ z_g - \sum_{g' \in G} \pi_{g'}.z_{g'} + \theta_g \geq 0, \quad \forall, g \in G, \\ z_g = c_g^{\mathrm{T}}.x_g + d_g^{\mathrm{T}} y, \quad \forall, g \in G, \\ A_g x_g + K_g y = b_g + \left(\xi_g^{+} - \xi_g^{-}\right), \quad \forall, g \in G, \\ Ry = q, \\ \xi_g^{+}, \xi_g^{-}, \theta_g \geq 0, \\ x_g, y \geq 0, \end{cases} \tag{34}$$

where ξ_g is deviation variable in each scenario, ω is the risk aversion coefficient, and λ is the weight of variance of the cost in different scenarios. The amount of ω in an example can be cost or penalty of shortage in a supply chain.

Constraint (31) is modified to constraint (34) to apply RSSP while facing disruption scenarios for satisfying demand:

$$\sum_{d} \mathrm{qdcg}_{\mathrm{pdcg}} = \mathrm{dem}_{\mathrm{cp}} + \left(\xi^{+}_{\mathrm{cpg}} - \xi^{-}_{\mathrm{cpg}}\right), \quad \forall, c \in C, p \in P, g \in G. \tag{35}$$

Supply chain capacity usually reduces under disruption scenarios, so $\xi^{+}_{\mathrm{cps}} = 0$ and just the value $\xi^{-}_{\mathrm{cps}} \geq 0$ in the constraint can make flexibility under disruption scenarios satisfying the demands. It is also assumed $\lambda = 0$, and just the mean performance of the system is taken into account. Model robustness is functioned with coefficient $\omega > 0$.

Three objective functions of the model defined by applying the RSSP approach are illustrated in the following equation:

$$\begin{cases} \mathrm{Min}\, Z_1^{\mathrm{RO}} = Z1 + \sum_{g \in G} p_g . Z1_g + \omega \sum_{g \in G} \sum_{c} \sum_{p} p_g . \xi^{-}_{\mathrm{cpg}}, \\ \mathrm{Max}\, Z_2^{\mathrm{RO}} = Z2 + \sum_{g \in G} p_g . Z2_g, \\ \mathrm{Max}\, Z_3^{\mathrm{RO}} = Z3 + \sum_{g \in G} p_g . Z3_g, \\ \mathrm{s.t.} \\ \mathrm{Eqs}\, 7 - 28, \\ \sum_{d} \mathrm{qdcg}_{\mathrm{pdcg}} = \mathrm{dem}_{\mathrm{cp}} + \left(\xi^{+}_{\mathrm{cpg}} - \xi^{-}_{\mathrm{cpg}}\right), \quad \forall, c \in C, p \in P, g \in G, \\ \mathrm{Eq}\, 30, \\ \xi^{-}_{\mathrm{cpg}} \geq 0, \\ \omega > 0. \end{cases} \tag{36}$$

3.2. ε-Constraint Method. Multiobjective problems include more than one objective function, chiefly conflicting objectives. As most real-world optimization problems should be modeled in multiobjective problems, this mathematical modelling area has been widely employed for decades. Many methods and approaches have been developed to tackle multiobjective problems [45].

The ε-constraint method is one of the most efficient methods which can be applied in multiobjective problems. Accordingly, one of the objective functions is optimized and other objectives are added to the constraints. The steps of this method are as follows:

(i) The payoff table is calculated for all objectives

(ii) One objective is selected to be optimized, and other objectives are considered as constraints

(iii) The ε values account for the constrained objective function

(iv) Efficient solutions (Pareto front) to the problem is achieved by the main objective optimization and the epsilon's parametric variation

Table 3: The amount of input parameters.

Parameter	Amount of parameter
$\mathbf{cap_{mpa}}$	A ~ U(500,900)
$\mathbf{car_{mr}}$	U(300,500)
$\mathbf{hcm_{r}}$	U(1,3)
$\mathbf{fcd_{db}}$	U(5000,7000)
$\mathbf{cad_{db}}$	B~(500,900)
$\mathbf{hcd_{pd}}$	U(2,4)
vlp	1
$\mathbf{dem_{cp}}$	U(200,300)
$\mathbf{trs_{rsm}}$	U(1,2)
$\mathbf{trm_{pmd}}$	U(1,2)
$\mathbf{trd_{pdc}}$	U(1,2)
$\mathbf{pcm_{mp}}$	U(2,3)
$\mathbf{vl_{r}}$	1
$\mathbf{cls_{cp}}$	U(3,6)
$\alpha_{\mathbf{rsg}}$	U(0.4,1)
$\mathbf{pc_{rs}}$	U(1,2)
$\mathbf{pcg_{rsg}}$	U(2,4)
$\mathbf{fc_{sr}}$	U(50,100)
$\mathbf{cas_{sr}}$	U(300,800)
$\mathbf{scs_{sr}}$	U(0.6,1)
$\mathbf{ens_{sr}}$	U(0.6,1)
$\mathbf{scs\prime_{s,r}}$	U(0.6,1)
$\mathbf{scm_{mp}}$	U(0.6,1)
$\mathbf{scd_{dp}}$	U(0.6,1)
$\mathbf{enm_{mp}}$	U(0.6,1)
$\mathbf{end_{dp}}$	U(0.6,1)
$\rho_{\mathbf{mp}}$	1
$\mathbf{fcm_{mpa}}$	U(7000,10000)
$\mathbf{fl_{srg}}$	U(0.1,0.5)

(v) Report the Pareto solutions

4. Numerical Example

A numerical example under different scenarios has been applied to check feasibility and evaluation of the performance of the model (Table 3). In this example, the supply chain network has four echelons, including suppliers, manufacturers, distributors, and customers. It is assumed that this supply chain has two manufacturers currently with two different capacity levels. There are three distribution centers with two different capacity levels and five demand centers as customers.

It is also assumed that two types of raw materials are needed to produce a product that is provided by four potential suppliers and two backup suppliers. In global disasters and pandemics, suppliers are one of the main parts of a supply chain that is vulnerable; hence, the four various disruption scenarios to prove the performance of the model are shown in Table 4. In this table, for example, the probability of scenario one is 0.75, and in this scenario, the disruption rate of supplier S1 for material r1 is 0.6.

The computations are run on an Intel® Core™ i5-2.5 GHz processor and 4 GB of Ram by GAMS 24.0.1 software using Bonmin solver.

In this part, the model is optimized to achieve results before any disruption. Every objective of the problem is optimized solely, and the value of other objectives is

Table 4: Disruption scenarios for suppliers.

No. of scenario	Probability of scenario (p_g)	Disrupted suppliers	Disruption rate of suppliers (s) for raw material (r) under scenarios (g)							
			S_1		S_2		S_3		S_4	
			r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2
1	0.5	$S1$, $S2$	0.6		0.4					
2	0.3	$S1$, $S2$	1	1	0.4	0.4				
3	0.15	$S1$, $S2$, $S4$	0.6		0.6				0.5	0.5
4	0.05	$S1$, $S2$, $S3$, $S4$		0.4	0.5		1	1	0.5	0.5

calculated based on the optimized objective to show the trade-off between objectives. For example, first, Z_1 (economic objective) is minimized, and the value of other objectives are achieved based on the optimized value of Z_1. Table 4 shows the values of the minimized costs and other objectives' amount such the environmental score before any disruption scenarios. Moreover, the amount of raw material and safety stock of raw material are illustrated in Table 5. Similarly, Tables 6 and 7, respectively, show the optimized value of social score (Z_2) and environmental score (Z_3).

The proposed model is a multiobjective model, so the ε-constraint method is used to convert the model to a single-objective model. Table 8 shows the payoff table of the three mentioned objectives.

To use the propose method, the first objective was optimized and the rest objectives were placed in the constraints by seven different values of ε to show conflicting of the objective function and reach the Pareto solutions. Table 9 shows Pareto solutions based on seven different ε values.

Additionally, Figure 2 illustrates a comparison of the values of objective functions Z_1-Z_2 and Z_1-Z_3 as an example to prove the conflict between two objectives.

It is evident from Figure 2 and Table 8 that, to cover the economic priorities as the first objective function, the social and environmental values in the supply chain are decreasing.

4.1. Results of Robust Scenario-Based Stochastic Programming. The outputs of the model in the normal condition are not suitable for disruption because the supply chain can lose its performance. Therefore, using the robust scenario-based stochastic programming approach helps to consider a summation of before and after disruption condition together. In the solved example, amount of $\omega = 1$. The value of each objective in this approach is shown in Table 10.

The results show that amount of total cost in this method increases in comparison to the normal condition. All suppliers are involved in providing the raw material and also the selected backup suppliers to deal with disruption. Hence, it is necessary to have sensitivity analysis on model robustness parameter (ω) before applying this method. For this reason, the effect of three different values of (ω) on objective function (Z_1^{RO}) and deviation variable from demand (ξ^{-}_{cpg}) is shown in Table 11.

Table 9 shows the changes in the total cost and the amount of unfulfilled demand (the shortage due to supply disruption) relative to changes in (ω). The amount of (ω) can be considered as a penalty of shortage in the supply chain. It is clear that amount of unfulfilled demand is decreasing by increasing the amount of penalty (ω). The model is completely robust with $\omega \geq 15$, and the amount of shortage equals 0. Figure 3 can prove that the model is entirely risk-averse when the total cost gets the maximum value, over 64621. Despite reducing the total cost, the total unfulfilled demand for two products increases significantly, and model robustness is intensely mitigated with <15.

By comparing, the results in Tables 5, 10, and 11 demonstrate that the supply chain has the lowest cost in the normal condition. So, if some disruptions happen, the supply chain is not prepared to meet the demands in the market. On the other hand, the total cost increases when the supply chain considers the disruption scenarios and is responsive to the demand. Hence, regarding the current situation in the world that pandemics and disasters impose a high cost on industries and supply chains, putting more investment into a resilient and sustainable supply chain is reasonable. The resilient and sustainable supply chain can mitigate risk and increase the responsiveness of SC in the future of the world.

By implementing the model on a numerical study, the numerical results show that the proposed model and approaches have the ability to solve the problem of sustainable-resilient supply chain design. Based on the numerical results on a hypothetical supply chain under normal conditions, if only the economic objective is considered, the minimum cost for the supply chain occurs, while the answer is debatable in two ways. First, other objectives may not be placed in good conditions, and social and environmental performance scores may be low in the supply chain; second, in any supply disruption scenario, supply chain performance is severely disrupted, and high costs, including shortages and nonsupply, occur. In other words, the answer is only economical, neither sustainable nor resilient. Similarly, the answer can only be optimal in normal conditions from an environmental or social perspective, while other objectives are not optimal, and the answer is not resilient. Also, supplier disruptions are shown with several possible scenarios to show different situations in the real world. In each

Table 5: Value of objective functions and supply portfolio by optimizing Z_1.

	Environmental score 4050	Social score 4246	Optimized value of Z_1 55990
Amount of raw material 1		Amount of raw material 1	Supplier
500		300	S1
173		500	S2
450		373	S3
450		400	S4
Amount of SS raw material 2		Amount of SS raw material 1	Manufacturer
11		38	M1
57		23	M2

Table 6: Value of objective functions and supply portfolio by optimizing Z_2.

	Environmental score 4095	Optimized value of social score (Z_2) 4527	Economic score 75350
Amount of raw material 2		Amount of raw material 1	Supplier
300		500	S1
500		500	S2
400		450	S3
400		450	S4
Amount of SS raw material 2		Amount of SS raw material 1	Manufacturer
27		62	M1
63		0	M2

Table 7: Value of objective functions and supply portfolio by optimizing Z_3.

	Optimized value of environmental score (Z_3) 439	Social score 4359	Economic score 76873
Amount of raw material 2		Amount of raw material 1	Supplier
300		500	S1
500		500	S2
400		450	S3
400		450	S4
Amount of SS raw material 2		Amount of SS raw material 1	Manufacturer
18		86	M1
180		18	M2

Table 8: Payoff table of three objectives.

Objective	Z_1	Z_2	Z_3
Z_1	55990	75350	76873
Z_2	4246	4527	4359
Z_3	4050	4095	4391

Table 9: Pareto solutions of objective functions.

Objective function	$\varepsilon 0$	$\varepsilon 1$	$\varepsilon 2$	$\varepsilon 3$	$\varepsilon 4$	$\varepsilon 5$	$\varepsilon 6$
Z_1	56283	56440	56623	56806	56988	57173	57420
Z_2	4359	4375	4392	4409	4426	4443	4459
Z_3	4095	4124	4154	4183	4213	4243	4272

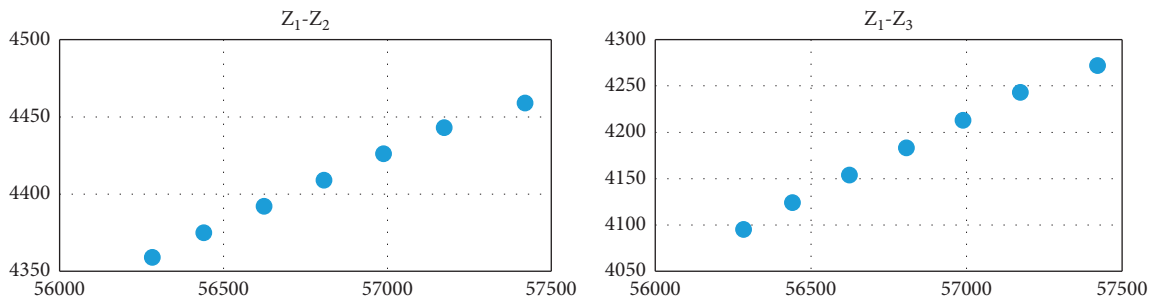


Figure 2: Pareto solutions of Z_1-Z_2 and Z_1-Z_3.

Table 10: Value of objective functions and supply portfolio using the SSP approach.

Mean of environmental score under normal and disruption Z_3^{RO}	Mean of social score under normal and disruption Z_2^{RO}		Mean of economic score under normal and disruption Z_1^{RO}
6603	7076		56671
Amount of raw material 2	Amount of raw material 1		Supplier
535	357	g_1	S1
492	357	g_2	S2
44	323	g_3	S3
500	535	g_4	S4
Amount of raw material 2	Amount of raw material 1	g_1	Backup supplier
500	0	g_2	S_1'
500	0	g_3	
463	0	g_4	
367	0	Scenario	
0	500	g_1	S_2'
0	500	g_2	
0	463	g_3	
0	191	g_4	

Table 11: Effect of different amounts of (ω) on objective function and deviation of demand.

			1				5				10				15		
ω	$\mathbf{Z_1^{RO}}$		56671				59840				63635				64621		
								Scenarios									
Customer	Product	g_1	g_2	g_3	g_4	g_1	g_2	g_3	g_4	g_1	g_2	g_3	g_4	g_1	g_2	g_3	g_4
C1	P1	62	44	5	0	100	100	100	100	0	0	95	6	0	0	0	0
	P2	150	150	150	150	150	150	150	150	150	150	150	150	0	0	0	0
C2	P1	120	120	0	120	0	0	27	120	0	120	120	0	0	0	0	0
	P2	0	0	0	100	0	0	0	0	31	0	0	0	0	0	0	0
C3	P1	100	100	100	10	0	36	100	100	100	100	15	100	0	0	0	0
	P2	0	0	0	1	120	0	0	111	22	55	0	110	0	0	0	0
C4	P1	0	110	110	110	22	110	110	0	22	55	0	100	0	0	0	0
	P2	100	100	100	100	100	100	100	100	95	100	100	100	0	0	0	0

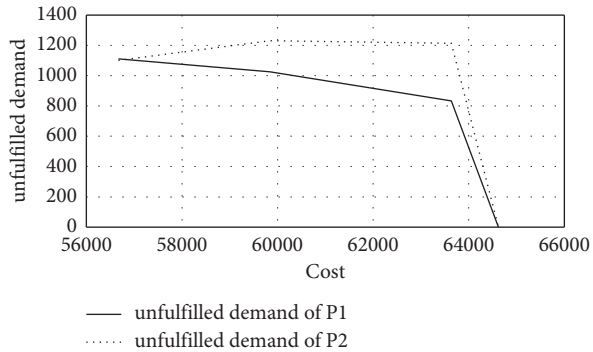


Figure 3: Trade-off between total cost, unfulfilled demand, and (ω).

scenario, it is determined which suppliers and how much of their supply capacity are lost. Then, considering disruption scenarios for suppliers as well as their social and environmental performance, the resilient-sustainable supply chain using the RSSP approach has been achieved.

5. Conclusion

In this paper, a multiobjective model for a sustainable and resilient supply chain network is formulated to deal with disruptions in disasters. The model optimizes a sustainable-resilient supply chain. The main objective of this model is to minimize supply chain costs, which is considered along with sustainable objectives that include maximizing social and environmental scores. In this regard, four strategies are proposed for a resilient supply chain based on redundancy practices. In this context, an RSSP method is applied to make the best of these strategies.

The ε-constraint method with different values of ε is utilized to convert the multiobjective model to the single-objective model, and it is solved by GAMS software. The performance of the model is evaluated by considering normal and disrupted conditions using a numerical example. Given the output of the numerical example and considering only the total cost objective, the minimum cost of the supply chain results in normal conditions. This solution is only economical and neither sustainable nor resilient. Therefore, under the supply disruption scenario, the supply chain performance is severely disrupted and many costs may be imposed, lack of supply. The model proves that the amount of cost objective increases when the disruption condition considers the objective function. However, it can be concluded that the supply chain has the lowest cost in the normal condition and the supply chain cannot tolerate any vulnerability. On the other hand, the sustainable and resilient supply chain can mitigate risk and take disruption by more investment.

One aspect that could be considered in future research is using a real case study instead of random numbers. The proposed model is NP-hard model; new metaheuristic algorithms are also proposed to solve this problem on larger scales. Finally, in disaster situations, most of the parameters are uncertain; using fuzzy number helps to improve the model.

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.

References

[1] A. Jabbarzadeh, B. Fahimnia, and F. Sabouhi, "Resilient and sustainable supply chain design: sustainability analysis under disruption risks," *International Journal of Production Research*, vol. 56, no. 17, pp. 5945–5968, 2018.

[2] M. Christopher and H. Peck, "Building the resilient supply chain," 2004.

[3] S. Rezapour, R. Z. Farahani, and M. Pourakbar, "Resilient supply chain network design under competition: a case study," *European Journal of Operational Research*, vol. 259, no. 3, pp. 1017–1035, 2017.

[4] K. B. Hendricks and V. R. Singhal, "The effect of supply chain glitches on shareholder wealth," *Journal of Operations Management*, vol. 21, no. 5, pp. 501–522, 2003.

[5] J.-H. Thun and D. Hoenig, "An empirical analysis of supply chain risk management in the German automotive industry," *International Journal of Production Economics*, vol. 131, no. 1, pp. 242–249, 2011.

[6] G. A. Zsidisin, "A grounded definition of supply risk," *Journal of Purchasing and Supply Management*, vol. 9, no. 5-6, pp. 217–224, 2003.

[7] https://www.nationalgeographic.com/history/reference/modern-history/why-we-evolved-to-feel-panic-anxiety/.

[8] M. A. Moktadir, A. Dwivedi, A. Rahman et al., "An investigation of key performance indicators for operational excellence towards sustainability in the leather products industry," *Business Strategy and the Environment*, vol. 29, no. 8, pp. 3331–3351, 2020.

[9] T. Linton and B. Vakil, "Coronavirus is proving we need more resilient supply chains," *Harward business review, March*, vol. 5, p. 2020, 2020.

[10] P.-I. Lee, Y.-L. Hu, P.-Y. Chen, Y.-C. Huang, and P.-R. Hsueh, "Are children less susceptible to COVID-19?" *Journal of Microbiology, Immunology, and Infection*, vol. 53, no. 3, pp. 371-372, 2020.

[11] L. Breen and C. Hannibal, "Learning from the Covid-19 pandemic: planning, controlling and driving change for greater resilience in supply chains: special issue call for

papers," *Supply Chain Management: International Journal*, vol. 17, 2020.

[12] D. Ivanov, "Viable supply chain model: integrating agility, resilience and sustainability perspectives—lessons from and thinking beyond the COVID-19 pandemic," *Annals of Operations Research*, vol. 33, pp. 1–21, 2020.

[13] E. Bagshaw and D. Powell, *Supermarkets Stockpile, Toilet Paper Production Runs 24 Hours*, The Sydney Morning Herald, New South Wales, Australia, 2020.

[14] P. Haren and D. Simchi-Levi, "How coronavirus could impact the global supply chain by mid-March," *Harvard Business Review*, vol. 28, 2020.

[15] T. K. Dasaklis, C. P. Pappis, and N. P. Rachaniotis, "Epidemics control and logistics operations: a review," *International Journal of Production Economics*, vol. 139, no. 2, pp. 393–410, 2012.

[16] M. Koyuncu and R. Erol, "Optimal resource allocation model to mitigate the impact of pandemic influenza: a case study for Turkey," *Journal of Medical Systems*, vol. 34, no. 1, pp. 61–70, 2010.

[17] L. M. Koonin, "Novel coronavirus disease (COVID-19) outbreak: now is the time to refresh pandemic plans," *Journal of Business Continuity & Emergency Planning*, vol. 13, no. 4, pp. 1–15, 2020.

[18] M. Kamalahmadi and M. M. Parast, "A review of the literature on the principles of enterprise and supply chain resilience: major findings and directions for future research," *International Journal of Production Economics*, vol. 171, pp. 116–133, 2016.

[19] B. Fahimnia and A. Jabbarzadeh, "Marrying supply chain sustainability and resilience: a match made in heaven," *Transportation Research Part E: Logistics and Transportation Review*, vol. 91, pp. 306–324, 2016.

[20] D. Marchese, E. Reynolds, M. E. Bates, H. Morgan, S. S. Clark, and I. Linkov, "Resilience and sustainability: similarities and differences in environmental management applications," *The Science of the Total Environment*, vol. 613, pp. 1275–1283, 2018.

[21] Z. Azarmand and E. Neishabouri, *Location Allocation Problem, Facility Location*, Springer, Heidelberg, Germany, 2009.

[22] L. V. Snyder, Z. Atan, P. Peng, Y. Rong, A. J. Schmitt, and B. Sinsoysal, "OR/MS models for supply chain disruptions: a review," *IIE Transactions (Institute of Industrial Engineers)*, vol. 48, no. 2, pp. 89–109, 2016.

[23] A. V. Thomas and B. Mahanty, "Dynamic assessment of control system designs of information shared supply chain network experiencing supplier disruption," *Operational Research*, vol. 27, 2018.

[24] R. Aldrighetti, D. Battini, D. Ivanov, and I. Zennaro, "Costs of resilience and disruptions in supply chain network design models: a review and future research directions," *International Journal of Production Economics*, vol. 24, Article ID 108103, 2021.

[25] M. S. Golan, B. D. Trump, J. C. Cegan, and I. Linkov, "Supply chain resilience for vaccines: review of modeling approaches in the context of the COVID-19 pandemic," *Industrial Management & Data Systems*, vol. 43, 2021.

[26] B. Scala and C. F. Lindsay, "Supply chain resilience during pandemic disruption: evidence from healthcare," *Supply Chain Management: International Journal*, vol. 26, 2021.

[27] A. Durmaz, H. Demir, and B. Sezen, "The role of negative entropy within supply chain sustainability," *Sustainable Production and Consumption*, vol. 34, 2021.

[28] M. Al-Saidi and H. Hussein, "The water-energy-food nexus and COVID-19: towards a systematization of impacts and responses," *The Science of the Total Environment*, vol. 779, Article ID 146529, 2021.

[29] C. W. Babbitt, G. A. Babbitt, and J. Oehman, "Behavioral impacts on residential food provisioning, use, and waste during the COVID-19 pandemic," *Sustainable Production and Consumption*, vol. 28, 2021.

[30] C. L. Karmaker, T. Ahmed, S. Ahmed, S. M. Ali, M. A. Moktadir, and G. Kabir, "Improving supply chain sustainability in the context of COVID-19 pandemic in an emerging economy: exploring drivers using an integrated model," *Sustainable production and consumption*, vol. 26, pp. 411–427, 2021.

[31] S. Yadav, S. Luthra, and D. Garg, "Modelling Internet of things (IoT)-driven global sustainability in multi-tier agri-food supply chain under natural epidemic outbreaks," *Environmental Science and Pollution Research*, vol. 28, no. 13, pp. 16633–16654, 2021.

[32] H. Gholizadeh, H. Fazlollahtabar, and M. Khalilzadeh, "A robust fuzzy stochastic programming for sustainable procurement and logistics under hybrid uncertainty using big data," *Journal of Cleaner Production*, vol. 258, Article ID 120640, 2020.

[33] Y. P. Fang, C. Fang, E. Zio, and M. Xie, "Resilient critical infrastructure planning under disruptions considering recovery scheduling," *IEEE Transactions on Engineering Management*, vol. 26, 2019.

[34] S. Elluru, H. Gupta, H. Kaur, and S. P. Singh, "Proactive and reactive models for disaster resilient supply chain," *Annals of Operations Research*, vol. 283, no. 1-2, pp. 199–224, 2019.

[35] A. A. Taleizadeh, A. Ghavamifar, and A. Khosrojerdi, *Resilient Network Design of Two Supply Chains under Price Competition: Game Theoretic and Decomposition Algorithm Approach*, Springer, Berlin, Germany, 2020.

[36] D. Ivanov, "Predicting the impacts of epidemic outbreaks on global supply chains: a simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case," *Transportation Research Part E: Logistics and Transportation Review*, vol. 136, Article ID 101922, 2020.

[37] B. Zahiri, J. Zhuang, and M. Mohammadi, "Toward an integrated sustainable-resilient supply chain: a pharmaceutical case study," *Transportation Research Part E: Logistics and Transportation Review*, vol. 103, pp. 109–142, 2017.

[38] D. Ivanov, "Revealing interfaces of supply chain resilience and sustainability: a simulation study," *International Journal of Production Research*, vol. 56, no. 10, pp. 3507–3523, 2018.

[39] V. D. Souza, J. Bloemhof-Ruwaard, and M. Borsato, "Exploring ecosystem network analysis to balance resilience and performance in sustainable supply chain design," *International Journal of Advanced Operations Management*, vol. 11, no. 1-2, pp. 26–45, 2019.

[40] A. Pavlov, D. Ivanov, D. Pavlov, and A. Slinko, "Optimization of network redundancy and contingency planning in sustainable and resilient supply chain resource management under conditions of structural dynamics," *Annals of Operations Research*, vol. 27, pp. 1–30, 2019.

[41] D. Ivanov and A. Dolgui, "Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak," *International Journal of Production Research*, vol. 58, no. 10, pp. 2904–2915, 2020.

[42] E.-H. Aghezzaf, C. Sitompul, and N. M. Najid, "Models for robust tactical planning in multi-stage production systems

with uncertain demands," *Computers & Operations Research*, vol. 37, no. 5, pp. 880–889, 2010.

[43] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios, "Robust optimization of large-scale systems," *Operations Research*, vol. 43, no. 2, pp. 264–281, 1995.

[44] C.-S. Yu and H.-L. Li, "A robust optimization model for stochastic logistic problems," *International Journal of Production Economics*, vol. 64, no. 1–3, pp. 385–397, 2000.

[45] K. Deb and K. Deb, "Multi-objective optimization," *Search Methodologies*, Springer, Boston, MA, USA, pp. 403–449, 2014.

Hindawi
Complexity
Volume 2021, Article ID 6642986, 11 pages
https://doi.org/10.1155/2021/6642986

Research Article

Proposition of New Metaphor-Less Algorithms for Reservoir Operation

Vartika Paliwal,[1] **Aniruddha D. Ghare**,[1] **Ashwini B. Mirajkar**,[1] **Neeraj Dhanraj Bokde**,[2] **and Zaher Mundher Yaseen**[3]

[1]*Department of Civil Engineering, Visvesvaraya National Institute of Technology, Nagpur 440010, Maharashtra, India*
[2]*Department of Engineering-Renewable Energy and Thermodynamics, Aarhus University, Aarhus 8000, Denmark*
[3]*Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam*

Correspondence should be addressed to Zaher Mundher Yaseen; zahermundheryaseen@duytan.edu.vn

Received 21 December 2020; Accepted 25 June 2021; Published 3 July 2021

Academic Editor: Abdelalim Elsadany

Copyright © 2021 Vartika Paliwal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on the current water crisis scenario, effective water resources management can play an essential role. Reservoir operation optimization is part of water resources management. Reservoir operation optimization is difficult as it involves a large number of variables and constraints to achieve this goal. The present study aims at exploring the performance of recently developed heuristic algorithms—Rao algorithms as applied to the reservoir operation studies for the first time. Rao algorithms are metaphor-less algorithms that require only basic parameters—population size and function evaluations. In the present study, Rao algorithms have been applied to two case studies: discrete four-reservoir operation system problem and continuous four-reservoir operation system problem (benchmark problems) for the assessment of their performance vis-à-vis other algorithms from the literature. The results showed that the Rao-1 algorithm provided the optimal solution with the least function evaluations when compared to Rao-2, Rao-3, and other algorithms applied in the past to the same benchmark problem. Consequently, the Rao-1 model is found to be superior to these approaches by taking less computational time. Hence, the Rao-1 algorithm can be considered suitable for application to reservoir operation optimization problems.

1. Introduction

The ever-rising population and change in regime towards the accelerated demand of water have a prerequisite for the complex optimization problems towards the global sustainability of the available water resources of the earth [1]. The sustenance for the ubiquitous natured water is very important for the attainment of ecological balance, and also to satisfy the rising need for water, it is important to utilize the available water optimally for its sustenance [2–5]. Thus, reservoir operation optimization is of prime significance in the current scenario, which overwhelms a huge number of variables and constraints. In general, water demands from the reservoir fulfilled are based on reservoir operating rules with available input variables and present water storage level along with the hydrological conditions [6–8]. Researchers are developing various optimization methods and applying newly developed approaches to achieve the best optimal solutions [9–11]. However, in recent decades, the field of populace-centered metaheuristic processes is engulfed with several 'new' algorithms based on the comparison of some natural phenomena or behavior of animals, fishes, insects, societies, cultures, planets, musical instruments, etc. [12, 13]. Optimization techniques have been evolved from traditional to evolutionary techniques. Sreenivasan and Vedula [14] applied chance-constrained Linear Programming (LP) to a multipurpose reservoir based on the reliability level the optimal solution was obtained. To optimize the nonlinear hydropower function using the LP method, the function was linearized and the solution was obtained within the tolerance limits. Kumar and Prakash [15] developed the Nonlinear Programming (NLP) model to analyze the operation of the multipurpose Koyna dam, India. It was analyzed for different dependable inflows and was found that after relaxing

the release, it can generate more hydropower as it was the vital importance. To incorporate the uncertainties due to inflow, Dynamic Programming was combined with fuzzy rule, and it resulted in the satisfying target performances for Dez and Karoon reservoirs in Iran [16].

One of the earliest review research conducted on the implementation of traditional models such as Discrete Differential Dynamic Programming (DDDP), LP, NLP, and stochastic models for the reservoir operation optimization and management, was carried out [17]. The study described the pros and cons of these techniques. The review findings evidenced the difficulty of obtaining a generalized model that can be applied for all real-world optimization problems. The traditional models have certain shortcomings like convergence depends upon the initial solution, inefficient in handling discrete search space, and sometimes get stuck in local optima. To overcome the shortcomings of the traditional approaches, evolutionary approaches were developed. Jalali et al. [18] proposed one of the well-established nature-inspired optimization algorithms called Ant Colony Optimization (ACO) to be statistically implemented at the Dez reservoir, Iran. The authors concluded that with proper tuning of parameters global optimal solution is achieved. Labadie 2015 [17] developed and implemented Bat Algorithm to Karoun-4 reservoirs and hypothetical systems and presented its excellence over the traditional approaches. Biogeography Based Optimization Algorithm was validated using a mathematical function and was further applied to the single and multireservoir system [19]. Asadieh and Afshar [20] presented a comparative analysis of the Charged System Search Algorithm (CSSA) with Particle Swarm Optimization (PSO), Genetic Algorithm (GA), gradient-based NLP, and ACO for benchmark problem and Dez reservoir in Iran. CSSA is found to be superior in comparison to other methods. Crow algorithm outperformed other techniques when applied to the multireservoir system in China [21]. The optimal cropping pattern for the Bilaspur project, Rajasthan, India, was developed using the Differential Evolution technique [22]. Garousi-Nejad et al. [23] tested the Firefly Algorithm on mathematical benchmark functions and operation optimization of the reservoir with irrigation and hydropower as the purposes. Firefly was found to be superior to GA in terms of convergence rate and variance. Rule curves for the Pechiparai reservoir, Tamil Nadu, India, were derived [24] using a reliability-based GA model. The harmony Search method was found to have potential when tested on the benchmark problem and effectively solved the flood management problem of the Narmab reservoir in Iran [25]. Afshar et al. [26] concluded that HBMO results comparable to LP and other well-developed optimization techniques. Hybrid Algorithm (HA) of Particle Swarm Optimization (PSO) Algorithm and Artificial Fish Swarm Algorithm (AFSA) was developed and implemented by Yaseen [27] for the analysis of the Karun-4 hydropower system. Hybridization overcomes the drawback of AFSA and PSO when assessed based on reliability, resilience, and vulnerability indices. Janga Reddy and Nagesh Kumar [28] applied Particle Swarm Optimization (PSO), Elitist Mutated PSO (EMPSO), and GA to the Bhadra reservoir system, India. EMPSO outperformed standard PSO and GA. Shark Algorithm (SA) was found to produce good results when applied to complex reservoir problems [29]. Another hybrid AI-based model was proposed in [30], which predicted water level prediction and uncertainty analysis at Urmia lake in Iran. This model was based on hybridization of improved adaptive neurofuzzy inference system (ANFIS) and multilayer perceptron (MLP) models are hybridized with a sunflower optimization (SO) algorithm and shown a significant improvement in improving lake's water level.

The applications of recently explored evolutionary algorithms, such as the Water Cycle Algorithm (WCA) [31], Weed Optimization Algorithm (WOA) [32], and Wolf Search Algorithm (WSA) [33], have been remarkably established over the past decade. The recent review research was conducted on the feasibility of evolutionary computing algorithms for reservoir operation modeling [34]. The review research confirmed the capacity of the evolutionary algorithms as advanced computer aid models owing to their capability to improve the stochastic complexity and for a better understanding of simulated reservoir operation. These approaches are adopted mostly from nature like Particle Swarm Optimization, Crow Algorithm, Weed Optimization, Shark Algorithm, and many more mimicking the behavior of particular species from nature; hence, are called Metaphor algorithms. These optimization algorithms need regulation of system explicit parameters. Subsequently, these are descriptions that proliferate the exertions in tweaking as well as the phase. The algorithm-specific parameter fewer optimization applications can be embarked upon by metaphor-less algorithm as introduced by Rao. The metaphor-less algorithm has an advantage over metaheuristic techniques in that it does not require algorithm-specific parameters to tune the algorithm. The metaphor-less algorithms applied so far in reservoir operation studies are Teaching Learning Based Optimization (TLBO) and Jaya Algorithm (JA). Kumar and Yadav [35] reported the satisfactory performances of TLBO and JA when applied to the benchmark studies. Paliwal et al. [36] tested JA on the benchmark problem and found it to result better than other approaches in the past and was also applied to a real case of Mula reservoir, Maharashtra, India. Chong et al. [37] applied JA for hydropower operation optimization to a reservoir system in Malaysia. In this study, the uncertainty of inflows is handled using the Thomas–Fiering model. Results obtained are compared with the results obtained from various metaheuristic approaches, and performance indices are calculated, which indicated JA is efficient in handling reservoir operation optimization problems. Motevali Bashi Naeini and Soltaninia [38] combined branch and bound (BB) with a hybrid of PSO-LP and applied it to benchmark problems to obtain a computationally efficient operation optimization algorithm at the dam design stage. Recently, Rao algorithms have been developed [39] and were tested on 23 benchmark functions along with 25 unconstrained and 2 constrained optimization problems. Wang et al. [40] applied the Rao-1 algorithm to parameter estimation of the photovoltaic cell model and found it to be suitable for such problems. Rao and Pawar [41] applied the Rao algorithm to mechanical system problems which are constrained in nature and found Rao algorithms to be superior to other algorithms.

The need for water resources management can be achieved by optimizing the existing reservoir operation [42, 43]. Various approaches have been adapted to achieve this goal in the past and still, researches are going on in view of achieving a better strategy [44, 45]. From the literature review, it is found that Rao algorithms have never been applied to the reservoir operation optimization problems, albeit having shown enough promise in the other areas of engineering optimization problems. This led to the thought of the application of these algorithms to the complex reservoir operation optimization problem to assess its applicability in such problems.

In this paper, this novel approach for reservoir operation optimization using the three Rao algorithms is presented. Rao algorithms are recently developed metaphor-less heuristic algorithms which just need mathematical operators in their equations and do not depend upon algorithm-specific parameters. Hence, it is considered highly suitable for problems such as reservoir operation that involve a large number of variables and constraints. To assess the potential of the proposed algorithms, it has been tested on two benchmark problems (discrete-four reservoir operation (DFRO) problem and continuous four-reservoir operation (CFRO) problem) from the literature. The performance comparison of proposed algorithms with the other existing optimization algorithms for reservoir operation optimization is also presented, referring to the past studies.

2. Materials and Methods

2.1. Description of the Rao Algorithms. Rao algorithms are metaphor-less algorithms and only need common control parameters like that in TLBO and JA. The update equation in Rao algorithms is inherited from JA. Similar to JA, they also require only mathematical operators to upgrade the solution based on the best and the worst solution. In JA, interactions were made between the candidate solution to be updated with the best and with the worst solution. In the Rao-1 algorithm, interaction is between the best and the worst values. In the other two algorithms, along with the interaction between the worst and the best value, there is random interaction between the candidate solutions based on their performances. The reservoir operation optimization process for the four-reservoir system problems using Rao algorithms has been demonstrated in Figure 1.

2.2. Methodology of the Proposed Algorithms: Rao Algorithms. The independent variables are initialized using minimum, maximum bounds of the particular variable and random number, as shown in the following equation:

$$X_{b,c} = (X_b) + r * ((X_b) - \min(X_b)), \tag{1}$$

where (X_b) is the minimum bound for the bth variable, r is a random number ([0, 1]), (X_b) is the maximum bound for the bth variable, $X_{b,c}$ = Value of bth variable for c^{th} candidate solution. Dependable variable values are generated using the values of independent variables. Then, the objective function is computed further, considering constraint violation. The function value is obtained considering penalties for the violation. Penalties are added to the objective function value for a minimization problem to obtain the function value and vice versa.

The best and the worst solutions are selected amongst the candidate solutions for the ath iteration. Let c be the candidate solution for a^{th} iteration, then the updated value of bth variable is obtained using equation (2a) for Rao-1, (2b) for Rao-2, and (2c) for Rao-3, respectively:

$$X'_{b,c,a} = X_{b,c,a} + r_{b,a}\left(X_{b,\text{best},a} - X_{b,\text{worst},a}\right), \tag{2a}$$

$$\begin{aligned} X'_{b,c,a} = {} & X_{b,c,a} + r_{1,b,a}\left(X_{b,\text{best},a} - X_{b,worst,a}\right) \\ & + r_{2,b,a}\left(\left|X_{b,c,a} \text{ or } X_{b,d,a}\right| - \left|X_{b,d,a} \text{ or } X_{b,c,a}\right|\right), \end{aligned} \tag{2b}$$

$$\begin{aligned} X'_{b,c,a} = {} & X_{b,c,a} + r_{1,b,a}\left(X_{b,\text{best},a} - \left|X_{b,\text{worst},a}\right|\right) \\ & + r_{2,b,a}\left(\left|X_{b,c,a} \text{ or } X_{b,d,a}\right| - \left|X_{b,d,a} \text{ or } X_{b,c,a}\right|\right), \end{aligned} \tag{2c}$$

where, $X_{b,c,a}$ = value of bth variable for c^{th} candidate solution for a^{th} iteration, $X_{b,\text{best},a}$ = value of bth variable for the best candidate solution for a^{th} iteration, $X_{b,\text{worst},a}$ = value of bth variable for the worst candidate solution for a^{th} iteration, $X_{b,c,a}{}'$ = updated value of bth variable for c^{th} candidate solution for a^{th} iteration, and $r_{1,b,a}$ and $r_{2,b,a}$ = random numbers for b^{th} variable during a^{th} iteration. In equations (2b) and (2c), the terms $X_{b,c,a}$ and $X_{b,d,a}$ represents the variable values corresponding to c^{th} and d^{th} candidate solution and random interaction between them. If the value corresponding to c^{th} is better than d^{th} then the term "$X_{b,c,a}$ or $X_{b,d,a}$" becomes $X_{b,c,a}$ and "$X_{b,d,a}$ or $X_{b,c,a}$" becomes $X_{b,d,a}$ and vice versa in the opposite case.

New function values that are computed using updated variable values are compared to the respective function values. The adopted function is the best one, and the worst was rejected. The best values for the respective candidate solution are now the preliminary set for the following iteration. The same process continues until the termination criterion is reached.

2.3. Case Study 1: Discrete Four-Reservoir Operation (DFRO) System Problem. A hypothetical discrete four-reservoir system introduced by Larson [46] has been used as the case study: a benchmark problem to test the potential of Rao algorithms. The schematic view of this case study is shown in Figure 2. This system was also used as a benchmark in past studies to test other optimization techniques in the field of reservoir operation studies. The system is a series and parallel combination of four reservoirs. Reservoirs 1–3 produce hydropower, while reservoir 4 is a multipurpose reservoir serving irrigation as well as hydropower production. For maximization of the profits from this system, a twelve-hour operating period is considered in the objective function.

Data for the benchmark problem are shown in Table 1.

The objective function (F) is the maximization of net profit obtained from all four reservoirs. Mathematically, it can be expressed as follows:

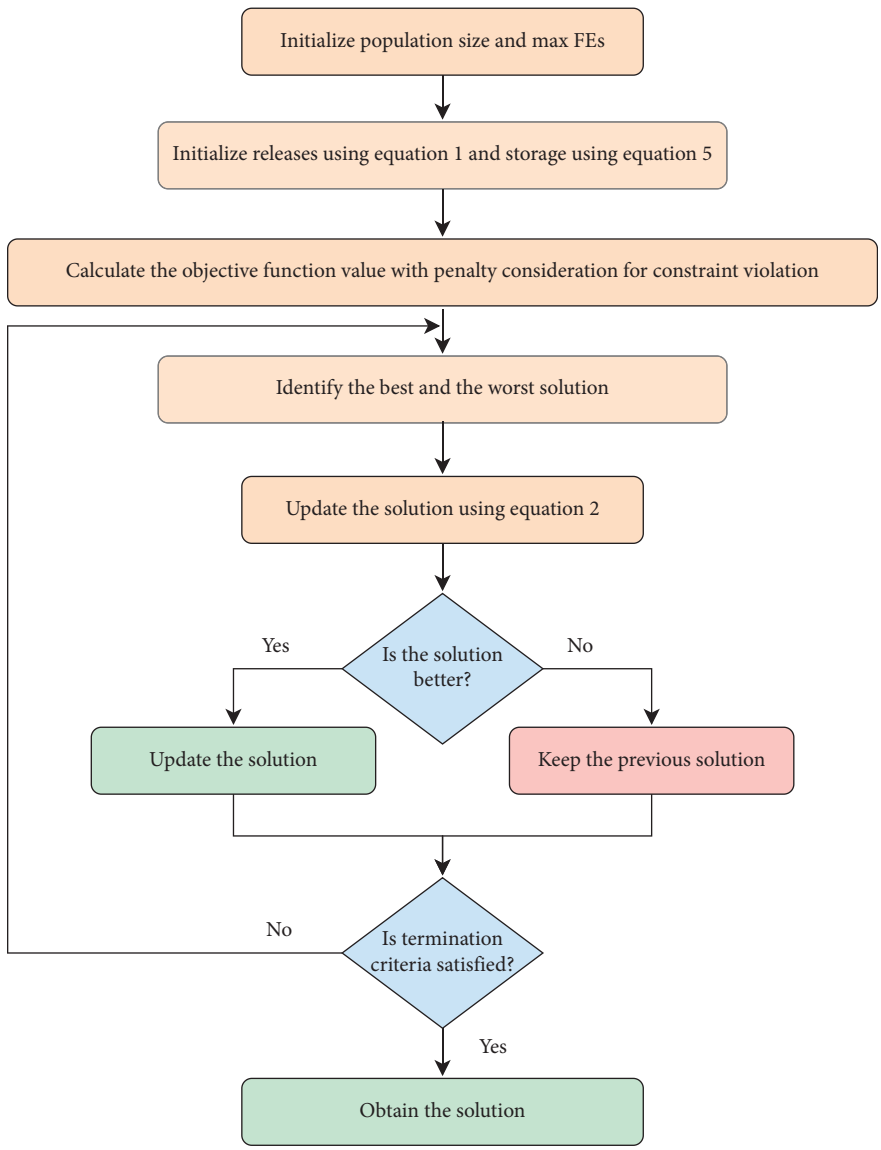


Figure 1: Flowchart of Rao-1 algorithm.

$$\text{Max } F = \sum_{i=1}^{4}\sum_{t=1}^{12} b_i(t)\cdot R_i(t) + \sum_{t=1}^{12} b_5(t)\cdot R_4(t), \tag{3}$$

where $b_i\ (t) = 4 * 12$ matrices of benefit from hydropower production from all the four reservoirs. The benefit function matrix is as follows:

$$b_i(t) = \begin{bmatrix} 1.1 & 1 & 1 & 1.2 & 1.8 & 2.5 & 2.2 & 2 & 1.8 & 2.2 & 1.8 & 1.4 \\ 1.4 & 1.1 & 1 & 1 & 1.2 & 1.8 & 2.5 & 2.2 & 2 & 1.8 & 2.2 & 1.8 \\ 1 & 1 & 1.2 & 1.8 & 2.5 & 2.2 & 2 & 1.8 & 2.2 & 1.8 & 1.4 & 1.1 \\ 1 & 1.2 & 1.8 & 2.5 & 2.2 & 2 & 1.8 & 2.2 & 1.8 & 1.4 & 1.1 & 1 \end{bmatrix}, \tag{4}$$

$b_5\ (t)$ is the benefit from irrigation for reservoir 4, and $b_5\ (t) = [1.61.71.81.92221.91.81.71.61.5]$.

The objective function is subjected to the following constraints.

2.3.1. Continuity Constraint. The continuity constraints for each reservoir over each operating period "t" are as follows:

$$S_i(t+1) = S_i(t) + I_i(t) + M.R_i(t), \tag{5}$$

where $S_i\ (t+1)$ denotes the reservoir storage at period "t" and for reservoirs $i = 1$ to 4. $S_i(t)$ presents the reservoir storage at the beginning of period "t" and for reservoirs $i = 1$ to 4. $I_i(t)$ indicates the reservoir inflows during the period "t" and for the reservoirs $i = 1$ to 4. $R_i(t)$ denotes the reservoirs releases during the period 't' and for the reservoirs $i = 1$ to 4. $M = 4 \times 4$ matrix of indices of reservoir connections, $M = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 \end{bmatrix}$.

2.3.2. End Storage Constraint. End storage ($S_i(13)$) should be at least 5, 5, 5, and 7 units for reservoirs 1, 2, 3, and 4, respectively, to maintain the continuity of the operation. The

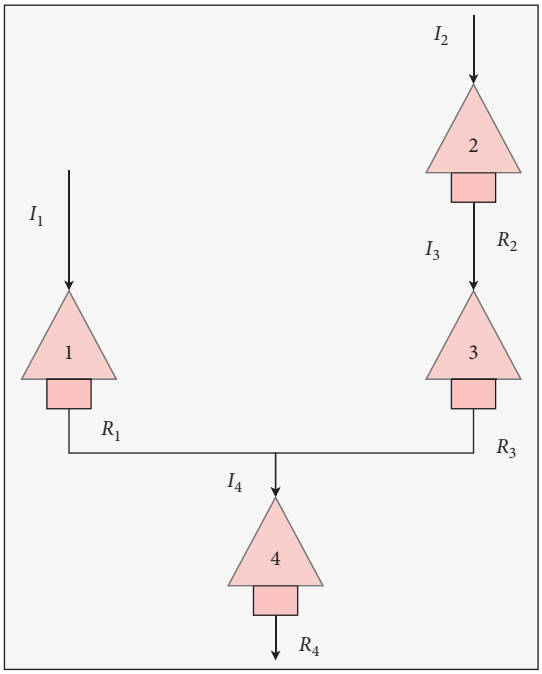


Figure 2: Four reservoir systems. 1, 2, 3, and 4: number of reservoirs; *I1*, *I2*, *I3*, and *I4*: inflows to reservoirs 1, 2, 3, and 4, respectively; *R1*, *R2*, *R3*, and *R4*: releases from reservoirs 1, 2, 3, and 4, respectively.

Table 1: Data for the DFRO system (benchmark problem) [46].

Parameters	Reservoir 1	Reservoir 2	Reservoir 3	Reservoir 4
Inflow	2	3	0	0
Minimum releases	0	0	0	0
Maximum releases	3	4	4	7
Minimum storage	0	0	0	0
Maximum storage	10	10	10	15
Initial storage	5	5	5	5

penalty function applied for this constraint violation is the same as that applied by the other researches and is as follows:

$$\begin{aligned} g_i\{S_i(13,d_i)\} &= 40\{S_i(13)-d_i\}^2, \quad \text{for } S_i(13)\le d_i, \\ g_i\{S_i(13,d_i)\} &= 0, \quad \text{for } S_i(13)>d_i. \end{aligned} \tag{6}$$

Hence, the modified function is represented as follows:

$$\begin{aligned} \text{Max } F = &\sum_{i=1}^{4}\sum_{t=1}^{12} b_i(t)\cdot R_i(t) \\ &+ \sum_{t=1}^{12} b_5(t)\cdot R_4(t) - g_i\{S_i(13,d_i)\}. \end{aligned} \tag{7}$$

2.4. Case Study 2: Continuous Four-Reservoir Operation (CFRO) System Problem. CFRO was introduced by Te Chow and Cortes Rivera [47] in 1974. The schematic sketch of the CFRO is shown in Figure 2. The CFRO problem has similarities with the DFRO problem. The difference is in the inflows and bounds. The connections between the reservoirs in the system and benefits functions are the same as DFRO. Hence, the continuity equation and objective function are the same.

The minimum releases for time period $t=1$ to 12 for reservoir 1, 2, 3, and 4 are 0.005, 0.0005, 0.0005, and 0.005, respectively. The maximum releases for periods $t=1$ to 12 for reservoirs 1, 2, 3, and 4 are 4, 4.5, 4.5, and 8, respectively. The initial and final storages for reservoirs 1, 2, 3, and 4 are 6, 6, 6, and 8, respectively. The minimum storages for $t>1$ to $t=12$ are 1 for all 4 reservoirs. The inflows for $t=1$ to 12 are shown in Table 2. The maximum storages for $t=2$ to 12 are given in Table 3. The penalty factor was 40 for DFRO and 13 for CFRO based on past studies.

3. Results and Discussion

3.1. DFRO. Rao algorithms were applied to the DFRO system optimization: the first considered benchmark problem. The population size is set to 50, 40, and 40 for Rao-1, Rao-2, and Rao-3 algorithms, respectively, after conducting the sensitivity analysis. The population size was initially set as 50 based on past studies (Kumar and Yadav [35]). Rao-1 resulted in the optimal solution for a population size of 50, while Rao-2 and Rao-3 did not. Rao-2 and Rao-3 yielded the optimal solutions for a population size of 40 in both cases. The global optimal solution with the objective function value of 401.3 has been achieved for the DFRO problem using the Rao-1 algorithm with a population size of 50 and maximum Function Evaluations (FEs) of 1,50,000. For Rao-2, the optimal value of 401.23 is achieved at a population size of 40 and Max FEs of 11,00,000 and that for Rao-3, the optimal value achieved is 401.4 at a population size of 40 and Maximum FEs of 12,51,000. Rao-3 led to a higher value of an objective function with a slight violation of constraints. According to the past studies also 401.3 is the global optimal solution for the provided benchmark problem without constraint violation, which is achieved in the case of the Rao-1 algorithm model. These algorithms result in three values-the best, the mean, and the worst for a particular solution. Runs represent the number of times the same model is operated for the same given set of conditions. Generally, it is preferred to be selected as 10. It can have a higher value also like 15 or 20, depending upon the variation observed in the results for the same set of conditions. In the present study, it has been adopted for runs as 10. For 10 runs, the best, the mean, and the worst values for these algorithms corresponding to their particular solution along with the standard deviation are shown in Table 4. Rao-1 showed the least standard deviation of 0.45, and the higher standard deviation is in the case of Rao-3. From Table 4, it can also be observed that there is less variation in the best and the worst values of the function. Hence, it can be said that Rao-1 has less standard deviation as it has confined the exploration of the updated value between the worst and the best. In the updated equation of Rao-1, it shows the random interaction between the best and the worst values rather than with the

Table 2: Inflows for the CFRO system.

Time	Reservoir 1	Reservoir 2	Reservoir 3	Reservoir 4
1	0.5	0.4	0	0
2	1	0.7	0	0
3	2	2	0	0
4	3	2	0	0
5	3.5	4	0	0
6	2.5	3.5	0	0
7	2	3	0	0
8	1.25	2.5	0	0
9	1.25	1.3	0	0
10	0.75	1.2	0	0
11	1.75	1	0	0
12	1	0.7	0	0

Table 3: Maximum storages for the CFRO system.

Time	Reservoir 1	Reservoir 2	Reservoir 3	Reservoir 4
2	12	15	8	15
3	12	15	8	15
4	10	15	8	15
5	9	12	8	15
6	8	12	8	15
7	8	12	8	15
8	9	15	8	15
9	10	17	8	15
10	10	18	8	15
11	12	18	8	15
12	12	18	8	15

Table 4: Best, mean, and worst objective function values obtained corresponding to the optimal solution using Rao algorithms for the DFRO system problem.

Algorithm	Runs	Best	Mean	Worst	Standard deviation
Rao-1	10	401.3	401.01	400.69	0.45
Rao-2	10	401.23	397.2	395.07	1.83
Rao-3	10	401.4	398.34	395.74	1.97

random variable value or specific value, thereby reducing the range. Hence, fewer changes of deviation in the solution, which may further lead to faster convergence in the case of the Rao-1 algorithm. On the other hand, Rao-2 and Rao-3 algorithms involve the interaction between the best and the worst along with random interactions between the variable of different candidate solutions, which may again need some more time (FEs) but involves a more random nature. This results in a higher standard deviation and the requirement of more number function evaluations. Figure 3 demonstrates the resulting release pattern for the four reservoirs using Rao-1, Rao-2, Rao-3, JA, and LP algorithms. The release pattern for reservoir 1 is found to be the same for Rao algorithms and JA but is different for LP at time step 1. Rao-1 and JA have 2 units' higher releases than LP, while at Step 4 of the period, LP has 2 units' higher releases for nearly the same benefit function in both cases, which shows the balance of net benefit for reservoir 1. For reservoir 2, Rao-2 and Rao-3 algorithm models have produced the same releases till 10 time steps and for the last 2 time steps showed different trends, thereby balancing the releases in this case. While Rao-1 synchronized with JA except for 3 time steps but both the models have led to an equal quantity of releases for this reservoir. Similarly, LP showed a different trend with these algorithms but has led to some amount of release for reservoir 2. In the case of reservoir 3, the releases obtained from the Rao-1 algorithm varied from those obtained from the other approaches; however, the net benefits achieved are the same, leading to the same optimal solution. Rao-1 and Rao-2 resulted in the optimal solution with releases varying from those of the other approaches at 1 and 2 time steps, respectively. Rao-2 released nearly the same quantity as that by other approaches while Rao-1 released a bit higher than the others. Since it is a multipurpose multireservoir operation, the releases for different reservoirs obtained from different approaches can be different, leading to the same optimal solution as the single objective function has been framed by using the benefit function associated with the releases from each reservoir.

The performance comparison of these algorithms with other approaches developed and applied to this system in the past has been shown in Table 5. From Table 5, it can be observed that Rao-1 results in the optimal value of 401.3 for 150,000 FEs while Rao-2 and Rao-3 resulted in the optimal value of 401.23 and 401.4 for the Max FEs of 1,100,000 and 1,251,000, respectively. From Table 5, it can be seen that JA has obtained the optimal value of 401.4 with FEs = 350,000 (Kumar and Yadav [35]) and 325,000 [36], while Rao-1 resulted in the optimal solution with FEs = 150,000, which is nearly 50% of the FEs in JA and a way better than Rao-2 and Rao-3 algorithms in terms of the number of function evaluations. Hence, it can be said that the Rao-1 algorithm model achieved the optimal solution in less computational time, indicating the faster convergence of the Rao-1 algorithm model. Thus, the Rao-1 algorithm model is found to be superior to other optimization algorithms and as well as to its parent algorithm for this case study.

3.2. CFRO. Rao algorithms have also been applied to the second case study, i.e., CFRO system optimization: another benchmark problem. The population size was initially set as 50, based on past studies (Kumar and Yadav [35]). The optimal solution with the objective function value of 308.8 has been achieved for the CFRO system problem using the Rao-1 algorithm with a population size of 50 and maximum Function Evaluations (FEs) of 155,000. For Rao-2, the optimal value of 304.64 is achieved at a population size of 50 and Max FEs of 725,000 and that for Rao-3, the optimal value achieved is 307 at a population size of 50 and Maximum FEs of 700,000. The LP model has obtained the value of 308.3 as the optimal objective function value. The objective function is the maximization of benefits which has better been achieved by the Rao-1 algorithm. Rao algorithms result in three values—the best, the mean, and the worst for a particular solution. Runs represent the number of times the same model is operated for the same given set of conditions.

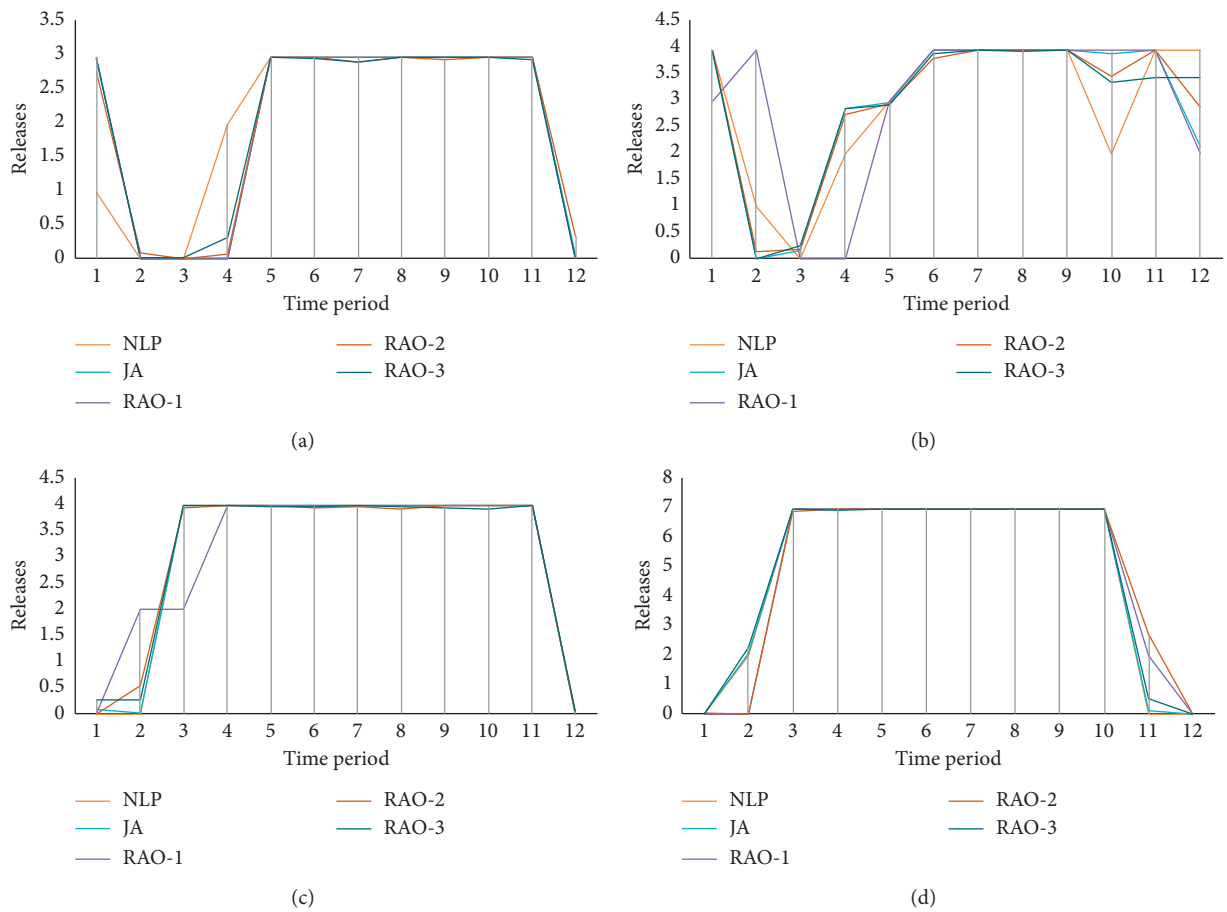


Figure 3: Release pattern for the DFRO system problem obtained through Rao algorithms. (a) Reservoir 1. (b) Reservoir 2. (c) Reservoir 3. (d) Reservoir 4.

Table 5: The benchmark models and the proposed new algorithm of the current research performance for the DFRO system problem.

Source	Model	The best objective function value	Function evaluations required
[46]	DPSA	401.30	NA
[48]	DDDP	401.3	NA
[49]	FDP	399.0	NA
[50]	GA	401.3	2,279,500
	PSO	399.7	748,000
	EMPSO	401.3	325,400
[32]	WOA	401.30	400,000
[35]	TLBO	401.3	350,000
	JA	401.4	350,000
[36]	JA	401.4	325,000
Present study	Rao-1	401.3	150,000
	Rao-2	401.23	1,100,000
	Rao-3	401.4	1,251,000

In the present study, it has been adopted for runs as 15. For 15 runs, the best, the mean, and the worst values for these algorithms corresponding to their particular solution along with the standard deviation are shown in Table 6. The highest standard deviation is observed for the Rao-3 algorithm. Figure 4 demonstrates the resulting release pattern for the CFRO using Rao-1, Rao-2, Rao-3, and LP algorithms. Releases obtained from Rao-1, LP, Rao-3 are the same for 12 time steps, while Rao-2 shows different releases at 2 time steps for reservoir 1. While, for reservoir 2, release patterns are different in the case of all the optimization algorithms. Variation in releases obtained from Rao algorithms and LP at few time steps is observed in the case of reservoirs 3 and 4.

The performance comparison of these algorithms with other approaches developed and applied to this system in the past has been shown in Table 7. From Table 5, it can be

Table 6: Best, mean, and worst objective function values obtained corresponding to the optimal solution using Rao algorithms for the CFRO system problem.

Algorithm	Runs	Best	Mean	Worst	Standard deviation
Rao-1	15	308.8	304.04	295.07	4.43
Rao-2	15	304.64	295.6	288.19	4.34
Rao-3	15	307	299.74	279.32	7.76

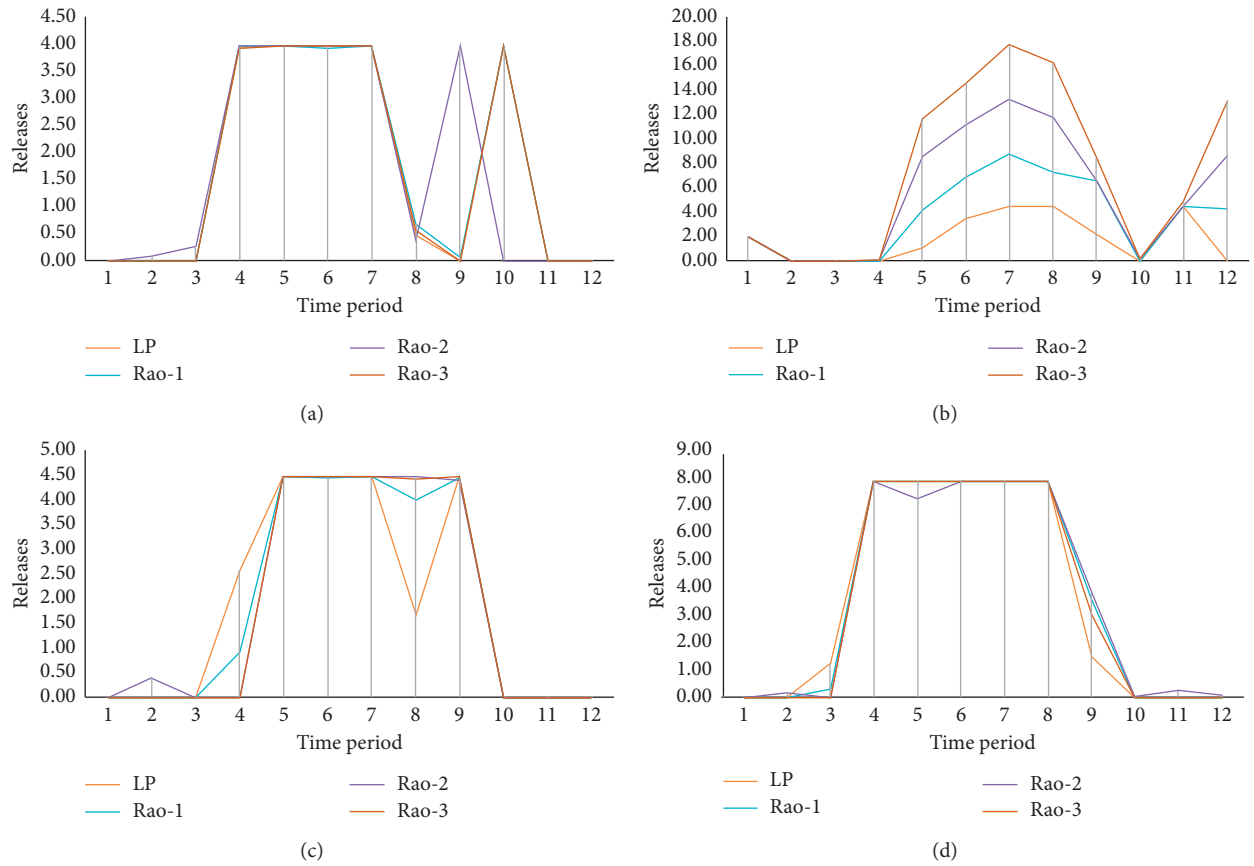


Figure 4: Release pattern for the CFRO system obtained through Rao algorithms. (a) Reservoir 1. (b) Reservoir 2. (c) Reservoir 3. (d) Reservoir 4.

observed that Rao-1 results in the optimal value of 308.8 for 155,000 FEs while Rao-2 and Rao-3 resulted in the optimal value of 304.64 and 307 for the Max FEs of 725,000 and 700,000, respectively. From Table 7, it can be seen that JA has obtained the optimal value of 308.4 withFEs = 350,000 (Kumar and Yadav [35]), which is the highest objective function value in the past, while Rao-1 resulted in the optimal solution of 308.8 with FEs = 155,000, which is nearly 50% of the FEs in JA and a way better than Rao-2 and Rao-3 algorithms in terms of the number of functions evaluations. The objective function is maximization which is better achieved by Rao-1 with slight constraint violation leading to the highest objective function value with a considerable range of violations and lesser FEs. Hence, it can be said that the Rao-1 algorithm model achieved the optimal solution in less computational time, indicating the faster convergence of the Rao-1 algorithm model. Thus, the Rao-1 algorithm model is found to be suitable for its application the complex reservoir operation problems.

4. Conclusions

The Rao Algorithms are similar to the Jaya Algorithm (JA) as these are also metaphor-less algorithms and do not need any algorithm-specific parameter. These algorithms just need mathematical operators in the update equation along with the best and the worst values. The complexity of JA has been examined empirically in terms of big-O notations using the GuessCompx tool [53, 54]. The complexity is compared with the metaheuristic technique (GA) and both methods showed linear complexity (Paliwal et al. [36]). Hence, these algorithms reduce the computational complexity. They are also found to be easy in application and efficient too. The parameters in the update equation of JA are shuffled to prepare

Table 7: The benchmark models and the proposed new algorithm of the current research performance for the CFRO system problem.

Source	Model	The best objective function value	Function evaluations required
[47]	LP	308.26	NA
[51]	DDP	308.23	NA
[49]	HBMO	308.07	1,100,000
[50]	LP	308.29	NA
	GA	302.42	100,000
	ICA	306.76	100,000
	COA	307.92	100,000
[32]	IBA	308.29	234,000
[32]	GA	300.47	1,600,000
	WOA	308.15	1,600,000
	LP	308.29	NA
[30]	PSO	307.54	—
	SHARK	308.3	—
[52]	NLP	308.29	50,000
	GA	307.12	50,000
	KRILL	307.54	50,000
	HYBRID	308.29	50,000
[35]	TLBO	308.3	350,000
	JA	308.4	350,000
Present study	LP	308.3	NA
	Rao-1	308.8	155,000
	Rao-2	304.64	725,000
	Rao-3	307	700,000

the update equations for Rao algorithms. In this study, outcomes from Rao Algorithms have been compared to other approaches applied in the past and also to its parental algorithm (JA). Rao-1 Algorithm is found to be superior to other algorithms for these case studies. Rao-1 Algorithm leads to the global optimal solution with the least FEs as compared to all other optimization models found in the literature for DFRO and higher objective function value with fewer FEs with considerable constraint violation for CFRO system problem. Rao-1 Algorithm model outperformed every other model in terms of computational time. Hence, it can be concluded that the Rao-1 Algorithm model can be utilized in the field of reservoir operation optimization as it can lead to near global optimal solution with much fewer function evaluations. Future research studies can be adopted on the investigation of the proposed optimization algorithm for other water resources management and operation.

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 no conflicts of interest.

Authors' Contributions

Vartika Paliwal, Aniruddha D. Ghare, Ashwini B. Mirajkar, and Neeraj Dhanraj Bokde conceptualized the study. Vartika Paliwal and Neeraj Dhanraj Bokde curated the data and provided software. Vartika Paliwal, Aniruddha D. Ghare, Ashwini B. Mirajkar, Neeraj Dhanraj Bokde, and Zaher Mundher Yaseen performed formal analysis. Vartika Paliwal, Aniruddha D. Ghare, Ashwini B. Mirajkar, and Neeraj Dhanraj Bokde investigated the study. Vartika Paliwal provided the methodology. Aniruddha D. Ghare, Ashwini B. Mirajkar, Neeraj Dhanraj Bokde and Zaher Mundher Yaseen contributed to project administration and supervised the study. Vartika Paliwal, Aniruddha D. Ghare, Ashwini B. Mirajkar, Neeraj Dhanraj Bokde, and Zaher Mundher Yaseen validated the study. Vartika Paliwal, Aniruddha D. Ghare, Ashwini B Mirajkar, and Zaher Mundher Yaseen contributed to visualization. Vartika Paliwal, Neeraj Dhanraj Bokde, and Zaher Mundher Yaseen wrote the original draft. Vartika Paliwal, Aniruddha D. Ghare, Ashwini B. Mirajkar, Neeraj Dhanraj Bokde, and Zaher Mundher Yaseen reviewed and edited the article.

References

[1] X. Yun, Q. Tang, J. Wang et al., "Impacts of climate change and reservoir operation on streamflow and flood characteristics in the Lancang-Mekong river basin," *Journal of Hydrology*, vol. 590, Article ID 125472, 2020.

[2] N. Suwal, X. Huang, A. Kuriqi, Y. Chen, K. P. Pandey, and K. P. Bhattarai, "Optimisation of cascade reservoir operation considering environmental flows for different environmental management classes," *Renewable Energy*, vol. 158, 2020.

[3] W. Zhang, X. Wang, X. Lei, P. Liu, X. Yan, and M. Feng, "Multicriteria decision-making model of reservoir operation considering balanced applicability in past and future: application to the Three Gorges reservoir," *Journal of Water Resources Planning and Management*, vol. 146, no. 6, Article ID 4020033, 2020.

[4] A. Gupta, S. Mishra, N. Bokde, and K. Kulat, "Need of smart water systems in India," *International Journal of Applied Engineering Research*, vol. 11, no. 4, 2016.

[5] A. D. Gupta, P. Pandey, A. Feijóo, Z. M. Yaseen, and N. D. Bokde, "Smart water technology for efficient water resource management: a review," *Energies*, vol. 13, no. 23, p. 6268, 2020.

[6] S. M. Awadh, H. Al-Mimar, and Z. M. Yaseen, "Groundwater availability and water demand sustainability over the upper mega aquifers of Arabian Peninsula and west region of Iraq," *Environment, Development and Sustainability*, vol. 23, 2020.

[7] R. Oliveira and D. P. Loucks, "Operating rules for multi-reservoir systems," *Water Resources Research*, vol. 33, no. 4, pp. 839–852, 1997.

[8] G. Yang and P. Block, "Transboundary water sharing policies conditioned on hydrologic variability to inform reservoir operations," *Hydrology and Earth System Sciences Discussions*, vol. 25, pp. 1–32, 2021.

[9] B. Dobson, T. Wagener, and F. Pianosi, "An argument-driven classification and comparison of reservoir operation optimization methods," *Advances in Water Resources*, vol. 128, pp. 74–86, 2019.

[10] M. Jahandideh-Tehrani, O. Bozorg-Haddad, and H. A. Loáiciga, "Application of non-animal-inspired evolutionary algorithms to reservoir operation: an overview," *Environmental Monitoring and Assessment*, vol. 191, no. 7, p. 439, 2019.

[11] Z. M. Yaseen, "A novel hybrid evolutionary data-intelligence algorithm for irrigation and power production management: application to multipurpose reservoir systems," *Sustainability*, vol. 11, 2019.

[12] A. S. Azad, M. S. A. Rahaman, J. Watada, P. Vasant, and J. A. G. Vintaned, "Optimization of the hydropower energy generation using meta-heuristic approaches: a review," *Energy Reports*, vol. 6, pp. 2230–2248, 2020.

[13] M. Kumar and A. J. Kulkarni, *Socio-Inspired Optimization Metaheuristics: A Review*, Springer, Berlin, Germany, 2019.

[14] K. R. Sreenivasan and S. Vedula, "Reservoir operation for hydropower optimization: a chance-constrained approach," *Sadhana*, vol. 21, no. 4, pp. 503–510, 1996.

[15] R. Arunkumar and V. Jothiprakash, "Optimal reservoir operation for hydropower generation using nonlinear programming model," *Journal of The Institution of Engineers (India): Series A*, vol. 93, no. 2, pp. 111–120, 2012.

[16] S. J. Mousavi, K. Ponnambalam, and F. Karray, "Reservoir operation using a dynamic programming fuzzy rule-based approach," *Water Resources Management*, vol. 19, no. 5, pp. 655–672, 2005.

[17] J. W. Labadie, "Optimal operation of multireservoir systems: state-of-the-art review," *Journal of Water Resources Planning and Management*, vol. 130, no. 2, pp. 93–111, 2004.

[18] M. R. Jalali, A. Afshar, and M. A. Mariño, "Reservoir operation by ant colony optimization algorithms," *Iranian Journal of Science and Technology, Transaction B: Engineering*, vol. 30, 2006.

[19] O. B. Haddad, S.-M. Hosseini-Moghari, and H. A. Loáiciga, "Biogeography-based optimization algorithm for optimal operation of reservoir systems," *Journal of Water Resources Planning and Management*, vol. 142, Article ID 4015034, 2015.

[20] B. Asadieh and A. Afshar, "Optimization of water-supply and hydropower reservoir operation using the charged system search algorithm," *Hydrology*, vol. 6, no. 1, p. 5, 2019.

[21] M. Ehteram, S. Binti Koting, H. A. Afan et al., "New evolutionary algorithm for optimizing hydropower generation considering multireservoir systems," *Applied Sciences*, vol. 9, no. 11, p. 2280, 2019.

[22] A. Vasan and K. S. Raju, "Optimal reservoir operation using differential evolution," in *Proceedings of the International Conference on Hydraulic Engineering: Research and Practice (ICON-HERP)*, Hangzhou, China, May 2004.

[23] I. Garousi-Nejad, O. Bozorg-Haddad, and H. A. Loáiciga, "Modified firefly algorithm for solving multireservoir operation in continuous and discrete domains," *Journal of Water Resources Planning and Management*, vol. 142, Article ID 4016029, 2016.

[24] V. Jothiprakash and G. Shanthi, "Single reservoir operating policies using genetic algorithm," *Water Resources Management*, vol. 20, no. 6, pp. 917–929, 2006.

[25] H. Bashiri-Atrabi, K. Qaderi, D. E. Rheinheimer, and E. Sharifi, "Application of harmony search algorithm to reservoir operation optimization," *Water Resources Management*, vol. 29, no. 15, pp. 5729–5748, 2015.

[26] A. Afshar, O. Bozorg Haddad, M. A. Mariño, and B. J. Adams, "Honey-bee mating optimization (HBMO) algorithm for optimal reservoir operation," *Journal of the Franklin Institute*, vol. 344, no. 5, pp. 452–462, 2007.

[27] Z. M. Yaseen, "Optimization of reservoir operation using new hybrid algorithm," *KSCE Journal of Civil Engineering*, vol. 22, pp. 1–13, 2018.

[28] M. Janga Reddy and D. Nagesh Kumar, "Multi-objective particle swarm optimization for generating optimal trade-offs in reservoir operation," *Hydrological Processes*, vol. 21, 2007.

[29] M. Ehteram, "Optimization of chain-reservoirs' operation with a new approach in artificial intelligence," *Water Resources Management*, vol. 31, 2017.

[30] M. Ehteram, A. Ferdowsi, M. Faramarzpour et al., "Hybridization of artificial intelligence models with nature inspired optimization algorithms for lake water level prediction and uncertainty analysis," *Alexandria Engineering Journal*, vol. 60, no. 2, pp. 2193–2208, 2021.

[31] O. B. Haddad, M. Moravej, and H. A. Loáiciga, "Application of the water cycle algorithm to the optimal operation of reservoir systems," *Journal of Irrigation and Drainage Engineering*, vol. 141, no. 5, Article ID 4014064, 2015.

[32] H.-R. Asgari, O. B. Haddad, M. Pazoki, and H. A. Loáiciga, "Weed optimization algorithm for optimal reservoir operation," *Journal of Irrigation and Drainage Engineering*, vol. 142, no. 2, 2016.

[33] E. Ahmadebrahimpour, "Optimal operation of reservoir systems using the wolf search algorithm (WSA)," *Water Supply*, vol. 19, no. 5, pp. 1396–1404, 2019.

[34] A. Ahmad, A. El-Shafie, S. F. M. Razali, and Z. S. Mohamad, "Reservoir optimization in water resources: a review," *Water Resources Management*, vol. 28, no. 11, pp. 3391–3405, 2014.

[35] V. Kumar and S. M. Yadav, "Optimization of reservoir operation with a new approach in evolutionary computation using TLBO algorithm and Jaya algorithm," *Water Resources Management*, vol. 32, no. 13, pp. 4375–4391, 2018.

[36] V. Paliwal, A. D. Ghare, A. B. Mirajkar, N. D. Bokde, and A. E. Feijóo Lorenzo, "Computer modeling for the operation optimization of Mula reservoir, upper Godavari basin, India, using the Jaya algorithm," *Sustainability*, vol. 12, no. 1, p. 84, 2019.

[37] K. L. Chong, S. H. Lai, A. N. Ahmed, W. Z. Wan Jaafar, and A. El-Shafie, "Optimization of hydropower reservoir operation based on hedging policy using Jaya algorithm," *Applied Soft Computing*, vol. 106, Article ID 107325, 2021.

[38] E. Motevali Bashi Naeini and S. Soltaninia, "Computationally efficient evolutionary optimisation for joint reservoir design-operations," *Proceedings of the Institution of Civil Engineers-Water Management*, pp. 1–16, 2021.

[39] R. V. Rao, "Rao algorithms: three metaphor-less simple algorithms for solving optimization problems," *International Journal of Industrial Engineering Computations*, vol. 11, pp. 107–130, 2020.

[40] L. Wang, Z. Wang, H. Liang, and C. Huang, "Parameter estimation of photovoltaic cell model with Rao-1 algorithm," *Optik*, vol. 210, Article ID 163846, 2020.

[41] R. V. Rao and R. B. Pawar, "Constrained design optimization of selected mechanical system components using Rao algorithms," *Applied Soft Computing*, vol. 89, Article ID 106141, 2020.

[42] A. Sharafati, Z. M. Yaseen, and S. Shahid, "A novel simulation-optimization strategy for stochastic-based designing of flood control dam: a case study of Jamishan dam," *Journal of Flood Risk Management*, vol. 14, no. 1, Article ID e12678, 2021.

[43] M. M. A. AL-Shammari, A. M. AL-Shamma'a, A. Al Maliki, H. M. Hussain, Z. M. Yaseen, and A. M. Armanuos, "Integrated water harvesting and aquifer recharge evaluation methodology based on remote sensing and geographical information system: case study in Iraq," *Natural Resources Research*, vol. 30, pp. 1–25, 2021.

[44] N. K. Biswas, F. Hossain, M. Bonnema, H. Lee, and F. Chishtie, "Towards a global reservoir assessment tool for predicting hydrologic impacts and operating patterns of existing and planned reservoirs," *Environmental Modelling & Software*, vol. 140, Article ID 105043, 2021.

[45] F. Giudici, D. Anghileri, A. Castelletti, and P. Burlando, "Descriptive or normative: how does reservoir operations modeling influence hydrological simulations under climate change?," *Journal of Hydrology*, vol. 595, Article ID 125996, 2021.

[46] R. E. Larson, "State increment dynamic programming," *IEEE Transactions on Automatic Control*, vol. 15, 1968.

[47] V. Te Chow and G. Cortes-Rivera, "Application of DDDP in water resources planning," Final report, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 1974.

[48] M. Heidari, V. T. Chow, P. V. Kokotović, and D. D. Meredith, "Discrete differential dynamic programing approach to water resources systems optimization," *Water Resources Research*, vol. 7, no. 2, pp. 273–282, 1971.

[49] D. N. Kumar and F. Baliarsingh, "Folded dynamic programming for optimal operation of multireservoir system," *Water Resources Management*, vol. 17, 2003.

[50] D. Nagesh Kumar and M. Janga Reddy, "Multipurpose reservoir operation using particle swarm optimization," *Journal of Water Resources Planning and Management*, vol. 133, no. 3, pp. 192–201, 2007.

[51] D. M. Murray and S. J. Yakowitz, "Constrained differential dynamic programming and its application to multireservoir control," *Water Resources Research*, vol. 15, no. 5, pp. 1017–1027, 1979.

[52] M. Ehteram, S.-F. Mousavi, H. Karami et al., "Fast convergence optimization model for single and multi-purposes reservoirs using hybrid algorithm," *Advanced Engineering Informatics*, vol. 32, pp. 287–298, 2017.

[53] M. Agenis-Nevers, N. D. Bokde, Z. M. Yaseen, and M. K. Shende, "An empirical estimation for time and memory algorithm complexities: newly developed R package," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2997–3015, 2021.

[54] M. Agenis and N. Bokde, "GuessCompx: empirically estimates algorithm complexity; R package version 1.0.3," 2019, https://cran.r-project.org/web/packages/GuessCompx/index.html.

Hindawi
Complexity
Volume 2021, Article ID 6664544, 11 pages
https://doi.org/10.1155/2021/6664544

Research Article

Government Supervision on Explosive Enterprises' Immoral Behaviors in E-Commerce Enterprises: An Evolutionary Game Analysis

Liang Shen,[1] **Yuanyuan Chen**,[1] **Runjie Fan**,[1] **and Yuyan Wang**[2]

[1]*School of Public Finance and Taxation, Shandong University of Finance and Economics, Jinan, Shandong 250014, China*
[2]*School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan, Shandong 250014, China*

Correspondence should be addressed to Yuyan Wang; wangyuyan1224@126.com

Received 15 December 2020; Revised 19 January 2021; Accepted 30 May 2021; Published 10 June 2021

Academic Editor: Baogui Xin

Copyright © 2021 Liang Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Explosive enterprises' immoral behaviors in the online shopping market are widespread and have not been effectively solved. Especially in developing countries, there is a direct relationship between massive immoral behaviors and the inefficiency of government supervision. Using an evolutionary game, this paper finds that immoral behavior is more likely to spread in online markets than in traditional markets. Only when government supervision and punishment are large enough and government's punishment for the illegal enterprise exceeds extra supervision costs that government pays, explosive immoral behaviors can be curbed. Additionally, consumer support is an essential factor in improving the efficiency of government supervision. This study sorts out the interactions between e-commerce market participants and the government, obtains a path to achieve efficient government regulation, and offers management insights. The findings can serve as a reference for ensuring order in the emerging online shopping market and can also provide theoretical references for future related research.

1. Introduction

With the development of the internet economy, online shopping has become a mainstream consumption method, which has increased the flexibility of shopping [1] and also reduced shopping time and shopping cost for consumers [2]. According to the 2017 World E-Commerce Report released by China International Electronic Commerce Center (CIECC), the global online retail transaction volume reached $3.5 trillion in 2019, accounting for 14.12% of the total global retail sales. In most developed countries, more than half of the population buys goods and services online. Among them, the growth rates of e-tailing in India, China, South Korea, and the United States were 31.9%, 27.3%, 18.1%, and 14.0%, respectively, and online shopping has become a globalized form of consumption (https://www.chyxx.com/industry/202010/902336.html). In recent years, internet and digital technologies have developed rapidly in China, and the online shopping market in China is also expanding rapidly. By the end of 2019, the scale of China's e-commerce transactions and online payment transactions reached CNY 34.81 trillion and CNY 249.88 trillion, respectively (https://www.chyxx.com/research/202009/897364.html). The scale of China's digital economy added value reached CNY 35.8 trillion, which has steadily ranked the second in the world. Some organizations predict that, with the impact of big data, cloud computing, artificial intelligence, blockchain, and other technologies on the online shopping market, online shopping will become the main shopping method in the future. The proportion of online shopping in total retail sales of consumer goods is expected to reach 20%–25% in 2020.

At the same time, the economic and ethical consequences of enterprises' immoral behaviors in network markets are more serious than those in traditional markets. Most directly, the massive expansion of enterprises' immoral behaviors will reduce national tax revenue on a large scale,

which leads to massive economic losses. For example, research by the UK's Center for Economic and Commercial Research showed that, in 2016 alone, counterfeit goods caused British to lose £17.3 billion, reducing employment opportunities by 72,000 (http://www.independent.co.uk/money/counterfeit-goods-tempting-danger-a7473751.html). It is easy to overlook that enterprises' immoral behaviors in network markets can destroy the ethical climate of the market [3], which can create a vicious cycle. On the one hand, enterprises' immoral behaviors can elicit immoral retaliation from consumers [4]. Consumers' perceptions of corporate behavior are fundamentally rooted in morality [5], and consumers who are subjected to enterprises' immoral behaviors can have a desire for revenge and retaliation [6], thus making the online shopping environment more chaotic. On the other hand, the process of spreading immoral behavior in network markets is similar to the process of virus spreading [7]. Once the fraudulent profiteering is not punished promptly, the negative incentive will spread contagiously in the society, forming "explosive immoral behaviors (EIBs)."

EIB is a phenomenon in which economic actors act to the detriment of others while maximizing their own utility [8]. There are many causes of EIBs, such as asymmetric information between supply and demand [9], corporate ethical failures [10], excessive corporate pressure [11], ineffective government regulation [12], and consumers' desire to purchase counterfeits [13]. Among them, information asymmetry is a prerequisite for the existence of such problems. However, due to the inaccessibility of online goods and the mixed nature of product information, the asymmetry of information between merchants and consumers in online marketplaces is exacerbated, ultimately resulting in market failures [14, 15]. This is demonstrated, on the one hand, by the fact that product quality problems have been generalized in the network market. In 2018, GAO staff in the United States conducted a study (the report can be retrieved from http://www.amz123.com/thread-48343.htm?sort = desc). They secretly purchased four branded goods (Nike Air Jordan sneakers, Yeti Mug, Urban Decay cosmetics, and UL-certified mobile phone chargers) from five popular e-commerce platforms, namely, Amazon, Walmart, Sears, Newegg Network, and eBay, and the results showed that 20 of the 47 products purchased were fakes. In a survey in the UK, 75% of respondents reported that they had bought fakes (the report can be retrieved from https://k.sina.cn/article_6474654027_181eb614b00100659f.html?http=fromhttp). A new study by Velocity MR, a market research firm in India, found that 30% of e-commerce products in India are fakes (the study can be retrieved from https://www.cifnews.com/article/34924). In China, according to the survey, 77.8% of e-commerce platforms have fakes on sale, and 98% of online customers have purchased fakes (the report can be retrieved from http://www.kjeport.com/detail/article/2015_1/1_26/2038796_1.shtml). The other hand of EIBs is that ethical misconduct in the network market has expanded, including distortion of goods display information, ethical misconduct of big data marketing, and large-scale tax avoidance. For example, in 2002, Amazon was exposed to the fact that the purchase price of the same item for members would be higher than the purchase price for nonmembers (the study can be retrieved from https://www.researchgate.net/publication/262493771_When_big_data_meets_dataveillance_The_hidden_side_of_analytics). And in 2019, a social survey in China showed that 88.32% of the respondents believed that "overcharging familiar customers" through big data was widespread, and 56.92% of the respondents said they had been "overcharged" (the report can be retrieved from http://news.timedg.com/2019-03/28/20821145.shtml). As seen above, the phenomenon of EIBs in the network market is intensifying. This will make the law-abiding enterprise management behavior unsustainable, which not only seriously affects the competitive environment of the online shopping market but also provides reasons for government regulation.

In 2017, UK police have shut down 28,000 websites selling counterfeit goods, more than 4,000 of which were set up by stealing customers' identity information and using their names. In 2019, the *E-Commerce Law of the People's Republic of China* came into effect to maintain order in the network market. In 2020, the US signed the first *Executive Order 19304 (E.O. 13904)* to stop the flow of counterfeit goods and other contraband into the US marketplace. The executive order focuses on combating counterfeit goods traded on third-party online e-commerce platforms, including Alibaba, Amazon, and eBay. However, the EIBs in the network market have not been stopped. This is because the attitude of enterprises toward government supervision may be disobedience or manipulation besides compliance [16]. Especially when the target of government supervision is not a minority group in the market, the regulatory authority will face an awkward situation of lawlessness, which more easily leads to the failure of regulation and enforcement. Meanwhile, government supervision can incur substantial costs, and ineffective or inefficient regulation may even reduce social welfare. Therefore, it is crucial to find the right and efficient path of governmental supervision to mitigate and curb EIBs in the network market.

The main contributions of this paper are as follows:

(i) This study elucidates the intrinsic logical relationship among the network market characteristics, the e-commerce enterprises' behaviors, and government regulation in order to better analyze the causes of current government supervision inefficiencies

(ii) This paper analyzes the evolutionary process and internal mechanism of EIB through evolutionary games, highlights the role of government supervision in curbing EIBs in online markets, and obtains a path to achieve efficient government regulation

(iii) This paper uses the Chinese government's regulation of EIBs in online markets as an example to verify the results obtained in the evolutionary game and to provide some managerial insights for countries where EIB regulation is inefficient or even ineffective

The study is structured as follows: the literature review is presented in Section 2. Section 3 gives a dynamic game

model for the choice of immoral strategies of large- and small-scaled enterprises to analyze the inherent relationship between enterprise behaviors and government supervision. In Section 4, the evolutionary game model is constructed, and an effective strategy for the government to regulate the network market is given. Finally, Section 5 gives conclusions and management insights.

2. Literature Review

The purpose of this paper is to clarify the inherent logical relationship between the characteristics of enterprise behaviors, network market characteristics, and government regulation and to explore the feasible conditions and implementation paths for government supervision of the network market. Therefore, in this section, whether EIBs in the network market require government supervision will be clarified, and the contribution and shortcomings of the literature on government supervision in the network market will be summarized.

Some scholars have focused on the managerial value of organizational ethics in response to enterprises' immoral behaviors in traditional markets [17, 18] and have insisted that the market is the best regulator of immoral behaviors rather than the government [19, 20]. However, online marketplaces differ from traditional markets. Mavlanova et al. [21] pointed out that the short duration of a single transaction when shopping online makes it more difficult for consumers to identify fraud signals. Meanwhile, Bergh et al. [22] argued that online sellers lack the awareness to recognize fraud signals. As a result, EIBs in the network market are more likely to arise and more insidious [23]. Another segment of scholars has noted the privacy risks of online marketplaces. Walsh et al. [24] concluded that the current federal and state laws in the United States are unable to protect the privacy of online buyers and highlighted the role of government regulation in protecting privacy. Mutimukwe et al. [25] suggested that the perceived privacy risk of e-marketplace users reduces their purchase utility. Anic et al. [26] showed that online shoppers want less privacy disclosure and that government regulation is very weak on privacy aspects of online marketplaces.

Previous studies generally agree that the emergence of online marketplaces has exacerbated EIBs of firms and caused market failures [27]. Calkins et al. [28] argued that although market regulators have more information about fraudulent transactions than government regulators, they have no incentive to proactively check for extensive online sellers who trade counterfeit goods. González [29] similarly believed that commercial interests also prevent the implementation of market regulation. To et al. [14] and Safaei and Thoben [7] showed that immoral behaviors in online shopping markets are more likely to arise and spread as information asymmetry between merchants and consumers is further amplified, which leads to EIBs and market failures. Thus, while market regulation shows great strengths and potential in combating fraudulent transactions, it is still exposed to the general weaknesses of commercial self-regulation. Only government intervention in the network market to regulate businesses is likely to create new constraints on profit-seeking behavior and thus mitigate EIBs [30].

In the research related to government regulation in the network market, some scholars believed that the law is a key factor in regulating immoral behaviors. According to the empirical analysis on data from 30 countries, Oxley and Yeung [31] found that the law can effectively promote the healthy development of e-commerce. Martinsons [32] believed that forceful regulation and explicit law are the important conditions of e-commerce development. Examining how e-commerce law should be developed in the United States, Ribstein and Kobayashi [33] argued that the e-commerce law should be enacted by states rather than federal. The flexibility and variability of state lawmaking and competition between state legislatures can contribute to government regulation. AlGhamdi and Drew [34] pointed out that the lack of explicit regulations and ineffective government supervision are key factors affecting the immoral behaviors of e-commerce enterprises. Agrawal et al. [35] suggested that the government needs to participate in the development of e-commerce rules within the legal framework. Kim [36] noted that legal changes should be used for the maintenance of e-marketplace stability and explored how the government can regulate the use and identification of electronic signatures. Additionally, some scholars have given other suggestions for the government to regulate immoral behaviors in the network market. In terms of how to improve government supervision efficiency, Weiser [37] believed that public institutions need to cooperate with private sectors with technological advantages to improve government supervision efficiency. Marsden [38] proposed that the public institutions, private sectors, and civil society institutions should jointly participate in the supervision of the network market to form a supervision mode of collaborative management.

In summary, for immoral behaviors in the network market, the extant literature mainly discussed the importance of government supervision and its key elements from a macroperspective. Few scholars probe into the inherent logical relationship between government supervision and EIBs and explore the specific implementation of the government supervision mechanism. It is believed that the current inefficiency of government supervision is related to the lack of analysis of the evolutionary process and mechanism of e-commerce enterprises' immoral behaviors from a microperspective. Therefore, this study uses the evolutionary game as an analysis tool to uncover the inherent mechanism of e-commerce enterprises' immoral behaviors and government supervision. A stable game strategy to improve the efficiency of government supervision and corresponding implementation measures are proposed.

3. Mechanism of Explosive Immoral Behaviors of E-Commerce Enterprises

Compared with traditional enterprises, EIBs are more likely to occur in e-commerce enterprises due to characteristics of the network market, such as virtual transaction entities and

the lower threshold of entry and exit. Consumers are faced with too much virtual data to make sensible choices, which makes the e-commerce enterprises' EIBs present the characteristics of universality, diversity, and difficulty to regulate.

The game process of government supervision and e-commerce enterprises is illustrated with figures. Assume that e-commerce enterprises can choose two strategies: one is to disclose true quality information of products, and this strategy is marked as MO; the other is to provide false quality information of products, that is, the enterprise adopts the immoral behavior, and this strategy is marked as IM. Suppose there are two competitive enterprises of $E1$ and $E2$ in the network market, and their scales and costs are different. Assume that $E1$ is large in scale and low in cost, and $E2$ is large in scale and high in cost. When the degree of government supervision is S, the strategic choices and profits of the two enterprises are shown in Figure 1.

Obviously, when MO is adopted by two enterprises, the profit of $E1$ is greater than that of $E2$. When both enterprises choose IM, the probability of $E1$ being discovered and penalized is greater which is illustrated by the larger slope in the profit curve of $E1$ than $E2$. When adopting immoral behaviors, small- and medium-sized enterprises (SMEs) are more difficult to supervise. As can be seen from Figure 1, SMEs will give up fakes only when $S > S^{**}$.

As can be seen from Figure 1, when the degree of government supervision is small $(0 < S < S^{*})$, two enterprises will choose IM, and the profit of $E1$ is greater than that of $E2$; as S increases, the profits of both enterprises adopting IM will decrease, and the profit of $E1$ decreases more. When $S \in (S^{*}, S^{**})$, it is more advantageous for $E1$ to choose MO, while $E2$ still adopts IM. When $S > S^{**}$, two enterprises will choose MO, and when the degree of government supervision is large enough $(S > S'')$,$E1$ with IM is more severely penalized, and its profit will be smaller than $E2$ with IM.

In order to further explain the government supervision strategy, a more detailed discussion is given to connect the government supervision and the immoral costs of different enterprises. It is assumed that there are e-commerce enterprise $L1$ and e-commerce enterprise $L2$ with different scales and costs. Enterprise $L1$ is small in scale and high in cost, whose cost is C_{L1} and profit is $\Delta\pi_{L1}$ when adopting IM. Enterprise $L2$ is large in scale and low in cost, whose cost is C_{L2} and profit is $\Delta\pi_{L2}$ when adopting IM. It is assumed that $\Delta\pi_{L2} > \Delta\pi_{L1}$ and $C_{L2} < C_{L1}$, and the degree of government supervision is S. The relationship among S, the profits, and the costs when enterprises adopt IM is shown in Figure 2.

When the cost and the profits of enterprises adopting IM are located in region I, there is $C_{L1} > \Delta\pi_{L1}$ and $C_{L2} > \Delta\pi_{L2}$. This means that the enterprise is not profitable when adopting IM. Even if the government does not supervise, the enterprises will not adopt IM. Region I is regarded as an invalid zone of IM.

Region II represents $C_{L1} < \Delta\pi_{L1}$, $C_{L2} > \Delta\pi_{L2}$, and $S < \Delta\pi_{L1} < \Delta\pi_{L2}$. This region is the invalid zone for $L2$ adopting IM, so $L2$ will adopt MO. Because S is weak, $L1$ can gain profit by IM, so $L1$ will adopt IM.

Region III represents $C_{L1} < \Delta\pi_{L1}$, $C_{L2} < \Delta\pi_{L2}$, and $S < \Delta\pi_{L1} < \Delta\pi_{L2}$. In this region, both $L1$ and $L2$ can gain profits by IM. Because S is weak, both $L1$ and $L2$ will adopt IM.

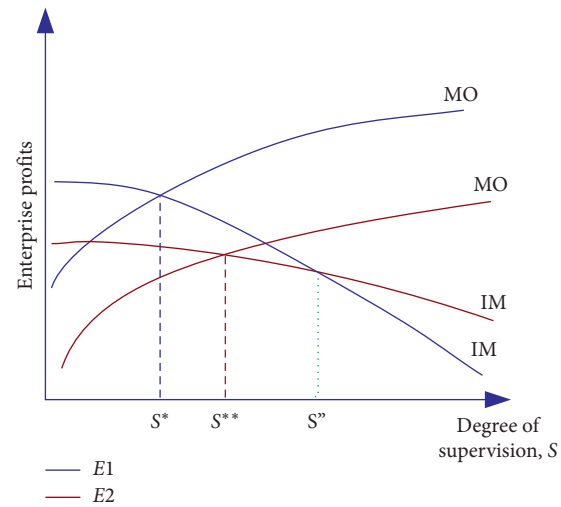


Figure 1: Profits of heterogeneous enterprises under government supervision.

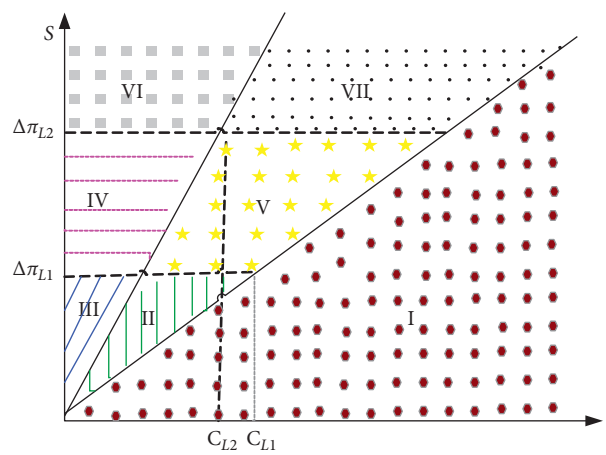


Figure 2: Relationship among S, profit, and cost under IM.

Region IV represents $C_{L1} < \Delta\pi_{L1}$, $C_{L2} < \Delta\pi_{L2}$, and $\Delta\pi_{L1} < S < \Delta\pi_{L2}$. In this region, both $L1$ and $L2$ can gain profits from IM, but S is effective for enterprise $L1$ and ineffective for enterprise $L2$. So, $L1$ will adopt MO and $L2$ will adopt IM.

Region V represents $C_{L1} < \Delta\pi_{L1}$, $C_{L2} > \Delta\pi_{L2}$, and $\Delta\pi_{L1} < S < \Delta\pi_{L2}$. This region is the invalid region of $L2$ adopting IM, so enterprise $L2$ will adopt MO. Although $L1$ can obtain profit by IM, $L1$ still adopts MO because S is strong. Therefore, this region is an effective region of government supervision.

Region VI represents $C_{L1} < \Delta\pi_{L1}$, $C_{L2} < \Delta\pi_{L2}$, and $\Delta\pi_{L1} < \Delta\pi_{L2} < S$. In this region, although both $L1$ and $L2$ can obtain profit by IM, $L1$ and $L2$ have to adopt MO due to the stronger government supervision S; that is, this region is also an effective region of government supervision.

Region VII represents $C_{L1} < \Delta\pi_{L1}$, $C_{L2} > \Delta\pi_{L2}$, and $\Delta\pi_{L1} < \Delta\pi_{L2} < S$. The region is the invalid region of $L2$ adopting IM. Although $L1$ can gain profit, government

supervision S is very strong, and both enterprises adopt MO. This region is also an effective region of government supervision.

Through the analysis of the models, it is indicated that when the competitive enterprise chooses IM, the enterprise choosing MO obtains a lower profit. When the declining speed of the enterprise's profit is slower by the increase of government supervision, the enterprise will adopt IM in a larger range $(0 < S < S^{**})$; otherwise, the enterprise will adopt IM in a smaller range $(0 < S < S^{*})$. That is to say, the competitor's choice of IM can stimulate other enterprises to adopt IM, thus inducing the EIB. Government supervision has become an important factor affecting enterprise strategy choices.

4. The Evolutionary Game between Government Supervision and Strategy Choices of E-Commerce Enterprises

In practice, government supervision on e-commerce enterprises needs costs, which directly affect supervision effectiveness. In the process of supervision, the government and the e-commerce enterprise constitute a dynamic and repeated game relationship, where two parties constantly correct their strategies through trial and error to maximize their own profits. Therefore, an evolutionary game model of government supervision and strategies of e-commerce enterprises is constructed to find effective equilibrium for government supervision.

In order to further discuss the effectiveness of government supervision, the degree of government supervision and enterprise profit are separated. For any e-commerce enterprise in the network market, it is assumed that the e-commerce enterprise has two strategies in the competitive market, namely, MO and IM. If e-commerce enterprise i adopts MO, the profit is π_i; if e-commerce enterprise i adopts IM, the profit can be increased by $\Delta\pi_i$; that is, the profit of adopting IM is $\pi_i + \Delta\pi_i$. For different choices, the government implements different degrees of supervision. The degree of supervision and supervision cost are generally corresponding; that is, the greater the degree, the greater the cost and the higher the probability of successful supervision. For the convenience of analysis, the supervision cost is used to reflect the degree of supervision, which does not affect the conclusions. Assume that when government supervision is successful, the supervision cost is S_g; when government supervision fails, the supervision cost is S_k. When government supervision is successful and all e-commerce enterprises adopt MO, the network market achieves the optimal operation, which will promote the development of the network economy, the increase of social welfare, and the enhancement of consumers' trust and support to the government. All these can be seen as the utility of government supervision. And suppose f is the utility of government supervision, namely, the degree of consumers' support to the government. When the government supervision is successful, if an enterprise adopts IM, the enterprise will be punished, and the punishment is $F\,(F > \Delta\pi_i)$.

Based on Hofbauer et al. [39] and Shen and Wang [40], the profit matrix of government-supervised e-commerce enterprises is shown in Table 1.

In an evolutionary game model, the strategies of participants are mutually influenced. The growth rate of a certain strategy depends on the fitness, i.e., the strategy that produces higher profit has a higher growth rate. As a result, the government and the e-commerce enterprise can increase their profits by mimicking the strategies that succeed.

It is assumed that the e-commerce enterprise can randomly and independently choose MO or IM and repeat the game in the market. Assume that the probability of e-commerce enterprise 1 choosing MO is p, and the probability of choosing IM is $1 - p$; the probability of successful government supervision is q, and the probability of supervision failure is $1 - q$. According to the Malthusian equation [41], the growth rate $\dot{p}/p$ of e-commerce enterprise i choosing MO should be equal to its fitness, $e_1 A\{q, 1-q\}^T$, minus its average fitness, $\{p, 1-p\}A\{q, 1-q\}^T$. $e_1 = [1, 0]$ represents the e-commerce enterprise choosing MO with 100% probability. Then, the profit matrix of e-commerce enterprise i is

$$A = \begin{bmatrix} \pi_i & \pi_i \\ \pi_i + \Delta\pi_i - F & \pi_i + \Delta\pi_i \end{bmatrix}. \tag{1}$$

So, $\dot{p} = p(1-p)\{1, -1\}A\{q, 1-q\}^T$, and it can be derived that $\dot{p} = p(1-p)(Fq - \Delta\pi_i)$.

Similarly, considering the growth rate of government's supervision success, $\dot{q}/q$, it can be derived that $\dot{q} = q(1-q)[(F - S_g + S_k) + (f - F)p]$. The local stability of the system at equilibrium is analyzed using the Jacobian matrix [42]:

$$J = \begin{bmatrix} \dfrac{\partial \dot{p}}{\partial p} & \dfrac{\partial \dot{p}}{\partial q} \\ \dfrac{\partial \dot{q}}{\partial p} & \dfrac{\partial \dot{q}}{\partial q} \end{bmatrix} = \begin{bmatrix} (1-2p)(Fq - \Delta\pi_i) & Fp(1-p) \\ (f-F)q(1-q) & (1-2q)\left[\left(F - S_g + S_k\right) + (f-F)p\right] \end{bmatrix}. \tag{2}$$

It can be derived that

Table 1: Profit matrix of government-supervised e-commerce enterprises.

		Government supervision	
		Success	Failure
Strategies of e-commerce enterprises	MO	$\pi_i, f - S_g$	$\pi_i, -S_k$
	IM	$\pi_i + \Delta\pi_i - F, F - S_g$	$\pi_i + \Delta\pi_i, -S_k$

$$\begin{aligned} \det J &= (1-2p)(1-2q)\left(Fq - \Delta\pi_i\right)\left[\left(F - S_g + S_k\right) + (f-F)p\right] - (f-F)Fpq(1-q)(1-p), \\ trJ &= (1-2p)\left(Fq - \Delta\pi_i\right) + (1-2q)\left[\left(F - S_g + S_k\right) + (f-F)p\right]. \end{aligned} \tag{3}$$

The evolutionarily stable strategy (ESS) of the system is analyzed as follows:

(i) When $f < S_g - S_k$, according to $\dot{p} = 0$ and $\dot{q} = 0$, the system equilibrium is $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, and $((F - S_g + S_k)/(F - f), \Delta\pi_i/F)$.

(1) When $F > F - S_g + S_k > 0$, $F > S_g - S_k > 0$. The ESS analysis is shown in Table 2.

(2) When $F - S_g + S_k < 0$, $0 < \Delta\pi_i < F < S_g - S_k$. The ESS analysis is shown in Table 3.

(ii) When $f > S_g - S_k$, according to $\dot{p} = 0$ and $\dot{q} = 0$, the system equilibrium is $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$.

(1) When $F > F - S_g + S_k > 0$, $F > S_g - S_k > 0$. The ESS analysis is shown in Table 4.

(2) When $F - S_g + S_k < 0$, $0 < \Delta\pi_i < F < S_g - S_k$. The ESS analysis is shown in Table 5.

From the analysis of Tables 2–5, the ESS can be reached only when the consumer support for government regulation satisfies $f > S_g - S_k$, However, it is worth noting that when $0 < \Delta\pi_i < F < S_g - S_k$, $(0,0)$, i.e., government supervision fails and the e-commerce enterprise adopts IM, is also the ESS. Actually, this situation is the other extreme of the network market with EIBs of e-commerce enterprises, which is caused by the weaker government punishment $F < S_g - S_k$. Therefore, to ensure the normal operation of the network market, $(1,1)$ must be regarded as the only ESS of the evolutionary game, that is, the following conditions must be met:

$$f > S_g - S_k, F > S_g - S_k > 0, F > \Delta\pi_i, S_g > S_k > 0. \tag{4}$$

(i) The government must punish enterprises that adopt IM. Only when the government's punishment is greater than the profit obtained by IM, i.e., $F > \Delta\pi_i$, enterprises will choose MO. Moreover, the higher the punishment imposed by the government, the greater the impact on the e-commerce enterprises and the fewer the enterprises adopting IM.

(ii) The degree of government supervision needs to be large, i.e., $S_g > S_k > 0$. Only when the government supervision reaches a certain level, supervision can be effective. If government supervision is weak and of no practical significance, the supervision is just a thankless task.

(iii) When the government supervision is enhanced, the increased supervision cost $S_g - S_k$ has to be compensated to maintain a high level of supervision. This requires the government to punish the enterprises that adopt IM at a punishment more than the increased cost, i.e., $F > S_g - S_k > 0$.

(iv) The consumers' support that must be met, $f > S_g - S_k$, is an indispensable condition for successful government supervision. The timely and comprehensive feedback of consumers not only directly affect enterprises' reputation but also provide evidence for authorities of the government supervision, reducing the supervision cost and improving supervision efficiency.

5. Practice and Discussion

In the previous section, the conditions for the government to effectively supervise EIBs in the network market are obtained through an evolutionary game: the government punishes enterprises with immoral behaviors strongly enough, the degree of government supervision is high enough, the government needs to be compensated for the supervision cost, and consumers should support government supervision through feedback. Previous studies suggest that governments in developing countries should strengthen the regulation of online shopping markets compared to developed countries [35]. In most developing countries, e-commerce markets lack clear regulations and effective government oversight [34]. Moreover, counterfeit products damage more the reputation of developing countries than the economy. Therefore, it is more important for governments in developing countries to take strict measures to regulate online counterfeiting and other EIBs. Studies have shown that market self-regulation of counterfeit products in China is inefficient [43]. Based on this, the Chinese government's supervision on EIBs in the network market is used as a typical example in the section to verify the results obtained from the evolutionary game and to provide some managerial insights for countries where EIBs' regulation is inefficient or even ineffective.

From the practice in China, e-commerce platforms play an important role in combating immoral behaviors, and government supervision is indispensable. Besides, e-commerce platforms play the due role only when government supervision is effective. For example, Alibaba, a giant

Table 2: The stability of the equilibrium in Case 1-(1).

Equilibrium point (p, q)	$\det J$		$\mathrm{tr} J$		Local stability
$(0,0)$	$-\Delta\pi_i(F - S_g + S_k)$	$-$	$-\Delta\pi_i + F - S_g + S_k$	Uncertain	Saddle point
$(0,1)$	$-(F - \Delta\pi_i)(F - S_g + S_k)$	$-$	$-\Delta\pi_i + S_g - S_k$	Uncertain	Saddle point
$(1,0)$	$\Delta\pi_i(f - S_g + S_k)$	$-$	$\Delta\pi_i + f - S_g + S_k$	Uncertain	Saddle point
$(1,1)$	$(F - \Delta\pi_i)(f - S_g + S_k)$	$-$	$-(F - \Delta\pi_i) - (f - S_g + S_k)$	Uncertain	Saddle point
$((F - S_g + S_k)(F - f), (\Delta\pi_i/F))$	$-\Delta\pi_i\ (F - S_g + S_k)(f - S_g + S_k)/(F - f)(1 - \Delta\pi_i/F)$	$+$	0		Uncertain point

Table 3: The stability of the equilibrium in Case 1-(2).

Equilibrium point (p, q)	$\det J$		$\mathrm{tr} J$		Local stability
$(0,0)$	$-\Delta\pi_i(F - S_g + S_k)$	$+$	$-\Delta\pi_i + F - S_g + S_k$	$-$	ESS
$(0,1)$	$-(F - \Delta\pi_i)(F - S_g + S_k)$	$+$	$-\Delta\pi_i + S_g - S_k$	$+$	Unstable point
$(1,0)$	$\Delta\pi_i(f - S_g + S_k)$		$\Delta\pi_i + f - S_g + S_k$	Uncertain	Saddle point
$(1,1)$	$(F - \Delta\pi_i)(f - S_g + S_k)$	$-$	$-(F - \Delta\pi_i) - (f - S_g + S_k)$	Uncertain	Saddle point
$((F - S_g + S_k)(F - f), (\Delta\pi_i/F))$	$-\Delta\pi_i\ (F - S_g + S_k)(f - S_g + S_k)/(F - f)(1 - \Delta\pi_i/F)$	$+$	0		Uncertain point

Table 4: The stability of the equilibrium in Case 2-(1).

Equilibrium point (p, q)	$\det J$		$\mathrm{tr} J$		Local stability
$(0,0)$	$-\Delta\pi_i(F - S_g + S_k)$	$-$	$-\Delta\pi_i + F - S_g + S_k$	Uncertain	Saddle point
$(0,1)$	$-(F - \Delta\pi_i)(F - S_g + S_k)$	$-$	$-\Delta\pi_i + S_g - S_k$	Uncertain	Saddle point
$(1,0)$	$\Delta\pi_i(f - S_g + S_k)$	$+$	$\Delta\pi_i + f - S_g + S_k$	$+$	Unstable point
$(1,1)$	$(F - \Delta\pi_i)(f - S_g + S_k)$	$+$	$-(F - \Delta\pi_i) - (f - S_g + S_k)$	$-$	ESS

Table 5: The stability of the equilibrium in Case 2-(2).

Equilibrium point (p, q)	$\det J$		$\mathrm{tr} J$		Local stability
$(0,0)$	$-\Delta\pi_i(F - S_g + S_k)$	$+$	$-\Delta\pi_i + F - S_g + S_k$	$-$	ESS
$(0,1)$	$-(F - \Delta\pi_i)(F - S_g + S_k)$	$+$	$-\Delta\pi_i + S_g - S_k$	$+$	Unstable point
$(1,0)$	$\Delta\pi_i(f - S_g + S_k)$	$+$	$\Delta\pi_i + f - S_g + S_k$	$+$	Unstable point
$(1,1)$	$(F - \Delta\pi_i)(f - S_g + S_k)$	$+$	$-(F - \Delta\pi_i) - (f - S_g + S_k)$	$-$	ESS

e-commerce enterprise in China, has a special antifaking office to deal with fakes in the network market after 2011, which is served by the deputy CFO as the administrator (the report is at https://tech.sina.com.cn/i/2016-07-01/doc-ifxtsatm1135104.shtml). More than 2,000 employees are on full time to deal with counterfeits, another 5,000 social volunteers participate in antifaking, and the direct cost is about CNY 1 billion every year (the data can be retrieved from https://www.kaitao.cn/article/20170811102329.htm). Alibaba also conducts random inspections of goods through big data algorithms and initiates about 100,000 random inspections every year. CNY 100 million is invested to purchase goods sold on the platform. Once fakes are found, the goods will be removed, and the e-commerce enterprise will even be driven out of the platform (the report can be retrieved from http://paper.ce.cn/jjrb/html/2018-08/29/content_371267.htm). Alibaba's antifaking measures were even appreciated by Doug Collins, the vice-chairman of the US House Judiciary Committee (http://www.sofreight.com/news_36233.html). Despite this, fakes on Taobao (http://www.taobao.com), Tmall (http://www.tmall.com), and other websites are still difficult to eliminate, which is directly related to the lack of rigorous government supervision.

(i) The government's punishment for enterprises with immoral behaviors is too weak, leading to the existence of EIBs in the network market.

Different from traditional brick-and-mortar enterprises, e-commerce enterprises directly open stores in the network market, most of which are SMEs [44]. These e-commerce SMEs just need to meet the entry and selling rules and obtain the qualifications of the e-commerce platforms, while the offline enterprises must face the supervision of different authorities such as the Industry and Commerce Bureau and Tax Bureau and have to undertake

multiple costs such as choosing site and decoration of brick-and-mortar stores. Some e-commerce SMEs even have only four or five employees to operate online stores. If e-commerce SMEs are investigated and forced to leave the platform due to immoral behaviors, they can change their names or move to another platform and then continue to sell products, and some of them even continue to adopt IM. On March 15, 2014, the State Administration for Industry and Commerce of China issued the *Administrative Measures on Internet Transactions*, which stipulates that online store owners will face fine of CNY 10,000 to CNY 30,000 (approximately USD 1450 to USD 4350) for providing false information about their products or services. With such a low punishment, it is difficult to curb the greed of counterfeiters. In response to this problem, during the *Two Sessions* (*Two Sessions* are the collective name for the National People's Congress of the People's Republic of China and the Chinese People's Political Consultative Conference held since 1959. Since the two conferences are basically coincident and the importance of the operation of the country is very high, it is referred to as *Two Sessions*. The significance of *Two Sessions* is to collect and organize the information and requirements of *Two Sessions'* representatives from the people and convey them to the central government.) of China in 2019, the head of China Market Supervisory Administration said that the government would impose huge punishments on fakes to greatly increase the cost of counterfeiters (the report is at https://news.sina.com.cn/o/2019-03-05/doc-ihrfqzkc1428622.shtml).

Moreover, the illegality cost of e-commerce platforms selling fakes is also low, and platforms bear little joint liability; even some e-commerce enterprises are removed from the platform due to EIBs. In European countries, in the judicial decisions of similar cases, the e-commerce platforms are generally required to undertake joint liability. For example, it was reported that in 2008, a court in Paris, France, ruled that eBay infringed on the rights of the LVHM Group and is enforced to compensate the company with $61 million for selling fakes (the report is at http://news.winshang.com/html/044/2510.html). Besides, the e-commerce platforms are the most profitable enterprises, and even they sell fakes. Although some counterfeit information is transferred to the platform staff, they even cover up the immoral behaviors to ensure their own profits.

(ii) The low-degree government supervision on EIBs leads to inefficiency.

According to statistics, in 2016, Alibaba Platform Management Department used big data and manual reexamination methods to find out 4,495 enterprises that produce and sell fakes; their sales were far superior to the standard to be sentenced to jail (CNY 50,000) (the report is at http://b2b.toocle.com/detail--6432092.html). Among them, the authorities accepted 1,184 cases and only can punish 469 cases under the current regulations. There were only 33 cases that can be confirmed as criminal judgments through public information, accounting for less than 1% of cases that can be punished. In the cases that have been sentenced, 37 of 47 were suspended, with a ratio of 79%. Why the degree of government supervision of online shopping anti-counterfeiting is so low? In addition to the hysteresis of the current law, this is also related to the performance evaluation of GDP growth and employment. All kinds of EIBs can meet the consumption needs of migrant workers and new residents in the process of urbanization. Moreover, the employment rate is increased by low-quality jobs due to the simplicity of producing and selling fakes. According to the data released by Zhiyan Consulting, the number of employees of Chinese WeChat shops exceeded 20 million in 2017 (the report is at http://www.bbtnews.com.cn/2019/0531/304162.shtml). It is estimated that, by 2024, the number will reach 52 million. The massive participation has prompted the development of e-commerce enterprises and also has caused many problems, such as tax evasion, sales of fakes, and disclosure of personal information. Many enterprises that produce and sell fakes are fully qualified and relatively large in scale, and some are even protected by the local governments. For example, in 2018, more than 100 enterprises producing condiments in a town in Tianjin, China, were closed. According to the survey, these enterprises were fraudulent for more than ten years, and the local government left them alone until they were exposed by the media (the report is at http://www.xinhuanet.com/fortune/2018-08/29/c_1123346277.htm).

(iii) The cost of government supervision cannot be compensated, leading to a loss of sustained motivation for supervision.

On April 14, 2016, the "QSIQ 12365 Complaint and Report Advisory Network" was launched in China. Through this website, consumers can check the information of enterprise product qualification and can also consult, complain, and report to the regional quality control departments. At the same time, to facilitate consumers to check the information of enterprise product qualification, Alibaba and other four e-commerce platforms were required to sign a cooperation agreement as a way to achieve the docking of e-commerce platforms and quality control website (the report is at https://www.sohu.com/a/69427874_350719). However, until today, e-commerce platforms have not realized the function of querying the product qualification, and the quality control website has not been effectively developed and used, which indicates that quality

control in China has not achieved the expected results. On the one hand, this is due to the lack of clarity in the definition of responsibilities of government departments and the imperfect regulatory support system. On the other hand, the high cost and low return of government regulation also make the government have no incentive to regulate. Similarly, on January 1, 2019, the Chinese government introduced the E-Commerce Law, which explicitly brings EIBs such as data falsification into the scope of legal regulation. However, due to the high cost, government regulation is hardly effective as expected. In contrast, China's Antimonopoly Law was put in effect in 2008 and has been enforced for 12 years, with relevant administrative precedents amounting to CNY 12 billion. This makes the regulatory costs of government regulation compensable, thus making the regulation effective and sustainable. It is proved by the Chinese experience that government regulation can only be effective and sustainable if the costs of government supervision are compensated.

(iv) The lack of consumer rights and responsibilities contributes to the prevalence of counterfeit goods.

At present, customers hate fake foods and medicines but prefer low prices when purchasing other kinds of goods. In particular, consumers who are willing to buy fakes that involve intellectual property rights do not take it as a serious behavior. According to China's annual survey of Day for Consumers' Rights and Interests in 2019, less than 60% of consumers choose to protect their rights after buying fakes, and 16.11% of consumers continue to use fakes (the report is at https://www.thepaper.cn/newsDetail_forward_3141915). In the above survey, 33.59% of consumers believed that the low price is the main reason why counterfeit goods are difficult to ban.

In response to the increasingly serious e-commerce EIBs, the Chinese government has taken measures to increase the degree of government supervision and punishment to regulate the network market, which has already achieved preliminary success. Since October 2017, the National Development and Reform Commission has issued four blacklists for special management of untrustworthiness in the field of e-commerce, involving a total of 1092 enterprises (the report is at http://field.10jqka.com.cn/20181106/c607973048.shtml). On May 14, 2018, the National Development and Reform Commission, the Central Network Information Office, the Ministry of Industry and Information Technology, etc., a total of 8 departments, jointly issued the *Notice on Strengthening Special Governance Work on Untrustworthiness in the E-Commerce Sector* to crack down illegal and untrustworthy behaviors in the field of e-commerce (the report is at http://www.sohu.com/a/273707363_100019209). In June 2018, the General Administration of China's Market Supervision, the National Development and Reform Commission, the Ministry of Industry and Information Technology, etc., a total of 11 departments, jointly issued the *Notice on Printing and Distributing the 2018 Network Market Supervision Special Action (Net Sword Action) Program*, focusing on rectifying online infringement and false propaganda, false illegal advertising, etc. *The Electronic Commerce Law of the People's Republic of China*, which was officially implemented on January 1, 2019, regulates the rights and obligations of e-commerce enterprises and e-commerce platforms. In 2016, Alibaba provided a total of 4495 enterprises with EIBs whose sales exceeded CNY 50,000 to the national public security organizations. In 2017 and 2018, these data dropped to 1910 and 1634 (he reports are at http://www.cicn.com.cn/zggsb/2018-01/18/cms103803article.shtml and https://baijiahao.baidu.com/s?id=1622256698847441370&wfr=spider&for=pc).

Among other countries, due to the lack of government supervision in the network market, Russia can only rely on several large-scaled platforms to spontaneously fight against fakes. Group-IB data showed that, in 2017, Russia's online counterfeit transactions exceeded USD 1.5 billion, a 23% increase from 2016 (the blog can be retrieved from https://blog.csdn.net/mozhe_/article/details/83105588). The US government has strengthened the government regulation on the network market. On April 3, 2019, President Trump signed a memorandum aiming at cracking down on online counterfeit transactions. It is recommended to enhance government supervision or legislative changes to better combat the sale of fakes. (the report is at https://www.stopfakes.gov/article?id=Presidential-Memorandum-on-Combating-Trafficking-in-Counterfeit-and-Pirated-Goods-Opportunity-for-Stakeholder-Comments). Germany, whose network market operates in a good order, has imposed severe legal sanctions on the sale of fakes and ordered enterprises to return the fakes unconditionally within 14 days. Meanwhile, the blacklist of online shops selling fakes is established, and the public can expose online shops for selling fakes on the supervision website, making shops to be investigated and punished by the judicial agency.

6. Conclusion and Policy Suggestion

From the aforementioned model analysis and case discussion, it is found that, in many countries, especially in developing countries, inefficient government supervision is one of the critical factors for e-commerce enterprises adopting immoral behaviors. When e-commerce enterprises gain profits more through IM than through MO, EIBs appear. Only when government supervision and punishment are strict enough and punishment for the enterprise adopting IM is higher than the extra cost of government supervision, EIBs will be curbed. Besides, consumers' support is an essential factor in improving the efficiency of government supervision.

The governance of EIBs is a systematic project, and it is not enough to rely solely on government supervision. It is necessary to construct a new mechanism of social cogovernance of "enterprise initiative, industry self-regulation, social supervision, and government supervision." Note that the government's attitude and behavior about EIBs in the network market play an important role in guiding and

guaranteeing the entire governance system. The following managerial implications can be summarized:

(i) There is an urgent need to improve the government supervision system for the network market. EIBs more easily occur in the online shopping market than in the traditional market. In the case of insufficient government supervision, the EIBs of e-commerce enterprises can cause great damage to the network market. Only by establishing a strict system of government supervision and making the punishment of IM higher than the increased profit by IM, the positive competition of the network market can be maintained.

(ii) Focus on supervising the EIBs of large enterprises. Large enterprises are the leader of the network market, whose probability of being found when adopting IM is much larger than that of SMEs. Government supervision on large enterprises can reduce the supervision cost and achieve effective supervision. Moreover, if large enterprises adopt MO and SMEs adopt IM, the living space of SMEs can be further compressed. However, there are a large number of SMEs in China, which make China's government supervision more difficult. This is also the subject of further research.

(iii) Encourage consumers to actively participate in the supervision of the network market. Consumers' support is of significance to reduce fakes, so protection laws and regulations for consumers should be improved to guide consumers to establish correct consumption concepts and consumption responsibilities. For example, India's newly revised *Consumer Protection Act* approved the establishment of the Central Consumer Protection Authority, which can authorize a centralized hotline for consumers to report any e-commerce fraudulent transaction. At present, less than 50% of countries have established online consumer protection laws.

(iv) Accelerate the establishment of the enterprise credit system, which can limit the behaviors of immoral enterprises. Establish a credit evaluation system for e-commerce platforms with department linkage, data sharing, and enterprise participation. Moreover, third-party organizations are encouraged to publish professional credit ratings for e-commerce enterprises and e-commerce platforms.

This paper finds that consumer participation in the regulation of network markets is crucial. In addition to this, the government should encourage the public, media, nongovernmental organizations, and other stakeholders to participate in EIB regulation. However, how to establish an effective hybrid governance system combining government supervision and informal regulation has not been explored in this paper, and this is the direction of our future research.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

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

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (no. 71971129), Shandong Province Higher Education Youth Innovation and Technology Support Program (no. 2019RWG017), and Shandong Provincial Natural Science Foundation (ZR201911160259). The authors would like to thank Dr. Rongyun Tang of the University of Tennessee for providing grammatical and syntactic corrections to the paper.

References

[1] W.-y. Kevin Chiang and G. E. Monahan, "Managing inventories in a two-echelon dual-channel supply chain," *European Journal of Operational Research*, vol. 162, no. 2, pp. 325–341, 2005.

[2] Z. Zhang, S. Liu, and B. Niu, "Coordination mechanism of dual-channel closed-loop supply chains considering product quality and return," *Journal of Cleaner Production*, vol. 248, Article ID 119273, 2020.

[3] D. Palmer, "Extending the process model of collective corruption," *Research in Organizational Behavior*, vol. 28, pp. 107–135, 2008.

[4] Y. Komarova Loureiro, K. L. Haws, and W. O. Bearden, "Businesses beware," *Journal of Service Research*, vol. 21, no. 2, pp. 184–200, 2018.

[5] Y. Komarova Loureiro, J. Bayuk, S. M. Tignor, G. Y. Nenkov, S. Baskentli, and D. Webb, "The case for moral consumption: examining and expanding the domain of moral behavior to promote individual and collective well-being," *Journal of Public Policy & Marketing*, vol. 35, no. 2, pp. 305–322, 2016.

[6] J. R. McColl-Kennedy, B. A. Sparks, and D. T. Nguyen, "Customer's angry voice: targeting employees or the organization?" *Journal of Business Research*, vol. 64, no. 7, pp. 707–713, 2011.

[7] M. Safaei and K. D. Thoben, "Measuring and evaluating of the network type impact on time uncertainty in the supply networks with three nodes," *Measurement*, vol. 56, pp. 121–127, 2014.

[8] P. A. Pavlou, H. Liang, and Y. Xue, "Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective," *MIS Quarterly*, vol. 31, no. 1, pp. 105–136, 2007.

[9] M. Rothschild and J. Stiglitz, "Equilibrium in competitive insurance markets: an essay on the economics of imperfect information," *The Quarterly Journal of Economics*, vol. 90, no. 4, pp. 629–649, 1976.

[10] D. Peterson, "Perceived leader integrity and ethical intentions of subordinates," *Leadership & Organization Development Journal*, vol. 25, no. 1, pp. 7–23, 2004.

[11] G. Labianca, J. F. Fairbank, G. Andrevski, and M. Parzen, "Striving toward the future: aspiration-performance

discrepancies and planned organizational change," *Strategic Organization*, vol. 7, no. 4, pp. 433–466, 2009.

[12] F. Gino, J. Gu, and C.-B. Zhong, "Contagion or restitution? When bad apples can motivate ethical behavior," *Journal of Experimental Social Psychology*, vol. 45, no. 6, pp. 1299–1302, 2009.

[13] S. Otim and V. Grover, "E-commerce: a brand name's curse," *Electronic Markets*, vol. 20, no. 2, pp. 147–160, 2010.

[14] P.-L. To, C. Liao, and T.-H. Lin, "Shopping motivations on Internet: a study based on utilitarian and hedonic value," *Technovation*, vol. 27, no. 12, pp. 774–787, 2007.

[15] Y. Wang, J. Yang, and L. Qi, "A game-theoretic model for the role of reputation feedback systems in peer-to-peer commerce," *International Journal of Production Economics*, vol. 191, pp. 178–193, 2017.

[16] C. Oliver, "Strategic responses to institutional processes," *Academy of Management Review*, vol. 16, no. 1, pp. 145–179, 1991.

[17] J. J. Kish-Gephart, D. A. Harrison, and L. K. Treviño, "Bad apples, bad cases, and bad barrels: meta-analytic evidence about sources of unethical decisions at work," *Journal of Applied Psychology*, vol. 95, no. 1, pp. 1–31, 2010.

[18] L. K. Treviño, G. R. Weaver, and S. J. Reynolds, "Behavioral ethics in organizations: a review," *Journal of Management*, vol. 32, no. 6, pp. 951–990, 2006.

[19] M. Faccio, "Politically connected firms," *American Economic Review*, vol. 96, no. 1, pp. 369–386, 2006.

[20] J.-J. Laffont and D. Martimort, "The design of transnational public good mechanisms for developing countries," *Journal of Public Economics*, vol. 89, no. 2-3, pp. 159–196, 2005.

[21] T. Mavlanova, R. Benbunan-Fich, and G. Lang, "The role of external and internal signals in E-commerce," *Decision Support Systems*, vol. 87, pp. 59–68, 2016.

[22] D. D. Bergh, B. L. Connelly, D. J. Ketchen, and L. M. Shannon, "Signalling theory and equilibrium in strategic management research: an assessment and a research agenda," *Journal of Management Studies*, vol. 51, no. 8, pp. 1334–1360, 2014.

[23] S. Román and P. J. Cuestas, "The perceptions of consumers regarding online retailers' ethics and their relationship with consumers' general internet expertise and word of mouth: a preliminary analysis," *Journal of Business Ethics*, vol. 83, no. 4, pp. 641–656, 2008.

[24] D. Walsh, J. M. Parisi, and K. Passerini, "Privacy as a right or as a commodity in the online world: the limits of regulatory reform and self-regulation," *Electronic Commerce Research*, vol. 17, no. 2, pp. 185–203, 2017.

[25] C. Mutimukwe, E. Kolkowska, and Å. Grönlund, "Information privacy in e-service: effect of organizational privacy assurances on individual privacy concerns, perceptions, trust and self-disclosure behavior," *Government Information Quarterly*, vol. 37, no. 1, Article ID 101413, 2020.

[26] I.-D. Anic, V. Škare, and I. Kursan Milaković, "The determinants and effects of online privacy concerns in the context of e-commerce," *Electronic Commerce Research and Applications*, vol. 36, Article ID 100868, 2019.

[27] C.-H. Wu, Z. Yan, S.-B. Tsai, W. Wang, B. Cao, and X. Li, "An empirical study on sales performance effect and pricing strategy for E-commerce: from the perspective of mobile information," *Mobile Information Systems*, vol. 2020, Article ID 7561807, 8 pages, 2020.

[28] M. M. Calkins, A. Nikitkov, and V. Richardson, "Mineshafts on treasure island: a relief map of the eBay fraud landscape," *Pittsburgh Journal of Technology Law and Policy*, vol. 8, 2008.

[29] A. G. González, "PayPal and eBay: the legal implications of the C2C electronic commerce model," in *Proceedings of the The 18th BILETA Conference: Controlling Information in the Online Environment*, London, UK, April 2003.

[30] E. L. Glaeser and A. Shleifer, "A reason for quantity regulation," *American Economic Review*, vol. 91, no. 2, pp. 431–435, 2001.

[31] J. E. Oxley and B. Yeung, "E-commerce readiness: institutional environment and international competitiveness," *Journal of International Business Studies*, vol. 32, no. 4, pp. 705–723, 2001.

[32] M. G. Martinsons, "Electronic commerce in China: emerging success stories," *Information & Management*, vol. 39, no. 7, pp. 571–579, 2002.

[33] L. E. Ribstein and B. H. Kobayashi, "State regulation of electronic commerce," *SSRN Electronic Journal*, vol. 51, no. 1, pp. 1–31, 2002.

[34] R. AlGhamdi and S. Drew, "Seven key drivers to online retailing in KSA," in *Proceedings of the IADIS International Conference on e-Society*, Avila, Spain, 2012.

[35] A. Agrawal, C. Catalini, and A. Goldfarb, "Some simple economics of crowdfunding," *Innovation Policy and the Economy*, vol. 14, no. 1, pp. 63–97, 2014.

[36] H. Kim, "Globalization and regulatory change: the interplay of laws and technologies in E-commerce in Southeast Asia," *Computer Law & Security Review*, vol. 35, no. 5, Article ID 105315, 2019.

[37] P. Weiser, "The future of internet regulation," *University of California Davis Law Review*, vol. 43, pp. 529–590, 2009.

[38] C. T. Marsden, "Beyond Europe: the Internet, regulation, and multistakeholder governance-Representing the consumer interest?" *Journal of Consumer Policy*, vol. 31, no. 1, pp. 115–132, 2008.

[39] J. Hofbauer, P. Schuster, and K. Sigmund, "A note on evolutionary stable strategies and game dynamics," *Journal of Theoretical Biology*, vol. 81, no. 3, pp. 609–612, 1979.

[40] L. Shen and Y. Wang, "Supervision mechanism for pollution behavior of Chinese enterprises based on haze governance," *Journal of Cleaner Production*, vol. 197, pp. 571–582, 2018.

[41] J. W. Weibull, *Evolutionary Game Theory*, Princeton Press, Princeton, NJ, USA, 1995.

[42] D. Friedman, "Evolutionary games in economics," *Econometrica*, vol. 59, no. 3, pp. 637–666, 1991.

[43] E. J. Balleisen, "Extracts from government and markets: toward a new theory of regulation," in *The Prospects for Effective Coregulation in the United States: A Historian's View for the Early Twenty-First Century*, E. J. Balleisen and A. David, Eds., pp. 450–453, Cambridge University Press, New York, NY, USA, 2010.

[44] R. Ramanathan, U. Ramanathan, and H.-L. Hsiao, "The impact of e-commerce on taiwanese SMEs: marketing and operations effects," *International Journal of Production Economics*, vol. 140, no. 2, pp. 934–943, 2012.

[45] D. J. Brass, K. D. Butterfield, and B. C. Skaggs, "Relationships and unethical behavior: a social network perspective," *Academy of Management Review*, vol. 23, no. 1, pp. 14–31, 1998.

[46] C. Chen, J. Zhang, and T. Delaurentis, "Quality control in food supply chain management: an analytical model and case study of the adulterated milk incident in China," *International Journal of Production Economics*, vol. 152, pp. 188–199, 2014.

[47] D. Rodrigues, R. Teixeira, and J. Shockley, "Inspection agency monitoring of food safety in an emerging economy: a multilevel analysis of Brazil's beef production industry," *International Journal of Production Economics*, vol. 214, pp. 1–16, 2019.

Hindawi
Complexity
Volume 2021, Article ID 5532672, 15 pages
https://doi.org/10.1155/2021/5532672

Research Article

Dynamic Stochastic Optimization of Emergent Blood Collection and Distribution from Supply Chain Perspective

Xiangyu Jin,[1,2] **Huajun Tang**,[2] **and Yuxin Huang**[2]

[1]*School of Logistics Management and Engineering, Zhuhai College of Science and Technology, Zhuhai, China*
[2]*School of Business, Macau University of Science and Technology, Macau, China*

Correspondence should be addressed to Huajun Tang; hjtang@must.edu.mo

Received 1 March 2021; Revised 5 April 2021; Accepted 12 April 2021; Published 24 April 2021

Academic Editor: Baogui Xin

Copyright © 2021 Xiangyu Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In response to emergencies, it is critical to investigate how to deliver emergency supplies efficiently and securely to disaster-affected areas and people. There is no doubt that blood is deemed one of the vital relief supplies, and ensuring smooth blood delivery may substantially alleviate subsequent impacts caused by the disaster. Taking red blood cell products as the research object, this work proposes a four-echelon blood supply chain model. Specifically, it includes blood donors, blood donation houses, blood centres, and hospitals. Furthermore, numerical analysis is provided to test the feasibility of blood collection and distribution schemes and conduct sensitivity analysis to test the impacts of the relevant parameters (e.g., apheresis donation proportion of red blood cells (RBCs), distance between blood donors and blood facilities, and times of blood donation) on the scheme. This research provides some scientific and reasonable support for decision makers and managerial implications for emergency departments and contributes to the study of emergent blood supply chain.

1. Introduction

As multidimensional conflicts in politics, economy, and environment become increasingly evident, the world has witnessed an increase in the number of natural and man-made disasters. For instance, the terrorist attack in New York City on 11 September 2001 shocked the entire world and inspired other countries to pay more attention on emergency response. This was followed by a number of similar events, such as Wenchuan Earthquake in China in 2008, the earthquake and tsunami in Japan in 2011, the Nice terrorist attack in France in 2016, and mountain fire in Greece in 2018. These occurrences cause considerable human casualties and economic losses and impose dramatic negative impacts on people's psychological states and lives. In this context, the response to such complex emergent events requires a normative response-guaranteed system. As one of the critical materials for emergency rescue, the delayed supply of blood in an emergency poses threats to people's lives and health, whereas an excessive supply leads to waste and even subsequent "blood famine." This suggests the importance of incorporating a timely, adequate, and low-cost blood supply into emergency response and management decisions.

The study of emergent blood supply is one hot branch in the field of emergent management. There is an increasing number of literatures in emergent blood supply, most of which mainly focused on the downstream blood supply chain [1] and less of which investigated the blood collection and distribution from the perspective of the supply chain [2].

The aim of this study is to explore the optimal operations of a multinode and four-echelon blood supply chain in emergency, based on the status quo of emergency rescue in China. Specifically, given the low hazard level of emergencies, a four-echelon blood supply chain is established, including blood donors, blood donation houses, blood centres, and hospitals. Through numerical analysis, an optimal blood collection and distribution scheme is acquired for the blood supply chain. Sensitivity of the parameters (e.g., proportion of RBC apheresis donation, average volume of blood supply, number of blood donation, and distance

between blood donors and blood facilities) is tested to explore the impacts on the optimal schemes.

The contribution of this work is to enrich the study of emergent blood collection and distribution in the field of emergent management and provide more scientific support for practical decision makers.

The remainder of this study is organized as follows. Section 2 summarizes relevant literatures on material supply and blood supply chain operations in response to emergencies. With respect to these literatures, it highlights that this work expands the previous research model of blood supply chain in several aspects, such as the selected targets, supply chain layers, dynamic stochastic demand, and blood collection and distribution models. In Section 3, a self-collection model is established for the blood supply chain in response to emergencies. Section 4 conducts the numerical analysis so as to validate the feasibility of the proposed model. Finally, Section 5 concludes this study and provides some managerial implications.

2. Literature Review

Previous studies on emergency management can have two main categories with respect to the evacuation direction, including victim evacuation [3–5] and supply of emergency relief materials. Blood product is one type of emergency relief. In the following, it reviews the aspects of general material supply and blood product supply, respectively.

2.1. General Material Supply in Emergencies. Materials scheduling during emergencies covers emergency relief material distribution and transhipment routing. A multi-period linear programming model was proposed for food aid layout optimization in Africa [6]. It aimed at minimising food transhipment and inventory costs. Goods delivery and victim evacuation were investigated under the circumstance that a limited number of transporting resources could be distributed to multiple destinations from diverse places [7]. This research involved three models with the goal of reducing the late deliveries. The earthquake in Turkey was taken for example to explore how the system cost could be minimised [8]. In the case where the supplied materials were limited and the demand for the material was known, they put forward an emergency supplies scheduling problem with a limited time window in emergency. Özdamar et al. [9] proposed a different model from traditional vehicle routing problems. Instead of returning vehicles that had completed their tasks to the place of departure and enabling them to wait there, the vehicles needed to wait for the next delivery command where they are. By combining network flow problems with vehicle routing issues, this research reduced the delay in scheduling emergency supplies to the greatest extent. Apart from that, a heuristic algorithm based on Lagrangian relaxation was adopted to solve the problem. To settle event response- and resource allocation-related issues in traffic accident management, a mathematical programming model with a probabilistic bound was developed. Yi and Özdamarb [10] was intended to investigate scheduling optimization problems in flood emergency rescue and elaborate and establish a flood emergency rescue programming framework, including demand, supply, inventory, and resource management. According to this framework, vehicles are considered as materials rather than variables. In subsequent studies, a heuristic algorithm of ant colony optimization was put forward to solve logistics problems in disaster-relief activities [11]. It decomposed the original emergency logistics into two stages. The first stage was to determine stochastic vehicle routing under the guidance of information elements, and in the second stage, the distribution of different vehicle types and various commodities were determined on the basis of network flow methods. Besides, a dynamic demand management model was proposed for emergency logistics operations under the circumstance that major natural disaster information conditions were incomplete [12]. Through fuzzy clustering, the disaster-affected area was divided into diverse groups. Then, the existing data were utilised to predict disaster-relief demands in different groups, and these demands were adopted as the basis of emergency supplies distribution. Yan et al. [13] determined that emergency supplies scheduling was correlated to disaster-affected road repair. For this reason, they selected seismic data from Taiwan in 1999 and took the randomness of vehicles' travel time into account. On this basis, a comprehensive optimization model was constructed for emergency supplies scheduling and disaster-affected road repair under uncertainty. It was expected that the cost incurred by accidental losses of after-calamity logistical support can be minimised.

Overall, emergency management is now proved challenging. Compared to conventional event management, emergency management needs to be studied further. The research on emergency management focuses on evacuation and supplies scheduling. In terms of emergency supplies scheduling, the existing literatures are primarily targeted at emergency scheduling of bulk supplies, but rarely related to perishable products, especially on blood products. Most of them did not take the characteristics of blood into account.

2.2. Blood Supply in Emergencies. Blood products fall into the category of important emergency relief supplies. This section presents a literature review of the studies associated with blood products. Blood supply chain includes blood collection, blood inspection and production, blood storage, and distributing the blood to those in need in a safe and timely manner. Previous studies focused on some of the following stages.

2.2.1. Collection. Blood collection, the first step in the whole blood supply chain operations, is one of the most fundamental tasks for mobile blood donation vehicles, blood donation houses, central blood stations, and blood centres. Red blood cells can be collected in two forms. One is whole blood collection and the other is apheresis collection, both of which are optional for donors.

Bosnes et al. [14] employed a logistic regression model to predict the arrival of blood donors in the blood bank. The

multiobjective random integer linear programming model was used to determine an appropriate combined strategy of blood collection [15]. In order to minimise the total cost and the number of donors required, the model took factors such as the compatibility of blood types and the availability of blood types and donors into consideration. Concerning blood donation appointment scheduling, Seda et al. [16] utilised mixed-integer linear programming (MILP) to pre-assign blood collection periods for different blood types with the expectation of striking a balance between collection and production of different blood types on various dates. On this basis, they provided a steady blood supply system. Williams and Masser [17] evaluated the blood donation motivation of 458 qualified donors and concluded that motivation negatively impacts willingness to donate blood. Goette and Stutzer [18] carried out a large-scale field test for 3 months and designed a follow-up period of 15 months. In this way, they conducted preliminary experimental research on the influence of material incentives on blood donation. It was found that distributing rewards (e.g., holding lotteries) play a positive role in elevating the blood donation rate. Such an effect was driven by less enthusiastic donors.

According to the literatures above, research studies associated with blood collection mainly focused on the behaviour and motivations of donors, the estimation of donors' arrival, and donation appointments. In this work, blood collection and operation schemes are taken into consideration although they are seldom investigated in the existing studies.

2.2.2. Inventory Management. Most of the existing literature on blood inventory management focused on macro qualitative issues, whereas this research offers a summary and analysis of previous papers combining both qualitative and quantitative approaches.

By selecting RBC product inventory/disposal databases from provincial hospitals as the research object, Heddle et al. [19] made a logistic regression analysis of RBC inventory-disposal data over 21 months in 156 hospitals. This aimed at clarifying the factors that affected RBC inventory updates and systematically confirming an optimal target level of RBC inventory updates. Dillon et al. [1] suggested a two-stage stochastic programming model that minimised cost and expiration quantity and considered vulnerability and demand uncertainty. Moreover, a blood supply chain was constructed based on a blood ordering and collection model according to two emergency replenishment strategies [20]. These strategies considered heterogeneous requirements of blood and analysed the impacts of different emergency replenishment strategies on results.

2.2.3. Blood Distribution and Scheduling. This stage includes the distribution, transportation, and transhipment of blood products. Perishable product (i.e., blood) distribution strategies were analysed for different places, such as a regional centre, proposing two general policies based on optimal approximation [21]. To be specific, the above two policies were comprised of a rotation strategy and a retention strategy. Based on the existing distribution system of the Red Crescent blood service centre in Turkey, Sahin et al. [22] considered the influence of decision making about position on the performance of blood centres, stations, and moving units and established a mathematical model to solve decision making about central-location-based allocation concerning issues of blood supply.

Speaking of transportation, apart from an allocation-routing programming model (IAR) built based on multiple vehicles, multiple supply sites, and multiple objectives, an analytic hierarchy process and MILP were selected to settle relevant problems [23]. In addition, an integer programming model was constructed to identify the number of vehicles needed to implement blood collection operations and to minimise vehicles' travel distance [24]. The aforementioned model also took the uncertainty of blood donation vehicle visits into consideration and determined an optimal route by virtue of the CPLEX solver and the branch-and-price algorithm. Since 2017, the number of articles about blood transportation route programming in emergency relief has been in decline. In general, these topics were investigated in combination with other stages, such as blood collection, distribution, and inventory.

As for transhipment, Wang and Ma [25] achieved an effective reduction in the blood system expiration rate by establishing a transport model based on inventory storage time. Dehghani and Abbasi [26] created a transport strategy based on the longest storage time for interhospital blood transportation to reduce inventory scrap cost.

2.2.4. Blood Supply Chain. In the last decade, there is increasing research on blood products, especially on blood product operations in the supply chain [2].

For instance, Sha and Huang [27], taking Beijing as the background, developed an emergency blood supply-scheduling model, covering "donors, temporary blood donation facilities, and blood centres." Fahimnia et al. [28] built a four-echelon (i.e., blood donors, mobile blood donation sites, blood centres, and demand sites) blood supply chain model to minimise cost and delivery time. Ramezanian and Behboodi [29] introduced a deterministic location model based on MILP. Involving blood donors, mobile blood donation facilities, and blood centres, the model considered parameters such as the distance between blood donors and blood facilities and the advertising budget of blood facilities, which were used to constitute the utility function to improve usage and motivate blood donors. To minimise the total cost of collection, inventory, and production, Özener et al. [30] constructed a blood supply chain model with donors and central blood stations as nodes. By considering three different blood products, the model made plans for donors to meet the demand for blood products within a given range. Hamdan and Diabat [31] proposed a blood supply chain model based on robust optimization and two-stage stochastic optimization to minimise the impact of disasters on the blood supply chain. A dual-objective framework of simultaneous minimisation system was adopted to reduce the cost and delivery time during possible

interruptions of the supply chain. According to the MILP model used by Samani et al. [32], various blood collection and separation methods and disruption scenarios were taken into consideration to make inventory and distribution decisions about the blood supply chain.

Literatures on blood supply chain operations with at least three stages during the last decade are listed in Tables 1 to 3.

As can be observed from Table 1, previous research on blood supply chain with at least three echelons is rather rare in the past 10 years. Therefore, this is still a new research area and direction in this aspect. Blood supply chain studies involving all nodes at four levels and above are even fewer [33–38]. This work targets at a blood supply chain including donors, the blood donation house, the blood centre, and the hospital. Clearly, it is concerned with the entire blood supply chain.

In terms of objective functions, all studies incorporate the cost, while some of the literatures consider objective functions of blood shortage penalty and transhipment time. Few literatures added weight coefficients in different objectives [39, 40]. In this study, the objective functions include the composite cost and the shortage penalty cost, both of which should be minimised; additionally, weight coefficients indicate the importance of different objectives. In the context of emergency, life safety is considered as the primary task in this work. Therefore, the cost incurred by expired products is excluded from the objective functions. Although the time factor is not counted as one of the objective functions, a rigid constraint for transhipment time is adopted as a constraint condition to ensure its urgency.

The present research is aimed at a four-echelon blood supply chain. Research on the operation of four-level blood supply chain can be traced back to 2017. By comparing different factors listed in Table 3, it is clear that almost all the literature on the four-level blood supply chain leaves various blood collection methods out of consideration. The research only refers to a simple method of whole blood collection. In this study, blood apheresis collection was also considered to explore how their variations affect results achieved based on a specific scheme. Compared with previous studies where blood donation houses were treated as nonadjustable fixed facilities, they are taken as adjustable candidate collection sites in this study and can be open or closed according to the actual situation. In the context of emergencies, this study also investigates how outcomes are affected by proportion of blood apheresis collection, the distance between blood donors and blood facilities, the average of the blood supply, and variations in the number of times of donation. A small amount of the existing literature involves sensitivity analysis but still attaches importance to the trade-off between objectives, the capacity of transport vehicles, and the variation of weight coefficients. This signifies that sensitivity analysis of the actual situation is still incomprehensive.

Overall, previous studies mainly focused on a single stage and rarely adopted the holistic perspective of supply chain. In this study, a blood supply chain model under emergencies is constructed, and the corresponding innovation can be summarized as follows.

(1) Most emergency relief supplies are materials in a broad sense and perishable products. Blood products are seldom investigated. However, this work is aimed at blood products and focused on their RBC products.

(2) From the perspective of the supply chain, studies on blood products turn out to be a single node or a supply chain of no more than three levels in most cases. The proposed model in this study consists of four stages. This expands the coverage of blood supply chain research and describes the status quo of blood supply chain operations.

(3) The present study takes optimization objectives into consideration comprehensively. To be specific, both the shortage cost and other composite operating costs are incorporated into the proposed model. Additionally, weight coefficients are introduced as the objective function to clarify importance of diverse objectives.

(4) Sensitivity analysis is conducted through altering the proportion of apheresis donation, times of blood donation, distance between blood donors and blood facilities and the average of the blood supply, etc. In this way, how parameters affect outcomes generated by the corresponding scheme can be explored.

(5) Compared with previous studies where blood donation houses were treated as nonadjustable fixed facilities, they are taken as adjustable candidate collection sites which can be open or closed according to the actual situation in this study.

3. Problem Description and Model Formulation

3.1. Problem Statement. Blood supply chain can be diverse in different countries or regions. In the following, China is taken as an example. A system of unified blood collection and supply institution planning and construction is implemented. Major units of blood collection, storage and inspection, and primary blood-use organisations are presented here.

(1) Blood centres are set up in provinces and municipalities directly under the Central Government, and central blood stations are set up in prefecture-level cities. As the two share similar functions, they are not distinguished in this work and are referred to collectively as "blood centres."

(2) Blood donation houses in each city are used for decentralised collection and preliminary testing of blood.

(3) As blood-use institutions, hospitals receive blood daily from the blood centres for clinical transfusion.

Including four stages of "blood donor-blood donation house-blood centre-hospital," a blood supply

Table 1: Supply chain echelon in some key literature.

Literature	Number of stages	Node				
		Donor groups	Blood donation houses	Blood donation vehicles	Blood centres	Demand points
Sha and Huang [27]		√	√	—	√	—
Jabbarzadeh et al. [33]		√	√	—	√	—
Arvan et al. [34]		√	—	—	√	√
Ramezanian et al. [29]		√	√	—	√	—
Osorio et al. [15]	3	√	—	—	√	√
Hamdan and Diabat [31]		—	—	√	√	√
Haeri et al. [35]		—	√	—	√	√
Liu et al. [36]		—	√	—	√	√
Wang and Chen [37]		—	√	—	√	√
Fahimnia et al. [28]		√	—	√	√	√
Zahiri and Pishvaee [38]		√	√	—	√	√
Attari and Jami [39]	4	√	—	√	√	√
Samani et al. [40]		√	√	√	√	√
Samani et al. [32]		√	√	—	√	√
This work	4	√	√	—	√	√

Note. Both blood donation houses and vehicles are designed for blood collection and preliminary blood tests only. They are deemed to be at the same level in this study.

Table 2: Objective functions in some key literature.

Literature	Number of stages	Objective functions				
		Cost	Shortage	Expiry	Time	Weighting
Sha and Huang [27]		√	√	—	—	—
Jabbarzadeh et al. [33]		√	—	—	—	—
Arvan et al. [34]		√	—	—	√	—
Ramezanian et al. [29]		√	√	—	—	—
Osorio et al. [15]	3	√	—	—	—	—
Hamdan and Diabat [31]		√	—	√	√	—
Haeri et al. [35]		√	—	—	—	—
Liu et al. [36]		√	√	—	—	—
Wang and Chen [37]		√	√	—	—	—
Fahimnia et al. [28]		√	—	—	√	—
Zahiri and Pishvaee [38]		√	√	—	—	√
Attari and Jami [39]	4	√	—	—	√	—
Samani et al. [40]		√	—	—	—	—
Samani et al. [32]		√	√	√	—	—
This work	4	√	√	—	√	√

Table 3: Factors considered in some key literature.

Literature	Sensitivity analysis	Different collection methods	Blood donation houses open or closed	Context of emergencies
Fahimnia et al. [28]	√	—	—	√
Zahiri and Pishvaee [38]	√	—	—	—
Attari and Jami [39]	—	—	—	—
Samani et al. [40]	—	—	—	√
Samani et al. [32]	—	√	—	√
This study	√	√	√	√

chain model is established to minimise comprehensive costs such as shortage cost of blood centre and opening cost of blood donation houses and so on. The model is expected to make the following decisions including

(1) Location of blood donation.
(2) Quantity of blood donated by donors at each facility.
(3) When and which blood donation house should be open.
(4) The quantity of blood assigned from blood centres to hospitals in the affected area. In addition, it also tests the impacts of the proportion of apheresis donation, times of blood donation, distance between blood donors and blood facilities, and the average of the blood supply chain on the decisions and the objective, respectively.

3.2. Model Assumptions

(1) Hospitals are assumed to be incapable of storing blood, and all blood for clinical use in these hospitals needs to be deployed by a blood centre.
(2) The location of blood donation houses is fixed, and the blood donation houses are in charge of transporting blood to blood centres in phases.
(3) Averages of both blood supply and demand are known in the affected area, and supply and demand within the research period both conform to Poisson distribution [25].
(4) A person can only donate one unit of blood (i.e., 200 ml of blood and referred to as 1U below) at a time.
(5) The blood inspection failure rate can be ignored.
(6) Automobiles within the affected area are assumed to be the vehicles for blood delivery.

3.3. Definition of Parameters and Variables. The sets, parameters, and variables involved in the model are defined as follows.

(1) Definition of sets

I: the set of donors, $i \in I$
J: the set of candidate sites for blood donation houses, $j \in J$
K: the set of blood centres, $k \in K$
H: the set of hospitals, $h \in H$
T: the set of rescue periods (one day is a period in this study), $t \in T$

(2) Definition of parameters

q: number of blood donation houses
U: penalty cost of shortage
OJ: unit collection and preliminary test cost of whole blood collection
OJ': unit collection and preliminary test cost of apheresis collection
OL: unit in-depth test and production cost of whole blood collection
OL': unit in-depth test and production cost of apheresis collection
TJ: unit transportation cost of vehicle
WH: unit blood inventory maintenance cost
D_{ht}: hospital demand in hospital h at time t
n: limit of blood donation times
b: maximum storage capacity of blood donation houses
cb: maximum storage capacity of blood centres
S_{it}: the maximum number of donor group i at time t
β: proportion of apheresis blood donation to total blood donation
A_{ij}: distance from donor group i to blood donation house j
A_{ik}: distance from donor group i to blood centre k
A_{jk}: distance from blood donation house j to blood centre k
A_{kh}: distance from blood centre k to hospital h
AC: maximum distance acceptable to the donor
LT_1: time requirement of blood delivery from blood centre to hospital
V_c: vehicle transport speed
N: large value
Rather than the individual behaviour of blood donation, this study focuses on the homogeneity of group blood donation after emergency, as such, the group is taken as the research object.

(3) Definition of variables

P_{jt}: 0-1 variable; it is 1 if the blood donation house j is open, and 0 if not.
Y_{ijt}: 0-1 variable; it is 1 if the donor group i goes to blood donation house j at time t for donation, and 0 if not.
Y_{ikt}: 0-1 variable; it is 1 if the donor group i goes to blood centre k at time t for donation, and 0 if not.
Y_{ijkt}: 0-1 variable; it is 1 if the donor group i goes to blood donation house j at time t for donation and delivers blood to blood centre k, and 0 if not.
Y_{kht}0-1 variable; it is 1 if blood centre k delivers blood to hospital h at time t, and 0 if not.
Q_{ijkt}: the amount of blood delivered to blood centre k at time t, donated by donor group i in blood donation hospital j.
Q_{ikt}: the amount of blood donated by donor group i at time t to blood centre k.
Q_{kht}: the amount of blood delivered from blood centre k to hospital h at time t.
IB_{kt}: the beginning inventory quantity of blood centre k at time t.

3.4. Model Formulation. The operation process of the blood supply chain model in this study is shown in Figure 1.

3.4.1. Objective Functions. In view of the role that blood products play in emergency management, the shortage of

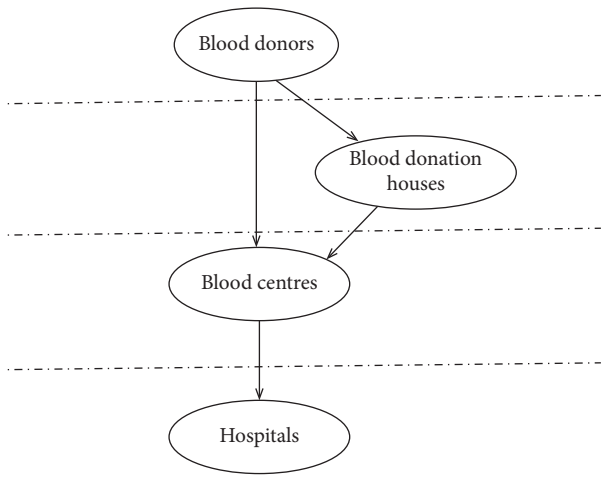


Figure 1: Operation chart of blood supply chain.

blood supply chain should be minimised. The shortage cost is shown in the following equation:

$$Z_1 = U * \sum_{h=1}^{H} \sum_{t=1}^{T} \left(D_{ht} - \sum_{k=1}^{K} Q_{kht} \right), \tag{1}$$

where U is the penalty cost of shortage; D_{ht} is the demand of hospital h at time t; Q_{kht} is the amount of blood delivered from blood centre k to hospital h at time; and $\sum_{h=1}^{H} \sum_{t=1}^{T} (D_{ht} - \sum_{k=1}^{K} Q_{kht})$ represents the amount of shortage of blood supply chain.

Moreover, blood supply chain operation should pay attention to cost control. The cost target is shown in the following equation:

$$Z_2 = \mathrm{PC} + \mathrm{OC} + \mathrm{TC} + \mathrm{IC}, \tag{2}$$

where PC is the opening cost of blood donation house; OC is the operating cost of blood collection and test; TC is the transportation cost; X is the upward rounding of value X in the symbol; and IC is the storage cost, and the formula is as follows.

$$\begin{aligned}
\mathrm{PC} &= \mathrm{OP} * \sum_{j=1}^{J} \sum_{t=1}^{T} P_{jt}, \mathrm{OC} = \frac{2\beta}{1+\beta} (OJ' + OL') \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{t=1}^{T} Q_{ijkt} + \sum_{i=1}^{I} \sum_{k=1}^{K} \sum_{t=1}^{T} Q_{ikt} \right) \\
&+ \frac{1-\beta}{1+\beta} (OJ + OL) \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{t=1}^{T} Q_{ijkt} + \sum_{i=1}^{I} \sum_{k=1}^{K} \sum_{t=1}^{T} Q_{ikt} \right), \\
\mathrm{TC} &= 2 * TJ \left[\sum_{k=1}^{K} \sum_{j=1}^{J} A_{jk} \left(\sum_{t=1}^{T} \frac{\sum_{i=1}^{I} Q_{ijkt}}{100} \right) + \sum_{k=1}^{K} \sum_{h=1}^{H} A_{kh} \left(\sum_{t=1}^{T} \frac{Q_{kht}}{100} \right) \right], \\
\mathrm{IC} &= WH \sum_{k=1}^{K} \sum_{t=1}^{T} \left(IB_{kt} - \sum_{h=1}^{H} Q_{kht} \right).
\end{aligned} \tag{3}$$

The overall objective function is shown in equation (4), where w_1 and w_2 are, respectively, the weight coefficients corresponding to the shortage cost target and other cost target, representing the weight emphasis of comprehensive measurement of each target.

$$\min Z = w_1 Z_1 + w_2 Z_2. \tag{4}$$

3.4.2. Model Constraints

$$\sum_{k=1}^{K} \left(\sum_{j=1}^{J} Q_{ijkt} + Q_{ikt} \right) \le 1 + \beta S_{it}, \quad \forall i \in I \, \forall t \in T, \tag{5}$$

$$\sum_{i=1}^{I} \sum_{k=1}^{K} Q_{ijkt} \le b, \quad \forall j \in J, \forall t \in T, \tag{6}$$

$$\sum_{j=1}^{J} Y_{ijt} + \sum_{k=1}^{K} Y_{ikt} \le 1, \quad \forall i \in I, \forall t \in T, \tag{7}$$

$$\sum_{t=1}^{T} \sum_{j=1}^{J} Y_{ijt} + \sum_{t=1}^{T} \sum_{k=1}^{K} Y_{ikt} \le n, \quad \forall i \in I, \tag{8}$$

$$\sum_{j=1}^{J} P_{jt} \le q, \quad \forall j \in J, \forall t \in T, \tag{9}$$

$$Y_{ijt} \le P_{jt}, \quad \forall i \in I, \forall j \in J, \forall t \in T, \tag{10}$$

$$Y_{ijkt} \le Q_{ijkt} \le N * Y_{ijkt}, \quad \forall i \in I, \forall j \in J, \forall k \in K, \forall t \in T, \tag{11}$$

$$Y_{ijkt} \le Y_{ijt} \; \forall i \in I \; \forall j \in J \; \forall k \in K \; \forall t \in T \tag{12}$$

$$Y_{ikt} \le Q_{ikt} \le N * Y_{ikt}, \quad \forall i \in I, \forall k \in K, \forall t \in T, \tag{13}$$

$$Y_{kht} \le Q_{kht} \le N * Y_{kht}, \quad \forall k \in K, \forall h \in H, \forall t \in T, \tag{14}$$

$$A_{ij} * Y_{ijt} \le AC, \quad \forall i \in I, \forall j \in J, \forall t \in T, \tag{15}$$

$$A_{ik} * Y_{ikt} \le AC, \quad \forall i \in I, \forall j \in J, \forall t \in T, \tag{16}$$

$$IB_{kt} - \sum_{h=1}^{H} Q_{kht} + \sum_{i=1}^{I}\left(\sum_{j=1}^{J} Q_{ijkt} + Q_{ikt}\right) = IB_{kt+1}, \quad \forall k \in K, \forall t \in T, \tag{17}$$

$$IB_{kt} \le cb, \quad \forall k \in K, \forall t \in T, \tag{18}$$

$$\frac{A_{kh} Y_{kht}}{V_c} \le LT_1, \quad \forall k \in K, \forall h \in H, \forall t \in T, \tag{19}$$

$$D_{ht} \ge \sum_{k=1}^{K} Q_{kht}, \quad \forall h \in H, \forall t \in T, \tag{20}$$

$$IB_{kt} \ge \sum_{h=1}^{H} Q_{kht}, \quad \forall k \in K, \forall t \in T. \tag{21}$$

Equation (5) indicates that the quantity of blood collected at time t should not exceed the maximum blood supply quantity limit at time t in the area. Equation (6) expresses that the total amount of blood collected at time t in the blood donation house will not exceed the volume of j. Equation (7) shows that group i should only choose one of the locations j or k for donation at the same time. Equation (8) calculates the limit of donation times. Equation (9) indicates the limit of the number of blood donation houses open. Equation (10) implies that only when the blood donation house j is open can a donor group i go for donation. Equations (11) and (12) suggest that if group i does not donate blood to a blood donation house j, no blood donated by group i will be sent from j to blood centre k. Equation (13) shows that group i can donate only after reaching the blood centre k. Equation (14) indicates that the blood will not be delivered from the blood centre without the assignment of the hospital. Equations (15) and (16) imply that group i will choose an institution within its acceptable distance for donation. Equation (17) demonstrates the constraint of blood centre inventory status update. Equation (18) indicates that the inventory of blood centre k will not exceed the maximum storage capacity of k. Equation (19) provides the time limit for the blood centre to deliver blood to the hospital. Equation (20) assumes that the hospital has no storage capacity. Equation (21) expresses that the amount of blood delivered by the blood centre to the hospital at time t will not exceed its inventory.

The above model is dynamic stochastic programming and needs specific data to test its feasibility in the next section.

4. Numerical Analysis

This section selects the 8.0 surface-wave magnitude (Ms) earthquake of Wenchuan in China for data analysis based on the model in Section 3, which caused severe damage to Wenchuan and its surrounding environment. As a major city in the vicinity, Chengdu undertook most of the medical treatment for victims. Hence, this research is based on relevant data of prefectures in Chengdu.

4.1. Data. According to the *2019 Statistical Yearbook of Chengdu* issued on the public information website of Chengdu Bureau of Statistics, there are 20 municipal districts and county-level cities in Chengdu, as shown in Figure 2. Through comparison, it is found that only minor differences lie in its 2008 and 2019 administrative division maps. In this paper, the latest version of the administrative division map was selected to perform relevant calculations.

4.1.1. Supply Data. Blood donors are divided into different groups based on Chengdu administrative divisions. As a result, 20 blood donation groups were identified, denoted as $i = 1, 2, \ldots,$ and 20, separately. Additionally, the administrative centre of each division was selected as the starting point for a group to donate blood.

After an emergency, all blood donation groups become active. They altruistically donate blood and help the disaster-affected people. According to statistics, the number of blood donors after an emergency is three times greater than usual. This work's research period lasted for 15 days, and corresponding investigations were made on a time frame of 1 day. After processing, the daily average number of blood donors since an emergency was obtained, that is, the mean values (60, 73, 80, 129, 80, 72, 44, 81, 48, 95, 125, 63, 54, 28, 33, 66, 85, 69, 70, 158) of blood donors in respective groups i. As assumed, the daily blood supply from a single group obeys the Poisson distribution. Hence, the blood volume donated can be calculated accordingly.

4.1.2. Demand Data. During the 12 May Wenchuan Earthquake, many hospitals in diverse districts and counties of Chengdu actively received and rescued patients transferred from the affected area. Among them, nine major hospitals were selected for this research, including Sichuan Provincial People's Hospital. Apart from numbering these hospitals, their mean values of demand are listed in Table 4 [41, 42].

It is assumed that the daily demand, D_{ht}, of each hospital abides by the Poisson distribution that considers the above data as the mean values. Under the circumstance that random numbers from the Poisson distribution can be generated based on these mean values, specific demands of the hospitals were acquired.

4.1.3. Distance Data. The Chengdu Blood Centre, denoted by k, is situated at Wuhou District, Chengdu City. It has multiple blood donation houses. To simplify calculations, 10

Figure 2: Administrative division map of Chengdu (the *2019 Statistical Yearbook of Chengdu*).

Table 4: Mean values of blood demand (unit: U).

Serial numbers of hospitals	1	2	3	4	5
Blood demand	289	126	52	48	46
Serial numbers of hospitals	6	7	8	9	
Blood demand	28	25	21	20	

blood donation houses were selected and numbered. Moreover, the distance from a blood donation house, j, to the Chengdu Blood Centre, k, was presented, where $j = 1, 2, \ldots,$ and 10, and $k = 1$ (Table 5).

20 blood donation groups, i, donate RBCs at the blood donation house, j, or the blood centre, k, and the corresponding distance is shown in Tables 6 and 7.

The other parameters were set as follows. The research period t of the proposed model in this study is defined as 15 days. The initial inventory of Phase 1, IB_{k1}, is equal to 600 U (1 $U = 200$ ml). The unit transportation cost of a vehicle is represented by TJ, and it is set as 0.8 CNY/100 units/km. When the blood volume is under 100 units (1 unit = 1 U), it is simply calculated as 100 units. WH is defined to be CNY 40/U; b is set as 600 U; cb means that storage capacity of the blood centre is 15,000 U; q, equal to 10, refers to the total number of blood donation houses; LT_1 is set as 2.5 h; AC signifies that the maximum acceptable distance from the centre to blood donation groups is 15 km; and V_c, 30 km/h, is the speed of a blood delivery vehicle.

4.2. Numerical Results and Discussion

4.2.1. Donation and Opening of Blood Donation Houses. Setting $\beta = 0.1$, $w_1 = 0.8$, and $w_2 = 0.2$, the Poisson random number is generated according to the given average group supply and hospital demand to get the daily supply quantity of each group and the daily demand quantity of each hospital. In the following numerical analysis, ratio of apheresis donation β is fluctuated and ranges from 0 to 1. Weight coefficients ($w_1 w_2$) indicate the importance of the two objectives. Life safety is considered as the most important task in this work. Hence, the weight coefficient of shortage penalty cost (w_1) is given a much higher value, while the coefficient of the other cost is relatively low. Lingo is introduced to solve the model to obtain the number of open blood donation houses and blood collection in each period, as shown in Table 8 (the number outside the bracket is the serial number of the open blood donation house, and the serial number of numbers in the bracket corresponds to the amount of blood collected in the blood donation house).

Table 5: Distance A_{jk} from the blood donation houses to the blood centres (unit: km).

Serial numbers of blood donation houses	1	2	3	4	5
Distance	25.5	63.8	63.4	36.2	26.5
Serial numbers of blood donation houses	6	7	8	9	10
Distance	24.2	44.3	82.5	45.3	59.9

When t ranges from 1 to 14, Group i will choose blood centre k or blood donation house j for blood donation, and blood donation house j will be open or closed. The blood collected at each stage will be added to the beginning inventory of the blood centre in the next stage.

As the results suggest, under the current data settings, not all the blood donation houses are required to be operational under the current data settings. If there is both a blood donation house and a blood centre within an acceptable distance, the blood centre will be preferred. Since the blood collected in the blood donation house has to be delivered to the blood centre, there are attendant transportation costs, in addition to the cost of opening the blood donation house.

4.2.2. Allocation Scheme. On account of the initial data settings in this section, the calculated allocation scheme is shown in Table 9.

Among them, the demands of 7 hospitals in period $t = 1$ are all met, while the demands of red blood cells of hospitals

Table 6: Distance A_{ij} from different groups to the blood donation house (unit: km).

	1	2	3	4	5	6	7	8	9	10
1	23.7	71.7	71.3	39.8	34.5	24.5	55	93.5	47.3	59.1
2	20.3	72.1	61.7	24.5	24.5	25.1	44.3	82.5	46.5	63.3
3	25.5	69.6	60.2	22.6	23.3	21.6	44.3	83.5	48.5	65.5
4	16	64.1	61.3	33.7	26.8	16.4	43.3	82	42.6	62.4
5	29	65.1	64.7	28.6	27.8	26.6	51.5	86	51.7	60.4
6	41.8	96.6	110	66.2	73	42	92.9	99.8	60.3	40.5
7	67	78.3	101	56.5	54.7	67.1	84.6	122.6	90.5	72.3
8	50	57.1	80.5	37	43.6	50	64.2	102.5	82.2	73.5
9	16.7	49.3	43.9	17.7	2.9	16.9	25.5	63.7	35	81.3
10	75.8	92.9	115.8	71	85	76	99	137.5	93.9	63
11	16	63.1	47	33.3	19.6	1	30.5	58.7	23.6	74.3
12	30	37.8	61	2.6	15.8	30	42.6	80.8	47.7	84.1
13	46.6	76.8	1	59.6	45.3	46.6	22.2	22	38	119.3
14	74.5	134.4	53.8	103.7	103	74.5	88	31.5	41.3	123.8
15	23.3	82.7	38.7	65.5	38.2	23.3	33.2	37.3	0.5	85.4
16	63.2	2.3	73.7	34.6	44.5	63.4	57	95	83	117.5
17	60.1	34.3	76.6	22.7	37.3	60.3	60	98.2	86	102
18	58	95	22.3	77.9	63.6	58.3	40.3	1	39.3	122.2
19	28.4	56.8	22.6	39.6	25.4	30	1	39.6	32.9	100.9
20	73.9	116.8	117.8	86.2	83.7	73.8	100.5	122.1	86.2	2.9

Table 7: Distance A_{ik} from different donation groups to the blood centre (unit: km).

	1	2	3	4	5	6	7	8	9	10
K	12.7	3.1	7.2	4.4	4.4	24.5	34.4	24	24.4	49.5
	11	12	13	14	15	16	17	18	19	20
K	25.8	25.6	64.5	83.3	45.3	62.6	43	82.8	44.5	59.1

Table 8: Number of blood donation houses open and blood collected in each period.

Period t	Opening status of blood donation houses		Number of blood donation houses
1	5 (56)	10 (153)	2
2	10 (180)	—	1
3	1 (115)	10 (103)	2
4	10 (170)	—	1
5	10 (147)	—	1
6	10 (177)	—	1
7	10 (163)	—	1
8	6 (127)	7 (71)	2
9	4 (88)	10 (158)	2
10	10 (173)	—	1
11	10 (163)	—	1
12	8 (70)	10 (177)	2
13	10 (185)	—	1
14	1 (143)	8 (72)	2

numbered 5 and 7 fail to be fully responded. In the period $t = 1$, the shortage of red blood cells is 27 U.

4.2.3. Optimal Distribution of Costs. Table 10 shows details of the optimal distribution of each cost in the blood supply chain.

Table 9: Hospital demand and allocation of red blood cells in period $t = 1$ (unit: U).

Serial number	1	2	3	4	5
Quantity demanded	269	137	46	42	46
Quantity allocated	269	137	46	42	42
Serial number	6	7	8	9	
Quantity demanded	26	23	21	17	
Quantity allocated	26	0	21	17	

Table 10: Details of the comprehensive costs of the blood supply chain (unit: CNY).

Shortage cost	Opening cost of blood donation houses	Operating cost
81,000	60,000	2,185,282
Transportation cost	Inventory cost	Optimal objective
4,094	5,160	515,707

4.3. Sensitivity Analysis. The following is an exploration of the sensitivity of the parameters on the optimal objective and optimal decisions.

4.3.1. Different Supply States. The mean of supply in the example is based on the expectation of a higher level of enthusiasm for donation in an emergency, whereas the reality might be more positive or negative than expected. The product of the mean of supply and the multiple can test the impacts of supply states on the scheme, where the multiple ranges from 0.1 to 1.5.

As shown in Figure 3, the shortage of red blood cells is negatively correlated with the average supply, while the opening cost of blood donation houses is also negatively correlated with the supply quantity, with a decline in the

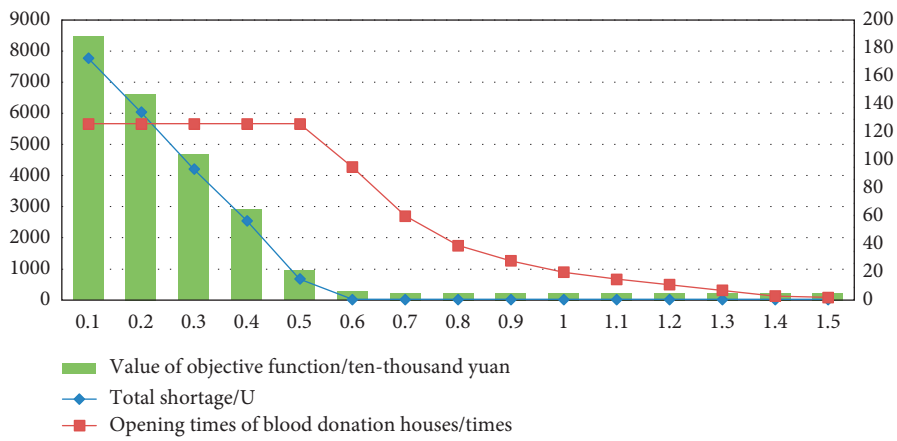


Figure 3: Shortage and opening status of blood donation houses.

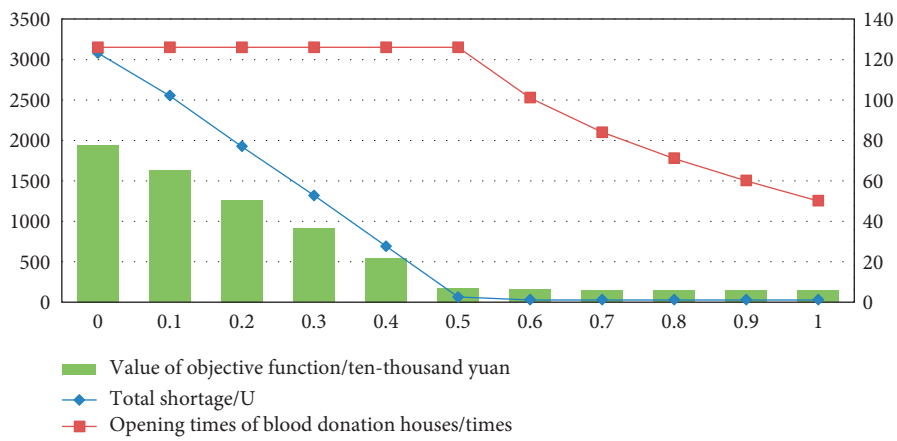


Figure 4: Impact of fluctuations in proportion of apheresis donation.

objective function value. This is because when the average supply is low, there is less supply and more shortage. As the mean value increases, the shortage of red blood cells decreases. Where the supply fails to meet the demand, a larger number of blood donation houses will be open for blood collection. As the supply increases, the need for blood donation houses reduces. Therefore, the opening times and costs of blood donation houses decrease with the increase of supply.

4.3.2. Fluctuations in Proportion of Apheresis Donation. The ratio of apheresis donation β is fluctuated, and the value of β ranges from 0 to 1. Changes in β demonstrate a slight effect on the scheme and the quantity in short, as calculated by the original supply and demand quantity. In this section, 0.4 times of the initial supply mean is taken as the new supply mean before generating the supply data through Poisson distribution. The demand data remain unchanged. The influence of changes in the value of β on the shortage quantity is as shown in Figure 4.

With the increase of β, there is a dramatic decrease in the number of red blood cell shortage. The reason is that the number of red blood cells donated by single donors in the form of apheresis is twice the number of red blood cells donated by whole blood. When β ranges from 0.6 to 1, the total amount of shortage remains 27U. This is because there will be no shortage since the second period, and the red blood cell shortage is related to the initial inventory setting of the first period when the supply quantity meets the demand. The opening times of blood donation houses remain unchanged when the value of β ranges from 0 to 0.5 and then decrease with the increase of the value of β. The corresponding reason is that the supply quantity meets the demand from the second period when the value of β ranges from 0.6 to 1. Since it is no longer necessary to open all blood donation houses to respond to blood collection demands, the total number of opening times of blood donation houses is reduced. The value of objective function decreases with the increase of the value of β.

Therefore, increasing the ratio of apheresis donation will in a way reduce shortage and the value of objective function.

4.3.3. Limit of Donation times. Based on the initial data settings, by adjusting the limit of donation times, the influence of changes in donation times on the scheme is

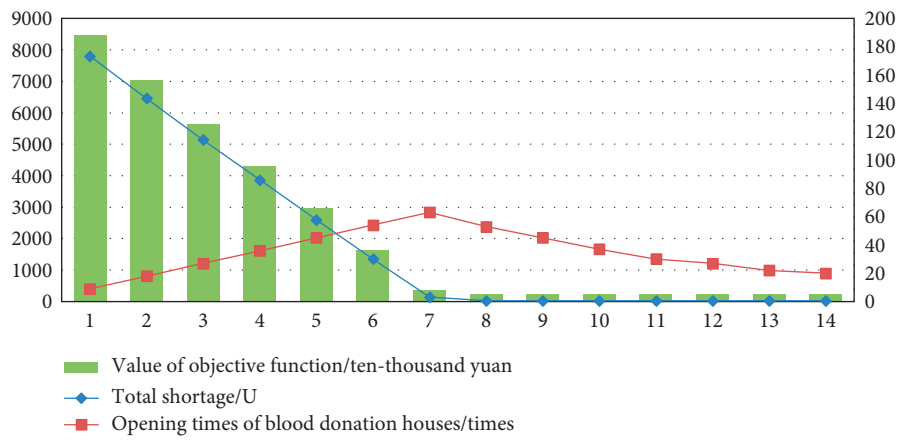


Figure 5: Results of different donation time limits.

Table 11: Impact of AC distance change on red blood cell shortage.

AC distance limits (km)	Total red blood cell shortage (U)	Opening times of blood donation houses (times)	Serial number of newly included group i	Value of objective function (CNY)
15	2551	126	—	6517748
20	2551	98	—	6500434
25	1100	70	6、8、17	3069429
30	1100	70	—	3069483
35	672	56	7、14	2054066
40	672	42	—	2045611
45	672	42	—	2045400
50	104	42	10	711645

discussed in this section. Figure 5 shows shortage of red blood cells and changes in the number of blood donation houses with different donation time limits.

As shown in Figure 5, there is a notable decrease in the shortage of red blood cells and the value of objective function with the relaxation of donation time limits for blood donors during the research period. The number of opening blood donation houses first increases and then decreases. The reason is that some people, at a great donation time limit, are not allowed to donate blood more than once in the research period, which leads to the unnecessary opening of blood donation houses and the decrease of opening times. However, with the relaxation of donation time limits, larger groups of people will have access to blood donation, resulting in an increase in the number of donations available and the opening times of blood donation houses. When the donation time limit is relaxed to such a level (for instance, seven times or more) that the blood supply meets demands, the number of blood donation houses will be reduced.

4.3.4. Changes in AC Distance. In this section, 0.4 times of the initial supply mean is taken as the new supply mean before generating the supply data through Poisson distribution. The demand data, however, remained unchanged.

Table 11 shows the number of red blood cell shortage in blood donation distance of 15–50 km, the maximum number of opening blood donation houses in each period, and the newly included blood donor groups after distance expansion.

As implied in Table 11, the shortage of red blood cells falls with the rise of people's acceptable distance after distance expansion. Some new blood donors are enrolled, and the number of blood donation houses required is decreasing. The cost of opening blood donation houses is reduced because some groups go to the same donation houses or to a blood centre within an acceptable distance after distance expansion.

5. Conclusions and Future Research

5.1. Conclusions. This study focuses on blood supply chain operation-related problems in emergencies. It proposed a four-echelon blood supply chain model consisting of blood donors, blood donation houses, blood centres, and hospitals. By virtue of the proposed model, this study attempts to minimise the composite costs including the shortage cost incurred in blood centres, the cost of opening a blood donation house, and the corresponding operating cost. Through numerical analysis, the feasibility of the model is verified and the following conclusions are drawn.

(1) From the holistic perspective of supply chain, a stochastic optimization-based blood supply chain model is constructed. Simultaneously, it takes blood

collection and supply problems in a blood supply chain into account. Lingo is introduced to solve the model to find the optimal corresponding operation scheme of the blood supply chain, such as the decisions on when and where blood donation houses should be open, how many blood donation houses should be open at a time, how much blood should be collected, how blood centres should distribute blood to hospitals in disaster areas, and so on. Furthermore, the sensitivity of the parameters of the blood supply chain (e.g., proportion of RBC apheresis donation, average of blood supply, times of blood donation, and distance between blood donors and blood facilities) is analysed to test the impacts of various parameter settings on the scheme. Thus, it is suggested that related decision makers need to get a feasible operational scheme by setting relevant parameters according to the actual situation.

(2) By changing the average blood supply, the blood shortage situation is alleviated due to an increase in the average supply. The opening times of blood donation houses and comprehensive costs are also negatively correlated with the supply quantity. Hence, the corresponding demands can be satisfied without opening more blood donation houses.

Therefore, blood donation should be encouraged through effective publicity and transparency in information sharing. In this regard, it is suggested that relevant decision makers should formulate contingency plans and publicity mechanisms for high-risk emergencies in the region prior to emergencies, take timely measures to enhance residents' willingness for blood donation in case of insufficient supply quantity, in order to reduce the threat of quantity shortage, and adjust the opening of blood donation houses and blood collection methods to reduce the overall cost.

(3) Based on fluctuations in the proportion of blood apheresis collection, it is assumed that an increase in this proportion results in a significant decrease of the total RBC shortage and the RBC shortage cost. If this proportion rises, the number of blood donation houses in service can be lowered somewhat. In other words, the blood demand can be satisfied without opening all blood donation houses. Additionally, blood supply shortages caused by overuses in the early phases of an emergency are also eased (donors who have donated blood are forbidden from donating again within a short time).

Considering this, there is a need to encourage the apheresis collection of RBCs. It may be important to continuously innovate the technology of red blood cell apheresis collection, appropriately raise the proportion of apheresis blood donation, in order to increase the number of red blood cells donated by unit donors and relieve blood shortage out of the requirement for donation interval, and increase blood collection in the short term to reduce the opening cost of blood donation houses.

(4) As shown by modifying the distance acceptable to donors, it is found that some new blood donors are enrolled to donate as the distance increases. If other parameters remain unchanged, an increase in this distance leads to a rise in the blood volume available. When the average blood supply is not high, encouraging blood donors (particularly loyal blood donation populations) to increase their acceptable distance to blood facilities may enable the blood supply to go up, thus easing the blood shortage situation. Considering that different blood donation populations may select the same blood donation house, the total number of opening times of blood donation houses is lowered as well. Alternatively, donors may donate blood in a blood centre within their acceptable distance rather than in a blood donation house. As a result, the cost incurred by opening a great number of blood donation houses is reduced.

In this case, emergency management and blood departments are expected to provide convenience for blood donors to expand their donation scope through special transfers or rewards, thereby promoting them to expand their donation distance, increasing blood supply within the jurisdiction to relieve the supply pressure, and reducing the opening times and overall costs of blood donation houses.

(5) Proper adjustment of the donation frequency of donor groups is necessary. Given the time interval requirement for blood donations, a group that made multiple blood donations over the research period implies a significant reduction in the number of subsequent donations. Nevertheless, secondary disasters associated with emergencies frequently bring about casualties, triggering a renewed rise in blood demand. A substantial decline in the number of donations within the early acceptable distances may result in a supply shortfall.

To increase the total amount of available blood and eliminate the risk of a blood shortage, it is recommended to encourage a greater number of donors to be involved in blood collection, conduct red blood cell collection in the form of apheresis processes, increase the ratio of apheresis collection through technical research and relevant investment and training, and promote donors to extend their donation distance to increase the number of available donor groups, thereby enhancing blood quantity and reducing the risk of a blood shortage. Moreover, it is suggested to properly relax the donation time limits for blood donors to alleviate a blood shortage due to excessive short-term donations. These four measures are expected to work best through a combination of them.

5.2. Future Research. There also exist some potential issues for the future study. First, mobile blood donation sites are

also important channels for blood collection and recommended to be incorporated in the future blood supply chain. Second, blood types and other indexes are suggested to be incorporated into subsequent investigations to explore blood supply chain operations further.

Data Availability

The data used to support the findings of this study are included within the article.

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 "Three Levels" Talent Construction Project of Zhuhai College of Science and Technology, the 2018 Discipline Co-Construction Project of the 13th Five-Year Plan of Philosophy and Social Sciences of Guangdong Province (grant no. GD18XTS06), the 13th Five-Year Plan of Philosophy and Social Sciences of Guangdong Province (grant no. GD18XGL24), and the Department of Science and Technology of Guangdong Province (grant no. 2020A0505090004).

References

[1] M. Dillon, F. Oliveira, and B. Abbasi, "A two-stage stochastic programming model for inventory management in the blood supply chain," *International Journal of Production Economics*, vol. 187, pp. 27–41, 2017.

[2] A. Pirabán, W. J. Guerrero, and N. Labadie, "Survey on blood supply chain management: models and methods," *Computers & Operations Research*, vol. 112, p. 104756, 2019.

[3] H. Tang, A. Elalouf, E. Levner, and T. C. E. Cheng, "Efficient computation of evacuation routes on a three-dimensional geometric network," *Computers & Industrial Engineering*, vol. 76, pp. 231–242, 2014.

[4] J. H. Zhang, H. Y. Liu, R. Zhu, and L. Yang, "Emergency evacuation of hazardous chemical accidents based on diffusion simulation," *Complexity*, vol. 2017, Article ID 4927649, 16 pages, 2017.

[5] Z. Liu, X. Li, and X. Chen, "Evacuation traffic management under diffusion of toxic gas based on an improved road risk level assessment method," *Complexity*, vol. 2019, Article ID 6768526, 11 pages, 2019.

[6] J. Ray, *A Multi-Period Linear Programming Modal for Optimally Scheduling the Distribution of Food-Aid in West Africa*, University of Tennessee, Knoxville, TN, USA, 1987.

[7] A. Rathi, R. Church, and R. Solanki, "Allocating resources to support a multicommodity flow with time windows," *Logistics & Transportation Review*, vol. 28, no. 2, pp. 167–188, 1992.

[8] A. Haghani and S.-C. Oh, "Formulation and solution of a multi-commodity, multi-modal network flow model for disaster relief operations," *Transportation Research Part A: Policy and Practice*, vol. 30, no. 3, pp. 231–250, 1996.

[9] L. Özdamar, E. Ekinci, and B. Küçükyazici, "Emergency logistics planning in natural disasters," *Annals of Operations Research*, vol. 129, no. 1-4, pp. 217–245, 2004.

[10] W. Yi and L. Özdamar, "A dynamic logistics coordination model for evacuation and support in disaster response activities," *European Journal of Operational Research*, vol. 179, no. 3, pp. 1177–1193, 2007.

[11] W. Yi and A. Kumar, "Ant colony optimization for disaster relief operations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 43, no. 6, pp. 660–672, 2007.

[12] J.-B. Sheu, "Dynamic relief-demand management for emergency logistics operations under large-scale disasters," *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, no. 1, pp. 1–17, 2010.

[13] S. Yan, C.-K. Lin, and S.-Y. Chen, "Logistical support scheduling under stochastic travel times given an emergency repair work schedule," *Computers & Industrial Engineering*, vol. 67, no. 1, pp. 20–35, 2014.

[14] V. Bosnes, M. Aldrin, and H. E. Heier, "Predicting blood donor arrival," *Transfusion*, vol. 45, no. 2, pp. 162–170, 2005.

[15] A. F. Osorio, S. C. Brailsford, and H. K. Smith, "Whole blood or apheresis donations? A multi-objective stochastic optimization approach," *European Journal of Operational Research*, vol. 266, no. 1, pp. 193–204, 2018.

[16] B. Seda, C. Giuliana, L. Ettore, and Y. Semih, "An appointment scheduling framework to balance the production of blood units from donation," *European Journal of Operational Research*, vol. 265, no. 3, pp. 1124–1143, 2018.

[17] L. A. Williams, J. Sun, and B. Masser, "Integrating self-determination theory and the theory of planned behaviour to predict intention to donate blood," *Transfusion Medicine*, vol. 29, no. S1, pp. 59–64, 2019.

[18] L. Goette and A. Stutzer, "Blood donations and incentives: evidence from a field experiment," *Journal of Economic Behavior & Organization*, vol. 170, pp. 52–74, 2020.

[19] N. M. Heddle, Y. Liu, R. Barty et al., "Factors affecting the frequency of red blood cell outdates: an approach to establish benchmarking targets," *Transfusion*, vol. 49, no. 2, pp. 219–226, 2009.

[20] L. Zheng and C. Xu, "Blood order and collection problems with two demand classes and emergency replenishment," *Journal of the Operational Research Society*, vol. 72, no. 3, pp. 501–518, 2021.

[21] G. P. Prastacos, "Optimal myopic allocation of a product with fixed lifetime," *Journal of the Operational Research Society*, vol. 29, no. 9, pp. 905–913, 1978.

[22] G. Sahin, S. Süral, and S. Meral, "Locational analysis for regionalization of Turkish red crescent blood services," *Computers & Operations Research*, vol. 34, no. 3, pp. 692–704, 2007.

[23] P. Sivakumar, K. Ganesh, and P. Parthiban, "Multi-phase composite analytical model for integrated allocation-routing problem – application of blood bank logistics," *International Journal of Logistics Economics and Globalisation*, vol. 1, no. 3/4, pp. 251–281, 2008.

[24] S. Gunpinar and G. Centeno, "An integer programming approach to the bloodmobile routing problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 86, pp. 94–115, 2016.

[25] K.-M. Wang and Z.-J. Ma, "Age-based policy for blood transshipment during blood shortage," *Transportation Research Part E: Logistics and Transportation Review*, vol. 80, pp. 166–183, 2015.

[26] M. Dehghani and B. Abbasi, "An age-based lateral-transshipment policy for perishable items," *International Journal of Production Economics*, vol. 198, pp. 93–103, 2018.

[27] Y. Sha and J. Huang, "The multi-period location-allocation problem of engineering emergency blood supply systems," *Systems Engineering Procedia*, vol. 5, no. 1, pp. 21–28, 2012.
[28] B. Fahimnia, A. Jabbarzadeh, A. Ghavamifar, and M. Bell, "Supply chain design for efficient and effective blood supply in disasters," *International Journal of Production Economics*, vol. 183, pp. 700–709, 2017.
[29] R. Ramezanian and Z. Behboodi, "Blood supply chain network design under uncertainties in supply and demand considering social aspects," *Transportation Research Part E: Logistics and Transportation Review*, vol. 104, pp. 69–82, 2017.
[30] O. Ö. Özener, A. Ekici, and E. Çoban, "Improving blood products supply through donation tailoring," *Computers & Operations Research*, vol. 102, pp. 10–21, 2019.
[31] B. Hamdan and A. Diabat, "A two-stage multi-echelon stochastic blood supply chain problem," *Computers & Operations Research*, vol. 101, pp. 130–143, 2019.
[32] M. R. G. Samani, S. M. Hosseini-Motlagh, and S. Homaei, "A reactive phase against disruptions for designing a proactive platelet supply network," *Transportation Research Part E: Logistics and Transportation Review*, vol. 140, Article ID 102008, 2020.
[33] A. Jabbarzadeh, B. Fahimnia, and S. Seuring, "Dynamic supply chain network design for the supply of blood in disasters: a robust model with real world application," *Transportation Research Part E: Logistics and Transportation Review*, vol. 70, pp. 225–244, 2014.
[34] M. Arvan, R. Tavakoli-Moghadam, and M. Abdollahi, "Designing a bi-objective and multi-product supply chain network for the supply of blood," *Uncertain Supply Chain Management*, vol. 3, no. 1, pp. 57–68, 2015.
[35] A. Haeri, S. M. Hosseini-Motlagh, M. R. Ghatreh Samani, and M. Rezaei, "A mixed resilient-efficient approach toward blood supply chain network design," *International Transactions in Operational Research*, vol. 27, no. 4, pp. 1962–2001, 2020.
[36] W. Q. Liu, G. Y. Ke, J. Chen, and L. M. Zhang, "Scheduling the distribution of blood products: a vendor-managed inventory routing approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 140, Article ID 101964, 2020.
[37] C. Wang and S. Chen, "A distributionally robust optimization for blood supply network considering disasters," *Transportation Research Part E: Logistics and Transportation Review*, vol. 134, Article ID 101840, 2020.
[38] B. Zahiri and M. S. Pishvaee, "Blood supply chain network design considering blood group compatibility under uncertainty," *International Journal of Production Research*, vol. 55, no. 7, pp. 2013–2033, 2017.
[39] M. Y. N. Attari and E. N. Jami, "Robust stochastic multi-choice goal programming for blood collection and distribution problem with real application," *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 2, pp. 2015–2033, 2018.
[40] M. R. G. Samani and S.-M. Hosseini-Motlagh, "An enhanced procedure for managing blood supply chain under disruptions and uncertainties," *Annals of Operations Research*, vol. 283, no. 1-2, pp. 1413–1462, 2019.
[41] Z. J. Ma, *Theory and Method of Blood Support for Unconventional Emergencies*, Science Press, Beijing, China, 2015.
[42] Z. J. Ma, *Investigation Of Blood Relief In "5.12" Wenchuan Earthquake*, Southwest Jiaotong University, Chengdu, China, 2010.

Hindawi
Complexity
Volume 2021, Article ID 6685951, 29 pages
https://doi.org/10.1155/2021/6685951

Research Article

Homogeneity Test of Many-to-One Risk Differences for Correlated Binary Data under Optimal Algorithms

Keyi Mou and Zhiming Li

College of Mathematics and System Sciences, Xinjiang University, Urumqi 834800, China

Correspondence should be addressed to Zhiming Li; zmli@xju.edu.cn

Received 3 January 2021; Revised 17 February 2021; Accepted 16 March 2021; Published 8 April 2021

Academic Editor: Baogui Xin

Copyright © 2021 Keyi Mou and Zhiming Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In clinical studies, it is important to investigate the effectiveness of different therapeutic designs, especially, multiple treatment groups to one control group. The paper mainly studies homogeneity test of many-to-one risk differences from correlated binary data under optimal algorithms. Under Donner's model, several algorithms are compared in order to obtain global and constrained MLEs in terms of accuracy and efficiency. Further, likelihood ratio, score, and Wald-type statistics are proposed to test whether many-to-one risk differences are equal based on optimal algorithms. Monte Carlo simulations show the performance of these algorithms through the total averaged estimation error, SD, MSE, and convergence rate. Score statistic is more robust and has satisfactory power. Two real examples are given to illustrate our proposed methods.

1. Introduction

Binary data are often encountered for paired organs (e.g., eyes and ears) in medical clinical studies. The responses of each patient are collected and recorded as paired data at the end of the study. The outcome can be none, unilateral, or bilateral cured. Data from all patients can be summarized in a contingency table. The correlation between responses from paired parts should be taken into account to avoid biased or misleading results.

Some probability models have been proposed for analyzing correlated paired data. Rosner introduced a constant R model under the assumption that the conditional probability of a response at one side of the paired body parts given response at the other side was R times the unconditional probability [1]. Under Rosner's model, asymptotic and exact tests were discussed [2–5]. However, Dallal pointed out that Rosner's model could give a poor fit if the characteristic was almost certain to occur bilaterally with widely varying group-specific prevalence [6]. He assumed that each group had a constant conditional probability γ and derived the likelihood ratio statistic to test prevalence equality. M'lan and Chen presented three objective Bayesian methods for bilateral data under Dallal's model [7]. Donner proposed an alternative model by assuming that the correlation coefficient was a fixed constant ρ in each of the groups [8]. Thompson proved that Donner's model could make full use of single and two-organ data to optimize the power of study [9]. Pei et al. applied the model into stratified paired data and assumed that the correlation coefficients of responses were the same to all subjects in two groups of each stratum [10].

Testing the homogeneity has received considerable attention for bilateral binary data. In ophthalmologic studies, Rosner [1] proposed two statistics to test the equality of rate difference in (in)dependence models. Tang et al. [2] developed exact and approximate unconditional procedures for the aforementioned statistics in small sample designs or sparse data structures. Further, Tang et al. [11] developed several statistics for testing the equality of cure rates, including likelihood ratio, score, and two Wald-type statistics in the (in)dependence models. Ma et al. [3] extended these tests to multigroup cases and investigated whether the response rates of the g groups $(g \geq 2)$ were identical under Rosner's model. From the above results, we note that it is crucial to derive the global and constrained MLEs under the hypotheses. However, there are usually no closed-form

solutions for maximum likelihood estimates (MLEs). Under Donner's model, Ma and Liu [12] used two-step algorithm to obtain MLEs and developed several tests for the proportion equality among g groups $(g \geq 2)$. Liu et al. [13] also used the method for constrained MLEs. Peng et al. [14] constructed confidence intervals (CIs) of proportion ratio under Rosner's model. They introduced Fisher scoring algorithm for constrained MLEs. Many algorithms were proposed to obtain MLEs for correlated binary data. However, there are few research studies on comparison of different algorithms for MLEs in multigroup binary design.

Under Donner's model, this paper aims to provide several algorithms for calculating global and constrained MLEs and extends the homogeneity tests of Tang et al. [11] to many-to-one case under optimal algorithms. Fisher scoring algorithm, two-step method, and generalized expectation-maximization (GEM) algorithm are taken into account, since they are widely used in calculating MLEs. Optimal algorithms for MLEs required by the objective test can be found through comparing these algorithms. The rest of this article is organized as follows. In Section 2, we review data structure and establish Donner's model for multigroup correlated binary data. Global and constrained MLEs are derived by various algorithms in Section 3. Based on the optimal algorithms, the likelihood ratio, score, and Wald-type statistics are constructed for testing the equality of many-to-one risk differences. The performance of algorithms is compared by the total averaged estimation error, SD of the averaged estimation error, MSE, and convergence rate in Section 4. Monte Carlo simulations show the empirical type I error rate and power of these tests. Two real examples are provided to illustrate the proposed methods in Section 5. Conclusions and further work are given in Section 6.

2. Preliminaries

Suppose there are g groups involving M individuals in the clinical trial, where the first group is control group and other $g-1$ groups are treatment groups. Let m_{li} be the number of patients with l responses $(l = 0, 1, 2)$ in the i-th group $(i = 1, 2, \ldots, g)$ and $N_i = \sum_{l=0}^{2} m_{li}$ be the total number of patients in the i-th group, which is assumed to be fixed. The data structure is shown in Table 1.

Let $p_{li}\,(l = 0, 1, 2, i = 1, 2, \ldots, g)$ be the probabilities of none, unilateral, and bilateral response(s) in the i-th group, where $p_{0i} + p_{1i} + p_{2i} = 1$ for any fixed i. Denote $\mathbf{m}_i = (m_{0i}, m_{1i}, m_{2i})^T$ and $\mathbf{p}_i = (p_{0i}, p_{1i}, p_{2i})^T$. For the i-th group, $\mathbf{m}_i$ follows a trinomial distribution. Thus, the probability density of $\mathbf{m}_i$ is expressed as follows:

$$f(\mathbf{p}_i|\mathbf{m}_i) = \frac{N_i!}{m_{0i}!m_{1i}!m_{2i}!} p_{0i}^{m_{0i}} p_{1i}^{m_{1i}} p_{2i}^{m_{2i}}, \quad i = 1, 2, \ldots, g. \tag{1}$$

Let Z_{hik} be an indicator of the k-th organ's response $(k = 1, 2)$ for the h-th patient $(h = 1, \ldots, N_i)$ in the i-th group. If there is a response, then $Z_{hik} = 1$, and 0 otherwise. Suppose that $\Pr(Z_{hik} = 1) = \pi_i\,(0 \leq \pi_i \leq 1)$, and $\mathrm{Corr}(Z_{hik}, Z_{hi(3-k)}) = \rho\,(0 \leq \rho \leq 1)$ under Donner's model. Thus, the probabilities p_{li} can be obtained by

$$\begin{aligned} p_{0i} &= \rho(1-\pi_i) + (1-\rho)(1-\pi_i)^2, \\ p_{1i} &= 2\pi_i(1-\rho)(1-\pi_i), \\ p_{2i} &= \rho\pi_i + (1-\rho)\pi_i^2, \end{aligned} \tag{2}$$

for $i = 1, 2, \ldots, g$. Based on the observed data $\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_g)$, the log-likelihood function can be given by

$$\begin{aligned} \ell(\boldsymbol{\pi}, \rho|\mathbf{m}) &= \sum_{i=1}^{g} l_i(\boldsymbol{\pi}_i, \rho|\mathbf{m}_i) + \log C \\ &= \sum_{i=1}^{g} \left\{ m_{0i}\log\left(\rho(1-\pi_i) + (1-\rho)(1-\pi_i)^2\right) + m_{1i}\log\left(2\pi_i(1-\rho)(1-\pi_i)\right) + m_{2i}\log\left(\rho\pi_i + (1-\rho)\pi_i^2\right) \right\} + \log C, \end{aligned} \tag{3}$$

where $C = \prod_{i=1}^{g} (N_i!/(m_{0i}!m_{1i}!m_{2i}!))$ is a constant.

Let $\delta^T = (\delta_2, \ldots, \delta_g)$, where $\delta_i = \pi_i - \pi_1\,(i = 2, \ldots, g)$ is the risk difference between the first group and the i-th group. We are interested to test the hypotheses below.

$$H_0\colon \delta_2 = \cdots = \delta_g \triangleq \delta \text{ versus } H_a\colon \delta_i \text{ is not all the same.} \tag{4}$$

Under H_0, that is $\pi_i = \pi_1 + \delta\,(i = 2, \ldots, g)$, the log-likelihood function can be rewritten as

$$\ell_0(\pi_1, \delta, \rho|\mathbf{m}) = l_{01}(\pi_1, \rho|\mathbf{m}_1) + \sum_{i=2}^{g} l_{0i}(\pi_1, \delta, \rho|\mathbf{m}_i) + \log C, \tag{5}$$

Table 1: Data structure of multigroup binary data.

Number of responses (l)	Group (i)				Total
	1	2	...	g	
0	$m_{01}(p_{01})$	$m_{02}(p_{02})$	...	$m_{0g}(p_{0g})$	S_0
1	$m_{11}(p_{11})$	$m_{12}(p_{12})$	...	$m_{1g}(p_{1g})$	S_1
2	$m_{21}(p_{21})$	$m_{22}(p_{22})$	...	$m_{2g}(p_{2g})$	S_2
Total	N_1	N_2	...	N_g	M

where

$$\begin{aligned} l_{01}(\pi_1,\rho|\mathbf{m}_1) &= m_{01}\log\left(\rho(1-\pi_1)+(1-\rho)(1-\pi_1)^2\right)+m_{11}\log\left(2\pi_1(1-\rho)(1-\pi_1)\right) \\ &\quad + m_{21}\log\left(\rho\pi_1+(1-\rho)\pi_1^2\right), \\ l_{0i}(\pi_1,\delta,\rho|\mathbf{m}_i) &= m_{0i}\log\left(\rho(1-\pi_1-\delta)+(1-\rho)(1-\pi_1-\delta)^2\right) \\ &\quad + m_{1i}\log\left(2(\pi_1+\delta)(1-\rho)(1-\pi_1-\delta)\right) \\ &\quad + m_{2i}\log\left(\rho(\pi_1+\delta)+(1-\rho)(\pi_1+\delta)^2\right), \quad i=2,\ldots,g, \end{aligned} \tag{6}$$

and $C=\prod_{i=1}^{g}(N_i!/(m_{0i}!m_{1i}!m_{2i}!))$ is a constant.

3. Test Methods

In this section, the global and constrained MLEs are first derived by various algorithms. Then, likelihood ratio, score, and Wald-type tests are constructed based on the optimal algorithms.

3.1. Global MLEs. Let $\widehat{\pi}_i$ and $\widehat{\rho}$ be the global MLEs of π_i and ρ. For the unknown parameters $\pi_i\,(i=1,2,\ldots,g)$ and ρ, their global MLEs are the solutions of the following equations:

$$\begin{aligned} \frac{\partial\ell}{\partial\pi_i} &= 0, \\ \frac{\partial\ell}{\partial\rho} &= 0, \quad i=1,2,\ldots,g \end{aligned} \tag{7}$$

where

$$\begin{aligned} \frac{\partial\ell}{\partial\pi_i} &= \frac{(2\pi_i-1)m_{1i}}{\pi_i(\pi_i-1)}+\frac{m_{2i}(\rho+2\pi_i-2\rho\pi_i)}{\pi_i(\rho+\pi_i-\rho\pi_i)}-\frac{m_{0i}(\rho+2\pi_i-2\rho\pi_i-2)}{(\pi_i-1)(\rho\pi_i-\pi_i+1)}, \\ \frac{\partial\ell}{\partial\rho} &= \sum_{i=1}^{g}\left\{\frac{m_{1i}}{(\rho-1)}-\frac{(\pi_i-1)m_{2i}}{(\rho+\pi_i-\rho\pi_i)}+\frac{\pi_i m_{0i}}{(\rho\pi_i-\pi_i+1)}\right\}. \end{aligned} \tag{8}$$

However, there are no closed-form solutions for the above equations. Thus, we need to obtain the global MLEs $\widehat{\pi}_i\,(i=1,2,\ldots,g)$ and $\widehat{\rho}$ by different algorithms.

3.1.1. Global MLEs Based on Fisher Scoring Algorithm. The initial values of π_i and ρ can be given by

$$\begin{aligned} \pi_i^{(0)} &= \frac{m_{1i}+2m_{2i}}{2N_i}, \quad i=1,2,\ldots,g, \\ \rho^{(0)} &= \frac{4S_0S_2-S_1^2}{(S_1+2S_0t(S_1+2S_2)}. \end{aligned} \tag{9}$$

The $(t+1)$-th approximates $\pi^{(t+1)}=(\pi_1^{(t+1)},\pi_2^{(t+1)},\ldots,\pi_g^{(t+1)})^T$ and $\rho^{(t+1)}$ can be obtained by Fisher scoring algorithm:

$$\begin{bmatrix}\boldsymbol{\pi}^{(t+1)}\\ \rho^{(t+1)}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\pi}^{(t)}\\ \rho^{(t)}\end{bmatrix}+I_1^{-1}(\boldsymbol{\pi},\rho)\times\begin{bmatrix}\dfrac{\partial\ell}{\partial\boldsymbol{\pi}}\\ \dfrac{\partial\ell}{\partial\rho}\end{bmatrix}\Bigg|_{\boldsymbol{\pi}=\boldsymbol{\pi}^{(t)},\rho=\rho^{(t)}}, \tag{10}$$

where $(\partial\ell/\partial\pi)=((\partial l_1/\partial\pi_1),(\partial l_2/\partial\pi_2),\ldots,(\partial l_g/\partial\pi_g))^T$ and I_1 is a $(g+1)\times(g+1)$ Fisher information matrix (see Appendix A). Repeat the process until the result converges.

3.1.2. Global MLEs Based on Two-Step Method. The two-step method is described by a third-order polynomial and Newton–Raphson algorithm. The detailed procedure is provided below.

(i) Take the initial value $\rho^{(0)}=((4S_0S_2-S_1^2)/((S_1+2S_0)(S_1+2S_2)))$. Moreover, equation (8) can be simplified as a third-order polynomial:

$$\left(4\rho-2\rho^2-2\right)N_i\pi_i^3+\left(3\rho^2N_i-\rho\left(5m_{0i}+6m_{1i}+7m_{2i}\right)+2m_{0i}+3m_{1i}+4m_{2i}\right)\pi_i^2 \\ +\left(\left(4\rho-\rho^2\right)N_i-2\rho m_{0i}-m_{1i}-2m_{2i}\right)\pi_i-\rho\left(m_{1i}+m_{2i}\right)=0,\quad i=1,2,\ldots,g. \tag{11}$$

Put $\rho^{(t)}$ $(t=0,1,2,\ldots)$ into the polynomial and solve its real root to obtain the t-th approximates of π_i $(i=1,2,\ldots,g)$, denoted by $\pi_i^{(t)}$.

(ii) Update the $(t+1)$-th approximate of ρ by Newton–Raphson algorithm:

$$\rho^{(t+1)}=\rho^{(t)}-\left(\frac{\partial^2\ell}{\partial\rho^2}\left(\boldsymbol{\pi}^{(t)},\rho^{(t)}\right)\right)^{-1}\frac{\partial\ell}{\partial\rho}\left(\boldsymbol{\pi}^{(t)},\rho^{(t)}\right) \tag{12}$$

where

$$\frac{\partial^2\ell}{\partial\rho^2}=-\sum_{i=1}^{g}\left\{\frac{\pi_i^2m_{0i}}{\left(\rho\pi_i-\pi_i+1\right)^2}+\frac{m_{1i}}{\left(\rho-1\right)^2}+\frac{\left(\pi_i-1\right)^2m_{2i}}{\left(\rho+\pi_i-\rho\pi_i\right)^2}\right\}, \tag{13}$$

and $\pi^{(t)}=(\pi_1^{(t)},\pi_2^{(t)},\ldots,\pi_g^{(t)})^T$. Repeat (i)-(ii) until convergence.

3.1.3. Global MLEs Based on GEM Algorithm. According to equation (1), we have

$$f\left(\pi_i,\rho|\mathbf{m}_i\right)\propto\left(\rho\left(1-\pi_i\right)+\left(1-\rho\right)\left(1-\pi_i\right)^2\right)^{m_{0i}}\left(2\pi_i\left(1-\rho\right)\left(1-\pi_i\right)\right)^{m_{1i}}\left(\rho\pi_i+\left(1-\rho\right)\pi_i^2\right)^{m_{2i}}. \tag{14}$$

Suppose patients with 0 response can be divided into two parts, whose responsibilities are $\rho(1-\pi_i)$ and $(1-\rho)(1-\pi_i)^2$, respectively. Let latent variables z_{1i} and $m_{0i}-z_{1i}$ be their total numbers. Observable variable m_{2i} can also be split into two latent variables z_{2i} and $m_{2i}-z_{2i}$. Suppose the probability of result happening in z_{2i} (or $m_{2i}-z_{2i}$) individuals is $\rho\pi_i$ (or $(1-\rho)\pi_i^2$). When $\mathbf{m}_i,\rho$, and π_i are given, z_{1i} and z_{2i} follow binomial distributions:

$$z_{1i}\sim B\left(m_{0i},\frac{\rho\left(1-\pi_i\right)}{\rho\left(1-\pi_i\right)+\left(1-\rho\right)\left(1-\pi_i\right)^2}\right), \\ z_{2i}\sim B\left(m_{2i},\frac{\rho\pi_i}{\rho\pi_i+\left(1-\rho\right)\pi_i^2}\right). \tag{15}$$

Denote $\mathbf{z}=(\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_g)$ and $\mathbf{z}_i=(z_{1i},z_{2i})^T$. Then, $(\mathbf{m},\mathbf{z})$ are complete data and observable data are incomplete data. Based on complete data, we have

$$f\left(\pi_i,\rho|\mathbf{m}_i,\mathbf{z}_i\right)\propto\left(\rho\left(1-\pi_i\right)\right)^{z_{1i}}\left(\left(1-\rho\right)\left(1-\pi_i\right)^2\right)^{m_{0i}-z_{1i}}\left(2\pi_i\left(1-\rho\right)\left(1-\pi_i\right)\right)^{m_{1i}} \\ \times\left(\rho\pi_i\right)^{z_{2i}}\left(\left(1-\rho\right)\pi_i^2\right)^{m_{2i}-z_{2i}}. \tag{16}$$

Thus, the log-likelihood function about complete data is

$$\ell\left(\boldsymbol{\pi},\rho|\mathbf{m},\mathbf{z}\right)\propto\sum_{i=1}^{g}l_i\left(\pi_i,\rho|\mathbf{m}_i,\mathbf{z}_i\right), \tag{17}$$

where

$$l_i\left(\pi_i,\rho|\mathbf{m}_i,\mathbf{z}_i\right)=z_{1i}\log\left(\rho\left(1-\pi_i\right)\right)+\left(m_{0i}-z_{1i}\right)\log\left(\left(1-\rho\right)\left(1-\pi_i\right)^2\right)+z_{2i}\log\left(\rho\pi_i\right) \\ +m_{1i}\log\left(2\pi_i\left(1-\rho\right)\left(1-\pi_i\right)\right)+\left(m_{2i}-z_{2i}\right)\log\left(\left(1-\rho\right)\pi_i^2\right). \tag{18}$$

The initial values of ρ,π_i are defined in equation (9). The process of GEM algorithm is described by the expectation (E) and maximization (M) steps as follows.

(i) *E Step*. Given $\mathbf{m}$ and the current approximates $\pi^{(t)}, \rho^{(t)}$. Denote

$$E_{z_{1i}}\left(z_{1i}|m_{0i}, \pi_i^{(t)}, \rho^{(t)}\right) = \frac{m_{0i}\rho^{(t)}\left(1-\pi_i^{(t)}\right)}{\rho^{(t)}\left(1-\pi_i^{(t)}\right)+\left(1-\rho^{(t)}\right)\left(1-\pi_i^{(t)}\right)^2},$$
$$E_{z_{2i}}\left(z_{2i}|m_{2i}, \pi_i^{(t)}, \rho^{(t)}\right) = \frac{m_{2i}\rho^{(t)}\pi_i^{(t)}}{\rho^{(t)}\pi_i^{(t)}+\left(1-\rho^{(t)}\right)\left(\pi_i^{(t)}\right)^2}. \tag{19}$$

Let $Q(\pi, \rho| \mid \mathbf{m}, \pi^{(t)}, \rho^{(t)})$ be the expected value of the log-likelihood function of π, ρ, with respect to the current conditional distribution of $\mathbf{z}$ as follows:

$$Q\left(\boldsymbol{\pi}, \rho|\mathbf{m}, \boldsymbol{\pi}^{(t)}, \rho^{(t)}\right) = E_{\mathbf{z}}\ell(\boldsymbol{\pi}, \rho|\mathbf{m}, \mathbf{z}) \propto \sum_{i=1}^{g} E_{\mathbf{z}_i} l_i(\pi_i, \rho|\mathbf{m}_i, \mathbf{z}_i) \tag{20}$$

where

$$\begin{aligned} E_{\mathbf{z}_i} l_i(\pi_i, \rho|\mathbf{m}_i, \mathbf{z}_i) &= E_{z_{1i}}\left(z_{1i}|m_{0i}, \pi_i^{(t)}, \rho^{(t)}\right)\log\left(\rho(1-\pi_i)\right) + m_{1i}\log\left(2\pi_i(1-\rho)(1-\pi_i)\right) \\ &+\left(m_{0i} - E_{z_{1i}}\left(z_{1i}|m_{0i}, \pi_i^{(t)}, \rho^{(t)}\right)\right)\log\left((1-\rho)(1-\pi_i)^2\right) \\ &+ E_{z_{2i}}\left(z_{2i}|m_{2i}, \pi_i^{(t)}, \rho^{(t)}\right)\log(\rho\pi_i) \\ &+\left(m_{2i} - E_{z_{2i}}\left(z_{2i}|m_{2i}, \pi_i^{(t)}, \rho^{(t)}\right)\right)\log\left((1-\rho)\pi_i^2\right). \end{aligned} \tag{21}$$

(ii) *M Step*. Update $\pi_1, \pi_2, \ldots, \pi_g$ and ρ in order to successively maximize expected value of the log-likelihood function in E step. The new approximate of parameters can be obtained by maximizing $Q(\pi, \rho|\mathbf{m}, \pi^{(t)}, \rho^{(t)})$ when other parameters are given as their latest approximates. Repeat E and M steps until the result converges.

3.2. Constrained MLEs. Let $\tilde{\pi}_1, \tilde{\delta}$ and $\tilde{\rho}$ be the constrained MLEs of π_1, δ, and ρ under H_0. Under H_0: $\delta_2 = \cdots = \delta_g \triangleq \delta$, it is obvious that $\pi_i = \pi_1 + \delta\,(i = 2, \ldots, g)$. Thus, the constrained MLEs satisfy the following equations:

$$\begin{aligned} \frac{\partial \ell_0}{\partial \pi_1} &= \frac{(2\pi_1-1)m_{11}}{\pi_1(\pi_1-1)} + \frac{m_{21}(\rho+2\pi_1-2\rho\pi_1)}{\pi_1(\rho+\pi_1-\rho\pi_1)} - \frac{m_{01}(\rho+2\pi_1-2\rho\pi_1-2)}{(\pi_1-1)(\rho\pi_1-\pi_1+1)} \\ &+ \sum_{i=2}^{g}\left\{\frac{m_{2i}(2(1-\rho)(\delta+\pi_1)+\rho)}{(\delta+\pi_1)^2-\rho(\delta+\pi_1)(\delta+\pi_1-1)} - \frac{m_{0i}(2(\rho-1)(\delta+\pi_1)+2-\rho)}{(\delta+\pi_1-1)((\delta+\pi_1)(1-\rho)-1)} + \frac{m_{1i}(2\delta+2\pi_1-1)}{(\delta+\pi_1)(\delta+\pi_1-1)}\right\} = 0, \\ \frac{\partial \ell_0}{\partial \delta} &= \sum_{i=2}^{g}\left\{\frac{m_{2i}(2(1-\rho)(\delta+\pi_1)+\rho)}{(\delta+\pi_1)^2-\rho(\delta+\pi_1)(\delta+\pi_1-1)} - \frac{m_{0i}((\rho-1)(2\delta+2\pi_1-1)+1)}{(\delta+\pi_1-1)((\delta+\pi_1)(1-\rho)-1)} + \frac{m_{1i}(2\delta+2\pi_1-1)}{(\delta+\pi_1)(\delta+\pi_1-1)}\right\} = 0, \\ \frac{\partial \ell_0}{\partial \rho} &= \frac{m_{11}}{(\rho-1)} - \frac{(\pi_1-1)m_{21}}{(\rho+\pi_1-\rho\pi_1)} + \frac{\pi_1 m_{01}}{(\rho\pi_1-\pi_1+1)} + \sum_{i=2}^{g}\left\{\frac{m_{11}}{(\rho-1)} - \frac{(\pi_1+\delta_i-1)m_{21}}{(\rho+(1-\rho)(\pi_1+\delta_i))} + \frac{(\pi_1+\delta_i)m_{01}}{((\rho-1)(\pi_1+\delta_i)+1)}\right\} = 0. \end{aligned} \tag{22}$$

However, their closed-form solutions are not given by these equations. Thus, we introduce Fisher scoring algorithm, two-stage procedure, and GEM algorithm to obtain the constrained MLEs $\widetilde{\delta}$, $\widetilde{\pi}_1$, and $\widetilde{\rho}$.

3.2.1. Constrained MLEs Based on Fisher Scoring Algorithm. The initial values of ρ, π_1 are defined in equation (9), and $\delta^{(0)} = 1$. The $(t+1)$-th updates of δ, π_1, and ρ can be calculated by Fisher scoring algorithm as follows:

$$\begin{bmatrix} \delta^{(t+1)} \\ \pi_1^{(t+1)} \\ \rho^{(t+1)} \end{bmatrix} = \begin{bmatrix} \delta^{(t)} \\ \pi_1^{(t)} \\ \rho^{(t)} \end{bmatrix} + I_2^{-1}(\delta, \pi_1, \rho) \times \begin{bmatrix} \dfrac{\partial \ell_0}{\partial \delta} \\ \dfrac{\partial \ell_0}{\partial \pi_1} \\ \dfrac{\partial \ell_0}{\partial \rho} \end{bmatrix}\Bigg|_{\delta=\delta^{(t)}, \pi_1=\pi_1^{(t)}, \rho=\rho^{(t)}}, \tag{23}$$

where I_2 is a 3×3 Fisher information matrix (see Appendix B).

3.2.2. Constrained MLEs Based on Two-Stage Procedure. The two-stage procedure is different from the two-step method in Section 3.1.2. Firstly, the MLE $\widetilde{\delta}$ of δ is given by Newton–Raphson algorithm. Then, $\widetilde{\pi}_1$ and $\widetilde{\rho}$ are obtained by Fisher scoring algorithm under given MLE $\widetilde{\delta}$. The detailed process is described as follows.

(i) The initial values $\pi_1^{(0)}$ and $\rho^{(0)}$ are defined in the equation (9), and $\delta^{(0)} = 1$. The $(t+1)$-th approximate $\delta^{(t+1)}$ of δ is obtained by Newton–Raphson algorithm:

$$\delta^{(t+1)} = \delta^{(t)} - \left(\frac{\partial^2 \ell_0}{\partial \delta^2}\right)^{-1} \left(\frac{\partial \ell_0}{\partial \delta}\right)\Bigg|_{\delta=\delta^{(t)}, \pi_1=\pi_1^{(t)}, \rho=\rho^{(t)}}, \tag{24}$$

where

$$\begin{aligned} \frac{\partial^2 \ell_0}{\partial \delta^2} = \sum_{i=2}^{g} \Bigg\{ & m_{0i} \Bigg(\frac{(\rho-1)(2\delta+2\pi_1-1)+1}{(\delta+\pi_1-1)^2((\delta+\pi_1)(1-\rho)-1)} \\ & - \frac{2(\rho-1)}{(\delta+\pi_1-1)((\delta+\pi_1)(1-\rho)-1)} - \frac{(\rho-1)((\rho-1)(2\delta+2\pi_1-1)+1)}{(\delta+\pi_1-1)((\delta+\pi_1)(1-\rho)-1)^2} \\ & + m_{2i} \Bigg(\frac{-(2(\delta+\pi_1)(1-\rho)+\rho)^2}{((\delta+\pi_1)^2(1-\rho)+\rho(\delta+\pi_1))^2} - \frac{2(\rho-1)}{(\delta+\pi_1)^2(1-\rho)+\rho(\delta+\pi_1)} \\ & + m_{1i} \Bigg(\frac{2}{(\delta+\pi_1)(\delta+\pi_1-1)} - \frac{2\delta+2\pi_1-1}{(\delta+\pi_1)(\delta+\pi_1-1)^2} - \frac{2\delta+2\pi_1-1}{(\delta+\pi_1)^2(\delta+\pi_1-1)}. \end{aligned} \tag{25}$$

(ii) Given $\delta = \widetilde{\delta}$, the $(t+1)$-th approximates of π_1 and ρ can be calculated as

$$\begin{bmatrix} \pi_1^{(t+1)} \\ \rho^{(t+1)} \end{bmatrix} = \begin{bmatrix} \pi_1^{(t)} \\ \rho^{(t)} \end{bmatrix} + I_3^{-1}(\pi_1, \rho) \times \begin{bmatrix} \dfrac{\partial \ell_0}{\partial \pi_1} \\ \dfrac{\partial \ell_0}{\partial \rho} \end{bmatrix}\Bigg|_{\delta=\widetilde{\delta}, \pi_1=\pi_1^{(t)}, \rho=\rho^{(t)}}, \tag{26}$$

where I_3 is a 2×2 Fisher information matrix of π_1 and ρ (see Appendix B).

3.2.3. Constrained MLEs Based on GEM Algorithm. Similar to global MLEs, the GEM algorithm is used to calculate constrained MLEs under H_0. The initial values of ρ, π_1 are defined in equation (9), and $\delta^{(0)} = \sum_{i=2}^{g}(\pi_i^{(0)} - \pi_1^{(0)})/(g-1)$. The detailed process is shown as follows.

(i) *E Step.* Let

$$\begin{aligned} Q\left(\pi_1, \delta, \rho | \mathbf{m}, \pi_1^{(t)}, \delta^{(t)}, \rho^{(t)}\right) &= E_{\mathbf{z}} \ell_0(\pi_1, \delta, \rho | \mathbf{m}, \mathbf{z}) \\ &= E_{\mathbf{z}_1} l_{01}(\pi_1, \rho | \mathbf{m}_1, \mathbf{z}_1) + \sum_{i=2}^{g} E_{\mathbf{z}_i} l_{0i}(\pi_1, \delta, \rho | \mathbf{m}_i, \mathbf{z}_i), \end{aligned} \tag{27}$$

where

$$
\begin{aligned}
&E_{\mathbf{z}_1} l_{01}(\pi_1,\rho|\mathbf{m}_1,\mathbf{z}_1) = E_{z_{11}}\left(z_{11}|m_{01},\pi_1^{(t)},\rho^{(t)}\right)\log\left(\rho\left(1-\pi_1\right)\right)\\
&+\left(m_{01}-E_{z_{11}}\left(z_{11}|m_{01},\pi_1^{(t)},\rho^{(t)}\right)\right)\log\left(\left(1-\rho\right)\left(1-\pi_1\right)^2\right)\\
&+m_{11}\log\left(2\pi_1\left(1-\rho\right)\left(1-\pi_1\right)\right)+E_{z_{21}}\left(z_{21}|m_{21},\pi_1^{(t)},\rho^{(t)}\right)\log\left(\rho\pi_1\right)\\
&+\left(m_{21}-E_{z_{21}}\left(z_{21}|m_{21},\pi_1^{(t)},\rho^{(t)}\right)\right)\log\left(\left(1-\rho\right)\pi_1^2\right),\\
&E_{\mathbf{z}_i} l_{0i}(\pi_1,\delta,\rho|\mathbf{m}_i,\mathbf{z}_i) = E_{z_{1i}}\left(z_{1i}|m_{0i},\pi_1^{(t)},\delta^{(t)},\rho^{(t)}\right)\log\left(\rho\left(1-\pi_1-\delta\right)\right)\\
&+\left(m_{0i}-E_{z_{1i}}\left(z_{1i}|m_{0i},\pi_1^{(t)},\delta^{(t)},\rho^{(t)}\right)\right)\log\left(\left(1-\rho\right)\left(1-\pi_1-\delta\right)^2\right)\\
&+m_{1i}\log\left(2\left(\pi_1+\delta\right)\left(1-\rho\right)\left(1-\pi_1-\delta\right)\right)\\
&+\left(m_{2i}-E_{z_{2i}}\left(z_{2i}|m_{2i},\pi_1^{(t)},\delta^{(t)},\rho^{(t)}\right)\right)\log\left(\left(1-\rho\right)\left(\pi_1+\delta\right)^2\right)\\
&+E_{z_{2i}}\left(z_{2i}|m_{2i},\pi_1^{(t)},\delta^{(t)},\rho^{(t)}\right)\log\left(\rho\left(\pi_1+\delta\right)\right),\quad i=2,\ldots,g,\\
&E_{z_{11}}\left(z_{11}|m_{01},\pi_1^{(t)},\rho^{(t)}\right)=\frac{m_{01}\rho^{(t)}\left(1-\pi_1^{(t)}\right)}{\rho^{(t)}\left(1-\pi_1^{(t)}\right)+\left(1-\rho^{(t)}\right)\left(1-\pi_1^{(t)}\right)^2},\\
&E_{z_{21}}\left(z_{21}|m_{21},\pi_1^{(t)},\rho^{(t)}\right)=\frac{m_{21}\rho^{(t)}\pi_1^{(t)}}{\rho^{(t)}\pi_1^{(t)}+\left(1-\rho^{(t)}\right)\left(\pi_1^{(t)}\right)^2},\\
&E_{z_{1i}}\left(z_{1i}|m_{0i},\pi_1^{(t)},\delta^{(t)},\rho^{(t)}\right)=\frac{m_{0i}\rho^{(t)}\left(1-\pi_1^{(t)}-\delta^{(t)}\right)}{\rho^{(t)}\left(1-\pi_1^{(t)}-\delta^{(t)}\right)+\left(1-\rho^{(t)}\right)\left(1-\pi_1^{(t)}-\delta^{(t)}\right)^2},\\
&E_{z_{2i}}\left(z_{2i}|m_{2i},\pi_1^{(t)},\delta^{(t)},\rho^{(t)}\right)=\frac{m_{2i}\rho^{(t)}\left(\pi_1^{(t)}+\delta^{(t)}\right)}{\rho^{(t)}\left(\pi_1^{(t)}+\delta^{(t)}\right)+\left(1-\rho^{(t)}\right)\left(\pi_1^{(t)}+\delta^{(t)}\right)^2},\quad i=2,\ldots,g.
\end{aligned}
\tag{28}
$$

(ii) *M Step*. Similar to Section 3.1.3, we choose parameters π_1,δ,ρ so that $Q(\pi_1,\delta,\rho|\mathbf{m},\pi_1^{(t)},\delta^{(t)},\rho^{(t)})$ increases. Repeat E and M steps until the result converges.

3.3. Likelihood Ratio Test. Likelihood ratio test statistic can be constructed through the global and constrained MLEs as follows:

$$
T_L = 2\left[\ell(\widehat{\pi},\widehat{\rho})-\ell_0\left(\widetilde{\pi}_1,\widetilde{\delta},\widetilde{\rho}\right)\right] = 2\left[\sum_{i=1}^{g} l_i(\widehat{\pi}_i,\widehat{\rho})-l_{01}(\widetilde{\pi}_1,\widetilde{\rho})-\sum_{i=2}^{g} l_{0i}\left(\widetilde{\pi}_1,\widetilde{\delta},\widetilde{\rho}\right)\right] \tag{29}
$$

where $\widehat{\pi}_i,\widehat{\rho}\,(i=1,2,\ldots,g)$ are the global MLEs and $\widetilde{\pi}_1,\widetilde{\delta},\widetilde{\rho}$ are the constrained MLEs.

3.4. Score Test. Note that H_0: $\delta_2=\cdots=\delta_g\triangleq\delta$ is equivalent to $\pi_2=\cdots=\pi_g$. The homogeneity test of many-to-one risk differences can be achieved by testing the equality of risks in treatment groups. Denote $\mathbf{U}=((\partial l_2/\partial\pi_2),\ \ldots,(\partial l_g/\partial\pi_g),0,0)$. Score test statistic is derived as

$$
T_{SC} = \mathbf{U}I_4\left(\pi_2,\ldots,\pi_g,\pi_1,\rho\right)^{-1}\mathbf{U}^T|_{\pi=\widetilde{\pi},\rho=\widetilde{\rho}}, \tag{30}
$$

where $\pi=(\pi_1,\pi_2,\ldots,\pi_g)^T$ and $\widetilde{\pi}=(\widetilde{\pi}_1,\widetilde{\pi}_2,\ldots,\widetilde{\pi}_g)^T$. $\widetilde{\pi}_i$ and $\widetilde{\rho}$ are constrained MLEs under H_0. I_4 is a $(g+1)\times(g+1)$ information matrix (see Appendix C for more information). Thus, T_{SC} can be simplified as

$$
T_{SC} = \sum_{i=2}^{g}\frac{u_i^2(\widetilde{\pi}_i,\widetilde{\rho})}{a_i(\widetilde{\pi}_i,\widetilde{\rho})}+\frac{\left(\sum_{i=2}^{g}\left(u_i(\widetilde{\pi}_i,\widetilde{\rho})b_i(\widetilde{\pi}_i,\widetilde{\rho})\right)/\left(a_i(\widetilde{\pi}_i,\widetilde{\rho})\right)\right)^2}{D(\widetilde{\boldsymbol{\pi}},\widetilde{\rho})-\sum_{i=1}^{g}\left(b_i^2(\widetilde{\pi}_i,\widetilde{\rho})\right)/\left(a_i(\widetilde{\pi}_i,\widetilde{\rho})\right)}, \tag{31}
$$

where

$$
\begin{aligned}
u_i(\pi_i,\rho) &= \frac{\partial l_i}{\partial \pi_i} = \frac{(2\pi_i-1)m_{1i}}{\pi_i(\pi_i-1)} + \frac{m_{2i}(\rho+2\pi_i-2\rho\pi_i)}{\pi_i(\rho+\pi_i-\rho\pi_i)} - \frac{m_{0i}(\rho+2\pi_i-2\rho\pi_i-2)}{(\pi_i-1)(\rho\pi_i-\pi_i+1)},\\
a_i(\pi_i,\rho) &= E\left(-\frac{\partial^2 l_i}{\partial \pi_i^2}\right) = \frac{N_i\left(-4\rho^2\pi_i^2+4\rho^2\pi_i-\rho^2+6\rho\pi_i^2-6\rho\pi_i+2\rho-2\pi_i^2+2\pi_i\right)}{\pi_i(1-\pi_i)(\rho+\pi_i-\rho\pi_i)(\rho\pi_i-\pi_i+1)},\\
b_i(\pi_i,\rho) &= E\left(-\frac{\partial^2 l_i}{\partial \pi_i \partial \rho}\right) = \frac{N_i\rho(2\pi_i-1)}{(\rho+\pi_i-\rho\pi_i)(\rho\pi_i-\pi_i+1)},\\
D(\boldsymbol{\pi},\rho) &= E\left(-\frac{\partial^2 \ell}{\partial \rho^2}\right) = \sum_{i=1}^{g}\frac{N_i\pi_i(\rho+1)(1-\pi_i)}{(1-\rho)(\rho+\pi_i-\rho\pi_i)(\rho\pi_i-\pi_i+1)}.
\end{aligned}
\tag{32}
$$

3.5. Wald-Type Test. Let $\beta = (\pi_2,\ldots,\pi_g,\pi_1,\rho)$ and

$$
\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & \ldots & \ldots & 0 & 0 & 0\\ 0 & 1 & -1 & & & 0 & 0 & 0\\ \vdots & \ddots & \ddots & \ddots & & \vdots & \vdots & \vdots\\ \vdots & & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots\\ 0 & \ldots & \ldots & 0 & 1 & -1 & 0 & 0 \end{bmatrix}_{(g-2)\times(g+1)}. \tag{33}
$$

The null hypothesis H_0: $\delta_2 = \cdots = \delta_g \triangleq \delta$ is equivalent to $\mathbf{C}\beta^T = 0$. Thus, Wald-type statistic is

$$
T_W = \left(\boldsymbol{\beta}\mathbf{C}^T\right)\left(\mathbf{C}I_4^{-1}\mathbf{C}^T\right)^{-1}\left(\mathbf{C}\boldsymbol{\beta}^T\right)\Big|_{\boldsymbol{\pi}=\widehat{\boldsymbol{\pi}},\rho=\widehat{\rho}}, \tag{34}
$$

where the Fisher information matrix I_4 is the same as that of score test (see Appendix D). It can be simplified as

$$
T_W = \sum_{i=1}^{g-2}\sum_{j=1}^{g-2}\left\{\left(\widehat{\pi}_{i+1}-\widehat{\pi}_{i+2}\right)\left(\widehat{\pi}_{j+1}-\widehat{\pi}_{j+2}\right)\sum_{k=1}^{g-2} r_{ki}(\widehat{\boldsymbol{\pi}},\widehat{\rho})r_{kj}(\widehat{\boldsymbol{\pi}},\widehat{\rho})\right\}, \tag{35}
$$

where

$$
\begin{aligned}
r_{ij}(\boldsymbol{\pi},\rho) &= \begin{cases} -\dfrac{1}{l_{ii}(\boldsymbol{\pi},\rho)}\displaystyle\sum_{k=j}^{i-1} l_{ik}(\boldsymbol{\pi},\rho)r_{kj}(\boldsymbol{\pi},\rho), & i>j,\\[2ex] \dfrac{1}{l_{ij}(\boldsymbol{\pi},\rho)}, & i=j,\\[2ex] 0, & i<j, \end{cases}\\
&= \begin{cases} \dfrac{f_{ij}(\boldsymbol{\pi},\rho)-\sum_{k=1}^{j-1} l_{ik}(\boldsymbol{\pi},\rho)l_{jk}(\boldsymbol{\pi},\rho)}{l_{jj}(\boldsymbol{\pi},\rho)}, & i>j,\\[2ex] \left(f_{ij}(\boldsymbol{\pi},\rho)-\displaystyle\sum_{k=1}^{j-1} l_{jk}(\boldsymbol{\pi},\rho)^2\right)^{(1/2)}, & i=j,\\[2ex] 0, & i<j, \end{cases}\\
&= \begin{cases} \dfrac{1}{a_{i+1}(\pi_{i+1},\rho)} + \dfrac{\left(b_{i+1}(\pi_{i+1},\rho)b_{j+1}(\pi_{j+1},\rho)\right)/\left(a_{i+1}(\pi_{i+1},\rho)a_{j+1}(\pi_{j+1},\rho)\right)}{D(\boldsymbol{\pi},\rho)-\sum_{i=1}^{g}\left(b_i^2(\pi_i,\rho)\right)/\left(a_i(\pi_i,\rho)\right)}, & i=j,\\[2ex] \dfrac{\left(b_{i+1}(\pi_{i+1},\rho)b_{j+1}(\pi_{j+1},\rho)\right)/\left(a_{i+1}(\pi_{i+1},\rho)a_{j+1}(\pi_{j+1},\rho)\right)}{D(\boldsymbol{\pi},\rho)-\sum_{i=1}^{g}\left(b_i^2(\pi_i,\rho)\right)/\left(a_i(\pi_i,\rho)\right)}, & i\neq j. \end{cases}
\end{aligned}
\tag{36}
$$

$$
f_{ij}(\boldsymbol{\pi},\rho) = \left(I_{ij}(\boldsymbol{\pi},\rho)-I_{(i+1)j}(\boldsymbol{\pi},\rho)\right)-\left(I_{i(j+1)}(\boldsymbol{\pi},\rho)-I_{(i+1)(j+1)(\boldsymbol{\pi},\rho)}\right), I_{ij}(\boldsymbol{\pi},\rho)
$$

Under H_0, test statistics T_L, T_{SC}, and T_W are asymptotically distributed as chi-square distribution with $g-2$ degrees of freedom. Thus, H_0 should be rejected if the value of test statistic is larger than $\chi^2_{g-2,1-\alpha}$ at the significant level α, where $\chi^2_{g-2,1-\alpha}$ is the $100(1-\alpha)$ percentile of the chi-square distribution with $g-2$ degrees of freedom.

4. Monte Carlo Simulations

In this section, the performance of several algorithms are compared with respect to average errors of MLEs, the number of iteration, and time cost. For convenience, we denote Fisher scoring algorithm, two-step method, and GEM algorithm for global MLEs as FSA, TSM, and GEM and Fisher scoring algorithm, two-stage procedure, and GEM algorithm for constrained MLEs as FSA, TSP, and GEM for tables and figures. Then, we investigate the type I error rates (TIEs) and power of the likelihood ratio, score, and Wald-type tests. In simulations, g and $\mathbf{N} = (N_1, N_2, \ldots, N_g)^T$ are arranged as shown in Table 2, where the scenarios 4, 8, and 12 are unbalanced designs.

Table 2: Scenarios of simulation.

g	Scenario	$N_1, N_2, \ldots, N_g$
3	1	30, 30, 30
	2	70, 70, 70
	3	100, 100, 100
	4	30, 70, 100
6	5	30, 30, 30, 30, 30, 30
	6	70, 70, 70, 70, 70, 70
	7	100, 100, 100, 100, 100, 100
	8	30, 40, 50, 60, 70, 80
9	9	30, 30, 30, 30, 30, 30, 30, 30, 30
	10	70, 70, 70, 70, 70, 70, 70, 70, 70
	11	100, 100, 100, 100, 100, 100, 100, 100, 100
	12	30, 40, 50, 60, 70, 80, 90, 100, 110

4.1. Selection of Algorithms. Under H_a or H_0, we randomly select 1000 sets of $\pi_i\,(i = 1, 2, \ldots, g)$ and ρ for each scenario in Table 2. Further, 10,000 samples are randomly produced for each parameter setting.

4.1.1. Evaluation of MLEs. Let $\overline{e}_i$ be averaged error among estimators and true values of parameters from the 10000 random samples under the i-th parameter setting. Denote $\overline{\overline{e}} = (1/1000)\sum_{i=1}^{1000}\overline{e}_i$ in each scenario, which is the total averaged error for 1000 parameter settings. The dispersion of $\overline{e}_i$ in each scenario can be reflected by standard deviation (SD) value of $\overline{e}_i$. The MSEs of the global MLEs can be evaluated by the differences from the true parameters to the corresponding estimated values under various parameter settings. The global MLEs can be calculated based on Fisher scoring algorithm, two-step method, and GEM algorithm. The constrained MLEs are obtained by Fisher scoring algorithm, two-stage procedure, and GEM algorithm. The random samples for the former are generated under H_a and the latter are generated under H_0. The convergence accuracy ϵ is defined by the differences from two close iterations and fixed as 1×10^{-5}. The MSEs of the three algorithms for global MLEs have no significant difference as shown in Tables 3–5 . That is to say, the global MLEs are identical by these algorithms. In Tables 6–8 , the values of $\overline{\overline{e}}$, SD, and MSEs from Fisher scoring algorithm are usually smaller than other two algorithms for constrained MLEs. So, Fisher scoring algorithm has higher accuracy for constrained MLEs. All MSEs become smaller and close to each other when sample size increases. Algorithms for MLEs have better MSEs in balanced designs than unbalanced designs.

Tables 9 and 10 show the number of failures for these algorithms to converge when $\varepsilon = 1 \times 10^{-5}, 1 \times 10^{-4}$ within $k = 100, 200$ iterations. Fisher scoring algorithm has lower failure rate for convergence when calculating global MLEs. GEM algorithm for global MLEs is sensitive to ϵ and k. Both reducing ϵ and adding k can markedly improve the convergence possibility of GEM algorithm for global MLEs. Fisher scoring and GEM algorithms for constrained MLEs hardly fail to converge within 100 iterations. As g increases, the number of failures becomes small.

4.1.2. Number of Iteration. Since the algorithms for global MLEs have identical accuracy in terms of MSEs, we further compare their efficiency by calculating the number of iteration and time cost when the convergence accuracy is $\epsilon = 1 \times 10^{-5}$. The average number of iteration is recorded for every parameter setting in Figure 1.

As shown in Figure 1, GEM algorithm takes more iterations to converge than Fisher scoring algorithm and two-step method. The number of iterations is most intensive between 20 and 30 and the upper limit increases when $N_i\,(i = 1, 2, \ldots, g)$ is bigger. For Fisher scoring algorithm, the number of iterations has slight superiority over two-step method.

4.1.3. Time Cost. The time required to achieve convergence can also reflect the convergence rate of an algorithm. Figure 2 shows that GEM algorithm has the worst performance in terms of time cost. Fisher scoring algorithm always takes less time, especially when $g = 6, 9$. As $N_i\,(i = 1, 2, \ldots, g)$ increases, the time distribution is more clustered for Fisher scoring algorithm and two-step method. When g is larger, difference of time cost is bigger between the two algorithms.

Based on the above results, Fisher scoring algorithm, two-step method, and GEM algorithm can produce the same global MLEs when the number of iterations is large enough. However, Fisher scoring algorithm for global MLEs has a better convergence rate in terms of number of iteration and time cost. Constrained MLEs from Fisher scoring algorithm has smaller MSEs. Therefore, it is advisable to choose Fisher scoring algorithms to obtain the global and constrained MLEs required by the homogeneity test in this paper.

4.2. Evaluation of Test Statistics. Since Fisher scoring algorithm has higher accuracy for global MLEs and is more efficient for constrained MLEs, we construct test statistics

Table 3: Errors of the global MLEs for $g = 3$.

Index	MLE	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
		FSA	TSM	GEM	FSA	TSM	GEM	FSA	TSM	GEM	FSA	TSM	GEM
$\bar{\bar{e}}$	$\hat{\pi}_1$	−0.000071	−0.000109	−0.000076	−0.000037	−0.000125	−0.000036	−0.000058	−0.000078	−0.000064	−0.000544	−0.001051	−0.000616
	$\hat{\pi}_2$	−0.000030	−0.000043	−0.000026	−0.000147	−0.000243	−0.000147	−0.000030	−0.000049	−0.000030	−0.000081	−0.000584	−0.000116
	$\hat{\pi}_3$	0.000025	0.000011	0.000028	0.000036	−0.000069	0.000036	−0.000047	−0.000061	−0.000054	−0.000373	−0.000824	−0.000385
	$\hat{\rho}$	−0.007755	−0.006677	−0.007174	−0.003747	−0.003360	−0.003342	−0.002615	−0.002376	−0.002213	−0.046909	−0.040828	−0.040980
SD of $\bar{e}_i$	$\hat{\pi}_1$	0.002403	0.002479	0.002454	0.001492	0.001615	0.001523	0.001254	0.001277	0.001281	0.002401	0.002466	0.002459
	$\hat{\pi}_2$	0.002347	0.002413	0.002402	0.001542	0.001680	0.001570	0.001251	0.001285	0.001274	0.001776	0.001842	0.001825
	$\hat{\pi}_3$	0.002362	0.002481	0.002426	0.001479	0.001701	0.001509	0.001255	0.001274	0.001284	0.001480	0.001560	0.001515
	$\hat{\rho}$	0.011982	0.015899	0.013870	0.006204	0.007919	0.007824	0.004785	0.005911	0.006477	0.008615	0.010931	0.010636
MSE	$\hat{\pi}_1$	0.005428	0.005438	0.005428	0.002290	0.002319	0.002289	0.001599	0.001605	0.001599	0.005327	0.005345	0.005327
	$\hat{\pi}_2$	0.005332	0.005339	0.005332	0.002286	0.002321	0.002286	0.001609	0.001616	0.001609	0.003192	0.003205	0.003192
	$\hat{\pi}_3$	0.005332	0.005338	0.005331	0.002278	0.002314	0.002278	0.001602	0.001608	0.001602	0.002289	0.002303	0.002289
	$\hat{\rho}$	0.009317	0.009317	0.009314	0.004099	0.004098	0.004097	0.002836	0.002836	0.002835	0.005716	0.005715	0.005714

Table 4: Errors of the global MLEs for $g = 6$.

Index	MLE	Scenario 5			Scenario 6			Scenario 7			Scenario 8		
		FSA	TSM	GEM	FSA	TSM	GEM	FSA	TSM	GEM	FSA	TSM	GEM
$\bar{\bar{e}}$	$\hat{\pi}_1$	0.000010	−0.000004	0.000011	0.000012	0.000008	0.000016	−0.000064	−0.000064	−0.000064	−0.000022	−0.000026	−0.000029
	$\hat{\pi}_2$	0.000009	−0.000007	0.000009	−0.000013	−0.000021	−0.000013	0.000067	0.000067	0.000066	−0.000027	−0.000029	−0.000027
	$\hat{\pi}_3$	0.000006	−0.000009	0.000007	−0.000030	−0.000036	−0.000030	−0.000038	−0.000039	−0.000039	−0.000096	−0.000097	−0.000095
	$\hat{\pi}_4$	0.000089	0.000080	0.000092	−0.000068	−0.000075	−0.000070	0.000077	0.000077	0.000077	0.000047	0.000045	0.000047
	$\hat{\pi}_5$	−0.000121	−0.000134	−0.000120	0.000056	0.000048	0.000056	0.000051	0.000051	0.000051	−0.000010	−0.000012	−0.000011
	$\hat{\pi}_6$	0.000062	0.000048	0.000063	−0.000011	−0.000019	−0.000015	0.000055	0.000055	0.000056	−0.000071	−0.000075	−0.000073
	$\hat{\rho}$	−0.011639	−0.011557	−0.011491	−0.005095	−0.005052	−0.004959	−0.003580	−0.003578	−0.003538	−0.006436	−0.006387	−0.006288
SD of $\bar{e}_i$	$\hat{\pi}_1$	0.002378	0.002395	0.002395	0.001539	0.001543	0.001554	0.001256	0.001256	0.001258	0.002298	0.002302	0.002313
	$\hat{\pi}_2$	0.002360	0.002378	0.002375	0.001490	0.001499	0.001500	0.001317	0.001318	0.001321	0.001984	0.001991	0.002004
	$\hat{\pi}_3$	0.002353	0.002367	0.002376	0.001522	0.001529	0.001538	0.001265	0.001266	0.001270	0.001757	0.001761	0.001777
	$\hat{\pi}_4$	0.002458	0.002470	0.002485	0.001473	0.001478	0.001485	0.001279	0.001279	0.001282	0.001571	0.001575	0.001583
	$\hat{\pi}_5$	0.002316	0.002333	0.002343	0.001461	0.001470	0.001470	0.001249	0.001250	0.001253	0.001499	0.001501	0.001512
	$\hat{\pi}_6$	0.002294	0.002308	0.002311	0.001495	0.001506	0.001503	0.001246	0.001246	0.001248	0.001380	0.001388	0.001395
	$\hat{\rho}$	0.004962	0.005520	0.005899	0.002608	0.002946	0.003562	0.001687	0.001711	0.001873	0.003362	0.003720	0.004300
MSE	$\hat{\pi}_1$	0.005353	0.005359	0.005353	0.002300	0.002303	0.002300	0.001604	0.001604	0.001604	0.005302	0.005302	0.005302
	$\hat{\pi}_2$	0.005387	0.005393	0.005387	0.002294	0.002297	0.002294	0.001597	0.001597	0.001597	0.004013	0.004014	0.004013
	$\hat{\pi}_3$	0.005331	0.005336	0.005330	0.002293	0.002295	0.002293	0.001614	0.001614	0.001614	0.003241	0.003242	0.003241
	$\hat{\pi}_4$	0.005379	0.005383	0.005379	0.002280	0.002283	0.002280	0.001597	0.001597	0.001597	0.002680	0.002681	0.002680
	$\hat{\pi}_5$	0.005373	0.005378	0.005372	0.002294	0.002296	0.002294	0.001608	0.001608	0.001608	0.002300	0.002301	0.002300
	$\hat{\pi}_6$	0.005302	0.005307	0.005302	0.002302	0.002304	0.002302	0.001617	0.001617	0.001617	0.002005	0.002006	0.002005
	$\hat{\rho}$	0.004907	0.004907	0.004906	0.002063	0.002063	0.002062	0.001430	0.001430	0.001430	0.002640	0.002640	0.002639

Table 5: Errors of the global MLEs for $g = 9$.

Index	MLE	Scenario 9			Scenario 10			Scenario 11			Scenario 12		
		FSA	TSM	GEM	FSA	TSM	GEM	FSA	TSM	GEM	FSA	TSM	GEM
$\bar{\bar{e}}$	$\hat{\pi}_1$	−0.000013	−0.000018	−0.000014	0.000008	0.000008	0.000010	−0.000043	−0.000043	−0.000045	−0.000006	−0.000006	−0.000003
	$\hat{\pi}_2$	−0.000106	−0.000106	−0.000103	0.000047	0.000047	0.000048	−0.000009	−0.000009	−0.000009	0.000008	0.000008	0.000010
	$\hat{\pi}_3$	−0.000043	−0.000050	−0.000048	−0.000065	−0.000066	−0.000065	−0.000029	−0.000029	−0.000027	0.000096	0.000096	0.000093
	$\hat{\pi}_4$	0.000018	0.000014	0.000019	0.000004	0.000004	0.000003	−0.000010	−0.000010	−0.000010	−0.000052	−0.000053	−0.000054
	$\hat{\pi}_5$	0.000030	0.000026	0.000032	−0.000021	−0.000021	−0.000021	−0.000024	−0.000024	−0.000024	−0.000024	−0.000024	−0.000022
	$\hat{\pi}_6$	−0.000012	−0.000016	−0.000013	−0.000021	−0.000021	−0.000021	0.000071	0.000071	0.000072	−0.000118	−0.000118	−0.000118
	$\hat{\pi}_7$	−0.000106	−0.000108	−0.000102	−0.000036	−0.000036	−0.000035	0.000077	0.000077	0.000077	−0.000019	−0.000020	−0.000019
	$\hat{\pi}_8$	−0.000151	−0.000155	−0.000151	−0.000005	−0.000006	−0.000005	0.000012	0.000013	0.000014	−0.000022	−0.000022	−0.000024
	$\hat{\pi}_9$	−0.000043	−0.000044	−0.000043	0.000058	0.000057	0.000058	−0.000008	−0.000008	−0.000008	−0.000018	−0.000018	−0.000017
	$\hat{\rho}$	−0.011963	−0.011923	−0.011885	−0.005189	−0.005188	−0.005158	−0.003624	−0.003622	−0.003582	−0.005150	−0.005138	−0.005082
SD of $\bar{e}_i$	$\hat{\pi}_1$	0.002344	0.002354	0.002357	0.001511	0.001511	0.001517	0.001355	0.001355	0.001358	0.002419	0.002420	0.002428
	$\hat{\pi}_2$	0.002344	0.002345	0.002353	0.001580	0.001580	0.001583	0.001293	0.001293	0.001295	0.002062	0.002063	0.002073
	$\hat{\pi}_3$	0.002272	0.002280	0.002279	0.001578	0.001577	0.001581	0.001281	0.001281	0.001286	0.001737	0.001738	0.001745
	$\hat{\pi}_4$	0.002354	0.002362	0.002362	0.001518	0.001518	0.001520	0.001270	0.001270	0.001274	0.001612	0.001612	0.001616
	$\hat{\pi}_5$	0.002344	0.002350	0.002350	0.001462	0.001463	0.001464	0.001259	0.001259	0.001260	0.001488	0.001488	0.001491
	$\hat{\pi}_6$	0.002288	0.002290	0.002298	0.001512	0.001513	0.001514	0.001284	0.001284	0.001286	0.001457	0.001457	0.001461
	$\hat{\pi}_7$	0.002308	0.002314	0.002320	0.001513	0.001513	0.001517	0.001263	0.001263	0.001265	0.001340	0.001340	0.001343
	$\hat{\pi}_8$	0.002247	0.002250	0.002257	0.001497	0.001497	0.001498	0.001232	0.001232	0.001234	0.001237	0.001238	0.001241
	$\hat{\pi}_9$	0.002415	0.002416	0.002425	0.001568	0.001568	0.001569	0.001268	0.001268	0.001275	0.001206	0.001206	0.001208
	$\hat{\rho}$	0.003264	0.003638	0.003896	0.001559	0.001575	0.001710	0.001313	0.001336	0.001572	0.001655	0.001766	0.002223
MSE	$\hat{\pi}_1$	0.005327	0.005329	0.005327	0.002327	0.002327	0.002327	0.001618	0.001618	0.001618	0.005370	0.005370	0.005370
	$\hat{\pi}_2$	0.005365	0.005366	0.005365	0.002358	0.002358	0.002358	0.001603	0.001603	0.001603	0.004034	0.004034	0.004034
	$\hat{\pi}_3$	0.005405	0.005407	0.005405	0.002322	0.002323	0.002322	0.001613	0.001613	0.001613	0.003205	0.003205	0.003205
	$\hat{\pi}_4$	0.005332	0.005333	0.005332	0.002312	0.002312	0.002312	0.001622	0.001622	0.001621	0.002705	0.002705	0.002705
	$\hat{\pi}_5$	0.005393	0.005394	0.005393	0.002307	0.002308	0.002307	0.001618	0.001618	0.001618	0.002325	0.002325	0.002325
	$\hat{\pi}_6$	0.005393	0.005394	0.005393	0.002339	0.002339	0.002339	0.001609	0.001609	0.001609	0.002015	0.002015	0.002015
	$\hat{\pi}_7$	0.005344	0.005345	0.005344	0.002314	0.002315	0.002314	0.001618	0.001618	0.001618	0.001811	0.001811	0.001811
	$\hat{\pi}_8$	0.005370	0.005371	0.005370	0.002325	0.002326	0.002325	0.001599	0.001599	0.001599	0.001627	0.001627	0.001627
	$\hat{\pi}_9$	0.005375	0.005376	0.005375	0.002324	0.002324	0.002324	0.001623	0.001623	0.001623	0.001471	0.001471	0.001471
	$\hat{\rho}$	0.003322	0.003322	0.003321	0.001355	0.001355	0.001355	0.000949	0.000949	0.000949	0.001367	0.001367	0.001367

Table 6: Errors of the constrained MLEs for $g = 3$.

Index	MLE	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
		FSA	TSP	GEM	FSA	TSP	GEM	FSA	TSP	GEM	FSA	TSP	GEM
$\bar{\bar{e}}$	$\tilde{\pi}_1$	0.000022	0.000010	0.000010	0.000002	−0.000002	−0.000002	−0.000019	−0.000017	−0.000017	−0.000007	−0.000005	−0.000005
	$\tilde{\delta}$	−0.000028	−0.000089	−0.000643	−0.000019	−0.000001	−0.000202	0.000019	0.000018	−0.001885	−0.000006	0.000012	0.000919
	$\tilde{\rho}$	−0.008981	0.024131	0.024078	−0.003995	0.025856	0.025855	−0.002805	0.028425	0.028424	−0.004202	0.015839	0.015835
SD of $\bar{e}_i$	$\tilde{\pi}_1$	0.001057	0.001057	0.001040	0.000536	0.000544	0.000540	0.000430	0.000435	0.000433	0.000966	0.000804	0.000804
	$\tilde{\delta}$	0.001492	0.003136	0.080522	0.000669	0.002452	0.078893	0.000554	0.002788	0.082119	0.001003	0.001908	0.081664
	$\tilde{\rho}$	0.001628	0.035285	0.035090	0.001006	0.033835	0.033819	0.000770	0.036703	0.036695	0.001087	0.022045	0.022035
MSE	$\tilde{\pi}_1$	0.005227	0.005264	0.005264	0.002218	0.002234	0.002234	0.001564	0.001575	0.001575	0.005168	0.005218	0.005218
	$\tilde{\delta}$	0.007836	0.008837	0.010014	0.003334	0.003391	0.007715	0.002349	0.002349	0.007794	0.006098	0.006143	0.009410
	$\tilde{\rho}$	0.009606	0.010420	0.010420	0.004217	0.005654	0.005654	0.002923	0.004798	0.004798	0.004424	0.004876	0.004876

Table 7: Errors of the constrained MLEs for $g = 6$.

Index	MLE	Scenario 5			Scenario 6			Scenario 7			Scenario 8		
		FSA	TSP	GEM	FSA	TSP	GEM	FSA	TSP	GEM	FSA	TSP	GEM
$\bar{\bar{e}}$	$\tilde{\pi}_1$	−0.000041	−0.000029	−0.000029	−0.000018	−0.000013	−0.000013	−0.000011	−0.000010	−0.000010	−0.000037	−0.000030	−0.000030
	$\tilde{\delta}$	0.000040	0.000125	0.002268	0.000020	0.000105	0.000982	0.000014	−0.000047	−0.000235	0.000040	0.000052	0.001381
	$\tilde{\rho}$	−0.004573	0.016995	0.016998	−0.001952	0.016597	0.016597	−0.001399	0.018878	0.018878	−0.002544	0.010935	0.010935
SD of $\bar{e}_i$	$\tilde{\pi}_1$	0.000960	0.000839	0.000835	0.000506	0.000492	0.000490	0.000426	0.000420	0.000419	0.000966	0.000781	0.000781
	$\tilde{\delta}$	0.001081	0.002351	0.040853	0.000563	0.001704	0.039240	0.000472	0.001902	0.041140	0.000989	0.001453	0.041265
	$\tilde{\rho}$	0.000985	0.025433	0.025429	0.000619	0.022597	0.022596	0.000475	0.025647	0.025647	0.000740	0.015697	0.015697
MSE	$\tilde{\pi}_1$	0.005182	0.005231	0.005231	0.002243	0.002264	0.002264	0.001576	0.001589	0.001589	0.005146	0.005202	0.005202
	$\tilde{\delta}$	0.006235	0.006223	0.006046	0.002695	0.002691	0.003429	0.001893	0.001889	0.003018	0.005672	0.005693	0.005703
	$\tilde{\rho}$	0.004862	0.005465	0.005465	0.002078	0.002741	0.002741	0.001459	0.002379	0.002379	0.002681	0.002927	0.002927

Table 8: Errors of the constrained MLEs for $g = 9$.

Index	MLE	Scenario 9			Scenario 10			Scenario 11			Scenario 12		
		FSA	TSP	GEM	FSA	TSP	GEM	FSA	TSP	GEM	FSA	TSP	GEM
$\bar{\bar{e}}$	$\tilde{\pi}_1$	0.000003	−0.000005	−0.000005	0.000009	0.000007	0.000007	−0.000003	−0.000007	−0.000007	−0.000032		−0.000010
													−0.000010
	$\tilde{\delta}$	0.000006	−0.000017	−0.000751	−0.000015	−0.000020	−0.000252	−0.000003	−0.000036	−0.000499	0.000035		0.000019
		0.001520	$\tilde{\rho}$	−0.003009	0.011971	0.011972	−0.001354	0.013596	0.013596	−0.000923	0.013603		0.013604
		−0.001333	0.006205		0.006205								
SD of $\bar{e}_i$	$\tilde{\pi}_1$	0.000969	0.000827	0.000825	0.000511	0.000491	0.000490	0.000424	0.000412	0.000412	0.001014		0.000807
		0.000807	$\tilde{\delta}$	0.001022	0.001715	0.026804	0.000546	0.001368	0.027170	0.000445	0.001238		0.026898
		0.001025	0.001114		0.027407								
	$\tilde{\rho}$	0.000810	0.017777	0.017775	0.000462	0.018045	0.018045	0.000384	0.017287	0.017287	0.000472		0.009352
		0.009352											
MSE	$\tilde{\pi}_1$	0.005190	0.005243	0.005243	0.002212	0.002235	0.002235	0.001547	0.001562	0.001562	0.005125		0.005184
		0.005184	$\tilde{\delta}$	0.005847	0.005860	0.005383	0.002491	0.002496	0.002724	0.001744	0.001748		0.002114
		0.005387	0.005428		0.005078								
	$\tilde{\rho}$	0.003205	0.003512	0.003512	0.001409	0.001852	0.001852	0.000977	0.001418	0.001418	0.001399		0.001489
		0.001489											

Table 9: Failures of algorithms for global MLEs converging within k iterations.

Algorithm	ϵ	k	$g = 3$			$g = 6$				$g = 9$				
			Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8	Scenario 9	Scenario 10	Scenario 11	Scenario 12
FSA	1×10^{-5}	100	1	0	0	0	0	0	0	0	0	0	0	0
	1×10^{-5}	200	1	0	0	0	0	0	0	0	0	0	0	0
	1×10^{-4}	100	1	0	0	0	0	0	0	0	0	0	0	0
TSM	1×10^{-5}	100	243317	135083	106116	125954	53483	25611	20143	36410	38960	7298	6831	6032
	1×10^{-5}	200	243317	135083	106116	125954	53483	25611	20143	36410	38960	7298	6831	6032
	1×10^{-4}	100	236336	133270	104018	124032	52662	25070	19915	35809	37998	7139	6742	5896
GEM	1×10^{-5}	100	233564	199442	182533	182567	94508	85098	65703	86451	49811	37771	31859	22437
	1×10^{-5}	200	42208	35038	33405	35059	21286	19705	16426	19154	12860	12575	11422	9941
	1×10^{-4}	100	616	118	46	94	1	0	0	1	0	0	0	0

Table 10: Failures of algorithms for constrained MLEs converging within k iterations.

Algorithm	ϵ	k	$g = 3$			$g = 6$					$g = 9$			
			Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8	Scenario 9	Scenario 10	Scenario 11	Scenario 12
FSA	1×10^{-5}	100	3	0	0	0	0	0	0	0	0	0	0	0
	1×10^{-5}	200	1	0	0	0	0	0	0	0	0	0	0	0
	1×10^{-4}	100	1	0	0	0	0	0	0	0	0	0	0	0
TSP	1×10^{-5}	100	133968	54065	39939	2255	50619	23128	15247	695	31434	15364	10550	404
	1×10^{-5}	200	133968	54065	39939	2255	50619	23128	15247	695	31434	15364	10550	404
	1×10^{-4}	100	133968	54065	39939	2255	50619	23128	15247	695	31434	15364	10550	404
GEM	1×10^{-5}	100	0	0	0	0	0	0	0	0	0	0	0	0
	1×10^{-5}	200	0	0	0	0	0	0	0	0	0	0	0	0
	1×10^{-4}	100	0	0	0	0	0	0	0	0	0	0	0	0

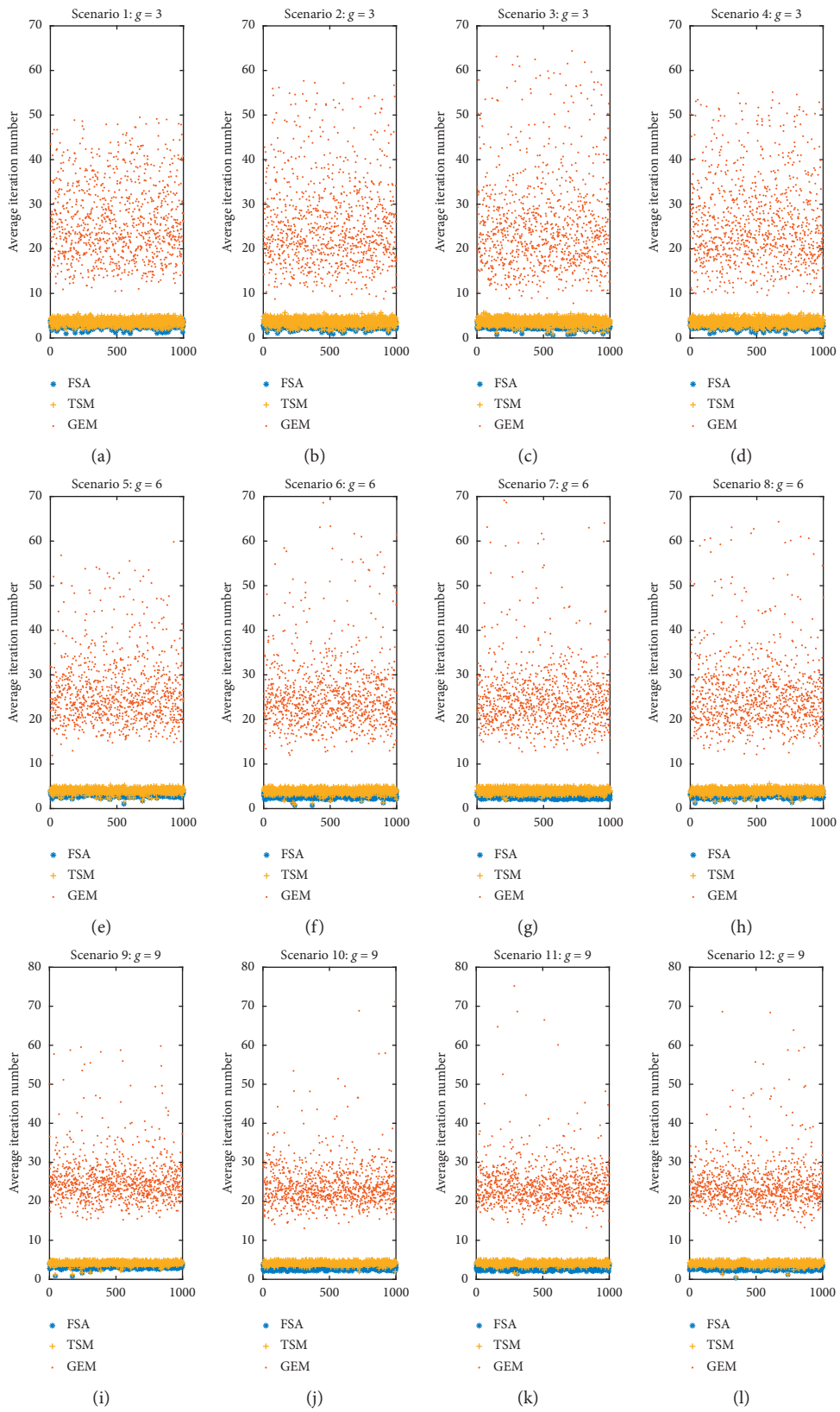


Figure 1: Number of iteration of algorithms for global MLEs from 1000 parameter settings.

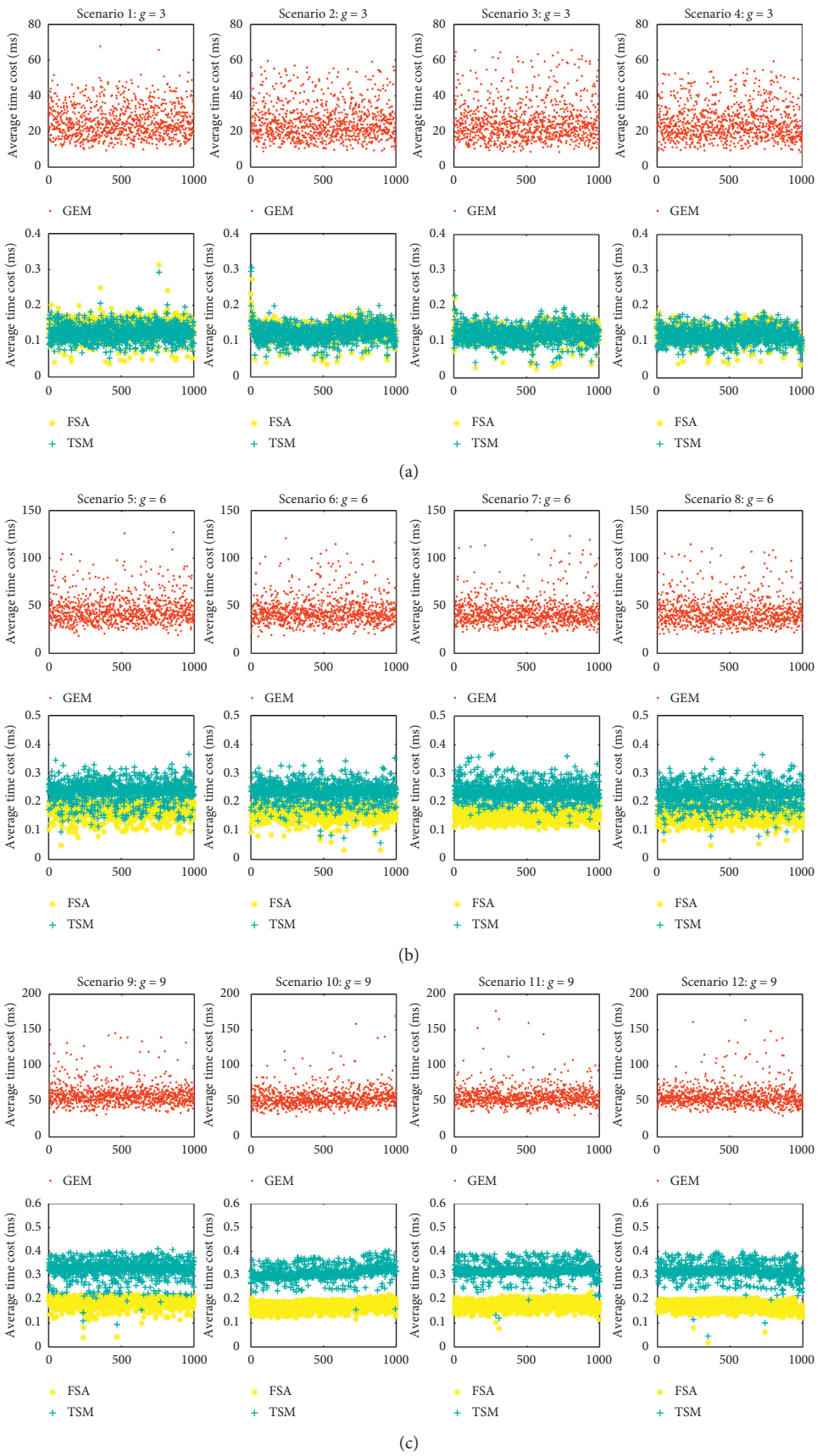


Figure 2: Time cost of algorithms for global MLEs from 1000 parameter settings.

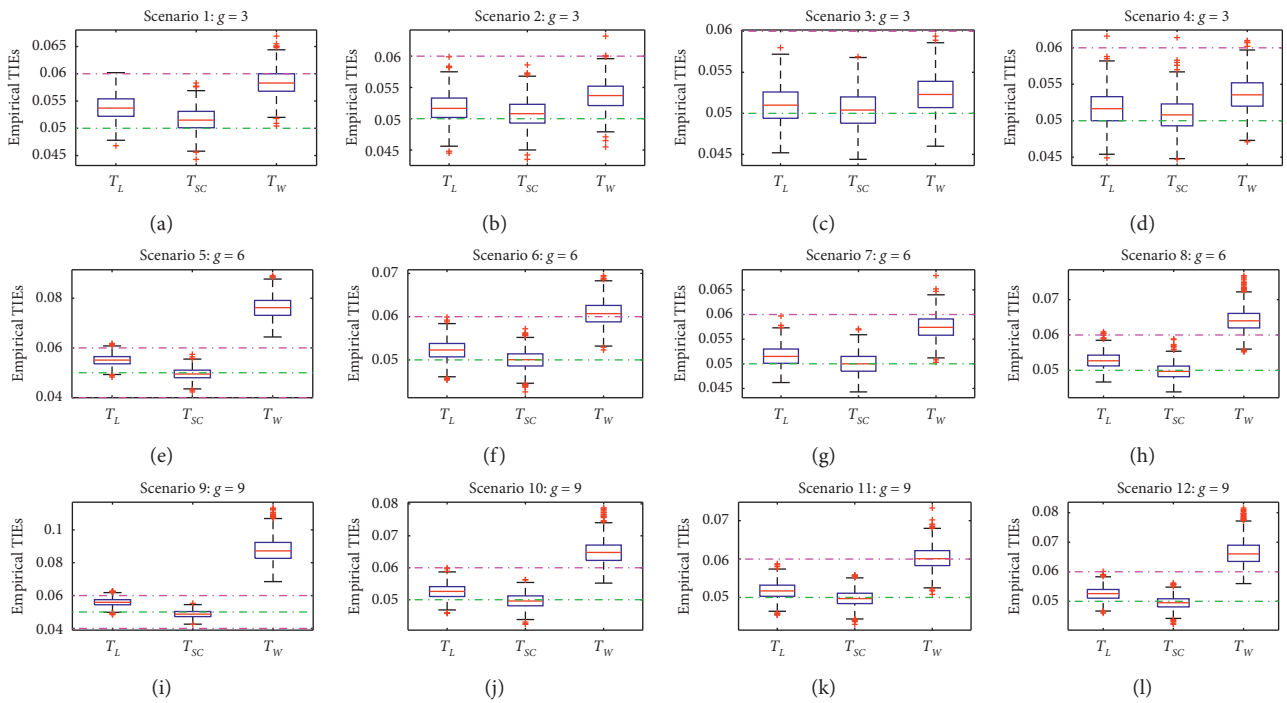
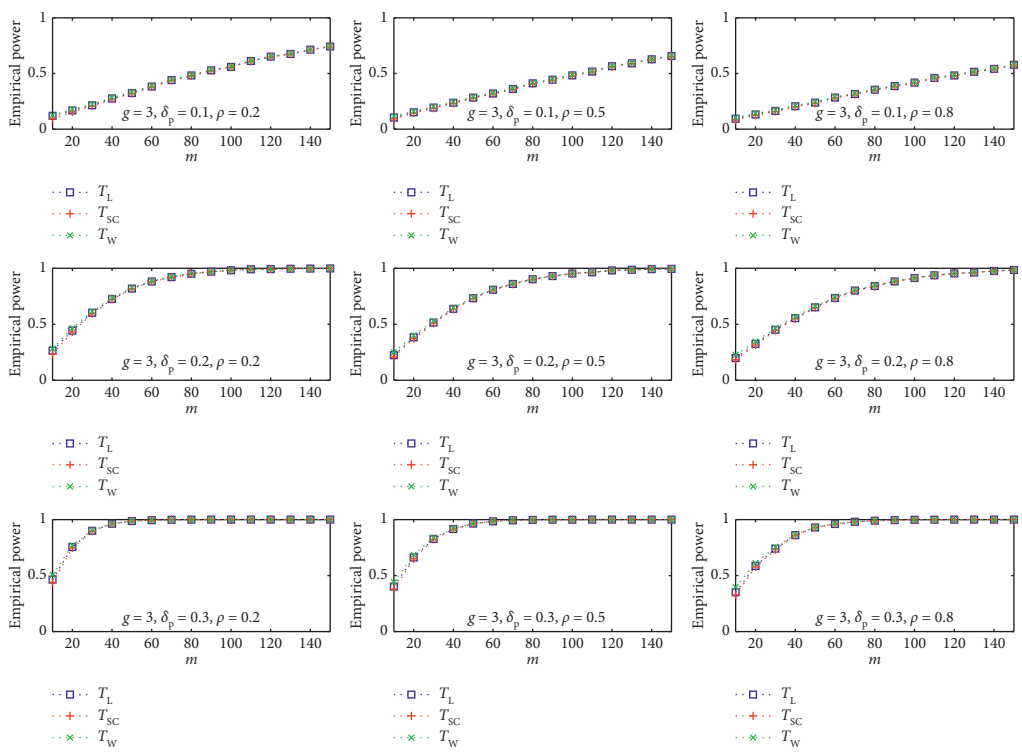


Figure 3: Empirical TIEs of test statistics.



(a)

Figure 4: Continued.

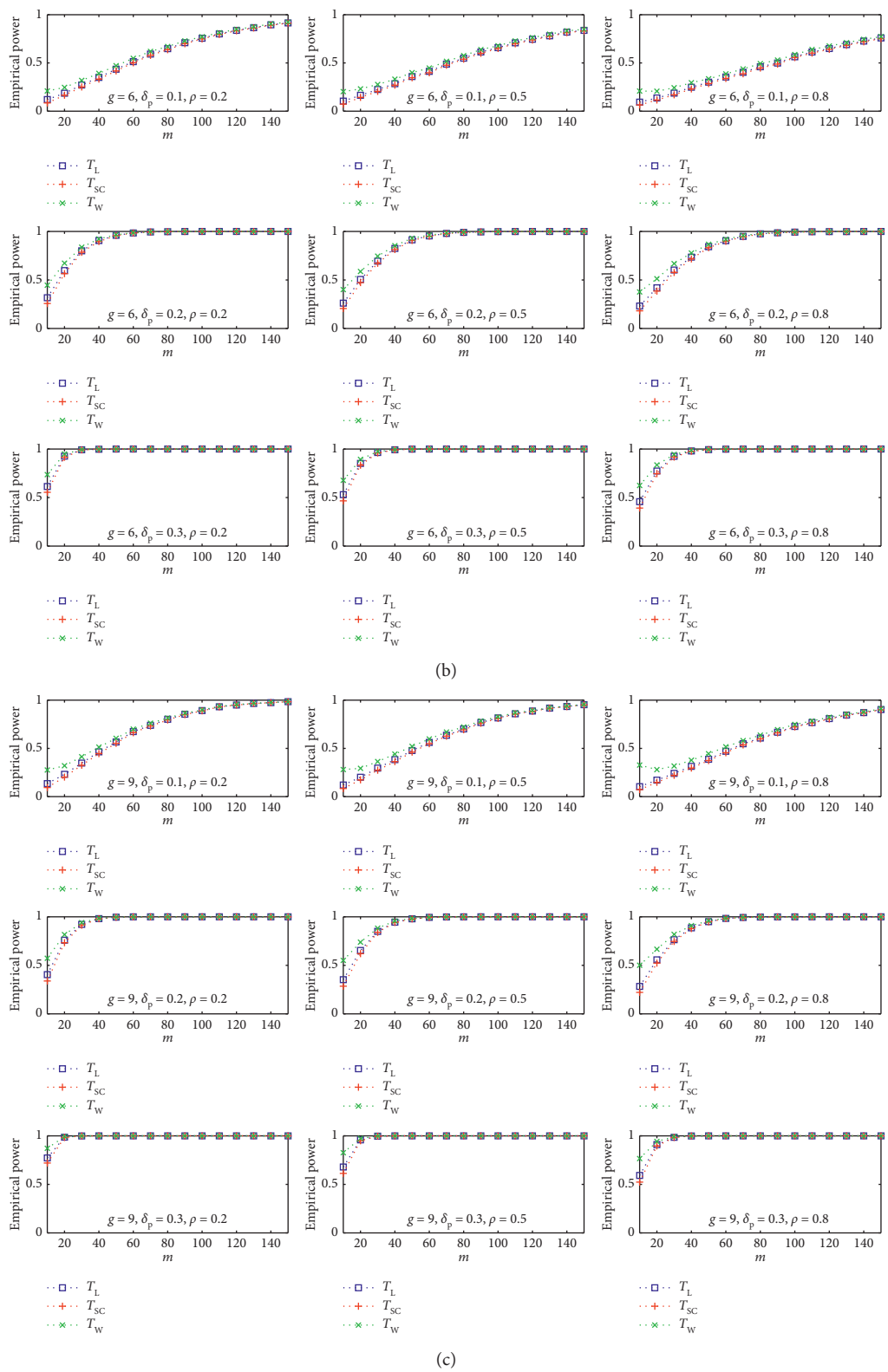


Figure 4: Empirical power of test statistics.

T_L, T_{SC}, and T_W based on global and constrained MLEs from the algorithm. The empirical TIEs and power of three test statistics can be demonstrated by Monte Carlo simulations. All tests are conducted at the significance level of $\alpha = 0.05$.

4.2.1. Type I Error Rate. TIE is the probability of rejecting a null hypothesis that is true. According to Tang et al. [2], a test is robust if its empirical TIE is between 0.04 and 0.06; liberal if it is greater than 0.06; and conservative if it is less than 0.04

Table 11: The number of affected eyes in each genetic type.

Number of affected eyes (l)	Genetic type			
	DOM	AR	SL	ISO
0	15	7	3	67
1	6	5	2	24
2	7	9	14	57

Table 12: Global and constrained MLEs.

Algorithm	$\widehat{\pi}_1$	$\widehat{\pi}_2$	$\widehat{\pi}_3$	$\widehat{\pi}_4$	$\widehat{\rho}$	$\widetilde{\pi}_1$	$\widetilde{\delta}$	$\widetilde{\rho}$
FSA	0.3625	0.5455	0.7926	0.4658	0.6416	0.3631	0.1449	0.6537
TSM/TSP	0.3625	0.5455	0.7926	0.4658	0.6416	0.3571	0.1509	0.6572
GEM	0.3625	0.5455	0.7926	0.4658	0.6416	0.3334	0.1779	0.6558

Table 13: Test statistics' values and p values.

Statistic	T_L	T_{SC}	T_W
Value	9.4332	8.7076	13.0135
p value	0.0089	0.0129	0.0015

Table 14: Prevalence of avoidable blindness from a sample population in Iran.

Blindness	Age group (unit: yrs)						
	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74	75 – 79	80+
0	964	541	469	257	242	127	104
1	23	17	18	16	32	30	29
2	2	8	4	5	3	9	10

Table 15: Global and constrained MLEs.

Algorithm	$\widehat{\pi}_1$	$\widehat{\pi}_2$	$\widehat{\pi}_3$	$\widehat{\pi}_4$	$\widehat{\pi}_5$	$\widehat{\pi}_6$	$\widehat{\pi}_7$	$\widehat{\rho}$	$\widetilde{\pi}_1$	$\widetilde{\delta}$	$\widetilde{\rho}$
FSA	0.0146	0.0262	0.0262	0.0449	0.0736	0.1446	0.1706	0.2701	0.0148	0.0419	0.2942
TSM/TSP	0.0146	0.0262	0.0262	0.0449	0.0736	0.1446	0.1706	0.2701	0.0137	0.0433	0.3024
GEM	0.0147	0.0262	0.0263	0.0450	0.0737	0.1448	0.1708	0.2739	0.0111	0.0462	0.3036

when the significance level is $\alpha = 0.05$. Power is the probability of correctly rejecting a null hypothesis when it is false in a statistical test. A good test should not only be robust but also make power as high as possible.

1,000 random parameter settings involving π_i and ρ $(i = 1, 2, \ldots, g)$ are generated under H_0, where g and $\mathbf{N}$ are shown in Table 2. 10,000 samples are randomly produced for every parameter setting. The empirical TIEs can be computed by dividing the number of times of rejecting H_0 with 10,000. Figure 3 further reflects the comprehensive performance of T_L, T_{SC}, and T_W. The empirical TIEs of T_{SC} are smaller and close to 0.05, which means score test is more robust. Wald-type test statistic tends to be liberal when $N_i (i = 1, 2, \ldots, g)$ is small (i.e., $N_i = 30, 70$) and unbalanced. Wald-type test statistic is more liberal when g is bigger. Empirical TIEs of three tests are closer to 0.05 as N_i increases. Balanced designs make test statistics have better performance than unbalanced designs.

4.2.2. Power. Let $N_1 = N_2 = \cdots = N_g = m$, $g = 3, 6, 9$, $\pi_1 = 0.2$, $\rho = 0.2, 0.5, 0.8$, $\delta_p = 0.1, 0.2, 0.3$, $\pi_i = \pi_1 + \delta_p$ when $i = 2k$, and $\pi_i = \pi_1$ when $i = 2k + 1\,(k = 1, 2, \ldots)$. 10,000 samples will be randomly generated under every parameter setting when m changes from 10 to 160 at intervals of 10. Empirical power can be computed by dividing the number of times of rejecting H_0 with 10,000. Figure 4 reflects how empirical power changes as m changes. T_W has higher power and T_L and T_{SC} have close power. Empirical power of three test statistics is close to each other as m increases. Increasing the number of groups can produce higher power. It means that all test statistics work well under multigroup cases. The moderately and highly relevant data (i.e., $\rho = 0.5$ and $\rho = 0.8$) have lower power than the mildly relevant data (i.e., $\rho = 0.2$).

According to the above results, score test is more robust and has satisfactory power. Thus, it should be recommended for the homogeneity test about many-to-one comparison of risk differences.

Table 16: Test statistics' values and p values.

Statistic	T_L	T_{SC}	T_W
Value	107.4057	127.2814	72.2784
p value	1.4474×10^{-21}	8.9818×10^{-26}	3.4369×10^{-14}

5. Real Examples

In this section, we introduce two real examples to illustrate the aforementioned methods. The first example was presented by Rosner [1] to illustrate the new proposed methods. As shown in Table 11, 216 persons aged 20–39 with retinitis pigmentosa (RP) were classified into four genetic types including autosomal dominant RP (DOM), autosomal recessive RP (AR), sex-linked RP (SL), and isolate RP (ISO). The results from the four groups were assessed by the Snellen visual acuity (VA). An eye was considered affected if VA was 20/50 or worse and normal if VA was 20/40 or better.

Through Fisher scoring algorithm, two-step method, and GEM algorithm, global MLEs $\widehat{\pi}_1, \widehat{\pi}_2, \widehat{\pi}_3, \widehat{\pi}_4$, and $\widehat{\rho}$ are listed in Table 12. We observe that three algorithms can produce the same results. According to global MLEs of proportions in AR, SL, and ISO groups, estimated risk differences can be calculated as $\widehat{\delta}_2 = 0.5455 - 0.3625 = 0.1830$, $\widehat{\delta}_3 = 0.7926 - 0.3625 = 0.4301$, and $\widehat{\delta}_4 = 0.4658 - 0.3625 = 0.1033$. Constrained MLEs $\widetilde{\pi}_1, \widetilde{\delta}, \widetilde{\rho}$ vary slightly in Fisher scoring algorithm, two-stage procedure, and GEM algorithm as shown in Table 12.

Through Fisher scoring algorithms, the values of T_L, T_{SC}, and T_W are 9.4332, 8.7076, and 13.0135, respectively, bigger than 95% percentile of the chi-square distribution with 2 degrees of freedom. Three tests all have p values smaller than 0.05 (see Table 13). Thus, there is strong evidence to reject the null hypothesis H_0: $\delta_2 = \delta_3 = \delta_4 \triangleq \delta$ at a significance level of 0.05.

The second example is a cross-sectional population-based study about avoidable blindness in Iran by Rajavi et al. [15]. Nearly 3000 persons were examined, where blindness is assessed for seven age groups presented in Table 14.

Table 15 provides global MLEs and constrained MLEs. The common risk differences δ from three algorithms are estimated to be 0.0419, 0.0433, and 0.0462. Table 16 shows that p values of test statistics T_L, T_{SC}, and T_W are 1.4474×10^{-21}, 8.9818×10^{-26}, and 3.4369×10^{-14}. So we have enough evidence to reject H_0: $\delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = \delta_7 \triangleq \delta$ at a significance level of 0.05.

6. Conclusion

This paper mainly studies the many-to-one comparison of risk differences for correlated binary data. Three algorithms, including Fisher scoring algorithm, two-step method, and GEM algorithm, are introduced for global MLEs. Constrained MLEs are calculated by Fisher scoring algorithm, two-stage procedure, and GEM algorithm. Simulations show that Fisher scoring algorithm, two-step method, and GEM algorithm can produce global MLEs with close MSEs. However, Fisher scoring algorithm for global MLEs behaves better in terms of convergence rate. Fisher scoring algorithm can obtain constrained MLEs with lower MSEs. Thus, it is recommended for getting constrained MLEs. Likelihood ratio, score, and Wald-type tests are constructed based on global and constrained MLEs from Fisher scoring algorithms. The results of simulations show that score test statistic is robust and has high empirical power.

In particular, Fisher scoring algorithm may not work well with high dimensional Fisher information matrix when the number of groups is massive. Then, the two-step method or two-stage procedure can be adopted to solve the high dimensional problems. As the modified Fisher scoring algorithm or Newton–Raphson algorithm, they only need to calculate a small part of parameters at every step to reduce dimension. GEM algorithm is also an alternative approach regardless of time cost. The biggest obstacle for GEM algorithm may occur in M step when the local optimal solution does not exist or cannot be found. Furthermore, the convergent performance of GEM algorithm is sensitive for the convergence accuracy and the iteration number.

In this article, we investigate the homogeneity test of many-to-one risk differences in multigroup design. If samples can be stratified by some control variables (e.g., age, gender, etc.), the treatment-by-stratum interaction should be considered. Thus, it is of great significance to develop many-to-one comparison in stratified correlated binary data. Other indexes such as risk ratio can be adopted for evaluating differences of proportions instead of risk difference.

Appendix

A. Derivation for Fisher Information Matrix I_1

Differentiating $(\partial l_i/\partial \pi_i)$ and $(\partial \ell/\partial \rho)$ with respect to π_i and ρ, we have

$$\frac{\partial^2 l_i}{\partial \pi_i^2} = -\frac{\left(2\pi_i^2 - 2\pi_i + 1\right)m_{1i}}{\pi_i^2\left(\pi_i - 1\right)^2} - \frac{\left(2\rho^2\pi_i^2 - 2\rho^2\pi_i + \rho^2 - 4\rho\pi_i^2 + 2\rho\pi_i + 2\pi_i^2\right)m_{2i}}{\pi_i^2\left(\rho + \pi_i - \rho\pi_i\right)^2}$$

$$-\frac{\left(2\rho^2\pi_i^2 - 2\rho^2\pi_i + \rho^2 - 4\rho\pi_i^2 + 6\rho\pi_i - 2\rho + 2\pi_i^2 - 4\pi_i + 2\right)m_{0i}}{\left(\pi_i - 1\right)^2\left(\rho\pi_i - \pi_i + 1\right)^2},$$

$$\frac{\partial^2 l_i}{\partial \pi_i \partial \rho} = \frac{m_{0i}}{\left(\rho\pi_i - \pi_i + 1\right)^2} - \frac{m_{2i}}{\left(\rho\pi_i - \pi_i - \rho\right)^2}, \quad i = 1, 2, \ldots, g, \tag{A.1}$$

$$\frac{\partial^2 \ell}{\partial \rho^2} = -\sum_{i=1}^{g}\left[\frac{\pi_i^2 m_{0i}}{\left(\rho\pi_i - \pi_i + 1\right)^2} + \frac{m_{1i}}{\left(\rho - 1\right)^2} + \frac{\left(\pi_i - 1\right)^2 m_{2i}}{\left(\rho + \pi_i - \rho\pi_i\right)^2}\right].$$

The expected values of the second-order differential equations can be calculated by substituting m_{0i}, m_{1i}, m_{2i} with $N_i p_{0i}, N_i p_{1i}, N_i p_{2i}$, respectively. So, we have

$$E\left(-\frac{\partial^2 l_i}{\partial \pi_i^2}\right) = \frac{N_i\left(-4\rho^2\pi_i^2 + 4\rho^2\pi_i - \rho^2 + 6\rho\pi_i^2 - 6\rho\pi_i + 2\rho - 2\pi_i^2 + 2\pi_i\right)}{\pi_i\left(1 - \pi_i\right)\left(\rho + \pi_i - \rho\pi_i\right)\left(\rho\pi_i - \pi_i + 1\right)},$$

$$E\left(-\frac{\partial^2 l_i}{\partial \pi_i \partial \rho}\right) = \frac{N_i\rho\left(2\pi_i - 1\right)}{\left(\rho + \pi_i - \rho\pi_i\right)\left(\rho\pi_i - \pi_i + 1\right)}, \tag{A.2}$$

$$E\left(-\frac{\partial^2 \ell}{\partial \rho^2}\right) = \sum_{i=1}^{g}\frac{N_i\pi_i\left(\rho + 1\right)\left(1 - \pi_i\right)}{\left(1 - \rho\right)\left(\rho + \pi_i - \rho\pi_i\right)\left(\rho\pi_i - \pi_i + 1\right)}.$$

Thus, I_1 can be derived as

$$I_1 = E\begin{bmatrix} -\frac{\partial^2 l_1}{\partial \pi_1^2} & 0 & \ldots & 0 & -\frac{\partial^2 l_1}{\partial \pi_1 \partial \rho} \\ 0 & -\frac{\partial^2 l_2}{\partial \pi_2^2} & \ldots & 0 & -\frac{\partial^2 l_2}{\partial \pi_2 \partial \rho} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & -\frac{\partial^2 l_g}{\partial \pi_g^2} & -\frac{\partial^2 l_g}{\partial \pi_g \partial \rho} \\ -\frac{\partial^2 l_1}{\partial \pi_1 \partial \rho} & -\frac{\partial^2 l_2}{\partial \pi_2 \partial \rho} & \ldots & -\frac{\partial^2 l_g}{\partial \pi_g \partial \rho} & -\frac{\partial^2 \ell}{\partial \rho^2} \end{bmatrix}. \tag{A.3}$$

B. Derivation for Fisher Information Matrixes I_2 and I_3

The second-order differential equations of δ, ρ, π_1 under H_0 are

$$\frac{\partial^2 \ell_0}{\partial \delta^2} = \sum_{i=2}^{g}\left[m_{0i}\left\{-\frac{2(\rho-1)}{(\delta+\pi_1-1)[(\delta+\pi_1)(1-\rho)-1]} + \frac{(\rho-1)(2\delta+2\pi_1-1)+1}{(\delta+\pi_1-1)^2[(\delta+\pi_1)(1-\rho)-1]}\right.\right.$$

$$\left.-\frac{(\rho-1)[(\rho-1)(2\delta+2\pi_1-1)+1]}{(\delta+\pi_1-1)[(\delta+\pi_1)(1-\rho)-1]^2}\right\} + m_{2i}\left\{\frac{-[2(\delta+\pi_1)(1-\rho)+\rho]^2}{\left[(\delta+\pi_1)^2(1-\rho)+\rho(\delta+\pi_1)\right]^2} - \frac{2(\rho-1)}{(\delta+\pi_1)^2(1-\rho)+\rho(\delta+\pi_1)}\right\}$$

$$+ m_{1i}\left\{\frac{2}{(\delta+\pi_1)(\delta+\pi_1-1)} - \frac{2\delta+2\pi_1-1}{(\delta+\pi_1)(\delta+\pi_1-1)^2} - \frac{2\delta+2\pi_1-1}{(\delta+\pi_1)^2(\delta+\pi_1-1)}\right\}, \frac{\partial^2 \ell_0}{\partial \rho^2} = -\frac{\pi_1^2 m_{01}}{(\rho\pi_1-\pi_1+1)^2} - \frac{m_{11}}{(\rho-1)^2}$$

$$-\frac{(\pi_1-1)^2 m_{21}}{(\rho+\pi_1-\rho\pi_1)^2} - \sum_{i=2}^{g}\left[\frac{(\pi_1+\delta)^2 m_{0i}}{[\rho(\pi_1+\delta)-\pi_1-\delta+1]^2} + \frac{m_{1i}}{(\rho-1)^2} + \frac{(\pi_1+\delta-1)^2 m_{2i}}{[\rho+\pi_1+\delta-\rho(\pi_1+\delta)]^2}\right], \frac{\partial^2 \ell_0}{\partial \pi_1^2}$$

$$= -\frac{(2\pi_1^2-2\pi_1+1)m_{11}}{\pi_1^2(\pi_1-1)^2} - \frac{(2\rho^2\pi_1^2-2\rho^2\pi_1+\rho^2-4\rho\pi_1^2+2\rho\pi_1+2\pi_1^2)m_{21}}{\pi_1^2(\rho+\pi_1-\rho\pi_1)^2}$$

$$= \frac{(2\rho^2\pi_1^2-2\rho^2\pi_1+\rho^2-4\rho\pi_1^2+6\rho\pi_1-2\rho+2\pi_1^2-4\pi_1+2)m_{01}}{(\pi_1-1)^2(\rho\pi_1-\pi_1+1)^2} + \frac{\partial^2 \ell_0}{\partial \delta^2}, \frac{\partial^2 \ell_0}{\partial \delta\, \partial \rho}$$

$$= \sum_{i=2}^{g}\left[m_{2i}\left\{\frac{(\delta+\pi_1)(\delta+\pi_1-1)[2(\delta+\pi_1)(1-\rho)+\rho]}{\left[(\delta+\pi_1)^2-\rho(\delta+\pi_1)(\delta+\pi_1-1)\right]^2} - \frac{1}{(\delta+\pi_1)(1-\rho)\rho} - \frac{\delta+\pi_1-1}{(\delta+\pi_1)^2-\rho(\delta+\pi_1)(\delta+\pi_1-1)}\right\}\right.$$

$$\left.+\frac{m_{0i}}{[(\delta+\pi_1)(1-\rho)-1]^2}\right], \frac{\partial^2 \ell_0}{\partial \delta\, \partial \pi_1} = \frac{\partial^2 \ell_0}{\partial \delta^2}, \frac{\partial^2 \ell_0}{\partial \pi_1 \partial \rho} = \frac{\partial^2 \ell_0}{\partial \delta\, \partial \rho}. \tag{B.1}$$

Then,

$$
\begin{aligned}
-E\left(\frac{\partial^2 \ell_0}{\partial \delta^2}\right) &= \sum_{i=2}^{g} N_i \left[\frac{2(\rho-1)(\delta+\pi_1)-\rho}{(\delta+\pi_1-1)} + \frac{[2(\delta+\pi_1)(1-\rho)+\rho]^2}{\left[(\delta+\pi_1)^2-\rho(\delta+\pi_1)(\delta+\pi_1-1)\right]} + \frac{2(\rho-1)(2\delta+2\pi_1-1)}{\delta+\pi_1}\right. \\
&\left. + \frac{(\rho-1)[(\rho-1)(2\delta+2\pi_1-1)+1]}{(\delta+\pi_1)(1-\rho)-1}\right], \\
-E\left(\frac{\partial^2 \ell_0}{\partial \rho^2}\right) &= \frac{N_1\pi_1(\rho+1)(1-\pi_1)}{(1-\rho)(\rho+\pi_1-\rho\pi_1)(\rho\pi_1-\pi_1+1)} + \sum_{i=2}^{g} N_i \left[\frac{2(\delta+\pi_1)(\delta+\pi_1-1)}{\rho-1} + \frac{(\delta+\pi_1)^2(\delta+\pi_1-1)}{(\delta+\pi_1)(1-\rho)-1}\right. \\
&\left. + \frac{(\delta+\pi_1)(\delta+\pi_1-1)^2}{(\delta+\pi_1)(1-\rho)+\rho}\right], \\
-E\left(\frac{\partial^2 \ell_0}{\partial \pi_1^2}\right) &= \frac{N_1\left(-4\rho^2\pi_1^2+4\rho^2\pi_1-\rho^2+6\rho\pi_1^2-6\rho\pi_1+2\rho-2\pi_1^2+2\pi_1\right)}{\pi_1(1-\pi_1)(\rho+\pi_1-\rho\pi_1)(\rho\pi_1-\pi_1+1)} - E\left(\frac{\partial^2 \ell_0}{\partial \delta^2}\right), \\
-E\left(\frac{\partial^2 \ell_0}{\partial \delta\, \partial \rho}\right) &= \sum_{i=2}^{g} N_i \left[2(2\delta+2\pi_1-1) - \frac{(\delta+\pi_1-1)[2(\delta+\pi_1)(1-\rho)+\rho]}{(\delta+\pi_1)(1-\rho)+\rho} + \frac{(\delta+\pi_1)[2(\delta+\pi_1)(1-\rho)+\rho-2]}{(\delta+\pi_1)(\rho-1)+1}\right], \\
-E\left(\frac{\partial^2 \ell_0}{\partial \delta\, \partial \pi_1}\right) &= -E\left(\frac{\partial^2 \ell_0}{\partial \delta^2}\right), \\
-E\left(\frac{\partial^2 \ell_0}{\partial \pi_1 \partial \rho}\right) &= -E\left(\frac{\partial^2 \ell_0}{\partial \delta\, \partial \rho}\right).
\end{aligned}
\tag{B.2}
$$

Finally, we have

$$
I_2 = -E\begin{bmatrix} \frac{\partial^2 \ell_0}{\partial \delta^2} & \frac{\partial^2 \ell_0}{\partial \delta\, \partial \pi_1} & \frac{\partial^2 \ell_0}{\partial \delta\, \partial \rho} \\ \frac{\partial^2 \ell_0}{\partial \delta\, \pi_1} & \frac{\partial^2 \ell_0}{\partial \pi_1^2} & \frac{\partial^2 \ell_0}{\partial \pi_1 \partial \rho} \\ \frac{\partial^2 \ell_0}{\partial \delta\, \partial \rho} & \frac{\partial^2 \ell_0}{\partial \pi_1 \partial \rho} & \frac{\partial^2 \ell_0}{\partial \rho^2} \end{bmatrix},
\tag{B.3}
$$

$$
I_3 = -E\begin{bmatrix} \frac{\partial^2 \ell_0}{\partial \pi_1^2} & \frac{\partial^2 \ell_0}{\partial \pi_1 \partial \rho} \\ \frac{\partial^2 \ell_0}{\partial \pi_1 \partial \rho} & \frac{\partial^2 \ell_0}{\partial \rho^2} \end{bmatrix}.
$$

C. Derivation for Score Test Statistic

Score test statistic is

$$
T_{SC} = \mathbf{U} I_4\left(\pi_2, \ldots, \pi_g, \pi_1, \rho\right)^{-1} \mathbf{U}^T \big|_{\boldsymbol{\pi}=\tilde{\boldsymbol{\pi}}, \rho=\tilde{\rho}},
\tag{C.1}
$$

where

$$
I_4 = -E\begin{bmatrix} \frac{\partial^2 l_2}{\partial \pi_2^2} & & & & \frac{\partial^2 l_2}{\partial \pi_2 \partial \rho} \\ & \ddots & & & \vdots \\ & & \frac{\partial^2 l_g}{\partial \pi_g^2} & & \frac{\partial^2 l_g}{\partial \pi_g \partial \rho} \\ & & & \frac{\partial^2 l_1}{\partial \pi_1^2} & \frac{\partial^2 l_1}{\partial \pi_1 \partial \rho} \\ \frac{\partial^2 l_2}{\partial \pi_2 \partial \rho} & \cdots & \frac{\partial^2 l_g}{\partial \pi_g \partial \rho} & \frac{\partial^2 l_1}{\partial \pi_1 \partial \rho} & \frac{\partial^2 \ell}{\partial \rho^2} \end{bmatrix},
\tag{C.2}
$$

$$
U = \left(\frac{\partial l_2}{\partial \pi_2}, \ldots, \frac{\partial l_g}{\partial \pi_g}, 0, 0\right),
$$

and the expected values of the second-order differential equations are given in Appendix A.

Let

$$
\begin{aligned}
I_4 &= \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & D \end{bmatrix}, \\
\mathbf{A} &= \begin{bmatrix} E\left(-\frac{\partial^2 l_2}{\partial \pi_2^2}\right) & & & \\ & \ddots & & \\ & & E\left(-\frac{\partial^2 l_g}{\partial \pi_g^2}\right) & \\ & & & E\left(-\frac{\partial^2 l_1}{\partial \pi_1^2}\right) \end{bmatrix} \\
&= \begin{bmatrix} a_2 & & & \\ & \ddots & & \\ & & a_g & \\ & & & a_1 \end{bmatrix}, \\
\mathbf{B} &= \left(E\left(-\frac{\partial^2 l_2}{\partial \pi_2 \partial \rho}\right), \ldots, E\left(-\frac{\partial^2 l_g}{\partial \pi_g \partial \rho}\right), E\left(-\frac{\partial^2 l_1}{\partial \pi_1 \partial \rho}\right)\right)^T \\
&= \left(b_2, \ldots, b_g, b_1\right)^T, \\
D &= E\left(-\frac{\partial^2 \ell}{\partial \rho^2}\right).
\end{aligned}
\tag{C.3}
$$

So, the inverse matrix of I_4 can be derived as

$$
I_4^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\left(D - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\left(D - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \\ -\left(D - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{A}^{-1} & \left(D - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix}, \tag{C.4}
$$

where

$$
\mathbf{A}^{-1} = \begin{bmatrix} a_2^{-1} & & & \\ & \ddots & & \\ & & a_g^{-1} & \\ & & & a_1^{-1} \end{bmatrix}. \tag{C.5}
$$

Let $\mathbf{U} = (u_2, \ldots, u_g, 0, 0)$. Thus,

$$\begin{aligned}\mathbf{U}I_4^{-1}\mathbf{U}^T &= \left(u_2, \ldots, u_g, 0\right)\left(\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\left(D - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{A}^{-1}\right)\left(u_2, \ldots, u_g, 0\right)^T \\ &= \left(u_2, \ldots, u_g, 0\right)\mathbf{A}^{-1}\left(u_2, \ldots, u_g, 0\right)^T + \frac{\left(u_2, \ldots, u_g, 0\right)\mathbf{A}^{-1}\mathbf{B}\mathbf{B}^T\mathbf{A}^{-1}\left(u_2, \ldots, u_g, 0\right)^T}{D - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}},\end{aligned}$$

$$\begin{aligned}\left(u_2, \ldots, u_g, 0\right)\mathbf{A}^{-1}\left(u_2, \ldots, u_g, 0\right)^T &= \sum_{i=2}^{g}\frac{u_i^2}{a_i}, \\ \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} &= \sum_{i=1}^{g}\frac{b_i^2}{a_i}, \\ \left(u_2, \ldots, u_g, 0\right)\mathbf{A}^{-1}\mathbf{B} &= \sum_{i=2}^{g}\frac{u_i b_i}{a_i}.\end{aligned} \tag{C.6}$$

We have score test statistic:

$$T_{SC} = \sum_{i=2}^{g}\frac{u_i^2\left(\widetilde{\pi}_i, \widetilde{\rho}\right)}{a_i\left(\widetilde{\pi}_i, \widetilde{\rho}\right)} + \frac{\left(\sum_{i=2}^{g}\left(u_i\left(\widetilde{\pi}_i, \widetilde{\rho}\right)b_i\left(\widetilde{\pi}_i, \widetilde{\rho}\right)\right)/a_i\left(\widetilde{\pi}_i, \widetilde{\rho}\right)\right)^2}{D\left(\widetilde{\boldsymbol{\pi}}, \widetilde{\rho}\right) - \sum_{i=1}^{g}\left(b_i^2\left(\widetilde{\pi}_i, \widetilde{\rho}\right)\right)/a_i\left(\widetilde{\pi}_i, \widetilde{\rho}\right)}. \tag{C.7}$$

D. Derivation for Wald-Type Statistic

Let $\beta = (\pi_2, \ldots, \pi_g, \pi_1, \rho)$ and

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & \ldots & \ldots & 0 & 0 & 0 \\ 0 & 1 & -1 & & & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & \ldots & \ldots & 0 & 1 & -1 & 0 & 0 \end{bmatrix}_{(g-2)\times(g+1)}. \tag{D.1}$$

Wald-type statistic is $T_W = (\beta\mathbf{C}^T)(\mathbf{C}I_4^{-1}\mathbf{C}^T)^{-1}(\mathbf{C}\beta^T)|_{\pi=\widehat{\pi},\rho=\widehat{\rho}}$.

The Fisher information matrix I_4 is the same as that of score test in Appendix C. Suppose

$$I_4^{-1} = \begin{bmatrix} I_{11} & I_{12} & \cdots & I_{1g} & I_{1(g+1)} \\ I_{21} & I_{22} & \cdots & I_{2g} & I_{2(g+1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ I_{(g+1)1} & I_{(g+1)2} & \cdots & I_{(g+1)g} & I_{(g+1)(g+1)} \end{bmatrix} = \left(I_{ij}\right), \tag{D.2}$$

and $F = CI_4^{-1}C^T = (f_{ij})$, $f_{ij} = (I_{ij} - I_{(i+1)j}) - (I_{i(j+i)} - I_{(i+1)(j+1)})$. Let $F = LL^T$ for convenience, where $L = (l_{ij})$ and

$$l_{ij} = \begin{cases} \dfrac{f_{ij} - \sum_{k=1}^{j-1} l_{ik}l_{jk}}{l_{jj}}, & i > j, \\ \left(f_{ij} - \sum_{k=1}^{j-1} l_{jk}^2\right)^{(1/2)}, i = j, 0, i < j. \end{cases} \tag{D.3}$$

So, the inverse of F is converted into $F^{-1} = (L^T)^{-1}L^{-1}$. We can obtain $L^{-1} = (r_{ij})$, where

$$r_{ij} = \begin{cases} -\dfrac{1}{l_{ii}}\sum_{k=j}^{i-1} l_{ik}r_{kj}, i > j, \dfrac{1}{l_{ij}}, i = j, 0, i < j. \end{cases} \tag{D.4}$$

Let $F^{-1} = (q_{ij})$. It is obvious that $q_{ij} = \sum_{k=1}^{g-2} r_{ki}r_{kj}$. Wald-type statistic can be simplified as

$$T_W = \sum_{i=1}^{g-2}\sum_{j=1}^{g-2}\left\{\left(\pi_{i+1} - \pi_{i+2}\right)\left(\pi_{j+1} - \pi_{j+2}\right)q_{ij}\right\}|_{\widehat{\boldsymbol{\pi}},\widehat{\rho}}. \tag{D.5}$$

Data Availability

The first example was presented by Rosner [1]. The second example is a cross-sectional population-based study about avoidable blindness in Iran by Rajavi et al. [15].

Conflicts of Interest

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

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant nos. 12061070 and 11661076) and the Science and Technology Department of Xinjiang Uygur Autonomous Region (grant no. 2018Q011).

References

[1] B. Rosner, "Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes," *Biometrics*, vol. 38, no. 1, pp. 105–114, 1982.

[2] M.-L. Tang, N.-S. Tang, and B. Rosner, "Statistical inference for correlated data in ophthalmologic studies," *Statistics in Medicine*, vol. 25, no. 16, pp. 2771–2783, 2006.

[3] C. X. Ma, G. Shan, and S. Liu, "Homogeneity test for correlated data in ophthalmologic studies," *PLoS One*, vol. 10, Article ID e0124337, 2015.

[4] G. Shan and C. Ma, "Exact methods for testing the equality of proportions for binary clustered data from otolaryngologic studies," *Statistics in Biopharmaceutical Research*, vol. 6, no. 1, pp. 115–122, 2014.

[5] X. Liu, G. Shan, L. Tian, and C.-X. Ma, "Exact methods for testing homogeneity of proportions for multiple groups of paired binary data," *Communications in Statistics—Simulation and Computation*, vol. 46, no. 8, pp. 6074–6082, 2016.

[6] G. E. Dallal, "Paired Bernoulli trials," *Biometrics*, vol. 44, no. 1, pp. 253–257, 1988.

[7] G. E. M'lan and M. H. Chen, "Objective Bayesian inference for bilateral data," *Bayesian Analysis*, vol. 10, pp. 139–170, 2015.

[8] A. Donner, "Statistical methods in ophthalmology: an adjusted chi-square approach," *Biometrics*, vol. 45, no. 2, pp. 605–611, 1989.

[9] J. R. Thompson, "The chi 2 test for data collected on eyes," *British Journal of Ophthalmology*, vol. 77, no. 2, pp. 115–117, 1993.

[10] Y. Pei, M.-L. Tang, W.-K. Wong, and N.-S. Tang, "Testing equality of correlations of two paired binary responses from two treated groups in a randomized trial," *Journal of Biopharmaceutical Statistics*, vol. 21, no. 3, pp. 511–525, 2011.

[11] N.-S. Tang, M.-L. Tang, and S.-F. Qiu, "Testing the equality of proportions for correlated otolaryngologic data," *Computational Statistics & Data Analysis*, vol. 52, no. 7, pp. 3719–3729, 2008.

[12] C.-X. Ma and S. Liu, "Testing equality of proportions for correlated binary data in ophthalmologic studies," *Journal of Biopharmaceutical Statistics*, vol. 27, no. 4, pp. 611–619, 2016.

[13] X. Liu, S. Liu, and C.-X. Ma, "Testing equality of correlation coefficients for paired binary data from multiple groups," *Journal of Statistical Computation and Simulation*, vol. 86, no. 9, pp. 1686–1696, 2016.

[14] X. Peng, C. Liu, S. Liu, and C.-X. Ma, "Asymptotic confidence interval construction for proportion ratio based on correlated paired data," *Journal of Biopharmaceutical Statistics*, vol. 29, no. 6, pp. 1137–1152, 2019.

[15] Z. Rajavi, M. Katibeh, H. Ziaei et al., "Rapid assessment of avoidable blindness in Iran," *Ophthalmology*, vol. 118, no. 9, pp. 1812–1818, 2011.

Hindawi
Complexity
Volume 2021, Article ID 6666699, 15 pages
https://doi.org/10.1155/2021/6666699

Research Article

Optimized Adaptive Neuro-Fuzzy Inference System Using Metaheuristic Algorithms: Application of Shield Tunnelling Ground Surface Settlement Prediction

Xinni Liu,[1,2] Sadaam Hadee Hussein,[3] Kamarul Hawari Ghazali,[2] Tran Minh Tung,[4] and Zaher Mundher Yaseen[4]

[1]*School of Information, Xi'an University of Finance and Economics, Xi'an, China*
[2]*Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, Pekan 26600, Malaysia*
[3]*Almaaref University College, Civil Engineering Department, Ramadi, Iraq*
[4]*Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam*

Correspondence should be addressed to Zaher Mundher Yaseen; yaseen@tdtu.edu.vn

Received 30 November 2020; Revised 20 February 2021; Accepted 24 February 2021; Published 12 March 2021

Academic Editor: Baogui Xin

Copyright © 2021 Xinni Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deformation of ground during tunnelling projects is one of the complex issues that is required to be monitored carefully to avoid the unexpected damages and human losses. Accurate prediction of ground settlement (GS) is a crucial concern for tunnelling problems, and the adequate predictive model can be a vital tool for tunnel designers to simulate the ground settlement accurately. This study proposes relatively new hybrid artificial intelligence (AI) models to predict the ground settlement of earth pressure balance (EPB) shield tunnelling in the Bangkok MRTA project. The predictive models were various nature-inspired frameworks, such as differential evolution (DE), particle swarm optimization (PSO), genetic algorithm (GA), and ant colony optimizer (ACO) to tune the adaptive neuro-fuzzy inference system (ANFIS). To obtain the accurate and reliable results, the modeling procedure is established based on four different dataset scenarios including (i) preprocessed and normalized (PPN), (ii) preprocessed and nonnormalized (PPNN), (iii) non-preprocessed and normalized (NPN), and (iv) non-preprocessed and nonnormalized (NPNN) datasets. Results indicated that PPN dataset scenario significantly affected the prediction models in terms of their perdition accuracy. Among all the developed hybrid models, ANOFS-PSO model achieved the best predictability performance. In quantitative terms, PPN-ANFIS-PSO model attained the least root mean square error value (RMSE) of 7.98 and a correlation coefficient value (CC) of 0.83. Overall, the attained results confirmed the superiority of the explored hybrid AI models as robust predictive model for ground settlement of earth pressure balance (EPB) shield tunnelling.

1. Introduction

The durability of underground excavation projects is normally dependent on accurate deformation prediction of the rock masses [1, 2]. Currently, the construction of a tunnel in the urban locations with small construction depth is facing a significant increase in growth in the complex geological formations and there is increase in risk conditions by external loading from the building [3]. However, the increase of budge and construction delay may exist when these conditions cannot be acknowledged before excavating the tunnel. Thus, to prevent project delays around the tunnel, a significant valid prediction is needed [4], particularly the structural failure and excessive deformation and structural failure that are forecasted by monitoring and data collection within the tunnel. Subsequent genuine and helpful actions are taken based on the feedback information [5]. Regardless of an increase in the experience and theoretical assessment obtained from the monitored data using several construction techniques, targeted and reliable techniques of available predictions are still absent [6]. The analytical and empirical techniques are not suitable for all geological cases because they cannot produce reliable results but only forecast a few numbers of geo-mechanical applying simplifications and parameters [4].

Despite the use of the earth pressure balance (EPB) shields over the years, the actual mechanism that governs shield-ground interaction is yet to be fully understood. The understanding of the EPB tunnelling-induced ground response mechanism is difficult as it requires both a reliable measurement of ground deformations in the field and the EPB shields' operational records. Few studies on EPB tunnelling are currently available but being that there has been a significant increase in the use of EPB tunnelling technique in recent times; especially in urban environments, there is a need for engineers to have a better understanding of the mechanisms of EPB tunnelling and how its parameters influence ground deformations. This will help to reduce the detrimental effects of this mechanism on the immediate surrounding.

Some of the engineering mechanics attributes of tunnel rock masses bothered on the failure and mechanism, meaning that engineers neither readily nor accurately forecasted because of the lack of clarity on the extent of rock mass support interaction, heterogeneity of the rock mass, and geotechnical environments before construction. Many years ago, some major studies had concentrated on the regularities and rock masses deformation and mechanism of ground surface settlements according to the in situ test data collected from the past projects and accumulated experience that have detected the durability of the tunnels.

There are three kinds of techniques, including the artificial intelligence and analytical and numerical methods that can be utilized during practice to evaluate the tunnel deformation. An analytical technique was performed on a shallow tunnel in the saturated ground according to [7] where they used two kinds of drainage conditions with and without full drainage at the ground-line interface. The results obtained showed that the solutions are narrowed to where ground deformations are small. Chou and Bobet [8] utilized 28 tunnels in determining the predictions from an analytical method in shallow tunnels on saturated ground. Based on this result, the difference between observation and predictions from real tunnels showed about 15% good agreement. Other different studies were carried out using the analytical methodologies [9–11].

Several AI techniques have been developed for addressing the challenges associated with rock and geotechnical engineering [12, 13]. Sou-Sen and Hsien-Chuang [14] used an artificial neural network- (ANN-) based regression model for the prediction of the influence of the ground surface settlement through intense excavation. They utilized case data that were obtained from just concluded deep excavation projects in Taiwan to develop a model. These results showed that the ANN-based forecasted models that can justifiably forecast the location and magnitude of maximum ground surface settlement influenced through deep excavation. An et al. [15] suggested an evolutionary neural network (ENN) model to simulate the ground settlement. The model was developed according to the mechanism which enables every part of the network structure, such as the learning parameters and several hidden nodes that can be developed via the genetic algorithms. These results showed a better achievement for the prediction of the ground settlement. Neaupane and Adhikari [16] established the ANN model for predicting ground movement throughout the tunnels. The predicted surface settlement above a horizontal ground movement and tunnel can be as a result of a tunnel construction through the assistance of input parameters that can cause direct physical significance. These results showed the ability of the ANN model to achieve positive outcomes and fairly and successfully forecasted the desired goal. Cheng et al. [17] developed an evolutionary fuzzy neuro-inference system (EFNIS) in facilitating geotechnical expert in decision making. The EFNIM consist of three separate AI technologies, such as genetic algorithm (GA), neural network (NN), and fuzzy logic (FL). Two case studies were considered, estimating slurry-wall duration that includes a selection of retaining wall construction techniques and estimating slurry-wall duration. The outcomes showed that EFNIS has an increased capability for the geotechnical challenges over the other classical AI models when the two references were compared. Santos and Celestino [18] confirmed the ANN model functionality by analysing tunnel settlement instrumentation. This study was centred on a settlement above shotcrete-supported tunnels on West Extension that were excavated in tertiary sediment utilizing the sequential excavation technique. The study has shown that the ANN model predicted accurately. Lee and Akutagawa [19] reported the ANN method as a rapid displacement prediction by utilizing the outcomes of field measurement in the NATM tunnels. They gathered data for NATM tunnels built on a coarse ground and analysed them based on the major tunnelling parameters, including tunnel displacement, support condition, and geometry discovered during constructions. The outcomes proposed that the ANN model could forecast tunnel deformations at the last phase before construction with reasonable increase in the level of accuracy of some information. Yao et al. [20] reported a multistep-ahead prediction model according to the SVM model of the tunnel that surrounds the rock displacement prediction. They used shuffled complex evolution algorithm (SCEUA) through some exponential transformations in improving the training capacity of the SVM. The outcomes demonstrated that SVM is better than the classical ANN and indicated that SVM can be an effective and feasible multistep technique for a tunnel which surrounds the rock displacement prediction. Pourtaghi and Lotfollahi-Yaghin [21] introduced the method that was based on a combination between the ANN model and wavelet theory for the prediction of maximum surface settlement due to tunnelling. The simulation results showed a reduction in estimated error values which represented the capacity to increase the activity approximation ability and wherefore exhibited outstanding learning capability as compared to the other activation activities. Wang et al. [22] utilized an easy relevancy vector machine with a wsRVM to module tunnelling-induced ground surface settlement establishment. Several conditions that affect settlement were examined, such as shield operational, geological, and geometrical parameters. The outcomes indicated that the prediction model works perfectly and the extension of the predictions could give a mean of

predicting uncertainty. Khamesi et al. [23] presented a numerical analysis, imperialistic competitive algorithm, and intelligent back analysis technique combining fuzzy systems to improve the capacity of intelligent back analysis in tunnelling. The results showed that PSO achieved the best parameters tuning performance while ICA showed great capacity for world searching on designed fuzzy system and it was seen that the fuzzy systems are equal to the inputs with given outputs for making back analysis feasible in a large worldwide space and to apply these methods to additional difficult engineering problems. Ahangari et al. [24] used new intelligent techniques in predicting subway settlement depended on gene expression programming (GEP) and ANFIS model in the settlement prediction. They obtained data from fifty-three tunnels all over the world, forty data sets were used in intelligent modeling, while thirteen data were utilized to evaluate its role. They deduced that the two intelligent techniques are approved to predict subway settlements. Hasanipanah et al. [25] revealed that hybrid ANN-PSO can be used to predict maximum surface settlement resulted from tunnelling. The model was developed based on 143 datasets. The results showed that the suggested PSO-ANN model can predict maximum surface settlements by producing an increased level of accuracy when compared to the ANN results.

The potential of the ANFIS model for modeling settlement problems has been documented in the literature [26]. Results indicated an optimistic research trend using the capability of the ANFIS model. However, there are certain problems with the ANFIS model and these are related with the tuning of the parameters of the membership function [27]. This problem is incorporated with the learning process of the model that at the first place influence the prediction accuracy. As a matter of fact, AI models are subjective to hyperparameters tuning [28]. The new era of computer aid application is advanced to the exploration of the hybridization of AI models with various nature inspired optimization algorithms such as ant colony (ACO), particle swarm optimization (PSO), genetic algorithm (GA), and differential evolution (DE) [29–32], which can be used to train AI models and improve their performance in addressing both high-dimensional and nonlinear problems. The introduced optimization algorithms have been evidenced their capacity in optimizing ANFIS model for modeling diverse prediction problems [33–38]. The hybridization of ANFIS model with nature inspired optimization algorithms demonstrated a remarkable improvement for prediction process of several engineering applications [39–43]. Hence, the main goal of the hybridization is to attain a stable and reliable learning process [44–46].

This work is driven by the need to study the feasibility of relatively new hybrid models based on the hybridization of ANFIS model with four different nature-inspired algorithms which are ACO, PSO, GA, and DE for modeling ground settlement of earth pressure balance (EPB) shield tunnelling. To attain an accurate and reliable predictability performance, the modeling procedure is established based on four different dataset scenarios where the data preprocessing and data normalization are incorporated. A comprehensive analysis and assessment are conducted on the achieved results. Several limitations are recognized and reported for possible future research devotion.

2. Underground Tunnel and Data Description

In the current research, the underground dataset used for the modeling development is belonged to the Bangkok Subway project. The project consisted of twenty kilometers length of twin tunnels. The entire project was initiated with eight EPB shields. The geological feature of the project is Chao Praya delta plain with topological around 1–0.5 meters above the sea level. During the excavation of the project, large number of surface settlement markers and array "within 50 meters depth" were installed to measure the ground settlement (see Figure 1). There are several factors effecting the surface settlement including the geometry of the tunnel (depth of the tunnel and the distance from shaft), geological parameters (e.g., the groundwater table, geology at the crown, and invert of the tunnel) and shield operation parameters (rate of penetration, face pressure, pitching angle, percentage tail void grout filling, and tail void grouping pressure) [47]. This study considered 13 related parameters to build the prediction matrix of the proposed hybrid predictive models.

3. Methodology Overview

3.1. Adaptive Neuro-Fuzzy Inference System (ANFIS). Neuro-fuzzy (NF) set approach represents an integrative soft computing method comprising of neural networks for patterns recognition for the immediate environment and a human expert-dependent fuzzy inference system for making solutions and differentiating solutions in a special field [48, 49]. The performance of these systems in decisions processing and explanation ought to mimic human-like expertise [50]. Additionally, the system has a fault tolerance capability that ensures the system will not be adversely affected during a deletion/amendment task. However, the applicability of soft computing approach is gaining momentum with the development of a firm foundation in various fields all around the world. Adaptive neuro-fuzzy inference system (ANFIS) was developed as a branch of the AI models whose mechanism of operation is based on Takagi–Sugeno fuzzy inference system [51]. ANFIS combines the advantages of neural networks and fuzzy logic in a single framework [52]. It is also equipped with a fuzzy inference system which can be trained; thus, it is taken as a system with better efficiency compared to systems with only a neuro-fuzzy system. This work used different input parameters and only one output f interference system. The rules of the first-order Sugeno fuzzy model [53] are as follows: $f_1, f_2, \ldots, f_n$, where n is the highest number of rules [54]:

$$\begin{aligned} &\text{Rule \#1: if } X \text{ is } A_1 \text{ and } Y \text{ is } B_1\text{, then } f_1 = p_1 x + q_1 y + r_1, \\ &\text{Rule \#2: if } X \text{ is } A_2 \text{ and } Y \text{ is } B_2\text{, then } f_2 = p_2 x + q_2 y + r_2, \end{aligned} \tag{1}$$

where A_1-A_2 and B_1– B_2 are the membership functions for multiple inputs including x and y. Figure 2 presents the main ANFIS model structure.

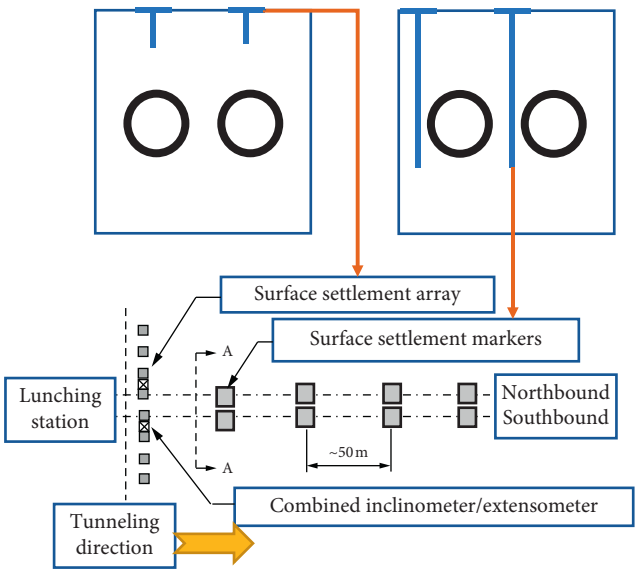


Figure 1: Typical instrumentation for ground surface settlement for the studied project.

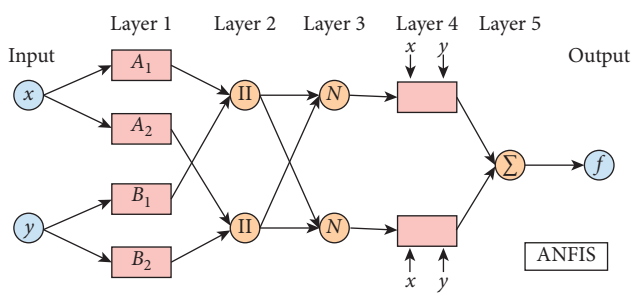


Figure 2: The structure of the adaptive neuro-fuzzy inference system model.

In an ANFIS model structure with multiple rules, several inputs, and one output, the nodes in the same layer perform the same role. The output of the ith node in the first layer is selected when the input membership functions (MFs) are contained in the first layer and the input values are delivered to the subsequent layer [55, 56]. The generalized bell function must have the best nonlinear parameters categorization capability for it to be used. In the first layer, every node is squared to initiate the membership grades. The input parameters are translated into linguistic terms:

$$O_i^1 = \mu_{A_i}(x), \quad i = 1, 2, \tag{2}$$

in which x = input parameters of node I, A_i = a linguistic term, and O_i^1 = the MF of A_i.

The output of the first layer is multiplied by the second layer (usually called a membership layer) to generate a new output. Each node in the second layer is considered a fixed node whose output depends on all the input values. The node i in the 3rd layer computes the firing strength ratio of the rules, with the outputs taken as the normalized weights [57]. The values of the inference rules-based outputs are provided by the fourth layer; the overall inputs of the previous layers are combined in this layer before converting the classification results into the final output. The ANFIS structure is identified by the applied learning algorithm. The functional forward pass signal in this algorithm continues until the fifth layer (the defuzzification layer).

Least squares estimate was used to identify the consequent parameters. The error rates propagate backward in the backward pass while during the gradient descent, the premise parameters are updated. The ANFIS models' membership function was later tuned using several nature inspired optimization algorithms including PSO, GA, and DE optimization algorithms to ensure a minimum solar radiation prediction error.

3.2. Particle Swarm Optimization (PSO). PSO is a computational framework put together by Eberhart and Kennedy [58] in 1995 for continuous and discontinues decision-making processes. It was based on the natural behavior of living species such as the schooling of fishes when searching for food sources. Being that in the PSO, each particle in the population is considered as a potential solution, the PSO is regarded as a population search-based method. During an active search, the optimal solution found by every particle is varied in a multidimensional space until there are no more computation limitations to be addressed. In the PSO, the major problems encountered during swarm optimization are associated with the position of N particles; this position is randomly assigned to the swarm in the D-dimensional space. Each solution in the swarm is associated with a position and each particle in the solution space is counted through a scoring function whose values explain the status of the problem. Several studies have applied PSO on several optimization problems [59], where it has been found that all the particles found the global best position in the solution space and achieved personal best positions. The new assigned position and velocity of the particles are updated by the following rules [60]:

$$p = p + v, \tag{3}$$

with

$$v = v + c_1.\text{rand.}\left(p_{\text{best}} - p\right) + c_2.\text{rand.}\left(g_{\text{best}} - p\right), \tag{4}$$

where p and v are the particle position and direction, c_1 is the local weight and c_2 is the global weight, p_{best} and $g_{\text{best}}$are the best positions for the particles and swarms, respectively, and rand is a random value. The operation of the PSO algorithm is depicted in Figure 3.

3.3. Genetic Algorithm (GA). The GA was first developed by [61] as an advanced optimization framework. As an AI framework, it uses vectors of 1's and 0's to represent complex structures. The GA was modeled after the concept of natural genetics as optimized functions are established by comparing 2 distinct approaches. The GA has a better capability of establishing global optimal solutions to huge combinatorial problem and this is its major advantage compared to

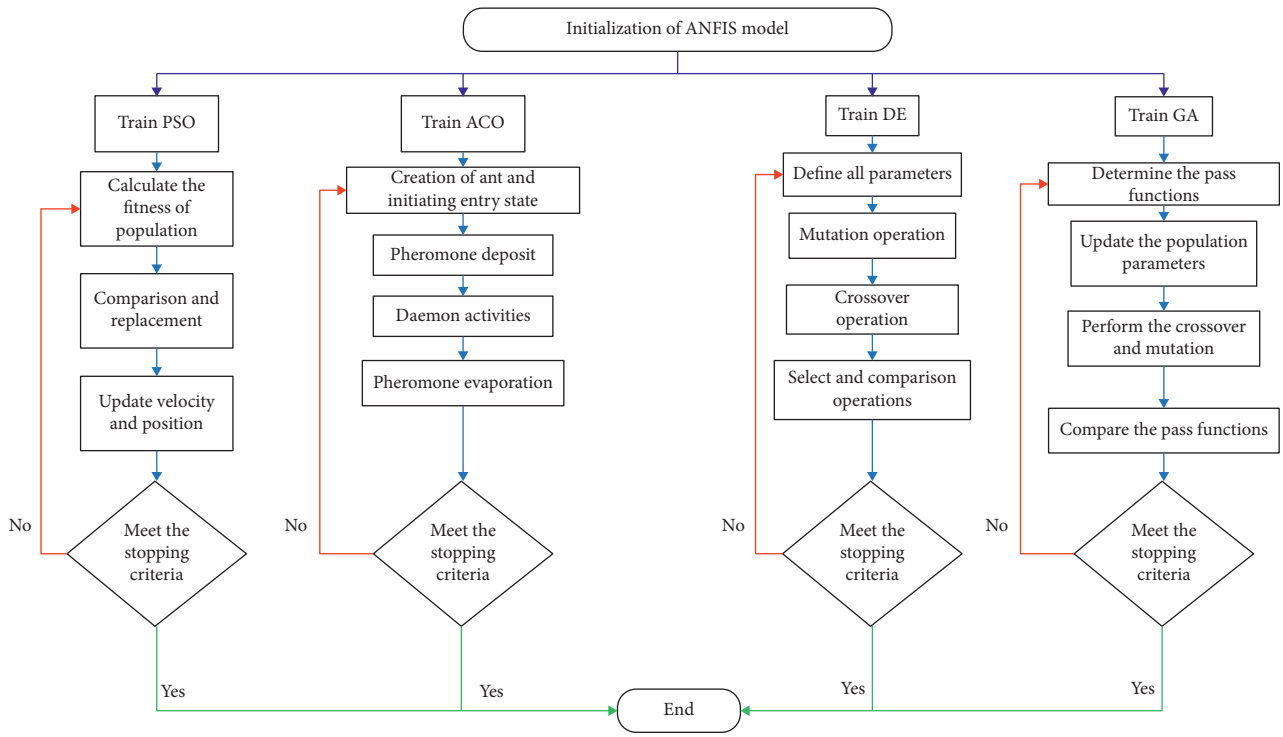


Figure 3: The flow chart of the proposed hybrid ANFIS models; the optimization procedure of each integrated nature inspired algorithm.

the other systems. Consequently, the GA is often used in multiobjective tasks optimization. The operation of the GA is similar to those of natural evolution processes; it relies on 3 operators which are selection, crossover, and mutation. The first optimization step using GA involves the evaluation of the fitness functions of the selected configuration (known as a chromosome). This step also involves sustaining a population of M solutions. If the annualized system cost of the evaluated chromosome is lower than the lowest known annualized system cost from the previous iterations, such a chromosome will be considered the optimal solution, and this would reduce the required number of problem iterations. Meanwhile, this optimum solution can also be exchanged with any better solution. Then, the crossover and mutation processes will be executed to select the best solution which will generate a new set of generations.

This process is repeated until an already specified convergence level is reached. Some precise parameters were used in the suggested GA model in this study; they are the size of the population, the rate of mutation, the crossover percentage, and the selection pressure (refer to Table 1). Figure 3 depicts the flow chart of the proposed GA.

3.4. Deferential Equation. The DE was developed as an intelligent framework for optimization tasks. Its operation mimics the basic optimized mutation, crossover, and selection operators. The DE relies on NP and D-dimensional parameter vectors because it is a parallel direct search method and is not impacted by minimization processes; hence, it is considered a population process when generating each generation G. In the DE, an initial population vector is randomly selected to cover all the parameter spaces; all the random choices are assigned a uniform probability distribution. DE generates new parameter vectors based on the available preliminary solution by generating the weight difference between 2 population vectors and a 3rd vector via a mutation operation as follows:

$$xi, G = [x1, i, G, x2, i, G, \ldots, xn, i, G], \quad i = 1, 2, \ldots, k, \tag{5}$$

where $x_1, G, i = 1, 2, 3, \ldots,$ NP are the generated mutant vectors using $v_{i,G+1}$, and r_1 , r_2, and r_3 are arbitrarily selected numbers $\in[1, 2, 3, \ldots,] \ldots$ NP. Note that NP must have these values, with i and F being actual that only differed from each other $\in[1, 2, 3, \ldots,] \ldots$ NP. The trial vector is established via a mixing process (crossover operation) that involves mixing the parameters of the mutated vector with the other parameters of the predetermined vectors as follows:

$$xLj \leq xj, i, 1 \leq xUj, \tag{6}$$

where $u_{i,G+1}$ is the trailer vector and $x_{i,G}$ is the target vector; rand $b(j)$ represents the Jth uniform random evaluation $\in\in[0.1]$, rnbr (i) is a random value index $\in\in[1, 2, 3, \ldots, d]$, and CR is a user-specified crossover constant. Lastly, the trial vector that gives the lowest cost function value compared to the target vector is used during the selection operation as the target value in the subsequent generation. Being that each population must serve as the target vector, NP tasks are considered as a one-generation procedure. A study by [62] has provided a deep description of the standard DE. Table 1 presents the proposed DE models' structure while its flow chart is depicted in Figure 3.

Table 1: The values of the models tuning parameters.

Models	Parameters	Values
Algorithm: PSO	Global learning coefficient	2
	Number of iterations	1200
	Number of populations	50
	Personal learning coefficient	0.9
	Inertia weight damping ratio	0.95
	Inertia weight	1
Algorithm: DE	Crossover probability	0.1
	Lower bound of scaling factor	0.3
	Upper bound of scaling factor	0.7
Algorithm: ACO	Number of populations	50
	Number of iterations	1200
	Deviation distance ratio	1
	Intensification factor	0.5
Algorithm: GA	Selection pressure	8
	Mutation percentage	0.5
	Number of mutants	50
	Mutation rate	0.1
	Number of iterations	1200
	Crossover percentage	0.7
	Number of populations	50
	Number of offspring	35
ANFIS	Train step size increase	1.15
	Train step size decrease	0.95
	Train initial step size	0.01
	Train error goal	0
	Train epoch	500

3.5. Ant Colony Optimization (ACO). ACO was put forward two decades ago by Dorigo [63] but has undergone several modifications since its development. ACO algorithms have a wide range of application due to their capability to solve both static and dynamic problems. Stignergy, a self-organization-enabling process, coordinates numerous activities in ant colonies; such activities include food hunting, labour partitioning, brood sorting, and cooperative transport in the ant colony. Ant colonies are known to contain simple individuals; yet, it is a complex but well-organized structure where all activities are coordinated by stignergy. Chemical deposits known as pheromone are deposited by the leading ants for the followers to trail while seeking the shortest route to food source. ACO relies on such procedure for establishing the optimal position in the solution space. The ants move in a forward and backward pattern and adopt a stepwise decision approach in finding the optimal solution to any problem [48–50]. Figure 3 shows the flowchart for ACO.

3.6. Modeling Development Scheme. The developed hybrid and standalone ANFIS models (ANFIS-PSO, ANFIS-DE, ANFIS-ACO, ANFIS-GA, and ANFIS) were constructed based on several geo-technical information. The information was used as numerical parameters to initiate the prediction matrix. Figure 4 presents the input/output parameters used in the current research. The dataset has a total of 49 observations and for the non-preprocessed datasets, the data were partitioned into 65% and 35% for training and testing purposes, respectively. For the preprocessed data, the modeling was executed with 38 observations based on the data partitioning percentage of 63%–35%. Several statistical measures, plots, and error levels between the experimental and predicted ground settlements were reported during the testing phase. The prediction process was conducted based on several scenarios associated with the data processing including (i) preprocessed and normalized (PPN) dataset, (ii) preprocessed and nonnormalized (PPNN) dataset, (iii) non-preprocessed and normalized (NPN) dataset, and (iv) non-preprocessed and nonnormalized (NPNN) dataset "with total 20 modeling investigations model 1-model 20." The statistical performance of all the developed predictive models are reported in Tables 2–5, respectively.

3.7. Modeling Performance Metrics. Several statistical metrics are computed to assess the predictability performance of the proposed predictive models including mean absolute error (MAE), root mean square error (RMSE), Legate and McCabe's index (LMI), PBIAS, Willmott's index (WI), correlation coefficient (CC), and relative root mean square error (RRMSE); these metrics were calculated, thus [64–68]

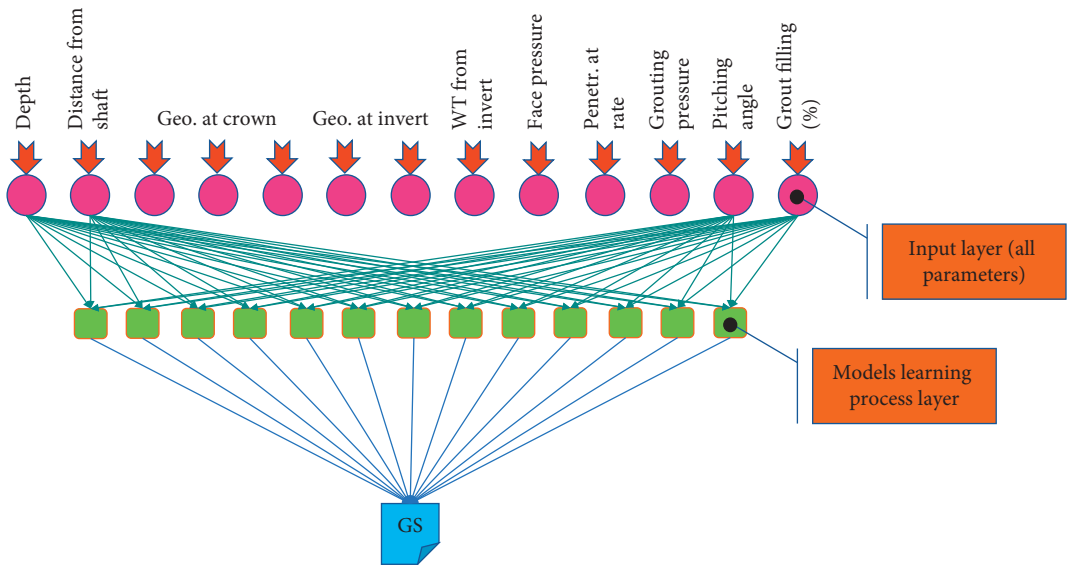


Figure 4: The related geo-science input parameters and the output ground settlement.

Table 2: The statistical evaluation of the developed hybrid and original ANFIS for the scenario of the preprocessed and normalized (PPN) dataset over the testing phase of the modeling.

Models order	Scenario models	RMSE	MAE	LMI	CC	WI	SRMSE
Model 1	PPN-ANFIS	11.57	8.77	−0.06	0.74	0.76	41.89
Model 2	PPN-ANFIS-PSO	7.99	6.22	0.25	0.83	0.82	28.90
Model 3	PPN-ANFIS-DE	8.61	6.96	0.16	0.66	0.57	31.16
Model 4	PPN-ANFIS-ACO	8.98	7.49	0.10	0.73	0.36	32.49
Model 5	PPN-ANFIS-GA	9.23	7.19	0.13	0.76	0.75	33.39

Table 3: The statistical evaluation of the developed hybrid and original ANFIS for the scenario of the preprocessed and nonnormalized (PPNN) dataset over the testing phase of the modeling.

Models order	Scenario models	RMSE	MAE	LMI	CC	WI	SRMSE
Model 6	PPNN-ANFIS	10.40	8.29	0.00	0.78	0.77	37.64
Model 7	PPNN-ANFIS-PSO	8.78	6.60	0.20	0.80	0.78	31.79
Model 8	PPNN-ANFIS-DE	9.08	7.51	0.09	0.41	0.60	32.86
Model 9	PPNN-ANFIS-ACO	10.04	8.44	−0.02	−0.48	0.04	36.34
Model 10	PPNN-ANFIS-GA	9.26	7.51	0.09	0.77	0.79	33.51

Table 4: The statistical evaluation of the developed hybrid and original ANFIS for the scenario of the non-preprocessed and normalized (NPN) dataset over the testing phase of the modeling.

Models order	Scenario models	RMSE	MAE	LMI	CC	WI	SRMSE
Model 11	NPN-ANFIS	12.96	10.10	−0.13	0.36	0.62	48.58
Model 12	NPN-ANFIS-PSO	10.82	8.14	0.09	0.29	0.59	40.54
Model 13	NPN-ANFIS-DE	8.99	7.44	0.17	0.49	0.59	33.70
Model 14	NPN-ANFIS-ACO	11.00	9.16	−0.03	−0.28	0.15	41.24
Model 15	NPN-ANFIS-GA	11.10	8.02	0.10	0.20	0.53	41.62

Table 5: The statistical evaluation of the developed hybrid and original ANFIS for the scenario of the non-preprocessed and nonnormalized (NPNN) dataset over the testing phase of the modeling.

Models order	Scenario models	RMSE	MAE	LMI	CC	WI	SRMSE
Model 16	NPNN-ANFIS	14.62	11.81	−0.32	0.30	0.56	54.79
Model 17	NPNN-ANFIS-PSO	10.07	8.11	0.09	0.38	0.63	37.73
Model 18	NPNN-ANFIS-DE	11.75	9.87	−0.11	−0.25	0.16	44.04
Model 19	NPNN-ANFIS-ACO	11.27	9.43	−0.06	−0.42	0.09	42.24
Model 20	NPNN-ANFIS-GA	10.67	8.55	0.04	0.38	0.62	40.01

$$\begin{aligned}
\mathrm{RMSE} &= \sqrt{\frac{1}{N_s}\sum_{j=1}^{N_s}\left((\mathrm{GS})_{\mathrm{exp}} - (\mathrm{GS})_{\mathrm{Sim}}\right)^2},\\
\mathrm{MAE} &= \frac{1}{N_s}\sum_{j=1}^{N_s}\left|(\mathrm{GS})_{\mathrm{exp}} - (\mathrm{GS})_{\mathrm{Sim}}\right|,\\
\mathrm{LMI} &= 1 - \left[\frac{\sum_{i=1}^{N_s}\left|(\mathrm{GS})_{\mathrm{exp}} - (\mathrm{GS})_{\mathrm{Sim}}\right|}{\sum_{i=1}^{N_s}\left|(\mathrm{GS})_{\mathrm{exp}} - \overline{(\mathrm{GS})_{\mathrm{exp}}}\right|}\right],\\
\mathrm{CC} &= \frac{\sum_{j=1}^{N_s}\left((\mathrm{GS})_{\mathrm{exp}} - \overline{(\mathrm{GS})_{\mathrm{exp}}}\right)\left((\mathrm{GS})_{\mathrm{Sim}} - \overline{(\mathrm{GS})_{\mathrm{Sim}}}\right)}{\sqrt{\sum_{j=1}^{N_s}\left((\mathrm{GS})_{\mathrm{exp}} - \overline{(\mathrm{GS})_{\mathrm{exp}}}\right)^2\sum_{j=1}^{N_T}\left((\mathrm{GS})_{\mathrm{Sim}} - \overline{(\mathrm{GS})_{\mathrm{Sim}}}\right)^2}},\\
\mathrm{WI} &= 1 - \left[\frac{\sum_{i=1}^{N_s}\left((\mathrm{GS})_{\mathrm{exp}} - (\mathrm{GS})_{\mathrm{Sim}}\right)^2}{\sum_{i=1}^{N_s}\left(\left|(\mathrm{GS})_{\mathrm{Sim}} - \overline{(\mathrm{GS})_{\mathrm{exp}}}\right| + \left|(\mathrm{GS})_{\mathrm{exp}} - \overline{(\mathrm{GS})_{\mathrm{exp}}}\right|\right)^2}\right],\\
\mathrm{SRMSE} &= \frac{\sqrt{(1/N_s)\sum_{j=1}^{N_s}\left((\mathrm{GS})_{\mathrm{exp}} - (\mathrm{GS})_{\mathrm{Sim}}\right)^2}}{\overline{(d_s/D)_{\mathrm{Obs}}}},
\end{aligned} \tag{7}$$

where the $(\mathrm{GS})_{\mathrm{exp}}$ and $(\mathrm{GS})_{\mathrm{Sim}}$ are the experimental and simulated ground settlements, $\overline{(\mathrm{GS})_{\mathrm{exp}}}$ and $\overline{(\mathrm{GS})_{\mathrm{Sim}}}$ are their mean values, and N_s is the sample size.

Among the proposed predictive models, superior model may be different in terms of different performance indices. This weakness can be solved using a new index called mean performance (MP) which is the integrate of all employed indices. To compute the MP value of each predictive model, it is necessary to convert the indices to standardized form in the range of [0 1] as following equations [69]:

$$\begin{aligned}
\widehat{\mathrm{RMSE}}_{\mathrm{Model}(i)} &= \frac{\left(\mathrm{RMSE}_{\mathrm{All}}^{\max} - \mathrm{RMSE}_{\mathrm{Model}(i)}\right)}{\left(\mathrm{RMSE}_{\mathrm{All}}^{\max} - \mathrm{RMSE}_{\mathrm{All}}^{\min}\right)},\\
\widehat{\mathrm{MAE}}_{\mathrm{Model}(i)} &= \frac{\left(\mathrm{MAE}_{\mathrm{All}}^{\max} - \mathrm{MAE}_{\mathrm{Model}(i)}\right)}{\left(\mathrm{MAE}_{\mathrm{All}}^{\max} - \mathrm{MAE}_{\mathrm{All}}^{\min}\right)},\\
\widehat{\mathrm{LMI}}_{\mathrm{Model}(i)} &= \frac{\left(\mathrm{LMI}_{\mathrm{Model}(i)} - \mathrm{LMI}_{\mathrm{All}}^{\min}\right)}{\left(\mathrm{LMI}_{\mathrm{All}}^{\max} - \mathrm{LMI}_{\mathrm{All}}^{\min}\right)},\\
\widehat{\mathrm{CC}}_{\mathrm{Model}(i)} &= \frac{\left(\mathrm{CC}_{\mathrm{Model}(i)} - \mathrm{CC}_{\mathrm{All}}^{\min}\right)}{\left(\mathrm{CC}_{\mathrm{All}}^{\max} - \mathrm{CC}_{\mathrm{All}}^{\min}\right)},\\
\widehat{\mathrm{WI}}_{\mathrm{Model}(i)} &= \frac{\left(\mathrm{WI}_{\mathrm{Model}(i)} - \mathrm{WI}_{\mathrm{All}}^{\min}\right)}{\left(\mathrm{WI}_{\mathrm{All}}^{\max} - \mathrm{WI}_{\mathrm{All}}^{\min}\right)},\\
\widehat{\mathrm{SRMSE}}_{\mathrm{Model}(i)} &= \frac{\left(\mathrm{SRMSE}_{\mathrm{All}}^{\max} - \mathrm{SRMSE}_{\mathrm{Model}(i)}\right)}{\left(\mathrm{SRMSE}_{\mathrm{All}}^{\max} - \mathrm{SRMSE}_{\mathrm{All}}^{\min}\right)},
\end{aligned} \tag{8}$$

where $\widehat{\mathrm{RMSE}}_{\mathrm{Model}(i)}$, $\widehat{\mathrm{MAE}}_{\mathrm{Model}(i)}$, $\widehat{\mathrm{LMI}}_{\mathrm{Model}(i)}$, $\widehat{\mathrm{CC}}_{\mathrm{Model}(i)}$, $\widehat{\mathrm{WI}}_{\mathrm{Model}(i)}$, and $\widehat{\mathrm{SRMSE}}_{\mathrm{Model}(i)}$ are the standardized values of the utilized performance metrices of ith model (hybrid ANFIS model), $\mathrm{RMSE}_{\mathrm{All}}^{\max}$, $\mathrm{MAE}_{\mathrm{All}}^{\max}$, $\mathrm{LMI}_{\mathrm{All}}^{\max}$, $\mathrm{CC}_{\mathrm{All}}^{\max}$, $\mathrm{WI}_{\mathrm{All}}^{\max}$, and $\mathrm{SRMSE}_{\mathrm{All}}^{\max}$ are the maximum values of indices among all predictive models while $\mathrm{RMSE}_{\mathrm{All}}^{\min}$, $\mathrm{MAE}_{\mathrm{All}}^{\min}$,

LMI_{All}^{min}, CC_{All}^{min}, WI_{All}^{min}, and $SRMSE_{All}^{min}$ are the minimum ones.

Then, the MP metric of ith model (hybrid ANFIS) can be expressed mathematically as follows [70]:

$$MP_{Model(i)} = \frac{\left(\widehat{RMSE}_{Model(i)} + \widehat{MAE}_{Model(i)} + \widehat{LMI}_{Model(i)} + \widehat{CC}_{Model(i)} + \widehat{WI}_{Model(i)} + \widehat{SRMSE}_{Model(i)}\right)}{6}. \quad (9)$$

From equation (9) it is clear that the MP value of each model is the mean value of its standardized performance indices which is in the range of [0 1]. However, the superior model has the highest value of MP compared to other ones.

4. Application Results and Analysis

This study was initiated to study the feasibility of using different variants of hybrid ANFIS models for the prediction of the magnitude of settlement in soil. In nature, soil structure behavior is influenced by several morphological and external parameters such as structural loading. However, the main concern is to quantify the exact possible settlement can be experienced to maintain the build sustainability and safety. Soil is usually characterized by nonstationary varying and nonlinearly pattern with the constituents and their properties. Empirical formulas are not strong enough for the simulation of the relationships between soil settlement and the other related parameters; hence, the use of strong and reliable predictive tools for the determination of the effects of the independent parameters on the dependent parameters was proposed in this study. Indeed, these reliable predictive models can contribute to the geo-science practical implementations.

4.1. Statistical Evaluation of the Developed Hybrid Models. Based on the results in Table 2 (PPN scenario), the best prediction performance was attained for the ANFIS-PSO with minimal RMSE and MAE values (7.99 and 6.22) and maximum CC and WI values (0.83 and 0.82). Table 3 reveals the modeling results of the second modeling scenario (PPNN). Based on the reported results in Table 3, the accurate prediction process was achieved using the same hybrid model developed for the first scenario (i.e., ANFIS-PSO). Model 7 indiacted minimum RMSE ≈ 8.78 and MAE ≈ 6.6, while the CC ≈ 0.80 and WI ≈ 0.78. Table 4 reported the third modeling scenario where the dataset was non-preprocessed and normalized. The hybridization of the differential evolution optimizer with ANFIS model attained the best predictability performance with computed statistical metrics (RMSE ≈ 8.99, MAE ≈ 7.44, LMI ≈ 0.17, CC ≈ 0.49, WI ≈ 0.59, and SRMSE ≈ 33.7). The results of the last modeling scenario are tabulated in Table 5. The best prediction results were obtained using the ANFIS-PSO model with statistical performance metrics (RMSE ≈ 10.07, MAE ≈ 8.11, LMI ≈ 0.09, CC ≈ 0.38, WI ≈ 0.63, and SRMSE ≈ 37.73). Generally, the variance of the prediction results achieved based on the developed classical and hybrid ANFIS models can be explained by the use of different learning processes, and AI models can attain during the train and test modeling phases.

The hybridization of ANFIS model with nature-inspired optimization algorithms yielded the maximum prediction accuracy as a result of the robust tuning of the membership function. In quantitative explanation and for the best input combination, these tables displayed the optimal prediction possibility using all the applied predictive models. Based on the results in Table 6, the minimal RMSE metric was achieved using PPN-ANFIS-PSO (RMSE ≈ 7.99) followed by PPNNANFIS-PSO (RMSE ≈ 8.78), NPN-ANFIS-DE (RMSE ≈ 8.99), and NPNN-ANFIS-PSO (RMSE ≈ 10.07).

4.2. Modeling Evaluation Based on Graphical Presentation. The proposed predictive models for the settlement quantification were examined using several graphical visualizations including mean performance index (MP), scatter plot, box plot, and Taylor diagram. Figure 5 demonstrates the MP of all the computed performance measures of the deployed predictive models and for all the examined scenarios over the testing modeling phase. Figures 5(a)–5(d) indicate similar prediction performance observed using the statistical metrics which are tabulated in Tables 2–5. ANFIS-PSO was the best predictive model for the 1st, 2nd, and 4th modeling scenarios with maximum MP = 1.0, whereas ANFIS-DE accomplished the best prediction results for the 3rd modeling scenario non-preprocessed and normalized (NPN) dataset with MP ≈ 0.99. Due to the diverse prediction performances exemplified in Figure 5, the best predictive models were abstracted and validated with respect to the applied modeling scenarios (Figure 6). Based on the visualization in Figure 6, the various variations in modeling predictability were avoided. Figure 6 evidenced the feasibility of the 1st modeling scenario which gave the best results with MP = 1 using the potential of the ANFIS-PSO model. This is clearly presenting the potential of the PSO tuning algorithm integrated with the standalone ANFIS for providing a robust and reliable predictive model for settlement prediction. The worst model was recognized for the non-preprocessed and nonnormalized dataset scenario with minimal MP = 0.03 using ANFIS-PSO.

The behavior of the best applied predictive models was drawn in the form of scatter plot in Figure 7. Scatter plot displays the variance between the experimental and predicted values of the settlement. The best predictive model was identified from the variance around the fitted line and the correlation coefficient magnitude. Figure 7 indicates that the preprocessed and normalized scenario using ANFIS-PSO model attained the best fit in accordance with the

Table 6: Summary of the best statistical performance of the developed hybrid ANFIS for all the examined modeling scenarios over the testing modeling phase.

Models order	Scenario models	RMSE	MAE	LMI	CC	WI	SRMSE
Model 2	PPN-ANFIS-PSO	7.99	6.22	0.25	0.83	0.82	28.90
Model 7	PPNN-ANFIS-PSO	8.78	6.60	0.20	0.80	0.78	31.79
Model 13	NPN-ANFIS-DE	8.99	7.44	0.17	0.49	0.59	33.70
Model 17	NPNN-ANFIS-PSO	10.07	8.11	0.09	0.38	0.63	37.73

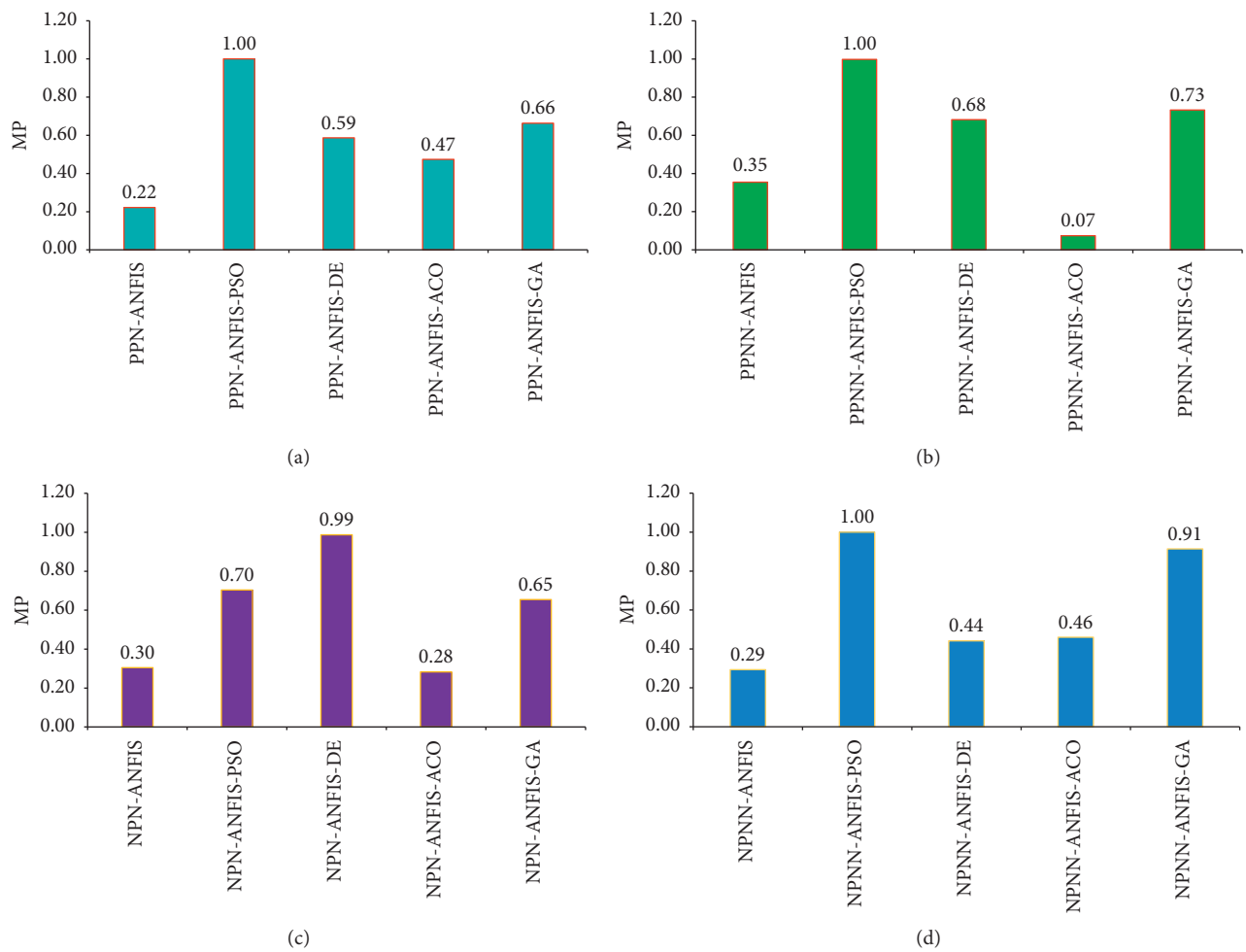


Figure 5: The mean performance index computed for the applied hybrid and classical ANFIS models over the test phase; (a) PPN scenario, (b) PPNN scenario, (c) NPN scenario, and (d) NPNN scenario.

determination coefficient scale ($R^2 = 0.69$), whereas the worst indicator was observed for the 4th scenario with ($R^2 = 0.14$) using ANFIS-PSO model. Based on the reported determination coefficient, the models performed slightly with less correlation. However, this is clearly indicating the high stochasticity of the simulated surface settlement owing to the huge number of the associated parameters.

The box plot computations for the applied predictive models are displayed in Figure 8. The degree of the spread in the predicted data and quartiles are 25, 50, 75, and the interquartile range (IQR). Based on the magnitudes of the lower (Q25%), median (Q50%), and upper (Q75%) quartiles, ANFIS-PSO with PPN scenario outperformed the other classical and hybrid ANFIS models. Based on the various statistical metrics and graphical representations, the proposed hybrid data-intelligence model is an excellent tool for predicting the settlement magnitudes.

The visualization of the 2-dimensional graphical examination of Taylor diagrams for the best prediction model (ANFIS-PSO) is shown in Figures 9(a) and 9(b) (processed and nonprocessed dataset, respectively). The figure presents a summary of the statistical performance of the actual and predicted tests in terms of the correlation coefficients, RMSE, and standard deviations. The Taylor curve is a graphical view of the similarity between the observed and predicted values. The results demonstrate that ANFIS-PSO

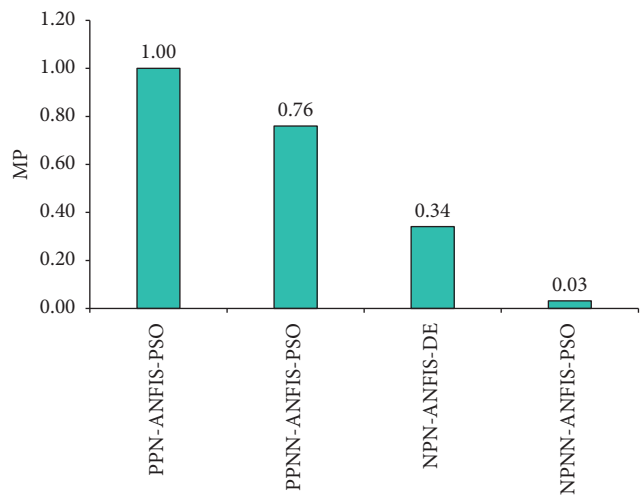


Figure 6: The best attained predictive models in accordance with the mean performance index over the testing phase.

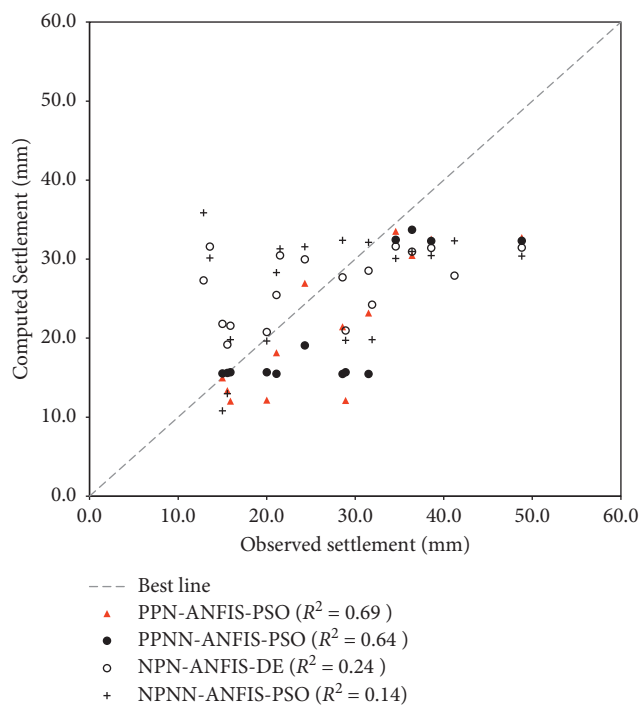


Figure 7: The variation between the observed and predicted settlement values over the testing phase and for the best predictive models.

using PPN scenario coordinated the nearest to the observed experimental tests. For the nonprocessed dataset, ANFIS-DE indicated the nearest coordination to the observed experimental tests.

4.3. Discussion and Possible Future Research Direction. In accordance to Figure 5, it was revealed that preprocessing the data by eliminating the redundant experimental test and scaling the data between (0-1) as normalized dataset could enhance the learning process of the applied hybrid predictive

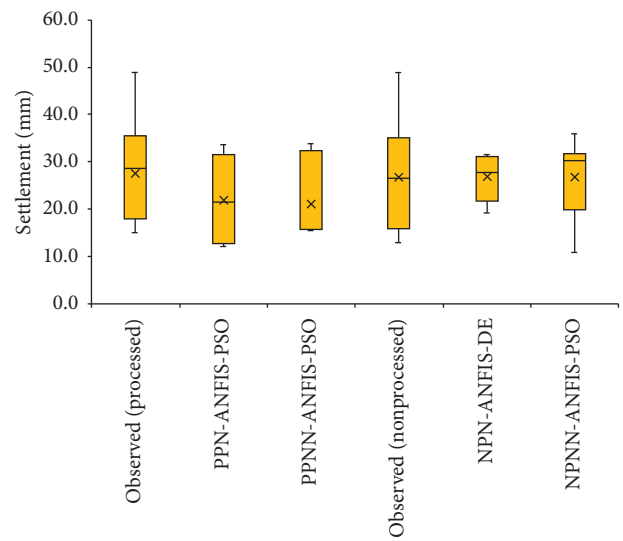


Figure 8: The box plots presentation of the observed and predicted settlement over the testing phase.

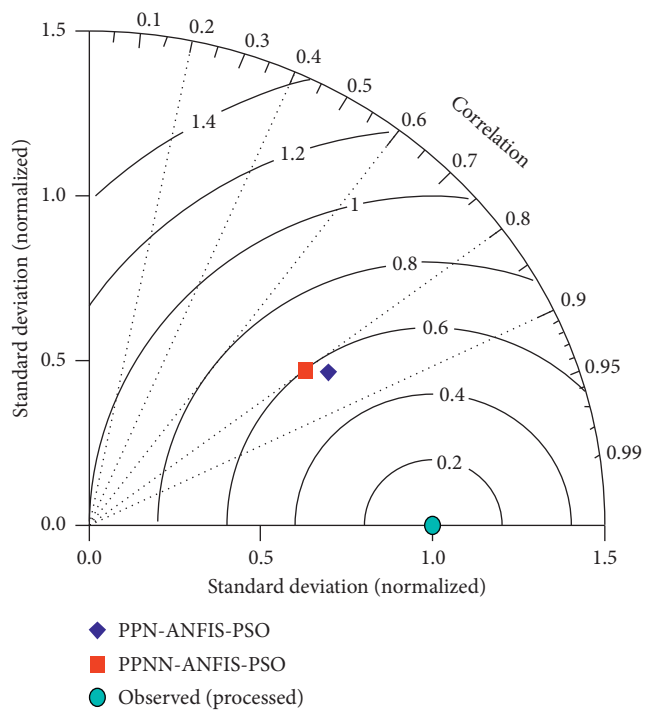


Figure 9: Normalized Taylor diagram visualization of the best applied predictive models over the testing phase.

models. This could be explained due to the exclusion of the vague attributes supplied to the prediction matrix. Hence, it is not always the case in providing more information attributes to fit data with higher accuracy. On the contrary, building the hybrid predictive model based on normal data information can offer insufficient learning process which leads to inaccurate prediction of result. The applications of AI models have been widely performed for several engineering problems and geo-technical is one of the distinguished research scopes experienced on the AI implications.

Consequently, the development of more robust and reliable AI models was based on their newly advanced versions in the motivation of engineers, designers, and scientists. The proposed hybrid ANFIS-PSO model provided an efficient and flexible methodology for computing ground settlement compared to the classical AI standalone models. In addition, the current research introduced a newly fusion of technology to eliminate boundaries between the digital and engineering spheres. The breakthroughs of novel AI models, such as the proposed one, are changing the solving manners of engineering problems. Hence, the proposed hybrid ANFIS tool and other intelligent models can be used in future automated and semiautomated design processes.

In this study, several hybrid ANFIS models were developed to predict ground settlement of earth pressure balance (EPB) shield tunnelling. The models were constructed based on various modeling data preprocessing scenarios. In accordance with the attained prediction performance, the potential of the hybrid ANFIS model demonstrated an acceptable prediction accuracy using preprocessed and normalized dataset. However, the prediction capacity is varied from one hybrid model to another and this definitely evidenced the existence of the modeling uncertainty, that is, one of the major limitations of the current research. Another reason, the influence of the dataset span on the prediction performance. In addition, one of the major limitations of ANFIS model is that when a large number of inputs are applied that causing a dimensionality issues and computational expense. Therefore, an optimizer is required to reduce these problems. In contrary, those metaheuristic algorithms also have their own limitations, such as unstable convergence for the DE algorithms and time consumption for the GA, while the PSO suffers from partial optimism. All in all, it is impossible to say that there is one metaheuristic method which can compete with other methods in all possible discrete functions. Thus, based on the reported modeling limitations, the following possible future researches can be further studied: (i) inspecting the associated uncertainties, either modeling uncertainty or input parameters uncertainty; (ii) extending the span of the modeled dataset with other related studies from the literature to give more informative details of the input attributes; and (iii) the data modeling division, various data division percentages can be examined for better learning process of the established models; (iv) the achieved modeling results accuracy level can be further improved by using several other nature-inspired algorithms such as nomadic algorithm [71], equilibrium algorithm [72], arithmetic algorithm [73], and black window algorithm [74].

5. Conclusion

Among several side effects during tunnelling excavation and initiating, ground settlement is one of the essential processes to be measured and monitored to avoid any unexpected damages and human losses. In general, ground settlement is a complex issue as there are several geo-technical parameters influencing the soil deformation. Accurate and reliable prediction of ground settlement is very important for tunnelling project management and maintenance and thus an adequate intelligent predictive model can contribute to the basic knowledge of designing tunnels. The current research was devoted on the assessment of relatively new hybrid AI models to predict the ground settlement of EPB shield tunnelling. The data collected from tunnel project were established in Thailand in the name of Bangkok MRTA project. The developed hybrid and standalone ANFIS models (i.e., ANFIS-PSO, ANFIS-DE, ANFIS-ACO, ANFIS-GA, and ANFIS) were constructed based on several related parameters to predict the GS. For the purpose of obtaining accurate and reliable investigation, the modeling procedure was conducted based on four different dataset scenarios including (i) preprocessed and normalized (PPN), (ii) preprocessed and nonnormalized (PPNN), (iii) non-preprocessed and normalized (NPN), and (iv) non-preprocessed and nonnormalized (NPNN). Modelling results revealed the superiority of the first modeling scenario that preprocessed and normalized dataset were incorporated. PPN dataset scenario demonstrated a significant impact on the perdition accuracy of the proposed models. In addition, among all the applied models, the hybrid ANFIS-PSO model accomplished the best predictability performance. The PPN-ANFIS-PSO model achieved minimum (RMSE = 7.98) and maximum (CC = 0.83).

Data Availability

Data used in this study will be provided upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] J. Shen, M. Karakus, and C. Xu, "A comparative study for empirical equations in estimating deformation modulus of rock masses," *Tunnelling and Underground Space Technology*, vol. 32, pp. 245–250, 2012.

[2] A. Kayabasi, C. Gokceoglu, and M. Ercanoglu, "Estimating the deformation modulus of rock masses: a comparative study," *International Journal of Rock Mechanics and Mining Sciences*, vol. 40, no. 1, pp. 55–63, 2003.

[3] G. Barla and S. Pelizza, "TBM tunnelling in difficult ground conditions," in *ISRM International Symposium*, Melbourne, Australia, November 2000.

[4] S. Mahdevari, S. R. Torabi, and M. Monjezi, "Application of artificial intelligence algorithms in predicting tunnel convergence to avoid TBM jamming phenomenon," *International Journal of Rock Mechanics and Mining Sciences*, vol. 55, pp. 33–44, 2012.

[5] A. N. Jiang, S. Y. Wang, and S. L. Tang, "Feedback analysis of tunnel construction using a hybrid arithmetic based on Support Vector Machine and Particle Swarm Optimisation," *Automation in Construction*, vol. 20, no. 4, pp. 482–489, 2011.

[6] L. Cantieni, G. Anagnostou, and R. Hug, "Interpretation of core extrusion measurements when tunnelling through

squeezing ground," *Rock Mechanics and Rock Engineering*, vol. 44, no. 6, p. 641, 2011.

[7] A. Bobet, "Analytical solutions for shallow tunnels in saturated ground," *Journal of Engineering Mechanics*, vol. 127, no. 12, pp. 1258–1266, 2001.

[8] W.-I. Chou and A. Bobet, "Predictions of ground deformations in shallow tunnels in clay," *Tunnelling and Underground Space Technology*, vol. 17, no. 1, pp. 3–19, 2002.

[9] K.-H. Park, "Analytical solution for tunnelling-induced ground movement in clays," *Tunnelling and Underground Space Technology*, vol. 20, no. 3, pp. 249–261, 2005.

[10] M.-W. Gui and S.-L. Chen, "Estimation of transverse ground surface settlement induced by DOT shield tunneling," *Tunnelling and Underground Space Technology*, vol. 33, pp. 119–130, 2013.

[11] J. Xu and Y. Xu, "Grey correlation-hierarchical analysis for metro-caused settlement," *Environmental Earth Sciences*, vol. 64, no. 5, pp. 1249–1256, 2011.

[12] S. Yagiz, C. Gokceoglu, E. Sezer, and S. Iplikci, "Application of two non-linear prediction tools to the estimation of tunnel boring machine performance," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 4-5, pp. 808–814, 2009.

[13] E. Ghasemi, "Particle swarm optimization approach for forecasting backbreak induced by bench blasting," *Neural Computing and Applications*, vol. 28, no. 7, pp. 1855–1862, 2017.

[14] L. Sou-Sen and L. Hsien-Chuang, "Neural-network-based regression model of ground surface settlement induced by deep excavation," *Automation in Construction*, vol. 13, no. 3, pp. 279–289, 2004.

[15] H. An, J. Sun, and X. Hu, "Study on intelligent method of prediction by small samples for ground settlement in shield tunnelling," in *Proceedings of the 30th ITA-AITES World Tunnel Congress C*, H. An, J. Sun, and X. Hu, Eds., Istanbul, Turkey, May 2004.

[16] K. M. Neaupane and N. R. Adhikari, "Prediction of tunneling-induced ground movement with the multi-layer perceptron," *Tunnelling and Underground Space Technology*, vol. 21, no. 2, pp. 151–159, 2006.

[17] M.-Y. Cheng, H.-C. Tsai, C.-H. Ko, and W.-T. Chang, "Evolutionary fuzzy neural inference system for decision making in geotechnical engineering," *Journal of Computing in Civil Engineering*, vol. 22, no. 4, pp. 272–280, 2008.

[18] O. J. Santos Jr and T. B. Celestino, "Artificial neural networks analysis of São Paulo subway tunnel settlement data," *Tunnelling and Underground Space Technology*, vol. 23, no. 5, pp. 481–491, 2008.

[19] J.-H. Lee and S. Akutagawa, "Quick prediction of tunnel displacements using Artificial Neural Network and field measurement results," *International Journal of the JCRM*, vol. 5, no. 2, pp. 53–62, 2009.

[20] B.-Z. Yao, C.-Y. Yang, J.-B. Yao, and J. Sun, "Tunnel surrounding rock displacement prediction using support vector machine," *International Journal of Computational Intelligence Systems*, vol. 3, no. 6, pp. 843–852, 2010.

[21] A. Pourtaghi and M. A. Lotfollahi-Yaghin, "Wavenet ability assessment in comparison to ANN for predicting the maximum surface settlement caused by tunneling," *Tunnelling and Underground Space Technology*, vol. 28, pp. 257–271, 2012.

[22] F. Wang, B. Gou, and Y. Qin, "Modeling tunneling-induced ground surface settlement development using a wavelet smooth relevance vector machine," *Computers and Geotechnics*, vol. 54, pp. 125–132, 2013.

[23] H. Khamesi, S. R. Torabi, H. Mirzaei-Nasirabad, and Z. Ghadiri, "Improving the performance of intelligent back analysis for tunneling using optimized fuzzy systems: case study of the Karaj Subway Line 2 in Iran," *Journal of Computing in Civil Engineering*, vol. 29, no. 6, p. 5014010, 2014.

[24] K. Ahangari, S. R. Moeinossadat, and D. Behnia, "Estimation of tunnelling-induced settlement by modern intelligent methods," *Soils and Foundations*, vol. 55, no. 4, pp. 737–748, 2015.

[25] M. Hasanipanah, M. Noorian-Bidgoli, D. Jahed Armaghani, and H. Khamesi, "Feasibility of PSO-ANN model for predicting surface settlement caused by tunneling," *Engineering with Computers*, vol. 32, no. 4, pp. 705–715, 2016.

[26] A. F. Cabalar, A. Cevik, and C. Gokceoglu, "Some applications of adaptive neuro-fuzzy inference system (ANFIS) in geotechnical engineering," *Computers and Geotechnics*, vol. 40, pp. 14–33, 2012.

[27] D. Karaboga and E. Kaya, "Training ANFIS by using an adaptive and hybrid artificial bee colony algorithm (aABC) for the identification of nonlinear static systems," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3531–3547, 2019.

[28] A. A. Al-Musawi, A. A. H. Alwanas, S. Q. Salih, Z. H. Ali, M. T. Tran, and Z. M. Yaseen, "Shear strength of SFRCB without stirrups simulation: implementation of hybrid artificial intelligence model," *Engineering with Computers*, vol. 36, no. 1, 2020.

[29] M. A. Jayaram, M. C. Nataraja, and C. N. Ravi Kumar, "Design of high performance concrete mixes through particle swarm optimization," *Journal of Intelligent Systems*, vol. 19, no. 3, pp. 249–264, 2010.

[30] M. Flint, S. Grünewald, and J. Coenders, "Ant colony optimization for ultra high performance concrete structures," *Designing and Building with UHPFRC*, 2013.

[31] G. Quaranta, A. Fiore, and G. C. Marano, "Optimum design of prestressed concrete beams using constrained differential evolution algorithm," *Structural and Multidisciplinary Optimization*, vol. 49, no. 3, 2014.

[32] C. a. Coello Coello, a. D. Christiansen, and F. S. Hernández, "A simple genetic algorithm for the design of reinforced concrete beams," *Engineering with Computers*, vol. 13, no. 4, 1997.

[33] K. Elbaz, S.-L. Shen, W.-J. Sun, Z.-Y. Yin, and A. Zhou, "Prediction model of shield performance during tunneling via incorporating improved particle swarm optimization into ANFIS," *IEEE Access*, vol. 8, pp. 39659–39671, 2020.

[34] P. Zhang, Z.-Y. Yin, Y.-F. Jin, and T. H. T. Chan, "A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest," *Engineering Geology*, vol. 265, Article ID 105328, 2020.

[35] M. Babanezhad, I. Behroyan, A. T. Nakhjiri, A. Marjani, M. Rezakazemi, and S. Shirazian, "High-performance hybrid modeling chemical reactors using differential evolution based fuzzy inference system," *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.

[36] H. Jing, H. N. Rad, M. Hasanipanah, D. J. Armaghani, and S. N. Qasem, "Design and implementation of a new tuned hybrid intelligent model to predict the uniaxial compressive strength of the rock using SFS-ANFIS," *Engineering with Computers*, pp. 1–18, 2020.

[37] L. Penghui, A. A. Ewees, B. H. Beyaztas et al., "Metaheuristic optimization algorithms hybridized with artificial intelligence model for soil temperature prediction: novel model," *IEEE Access*, vol. 8, pp. 51884–51904, 2020.

[38] S. Zhang, X.-N. Bui, N.-T. Trung, H. Nguyen, and H.-B. Bui, "Prediction of rock size distribution in mine bench blasting using a novel ant colony optimization-based boosted regression tree technique," *Natural Resources Research*, vol. 29, no. 2, pp. 867–886, 2020.

[39] Q.-T. Bui, M. Van Pham, Q.-H. Nguyen, L. X. Nguyen, and H. M. Pham, "Whale Optimization Algorithm and Adaptive Neuro-Fuzzy Inference System: a hybrid method for feature selection and land pattern classification," *International Journal of Remote Sensing*, pp. 1–16, 2019.

[40] A. Jaafari, E. K. Zenner, M. Panahi, and H. Shahabi, "Hybrid artificial intelligence models based on a neuro-fuzzy system and metaheuristic optimization algorithms for spatial prediction of wildfire probability," *Agricultural and Forest Meteorology*, vol. 266-267, pp. 198–207, 2019.

[41] K. Elbaz, S.-L. Shen, A. Zhou, D.-J. Yuan, and Y.-S. Xu, "Optimization of EPB shield performance with adaptive neuro-fuzzy inference system and genetic algorithm," *Applied Sciences*, vol. 9, no. 4, p. 780, 2019.

[42] P. A. Sari, M. Suhatril, N. Osman et al., "Developing a hybrid adoptive neuro-fuzzy inference system in predicting safety of factors of slopes subjected to surface eco-protection techniques," *Engineering with Computers*, vol. 36, no. 6, pp. 1–8, 2019.

[43] Z. M. Yaseen, W. H. Melini, H. W. Mohtar et al., "Implementation of Univariate Paradigm for streamflow simulation using hybrid data-driven model: case study in tropical region," *IEEE Access*, vol. 7, 2019.

[44] M. J. Varnamkhasti, "A hybrid of adaptive neuro-fuzzy inference system and genetic algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 25, no. 3, pp. 793–796, 2013.

[45] B. Mohammad and J. Varnamkhasti, "ANFISGA-adaptive neuro-fuzzy inference system genetic algorithm," *Global Journal of Computer Science and Technology*, vol. 11, no. 1, 2011.

[46] Z. Yuan, L. N. Wang, and X. Ji, "Prediction of concrete compressive strength: research on hybrid models genetic based algorithms and ANFIS," *Advances in Engineering Software*, 2014.

[47] S. Suwansawat and H. H. Einstein, "Artificial Neural Networks for Predicting the Maximum Surface Settlement Caused by EPB Shield Tunneling," *Tunnelling and Underground Space Technology*, 2006.

[48] J. S. R. Jang and C. T. Sun, "Neuro-fuzzy modeling and control," in *Proceedings of the IEEE*, vol. 83, no. 3, pp. 378–406, 1995.

[49] L. Rutkowski, K. Cpałka, R. Nowicki, A. Pokropińska, and R. Scherer, "Neuro-fuzzy systems," in *Proceedings of the Computational Complexity: Theory, Techniques, and Applications,*, 2012.

[50] J. Vieira, F. Dias, and A. Mota, "Neuro-fuzzy systems: a survey," in *Proceedings of the 5th WSEAS NNA International Conference on Neural Networks and Applications*, Udine, Italy, March 2004.

[51] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics*, 1993.

[52] N. Walia, H. Singh, and A. Sharma, "ANFIS: adaptive neuro-fuzzy inference system—a survey," *International Journal of Computer Applications*, 2015.

[53] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets and Systems*, vol. 28, no. 1, pp. 15–33, 1988.

[54] V. Nikolić, D. Petković, S. Shamshirband, and Ž. Ćojbašić, "Adaptive neuro-fuzzy estimation of diffuser effects on wind turbine performance," *Energy*, vol. 89, no. C, pp. 324–333, 2015.

[55] H. Tao, A. A. Ewees, A. O. Al-Sulttani et al., "Global solar radiation prediction over North Dakota using air temperature: development of novel hybrid intelligence model," *Energy Reports*, vol. 7, pp. 136–157, 2021.

[56] S. Maroufpoor, E. Maroufpoor, O. Bozorg-Haddad, J. Shiri, and Z. Mundher Yaseen, "Soil moisture simulation using hybrid artificial intelligent model: hybridization of adaptive neuro fuzzy inference system with grey wolf optimizer algorithm," *Journal of Hydrology*, vol. 575, 2019.

[57] G. Landeras, J. J. López, O. Kisi, and J. Shiri, "Comparison of gene expression programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain)," *Energy Conversion and Management*, 2012.

[58] J. Kennedy and R. Eberhart, "A new optimizer using particles swarm theory," in *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, Nagoya, Japan, October 1995.

[59] M. A. Mohandes, "Modeling global solar radiation using Particle Swarm Optimization (PSO)," *Solar Energy*, vol. 86, no. 11, pp. 3137–3145, 2012.

[60] J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting," *IEEE Transactions on Power Systems*, 2011.

[61] J. H. Holland, "Adaptation in natural and artificial systems: an introductory analysis with applications to biology," *Control and Artificial Intelligence*, MIT Press, Cambridge, MA, USA, 1975.

[62] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, 1997.

[63] M. Dorigo and G. Di Caro, "Ant colony optimization: a new meta-heuristic," in *Proceedings of the 1999 Congress on Evolutionary Computation, CEC 1999*, Washington, DC, USA, July 1999.

[64] A. A. H. Alwanas, A. A. Al-Musawi, S. Q. Salih, H. Tao, M. Ali, and Z. M. Yaseen, "Load-carrying capacity and mode failure simulation of beam-column joint connection: application of self-tuning machine learning model," *Engineering Structures*, vol. 194, 2019.

[65] B. Keshtegar, M. Bagheri, and Z. M. Yaseen, "Shear strength of steel fiber-unconfined reinforced concrete beam simulation: application of novel intelligent model," *Composite Structures*, vol. 212, 2019.

[66] Z. M. Yaseen, S. M. Awadh, A. Sharafati, and S. Shahid, "Complementary data-intelligence model for river flow simulation," *Journal of Hydrology*, vol. 567, 2018.

[67] Z. Yaseen, M. Ehteram, A. Sharafati, S. Shahid, N. Al-Ansari, and A. El-Shafie, "The integration of nature-inspired algorithms with least square support vector regression models: application to modeling river dissolved oxygen concentration," *Water*, vol. 10, no. 9, p. 1124, 2018.

[68] Z. A. Al-Sudani, S. Q. Salih, and Z. M. Yaseen, "Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation," *Journal of Hydrology*, vol. 573, pp. 1–12, 2019.

[69] A. Sharafati, K. Khosravi, P. Khosravinia et al., "The potential of novel data mining models for global solar radiation prediction," *International Journal of Environmental Science and Technology*, vol. 16, no. 11, pp. 7147–7164, 2019.

[70] A. Sharafati, R. Yasa, and H. M. Azamathulla, "Assessment of stochastic approaches in prediction of wave-induced pipeline scour depth," *Journal of Pipeline Systems Engineering and Practice*, vol. 9, no. 4, Article ID 04018024, 2018.

[71] S. Q. Salih and A. A. Alsewari, "A new algorithm for normal and large-scale optimization problems: nomadic people optimizer," *Neural Computing and Applications*, vol. 32, no. 14, pp. 1–28, 2019.

[72] A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, "Equilibrium optimizer: a novel optimization algorithm," *Knowledge-Based Systems*, vol. 191, Article ID 105190, 2020.

[73] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The arithmetic optimization algorithm," *Computer Methods in Applied Mechanics and Engineering*, vol. 376, p. 113609, 2021.

[74] V. Hayyolalam and A. A. P. Kazem, "Black widow optimization algorithm: a novel meta-heuristic approach for solving engineering optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 87, Article ID 103249, 2020.

Hindawi
Complexity
Volume 2021, Article ID 6656261, 9 pages
https://doi.org/10.1155/2021/6656261

Research Article

Optimization for Due-Date Assignment Single-Machine Scheduling under Group Technology

Li-Yan Wang,[1] **Mengqi Liu**,[2] **Ji-Bo Wang**,[1] **Yuan-Yuan Lu**,[3] **and Wei-Wei Liu**[4]

[1]*School of Science, Shenyang Aerospace University, Shenyang 110136, China*
[2]*Business School, Hunan University, Changsha 410082, Hunan, China*
[3]*College of Mathematics, Jilin Normal University, Siping, Jilin, China*
[4]*Department of Science, Shenyang Sport University, Shenyang 110102, China*

Correspondence should be addressed to Mengqi Liu; liumengqi163@126.com

Received 9 November 2020; Revised 20 January 2021; Accepted 1 February 2021; Published 5 March 2021

Academic Editor: Lei Xie

Copyright © 2021 Li-Yan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, the single-machine scheduling problem is studied by simultaneously considering due-date assignment and group technology (GT). The objective is to determine the optimal sequence of groups and jobs within groups and optimal due-date assignment to minimize the weighted sum of the absolute value in lateness and due-date assignment cost, where the weights are position dependent. For the common (CON) due-date assignment, slack (SLK) due-date assignment, and different (DIF) due-date assignment, an $O(n \log n)$ time algorithm is proposed, respectively, to solve the problem, where n is the number of jobs.

1. Introduction

In the manufacturing industry, it is well known that firms can increase production efficiency by adopting the group technology (GT). The group technology is an approach to manufacturing that seeks to improve efficiency in high-volume production by exploiting the similarities of different products and activities in their production (Neufeld et al. [1] and Yang et al. [2]). Through decades of application, people have found many advantages of the group technology. For example, changeover between different jobs in the same group is simplified, reducing the costs or time involved; jobs in the same group spend less time waiting, which results in less work-in-process inventory; jobs in the same group tend to move through production in a direct route, reducing the manufacturing lead time (see the work of Yang and Yang [3], Lu et al. [4], Yin et al. [5], Wang and Liu [6], Wang and Wang [7], Qin et al. [8], Lu et al. [9], Zhang et al. [10], Liu et al. [11], and Wang and Liang [12]).

In recent years, the problem of the due-date assignment has been closely focused on by scholars (see the work of Yin et al. [13, 14], Wang et al. [15], and Shabtay [16]). Due to the increasing interest in the Just-In-Time (JIT) system, the issue of schedule allocation is becoming more and more important in practical applications. In the classical scheduling problem, the due date is usually a given constant, while in the actual application of life, the duration of the job is not a fixed constant, but a decision variable. In order to strengthen the global competition and improve the service level for customers, the jobs are required to be processed too early or too late in the JIT system, which leads to the trouble of scheduling problems involving advance and delay costs and the expiration date of the construction period. For the early completion of the job, it means that we have to bear a certain inventory cost, while for the delayed completion of the job, we have to bear the contract penalty, and the customer's goodwill is damaged (see the work of Li et al. [17], Liu et al. [18], Wang et al. [19]). A lot of literature deals with the problems such as the CON, SLK, and some other due-date assignment methods considering jobs. However, under the group technology, there are relatively few studies on the problem of the assignment of jobs. Li et al. [17] considered three due-date assignment methods under group technology. The objective function is to minimize the weighted sum of earliness, tardiness, due-date assignment, and completion time. For the CON, SLK, and different due-date assignments

(DIFs), they proved that the problem can be solved in polynomial time, respectively. Ji et al. [20] studied the single-machine slack due-date window assignment scheduling with group technology. The objective function is to minimize the total cost including the earliness, tardiness, due window starting time, and due window size. They proved that the problem can be solved in $O(n \log n)$ time.

Brucker [21] considered the CON due-date scheduling problem of minimizing the total cost comprising the total weighted absolute lateness value and common due-date (CON) cost, where the weight is position dependent. He proved that the problem can be solved in a time $O(n \log n)$. Liu et al. [18] considered the SLK due-date assignment scheduling, and the goal is to minimize the total cost comprising the total weighted absolute lateness value and common flow allowance (SLK) cost, where the weight is a position-dependent weight. They proved that the problem can be solved in a time $O(n \log n)$. Wang et al. [19] studied the scheduling problems of single machine resource allocation in job-dependent learning effects. Under the linear and convex resource consumption functions, they proved that the CON and SLK due-date assignment problems can be solved in polynomial time, respectively. Sun et al. [22] considered single-machine scheduling problems on resource allocation, group technology, and learning effect. Under the linear and convex resource consumption functions, they proved that the SLK due-date assignment problem can be solved in polynomial time.

According to this study, we consider due-date assignment and scheduling problems with group technology. Under the assumption of group technology, the jobs are classified into groups by exploiting the similarities of different products and activities in their production. In this paper, we proceed to the study Brucker [21], Liu et al. [18], and Wang et al. [19], which is an extension of their work from considering the CON, SLK, and DIF due-date assignment scheduling problems with group technology and position-dependent weights. We organize the rest of the paper as follows: the problem is formulated in Section 2, several results of the optimal solution are introduced in Section 3, and the conclusions are summarized in Section 4.

2. Problem Formulation

In the study, the problem can be formally described as follows: there are n independent non-preemption jobs grouped into f groups, i.e., $\{G_1, G_2, \ldots, G_f\}$. A single machine and all the jobs are available at time zero, and the single machine can handle one job at a time and preemption is not allowed. Let the number of jobs in group G_i be n_i, i.e., $\{J_{i,1}, J_{i,2}, \ldots, J_{i,n_i}\}$, where $J_{i,j}$ denotes the job J_j of group G_i, $i = 1, 2, \ldots, f$, $j = 1, 2, \ldots, n_i$ and $n_1 + n_2 + \cdots + n_f = n$. Jobs in the same group are processed consecutively and do not need setup times. Let p_{ij} denote the processing time of job $J_{i,j}$ and s_i be the sequence-independent machine setup time incurred before the process of the first job of group G_i. Each job $J_{i,j}$ $(i = 1, 2, \ldots, f; j = 1, 2, \ldots, n_i)$ has a due date $d_{i,j}$, which is assignable according to one of the following three due-date assignment methods:

(1) The common (CON) due-date assignment in which all jobs of group G_i are assigned the same due date, i.e., $d_{i,j} = d_i^{\text{opt}}$ for $i = 1, 2, \ldots, f$ and $j = 1, 2, \ldots, n_i$

(2) The common flow allowance (slack, SLK) assignment in which the due-date $d_{i,j}$ for all jobs of group G_i are assigned an equal flow allowance that is equal to its processing time plus the common flow allowance, i.e., $d_{i,j} = p_{i,j} + q_i^{\text{opt}}$ for $i = 1, 2, \ldots, f$ and $j = 1, 2, \ldots, n_i$

(3) The different (DIF) due-date assignments in which the due date $d_{i,j}$ for all jobs of group G_i are assigned a different due date with no restrictions ($i = 1, 2, \ldots, f$ and $j = 1, 2, \ldots, n_i$)

Let $C_{i,j}$ be the completion time of job $J_{i,j}$. The goal is to determine the due date $d_{i,j}$ (i.e., the CON d_i^{opt}, the SLK q_i^{opt}, and the DIF $d_{i,j}$) and an optimal sequence π^* such that the following objective functions are to be minimized:

$$\begin{aligned} Z_1(\pi, D(\pi)) &= \sum_{i=1}^{f}\left(\sum_{j=1}^{n_i} \omega_{ij}\left|L_{\pi(i,j)}\right| + \omega_{i,0} d_i^{\text{opt}}\right), \\ Z_2(\pi, D(\pi)) &= \sum_{i=1}^{f}\left(\sum_{j=1}^{n_i} \omega_{i,j}\left|L_{\pi(i,j)}\right| + \omega_{i,0} q_i^{\text{opt}}\right), \\ Z_3(\pi, D(\pi)) &= \sum_{i=1}^{f}\sum_{j=1}^{n_i}\left(\omega_{i,j}\left|L_{\pi(i,j)}\right| + \omega_{i,0} d_{\pi(i,j)}\right), \end{aligned} \tag{1}$$

where $\omega_{i,j}$ is the nonnegative weight of the jth position in group G_i (i.e., position-dependent weight, $i = 1, 2, \ldots, f; j = 1, 2, \ldots, n_i$), $\omega_{i,0}$ denotes the unit cost of d_i^{opt} (q_i^{opt} and $d_{\pi(i,j)}$), $i = 1, 2, \ldots, f$, $D(\pi) = (d_{1,1}(\pi), \ldots, d_{1,n_1}(\pi); \ldots; d_{f,1}(\pi), \ldots, d_{f,n_f}(\pi))$ is the due-date assignment vector under schedule π, and $L_{i,j} = C_{i,j} - d_{i,j}$ is the lateness of job $J_{i,j}$. Note that $|L_{\pi(i,j)}| = E_{\pi(i,j)} + T_{\pi(i,j)}$, where $E_{\pi(i,j)} = \max\{0, d_{\pi(i,j)} - C_{\pi(i,j)}\}$ is the earliness of job J_{ij}, $T_{\pi(i,j)} = \max\{0, C_{\pi(i,j)} - d_{\pi(i,j)}\}$ is the tardiness of job J_{ij}, and $\pi(i, j)$ represents the job that is in the jth position of group G_i in π for $1 \le k \le n_i$. By using the three-field notation (see Graham et al. [23]), the problems can be denoted by

$$\begin{aligned} &1|\text{GT, CON}|\sum_{i=1}^{f}\left(\sum_{j=1}^{n_i} \omega_{ij}\left|L_{\pi(i,j)}\right| + \omega_{i,0} d_i^{\text{opt}}\right), \\ &1|\text{GT, SLK}|\sum_{i=1}^{f}\left(\sum_{j=1}^{n_i} \omega_{i,j}\left|L_{\pi(i,j)}\right| + \omega_{i,0} q_i^{\text{opt}}\right), \\ &1|\text{GT, DIF}|\sum_{i=1}^{f}\sum_{j=1}^{n_i}\left(\omega_{i,j}\left|L_{\pi(i,j)}\right| + \omega_{i,0} d_{\pi(i,j)}\right), \end{aligned} \tag{2}$$

respectively, where GT denotes the group technology.

3. Main Results

Lemma 1. *For each of the three due-date assignment methods, there exists an optimal schedule with zero machine idle times.*

Proof. It is omitted due to simplicity. □

Therefore, it is convenient to introduce a dummy job $J_{i,0}$ of group G_i for $i = 1, 2, \ldots, f$ with processing time $p_{i,0} = 0$ and weight $\omega_{i,0}$ which is always scheduled at time 0, i.e., $\pi(i, 0) = 0$. Consequently, for any given job schedule π, the completion times can be calculated by the following equation:

$$C_{\pi(i,j)} = S_i(\pi) + s_i + \sum_{k=1}^{j} p_{\pi(i,k)}, \tag{3}$$

where $S_i(\pi)$ denotes the starting time of group G_i in schedule π for $i = 1, 2, \ldots, f$.

Lemma 2. *If the sequence of groups is fixed, for the CON and SLK due-date assignment methods, there exists an optimal sequence with the property that d_i^{opt} (q_i^{opt}) coincides with the job completion times of group G_i for $i = 1, 2, \ldots, f$.*

Proof. Let the sequence of groups be fixed. Consider an optimal sequence π in G_i.

(a) First, we prove that the common due-date d_i^{opt} of group G_i coincides with the completion time of a job of group G_i. Assume that $C_{\pi(i,k_i)} < d_i^{\text{opt}} < C_{\pi(i,k_i+1)}$; therefore,

$$\begin{aligned} Z_{1i}(\pi, D_i(\pi)) &= \sum_{j=1}^{n_i} \omega_{i,j} \left| L_{\pi(i,j)} \right| + \omega_{i,0} d_i^{\text{opt}}, \\ &= \sum_{j=1}^{n_i} \omega_{i,j} \left| C_{\pi(i,j)} - d_i^{\text{opt}} \right| + \omega_{i,0} d_i^{\text{opt}}, \\ &= \sum_{j=1}^{k_i} \omega_{i,j} \left(d_i^{\text{opt}} - C_{\pi(i,j)} \right) \\ &\quad + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(C_{\pi(i,j)} - d_i^{\text{opt}} \right) + \omega_{i,0} d_i^{\text{opt}}. \end{aligned} \tag{4}$$

If $d_i^{\text{opt}} = C_{\pi(i,k_i)}$, we have

$$\begin{aligned} Z_{1i}^1(\pi, D_i(\pi)) &= \sum_{j=1}^{k_i} \omega_{i,j} \left(C_{\pi(i,k_i)} - C_{\pi(i,j)} \right) \\ &\quad + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(C_{\pi(i,j)} - C_{\pi(i,k_i)} \right) \\ &\quad + \omega_{i,0} C_{\pi(i,k_i)}. \end{aligned} \tag{5}$$

If $d_i^{\text{opt}} = C_{\pi(i,k_i+1)}$, we have

$$\begin{aligned} Z_{1i}^2(\pi, D_i(\pi)) &= \sum_{j=1}^{k_i} \omega_{i,j} \left(C_{\pi(i,k_i+1)} - C_{\pi(i,j)} \right) \\ &\quad + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(C_{\pi(i,j)} - C_{\pi(i,k_i+1)} \right) \\ &\quad + \omega_{i,0} C_{\pi(i,k_i+1)}. \end{aligned} \tag{6}$$

Let $x = d_i^{\text{opt}} - C_{\pi(i,k_i)} > 0$, $y = C_{\pi(i,k_i+1)} - d_i^{\text{opt}} > 0$, then we can obtain that

$$\begin{aligned} &Z_{1i}(\pi, D_i(\pi)) - Z_{1i}^1(\pi, D_i(\pi)), \\ &\quad = \sum_{j=1}^{k_i} \omega_{i,j} \left(d_i^{\text{opt}} - C_{\pi(i,k_i)} \right) + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(C_{\pi(i,k_i)} - d_i^{\text{opt}} \right) + \omega_{i,0} \left(d_i^{\text{opt}} - C_{\pi(i,k_i)} \right), \\ &\quad = \sum_{j=0}^{k_i} \omega_{i,j} \left(d_i^{\text{opt}} - C_{\pi(i,k_i)} \right) + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(C_{\pi(i,k_i)} - d_i^{\text{opt}} \right), \\ &\quad = x \left(\sum_{j=0}^{k_i} \omega_{i,j} - \sum_{j=k_i+1}^{n_i} \omega_{i,j} \right), \end{aligned} \tag{7}$$

$$\begin{aligned} &Z_{1i}(\pi, D_i(\pi)) - Z_{1i}^2(\pi, D_i(\pi)), \\ &\quad = \sum_{j=1}^{k_i} \omega_{i,j} \left(d_i^{\text{opt}} - C_{\pi(i,k_i+1)} \right) + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(C_{\pi(i,k_i+1)} - d_i^{\text{opt}} \right) + \omega_{i,0} \left(d_i^{\text{opt}} - C_{\pi(i,k_i+1)} \right), \\ &\quad = \sum_{j=0}^{k_i} \omega_{i,j} \left(d_i^{\text{opt}} - C_{\pi(i,k_i+1)} \right) + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(C_{\pi(i,k_i+1)} - d_i^{\text{opt}} \right), \\ &\quad = -y \left(\sum_{j=0}^{k_i} \omega_{i,j} - \sum_{j=k_i+1}^{n_i} \omega_{i,j} \right). \end{aligned} \tag{8}$$

Thus, $Z_{1i}^{1}(\pi, D_i(\pi)) \le Z_{1i}(\pi, D_i(\pi))$ if $\sum_{j=0}^{k_i} \omega_{i,j} \ge \sum_{j=k_i+1}^{n_i} \omega_{i,j}$ and $Z_{1i}^{2}(\pi, D_i(\pi)) \le Z_{1i}(\pi, D_i(\pi))$ otherwise.

(b) Assume that $C_{\pi(i,l_i)} < q_i^{\text{opt}} < C_{\pi(i,l_i+1)}$; therefore,

$$\begin{aligned} Z_{2i}(\pi) &= \sum_{j=1}^{n_i} \omega_{i,j} \left| L_{\pi(i,j)} \right| + \omega_{i,0} q_i^{\text{opt}}, \\ &= \sum_{j=1}^{n_i} \omega_{i,j} \left| C_{\pi(i,j)} - p_{\pi(i,j)} - q_i^{\text{opt}} \right| + \omega_{i,0} q_i^{\text{opt}}, \\ &= \sum_{j=1}^{n_i} \omega_{i,j} \left| C_{\pi(i,j-1)} - q_i^{\text{opt}} \right| + \omega_{i,0} q_i^{\text{opt}}, \\ &= \sum_{j=1}^{l_i+1} \omega_{i,j} \left(q_i^{\text{opt}} - C_{\pi(i,j-1)} \right) + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(C_{\pi(i,j-1)} - q_i^{\text{opt}} \right) + \omega_{i,0} q_i^{\text{opt}}, \end{aligned} \tag{9}$$

where $C_{\pi(i,0)} = S_i(\pi) + s_i$.

If $q_i^{\text{opt}} = C_{\pi(i,l_i)}$, we have

$$\begin{aligned} Z'_{2i}(\pi) &= \sum_{j=1}^{l_i+1} \omega_{i,j} \left(C_{\pi(i,l_i)} - C_{\pi(i,j-1)} \right) \\ &\quad + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(C_{\pi(i,j-1)} - C_{\pi(i,l_i)} \right) + \omega_{i,0} C_{\pi(i,l_i)}. \end{aligned} \tag{10}$$

If $q_i^{\text{opt}} = C_{\pi(i,l_i+1)}$, we have

$$\begin{aligned} Z''_{2i}(\pi) &= \sum_{j=1}^{l_i+1} \omega_{i,j} \left(C_{\pi(i,l_i+1)} - C_{\pi(i,j-1)} \right) \\ &\quad + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(C_{\pi(i,j-1)} - C_{\pi(i,l_i+1)} \right) + \omega_{i,0} C_{\pi(i,l_i+1)}. \end{aligned} \tag{11}$$

Let $x' = q_i^{\text{opt}} - C_{\pi(i,l_i)} > 0$ and $y' = C_{\pi(i,l_i+1)} - q_i^{\text{opt}} > 0$; then, we have

$$\begin{aligned} &Z_{2i}(\pi, D_i(\pi)) - Z'_{2i}(\pi, D_i(\pi)), \\ &\quad = \sum_{j=1}^{l_i+1} \omega_{i,j} \left(q_i^{\text{opt}} - C_{\pi(i,l_i)} \right) + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(C_{\pi(i,l_i)} - q_i^{\text{opt}} \right) + \omega_{i,0} \left(q_i^{\text{opt}} - C_{\pi(i,l_i)} \right), \\ &\quad = \sum_{j=0}^{l_i+1} \omega_{i,j} \left(q_i^{\text{opt}} - C_{\pi(i,l_i)} \right) + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(C_{\pi(i,l_i)} - q_i^{\text{opt}} \right), \\ &\quad = x' \left(\sum_{j=0}^{l_i+1} \omega_{i,j} - \sum_{j=l_i+2}^{n_i} \omega_{i,j} \right), \\ &Z_{2i}(\pi, D_i(\pi)) - Z''_{2i}(\pi, D_i(\pi)), \\ &\quad = \sum_{j=1}^{l_i+1} \omega_{i,j} \left(q_i^{\text{opt}} - C_{\pi(i,l_i+1)} \right) + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(C_{\pi(i,l_i+1)} - q_i^{\text{opt}} \right) + \omega_{i,0} \left(q_i^{\text{opt}} - C_{\pi(i,l_i+1)} \right), \\ &\quad = \sum_{j=0}^{l_i+1} \omega_{i,j} \left(q_i^{\text{opt}} - C_{\pi(i,l_i+1)} \right) + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(C_{\pi(i,l_i+1)} - q_i^{\text{opt}} \right), \\ &\quad = -y' \left(\sum_{j=0}^{l_i+1} \omega_{i,j} - \sum_{j=l_i+2}^{n_i} \omega_{i,j} \right). \end{aligned} \tag{12}$$

Thus, $Z'_{2i}(\pi, D_i(\pi)) \le Z_{2i}(\pi, D_i(\pi))$ if $\sum_{j=0}^{l_i+1} \omega_{i,j} \ge \sum_{j=l_i+2}^{n_i} \omega_{i,j}$ and $Z''_{2i}(\pi, D_i(\pi)) \le Z_{2i}(\pi, D_i(\pi))$ otherwise.

From (a) and (b), this implies that d_i^{opt} (q_i^{opt}) is equal to the job completion times of group G_i for $i = 1, 2, \ldots, f$. □

Lemma 3. *For the CON due-date assignment, if the sequence of groups is fixed, for a given schedule π, there exists an optimal schedule of group G_i in which $d_i^{opt} = C_{\pi(i,k_i)}$, where k_i is a median for the sequence $\omega_{i,0}, \omega_{i,1}, \ldots, \omega_{i,n_i}$ and $\sum_{j=0}^{k_i-1} \omega_{i,j} \le \sum_{j=k_i}^{n_i} \omega_{i,j}$ and $\sum_{j=0}^{k_i} \omega_{i,j} \ge \sum_{j=k_i+1}^{n_i} \omega_{i,j}$.*

Proof. By using the technique of small perturbations, from Lemma 2, we assume that $d_i^{\text{opt}} = C_{\pi(i,k_i)}$. Applying (7) and (8) to the cases of moving the common due-date x units of time to the left (right), we have $\sum_{j=k_i}^{n_i} \omega_{i,j} - \sum_{j=0}^{k_i-1} \omega_{i,j} \ge 0$ and $\sum_{j=0}^{k_i} \omega_{i,j} - \sum_{j=k_i+1}^{n_i} \omega_{i,j} \ge 0$. □

Lemma 4. *For the SLK due-date assignment, if the sequence of groups is fixed, for a given schedule π, there exists an optimal schedule of group G_i in which $q_i^{opt} = C_{\pi(i,l_i)}$, where l_i is a median for the sequence $\omega_{i,0}, \omega_{i,1}, \ldots, \omega_{i,n_i}$ and $\sum_{j=0}^{l_i} \omega_{i,j} \le \sum_{j=l_i+1}^{n_i} \omega_{i,j}$ and $\sum_{j=0}^{l_i+1} \omega_{i,j} \ge \sum_{j=l_i+2}^{n_i} \omega_{i,j}$.*

Proof. It is similar to the proof of Lemma 3. □

Lemma 5. *For the DIF due-date assignment method, there exists an optimal sequence such that $d_{\pi(i,j)} \le C_{\pi(i,j)}$.*

Proof. We consider the case that contradicts this optimal property. Consider an optimal sequence π in G_i, if $C_{\pi(i,j)} < d_{\pi(i,j)}$, and then, the total cost for job $J_{\pi(i,j)}$ is

$$Z_{\pi(i,j)} = \omega_{i,j}\left(d_{\pi(i,j)} - C_{\pi(i,j)}\right) + \omega_{i,0} d_{\pi(i,j)}. \tag{13}$$

We shift $d_{\pi(i,j)}$ to the left such that $d_{\pi(i,j)} = C_{\pi(i,j)}$, and we have

$$\widetilde{Z}_{\pi(i,j)} = \omega_{i,0} d_{\pi(i,j)} < Z_{\pi(i,j)}. \tag{14}$$

Hence, the case $C_{\pi(i,j)} < d_{\pi(i,j)}$ is not an optimal due-date assignment. Lemma 5 is proved. □

Lemma 6. *For the DIF due-date assignment method, let π be a given sequence, and the optimal due-date $d_{\pi(i,j)}$ for job $J_{\pi(i,j)}$ can be obtained as follows:*

$$d_{\pi(i,j)} = \begin{cases} 0, & \omega_{i,j} \le \omega_{i,0}, \\ C_{\pi(i,j)}, & \omega_{i,j} > \omega_{i,0}. \end{cases} \tag{15}$$

Proof. From Lemma 5, for job $J_{\pi(i,j)}$, we have

$$Z_{\pi(i,j)} = \omega_{i,j}\left(C_{\pi(i,j)} - d_{\pi(i,j)}\right) + \omega_{i,0} d_{\pi(i,j)} = \omega_{i,j} C_{\pi(i,j)} + \left(\omega_{i,0} - \omega_{i,j}\right) d_{\pi(i,j)}. \tag{16}$$

Obviously, if $\omega_{i,j} \le \omega_{i,0}$, $d_{\pi(i,j)}$ should be 0, if $\omega_{i,j} > \omega_{i,0} d_{\pi(i,j)}$ should be $C_{\pi(i,j)}$. Hence, the total cost for job $J_{\pi(i,j)}$ is

$$Z_{\pi(i,j)} = \eta_{i,j} C_{\pi(i,j)}, \tag{17}$$

where $\eta_{i,j} = \min\{\omega_{i,j}, \omega_{i,0}\}$. □

For the CON due-date assignment, obviously, the value of k_i ($i = 1, 2, \ldots, f$) given in Lemma 3 is independent of the job processing times and the job sequence. Therefore, it is optimal for any job sequence within each group. For a given schedule π, $d_i^{\text{opt}} = C_{\pi(i,k_i)}$, and the total cost of all the jobs within G_i for $i = 1, \ldots, m$ is given by

$$\begin{aligned}
Z_{1i}(\pi) &= \sum_{j=1}^{n_i} \omega_{i,j} \left| L_{\pi(i,j)} \right| + \omega_{i,0} d_i^{\text{opt}}, \\
&= \sum_{j=1}^{n_i} \omega_{i,j} \left| C_{\pi(i,j)} - d_i^{\text{opt}} \right| + \omega_{i,0} d_i^{\text{opt}}, \\
&= \omega_{i,0} C_{\pi(i,k_i)} + \sum_{j=1}^{k_i} \omega_{i,j} \left(C_{\pi(i,k_i)} - C_{\pi(i,j)} \right) + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(C_{\pi(i,j)} - C_{\pi(i,k_i)} \right), \\
&= \omega_{i,0} (S_i + s_i) + \sum_{j=0}^{k_i} \omega_{i,j} \left(\sum_{h=j+1}^{k_i} p_{\pi(i,h)} \right) + \sum_{j=k_i+1}^{n_i} \omega_{i,j} \left(\sum_{h=k_i+1}^{j} p_{\pi(i,h)} \right), \\
&= \sum_{h=1}^{k_i} p_{\pi(i,h)} \left(\sum_{j=0}^{h-1} \omega_{i,j} \right) + \sum_{h=k_i+1}^{n_i} p_{\pi(i,h)} \left(\sum_{j=h}^{n_i} \omega_{i,j} \right) + \omega_{i,0} (S_i + s_i), \\
&= \sum_{j=1}^{n_i} W_{i,j} p_{\pi(i,j)} + \omega_{i,0} (S_i + s_i),
\end{aligned} \tag{18}$$

where

$$W_{i,j} = \begin{cases} \sum_{h=0}^{j-1} \omega_{i,h}, & j = 1, 2, \ldots, k_i, \\ \sum_{h=j}^{n_i} \omega_{i,h}, & j = k_i + 1, \ldots, n_i. \end{cases} \tag{19}$$

For the SLK due-date assignment, obviously, the value of l_i $(i = 1, 2, \ldots, f)$ given in Lemma 4 is independent of the job processing times and the job sequence. Therefore, it is optimal for any job sequence within each group. For a given schedule π, $q_i^{\text{opt}} = C_{\pi(i,l_i)}$, and the total cost of all the jobs within G_i for $i = 1, 2, \ldots, f$ is given by

$$\begin{aligned} Z_{2i}(\pi) &= \sum_{j=1}^{n_i} \omega_{i,j} \left| L_{\pi(i,j)} \right| + \omega_{i,0} q_i^{\text{opt}}, \\ &= \sum_{j=1}^{n_i} \omega_{i,j} \left| C_{\pi(i,j)} - p_{\pi(i,j)} - q_i^{\text{opt}} \right| + \omega_{i,0} q_i^{\text{opt}}, \\ &= \sum_{j=1}^{n_i} \omega_{i,j} \left| C_{\pi(i,j-1)} - q_i^{\text{opt}} \right| + \omega_{i,0} q_i^{\text{opt}}, \\ &= \sum_{j=1}^{n_i} \omega_{i,j} \left| C_{\pi(i,j-1)} - C_{\pi(i,l_i)} \right| + \omega_{i,0} C_{\pi(i,l_i)}, \\ &= \sum_{j=1}^{l_i+1} \omega_{i,j} \left(C_{\pi(i,l_i)} - C_{\pi(i,j-1)} \right) + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(C_{\pi(i,j-1)} - C_{\pi(i,l_i)} \right) + \omega_{i,0} C_{\pi(i,l_i)}, \\ &= \sum_{j=0}^{l_i+1} \omega_{i,j} \left(\sum_{h=j}^{l_i} p_{\pi(i,h)} \right) + \sum_{j=l_i+2}^{n_i} \omega_{i,j} \left(\sum_{h=l_i+1}^{j-1} p_{\pi(i,h)} \right) + \omega_{i,0} (S_i + s_i), \\ &= \sum_{j=1}^{l_i} p_{\pi(i,j)} \left(\sum_{h=0}^{j} \omega_{i,h} \right) + \sum_{j=l_i+1}^{n_i-1} p_{\pi(i,j)} \left(\sum_{h=j+1}^{n_i} \omega_{i,h} \right) + \omega_{i,0} (S_i + s_i), \\ &= \sum_{j=1}^{n_i} V_{i,j} p_{\pi(i,j)} + \omega_{i,0} (S_i + s_i), \end{aligned} \tag{20}$$

where

$$V_{i,j} = \begin{cases} \sum_{h=0}^{j-1} \omega_{i,h}, & j = 1, 2, \ldots, l_i, \\ \sum_{h=j+1}^{n_i} \omega_{i,h}, & j = l_i + 1, \ldots, n_i - 1, \\ 0, & j = n_i. \end{cases} \tag{21}$$

For the DIF due-date assignment, from Lemma 6, the total cost of all the jobs within G_i for $i = 1, 2, \ldots, f$ is given by

$$\begin{aligned} Z_{3i}(\pi) &= \sum_{j=1}^{n_i} \left(\omega_{i,j} \left| L_{\pi(i,j)} \right| + \omega_{i,0} d_{\pi(i,j)} \right), \\ &= \sum_{j=1}^{n_i} \eta_{i,j} C_{\pi(i,j)}, \\ &= \sum_{j=1}^{n_i} \eta_{i,j} \left(S_i + s_i + \sum_{h=1}^{j} p_{\pi(i,h)} \right), \\ &= \sum_{j=1}^{n_i} U_{i,j} p_{\pi(i,j)} + (S_i + s_i) \sum_{j=1}^{n_i} \omega_{i,j}, \end{aligned} \tag{22}$$

where

$$U_{i,j} = \sum_{h=j}^{n_i} \eta_{i,h}, \quad h = 1, 2, \ldots, n_i. \tag{23}$$

Obviously, from (18), (20), and (22), the term $\sum_{j=1}^{n_i} W_{i,j} p_{\pi(i,j)}$ ($\sum_{j=1}^{n_i} V_{i,j} p_{\pi(i,j)}$ and $\sum_{j=1}^{n_i} U_{i,j} p_{\pi(i,j)}$) is only concerned with the job processing sequence within group G_i and can be minimized by the HLP rule (see the work of Hardy et al. [24]), i.e., the optimal job sequence within group G_i $(i = 1, 2, \ldots, f)$ can be obtained by arranging the elements of the $W_{i,j}$ ($V_{i,j}$ and $U_{i,j}$) and $p_{i,j}$ vectors in opposite orders. The term $\omega_{i,0}(S_i + s_i)$ $((S_i + s_i) \sum_{j=1}^{n_i} \omega_{i,j})$ is only dependent on the starting time of the group G_i, and the optimal sequence of the groups $\{G_1, G_2, \ldots, G_f\}$ can be obtained by the following lemma.

Lemma 7. *For the problems* $1|\text{GT}, \text{CON}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} d_i^{\text{opt}})$ *and* $1|\text{GT}, \text{SLK}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{i,j} |L_{\pi(i,j)}| + \omega_{i,0} q_i^{\text{opt}})$, *the optimal group sequence can be obtained by arranging the groups in nondecreasing order of* $\theta_i = (s_i + \sum_{j=1}^{n_i} p_{i,j}) \omega_{i,0}$, $i = 1, 2, \ldots, f$, *respectively. For the problem* $1|\text{GT}, \text{DIF}| \sum_{i=1}^{f} \sum_{j=1}^{n_i} (\omega_{i,j} |L_{\pi(i,j)}| + \omega_{i,0} d_{\pi(i,j)})$, *the optimal group sequence can be obtained by*

arranging the groups in nondecreasing order of $\theta_i = (s_i + \sum_{j=1}^{n_i} p_{i,j}) \sum_{j=1}^{n_i} \omega_{i,j}$, $i = 1, 2, \ldots, f$.

Proof. For the problem $1|\mathrm{GT}, \mathrm{CON}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} d_i^{\mathrm{opt}})$, let π and π' be two sequences where the difference between π and π' is a pairwise interchange of two adjacent groups G_k and G_l, that is, $\pi = [A, G_k, G_l, B]$ and $\pi' = [A, G_l, G_k, B]$, where A and B are partial sequences. To show π dominates π', it suffices to show that $Z_1(\pi, D(\pi)) \le Z_1(\pi', D(\pi'))$. We assume that S denotes the completion time of the last job in A, and we have

$$\begin{aligned} S_l(\pi) &= S + s_k + \sum_{j=1}^{n_k} p_{k,j}, \\ S_k(\pi') &= S + s_l + \sum_{j=1}^{n_l} p_{l,j}. \end{aligned} \tag{24}$$

Suppose that

$$\frac{s_k + \sum_{j=1}^{n_k} p_{k,j}}{\omega_{k,0}} \le \frac{s_l + \sum_{j=1}^{n_l} p_{l,j}}{\omega_{l,0}}, \tag{25}$$

and from (18), we have

$$\begin{aligned} &Z_1(\pi, D(\pi)) - Z_1(\pi', D(\pi')), \\ &\quad = \omega_{k,0}(S + s_k) + \omega_{l,0}\left(S + s_k + \sum_{j=1}^{n_k} p_{kj} + s_l\right) \\ &\quad - \omega_{l,0}(S + s_l) - \omega_{k,0}\left(S + s_l + \sum_{j=1}^{n_l} p_{lj} + s_k\right), \\ &\quad = \omega_{l,0}\left(s_k + \sum_{j=1}^{n_k} p_{k,j}\right) - \omega_{k,0}\left(s_l + \sum_{j=1}^{n_l} p_{l,j}\right) \\ &\quad \le 0. \end{aligned} \tag{26}$$

Therefore, $Z_1(\pi, D(\pi)) \le Z_1(\pi', D(\pi'))$. This completes the proof of the problem $1|\mathrm{GT}, \mathrm{CON}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} d_i^{\mathrm{opt}})$.

For the problem $1|\mathrm{GT}, \mathrm{SLK}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} q_i^{\mathrm{opt}})$, the proof can be obtained similarly.

For the problem $1|\mathrm{GT}, \mathrm{DIF}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} d_i^{\mathrm{opt}})$, stemming from the proof of the problem $1|\mathrm{GT}, \mathrm{CON}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j))}| + \omega_{i,0} d_i^{\mathrm{opt}})$, we suppose that

$$\frac{s_k + \sum_{j=1}^{n_i} p_{k,j}}{\sum_{j=1}^{n_k} \omega_{k,j}} \le \frac{s_l + \sum_{j=1}^{n_i} p_{l,j}}{\sum_{j=1}^{n_l} \omega_{l,j}}, \tag{27}$$

and from (22), we have

$$\begin{aligned} &Z_3(\pi, D(\pi)) - Z_3(\pi', D(\pi')), \\ &\quad = (S + s_k) \sum_{j=1}^{n_k} \omega_{k,j} + \left(S + s_k + \sum_{j=1}^{n_k} p_{kj} + s_l\right) \sum_{j=1}^{n_l} \omega_{l,j} - (S + s_l) \sum_{j=1}^{n_l} \omega_{l,j} - \left(S + s_l + \sum_{j=1}^{n_l} p_{lj} + s_k\right) \sum_{j=1}^{n_k} \omega_{k,j}, \\ &\quad = \sum_{j=1}^{n_l} \omega_{l,j}\left(s_k + \sum_{j=1}^{n_k} p_{k,j}\right) - \sum_{j=1}^{n_k} \omega_{k,j}\left(s_l + \sum_{j=1}^{n_l} p_{l,j}\right) \\ &\quad \le 0. \end{aligned} \tag{28}$$

Therefore, $Z_1(\pi, D(\pi)) \le Z_1(\pi', D(\pi'))$. This completes the proof of the problem $1|\mathrm{GT}, \mathrm{CON}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} d_i^{\mathrm{opt}})$. □

Based on the abovementioned analysis, we can present the following algorithm to solve the problems $1|\mathrm{GT}, \mathrm{CON}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} d_i^{\mathrm{opt}})$, $1|\mathrm{GT}, \mathrm{SLK}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{i,j} |L_{\pi(i,j)}| + \omega_{i,0} q_i^{\mathrm{opt}})$, and $1|\mathrm{GT}, \mathrm{DIF}| \sum_{i=1}^{f} \sum_{j=1}^{n_i} (\omega_{i,j} |L_{\pi(i,j)}| + \omega_{i,0} d_{\pi(i,j)})$.

Theorem 1. *The problems* $1|\mathrm{GT}, \mathrm{CON}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} d_i^{\mathrm{opt}})$, $1|\mathrm{GT}, \mathrm{SLK}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{i,j} |L_{\pi(i,j)}| + \omega_{i,0} q_i^{\mathrm{opt}})$, *and* $1|\mathrm{GT}, \mathrm{DIF}| \sum_{i=1}^{f} \sum_{j=1}^{n_i} (\omega_{i,j} |L_{\pi(i,j)}| + \omega_{i,0} d_{\pi(i,j)})$ *can be solved in* $O(n \log n)$ *time, respectively.*

Proof. Step 1 and Step 4 need time $O(n)$, respectively. Step 2 needs $O(\sum_{i=1}^{f} n_i \log n_i) \le O(n \log n)$ time. Step 3 needs $O(f \log f) \le O(n \log n)$ time ($\sum_{i=1}^{f} n_i = n$ and $f < n$). Thus, the total time complexity of Algorithm 1 is $O(n \log n)$. □

Example 1. We only consider the problem $1|\mathrm{GT}, \mathrm{CON}| \sum_{i=1}^{f} (\sum_{j=1}^{n_i} \omega_{ij} |L_{\pi(i,j)}| + \omega_{i,0} d_i^{\mathrm{opt}})$. Consider $n = 12, f = 3$, G_1: $[J_{1,1}, J_{1,2}, J_{1,3}]$, $p_{1,1} = 2, p_{1,2} = 4, p_{1,3} = 6$, $\omega_{1,1} = 2$, $\omega_{1,2} = 4, \omega_{1,3} = 3, \omega_{1,0} = 2, s_1 = 2$; G_2: $[J_{2,1}, J_{2,2}, J_{2,3}, J_{2,4}]$, $p_{2,1} = 5, p_{2,2} = 8, p_{2,3} = 4, p_{2,4} = 3$, $\omega_{2,1} = 3, \omega_{2,2} = 2$, $\omega_{2,3} = 2$, $\omega_{2,4} = 3, \omega_{2,0} = 1, s_2 = 4$; and G_3: $[J_{,}, J_{3,2}, J_{3,3}, J_{3,4}, J_{3,5}]$, $p_{3,1} = 9, p_{3,2} = 4, p_{3,3} = 2, p_{3,4} = 3, p_{3,5} = 7$, $\omega_{3,1} = 4, \omega_{3,2} = 2$, $\omega_{3,3} = 5, \omega_{3,4} = 3, \omega_{3,5} = 7, \omega_{3,0} = 5, s_3 = 3$.

The solution is as follows:

Step 1: we calculate $k_1 = 2, W_{1,1} = 2, W_{1,2} = 4, W_{1,3} = 3$; $k_2 = 2, W_{2,1} = 1, W_{2,2} = 4, W_{2,3} = 5, W_{2,4} = 3$; and $k_3 = 3$, $W_{3,1} = 5, W_{3,2} = 9, W_{3,3} = 11, W_{3,4} = 10, W_{3,5} = 7$

Step 2: sequence of jobs within each group: G_1: $[J_{1,3} \longrightarrow J_{1,1} \longrightarrow J_{1,2}]$, G_2: $[J_{2,2} \longrightarrow J_{2,3} \longrightarrow J_{2,4} \longrightarrow J_{2,1}]$, and G_3: $[J_{3,1} \longrightarrow J_{3,2} \longrightarrow J_{3,3} \longrightarrow J_{3,4} \longrightarrow J_{3,5}]$

Step 1: for the problem $1|\mathrm{GT}, \mathrm{CON}|\sum_{i=1}^{f}(\sum_{j=1}^{n_i}\omega_{ij}|L_{\pi(i,j)}| + \omega_{i,0}d_i^{\mathrm{opt}})$, we calculate k_i of each group according to Lemma 3 and calculate $W_{i,j}$ according to equation (19). For the problem $1|\mathrm{GT}, \mathrm{SLK}|\sum_{i=1}^{f}(\sum_{j=1}^{n_i}\omega_{i,j}|L_{\pi(i,j)}| + \omega_{i,0}q_i^{\mathrm{opt}})$, we calculate l_i of each group according to Lemma 4 and calculate $V_{i,j}$ according to equation (21). For the problem $1|\mathrm{GT}, \mathrm{DIF}|\sum_{i=1}^{f}\sum_{j=1}^{n_i}(\omega_{i,j}|L_{\pi(i,j))}| + \omega_{i,0}d_{\pi(i,j)})$, we calculate $U_{i,j}$ according to equation (23).
Step 2: we assign the smallest $W_{i,j}$ ($V_{i,j}$ and $U_{i,j}$) value to the job with the largest $p_{i,j}$ value, the second smallest $W_{i,j}$ ($V_{i,j}$ and $U_{i,j}$) value to the job with the second largest $p_{i,j}$ value, and so on, with ties broken arbitrarily, and then obtain the internal job sequence of each group.
Step 3: we arrange the groups in nondecreasing order of θ_i by Lemma 7.
Step 4: according to Lemma 3, we calculate d_i^{opt} for the problem $1|\mathrm{GT}, \mathrm{CON}|\sum_{i=1}^{f}(\sum_{j=1}^{n_i}\omega_{ij}|L_{\pi(i,j)}| + \omega_{i,0}d_i^{\mathrm{opt}})$. According to Lemma 4, we calculate q_i^{opt} for the problem $1|\mathrm{GT}, \mathrm{SLK}|\sum_{i=1}^{f}(\sum_{j=1}^{n_i}\omega_{i,j}|L_{\pi(i,j)}| + \omega_{i,0}q_i^{\mathrm{opt}})$. According to Lemma 6, we calculate $d_{i,j}$ for the problem $1|\mathrm{GT}, \mathrm{DIF}|\sum_{i=1}^{f}\sum_{j=1}^{n_i}(\omega_{i,j}|L_{\pi(i,j)}| + \omega_{i,0}d_{\pi(i,j)})$.

Algorithm 1: Optimal algorithm.

Step 3: $\theta_1 = 7$, $\theta_2 = 24$, $\theta_3 = 28/5$, and $\theta_3 < \theta_1 < \theta_2$; hence, the optimal group sequence is $[G_3, G_1, G_2]$
Step 4: $d_3^{\mathrm{opt}} = s_3 + \sum_{j=1}^{k_3} p_{3j} = 18$, $d_1^{\mathrm{opt}} = S_1 + s_1 + \sum_{j=1}^{k_1} p_{1j} = 28 + 2 + 8 = 38$, and $d_2^{\mathrm{opt}} = S_2 + s_2 + \sum_{j=1}^{k_2} p_{2j} = 42 + 4 + 12 = 58$

4. Conclusions

In this paper, we studied the single-machine scheduling problem involving the due-date assignment and job scheduling under the group technology. The due dates are assignable according to one of the following three due-date assignment methods: CON, SLK, and DIF due-date assignment. The objective is to find the optimal due dates of jobs, a sequence for groups, and jobs to minimize a total cost function. We show that the problem can be solved in polynomial time. In future study, we can consider the group scheduling models associated with learning and deteriorating effects. In addition, we can further study the group scheduling with CON and SLK due-date assignment in the flow shop setting.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (grant nos. 71871091 and 71471057) and the Natural Science Foundation of Liaoning Province in China (grant no. 2020-MS-233).

References

[1] J. S. Neufeld, J. N. D. Gupta, and U. Buscher, "A comprehensive review of flowshop group scheduling literature," *Computers & Operations Research*, vol. 70, pp. 56–74, 2016.
[2] L. Yang, Y. Zhao, and X. Ma, "Group maintenance scheduling for two-component systems with failure interaction," *Applied Mathematical Modelling*, vol. 71, pp. 118–137, 2019.
[3] S.-J. Yang and D.-L. Yang, "Single-machine scheduling simultaneous with position-based and sum-of-processing-times-based learning considerations under group technology assumption," *Applied Mathematical Modelling*, vol. 35, no. 5, pp. 2068–2074, 2011.
[4] Y.-Y. Lu, J.-J. Wang, and J.-B. Wang, "Single machine group scheduling with decreasing time-dependent processing times subject to release dates," *Applied Mathematics and Computation*, vol. 234, pp. 286–292, 2014.
[5] N. Yin, L. Kang, and X.-Y. Wang, "Single-machine group scheduling with processing times dependent on position, starting time and allotted resource," *Applied Mathematical Modelling*, vol. 38, no. 19-20, pp. 4602–4613, 2014.
[6] J.-J. Wang and Y.-J. Liu, "Single-machine bicriterion group scheduling with deteriorating setup times and job processing times," *Applied Mathematics and Computation*, vol. 242, pp. 309–314, 2014.
[7] J.-B. Wang and J.-J. Wang, "Single machine group scheduling with time dependent processing times and ready times," *Information Sciences*, vol. 275, pp. 226–231, 2014.
[8] H. Qin, Z.-H. Zhang, and D. Bai, "Permutation flowshop group scheduling with position-based learning effect," *Computers & Industrial Engineering*, vol. 92, pp. 1–15, 2016.
[9] Y.-Y. Lu, J.-B. Wang, P. Ji, and H. He, "A note on resource allocation scheduling with group technology and learning effects on a single machine," *Engineering Optimization*, vol. 49, no. 9, pp. 1621–1632, 2017.
[10] X. Zhang, L. Liao, W. Zhang, T. C. E. Cheng, Y. Tan, and M. Ji, "Single-machine group scheduling with new models of position-dependent processing times," *Computers & Industrial Engineering*, vol. 117, pp. 1–5, 2018.
[11] F. Liu, J. Yang, and Y.-Y. Lu, "Solution algorithms for single-machine group scheduling with ready times and deteriorating jobs," *Engineering Optimization*, vol. 51, no. 5, pp. 862–874, 2019.
[12] J.-B. Wang and X.-X. Liang, "Group scheduling with deteriorating jobs and allotted resource under limited resource availability constraint," *Engineering Optimization*, vol. 51, no. 2, pp. 231–246, 2019.
[13] Y. Yin, T. C. E. Cheng, C.-C. Wu, and S.-R. Cheng, "Single-machine common due-date scheduling with batch delivery costs and resource-dependent processing times," *International Journal of Production Research*, vol. 51, no. 17, pp. 5083–5099, 2013.

[14] Y. Yin, M. Liu, T. C. E. Cheng, C.-C. Wu, and S.-R. Cheng, "Four single-machine scheduling problems involving due date determination decisions," *Information Sciences*, vol. 251, pp. 164–181, 2013.
[15] D.-J. Wang, Y. Yin, J. Xu, W.-H. Wu, S.-R. Cheng, and C.-C. Wu, "Some due date determination scheduling problems with two agents on a single machine," *International Journal of Production Economics*, vol. 168, pp. 81–90, 2015.
[16] D. Shabtay, "Optimal restricted due date assignment in scheduling," *European Journal of Operational Research*, vol. 252, no. 1, pp. 79–89, 2016.
[17] S. Li, C. T. Ng, and J. Yuan, "Group scheduling and due date assignment on a single machine," *International Journal of Production Economics*, vol. 130, no. 2, pp. 230–235, 2011.
[18] W. Liu, X. Hu, and X. Wang, "Single machine scheduling with slack due dates assignment," *Engineering Optimization*, vol. 49, no. 4, pp. 709–717, 2017.
[19] J.-B. Wang, X.-N. Geng, L. Liu, J.-J. Wang, and Y.-Y. Lu, "Single machine CON/SLK due date assignment scheduling with controllable processing time and job-dependent learning effects," *The Computer Journal*, vol. 61, no. 9, pp. 1329–1337, 2018.
[20] M. Ji, K. Chen, J. Ge, and T. C. E. Cheng, "Group scheduling and job-dependent due window assignment based on a common flow allowance," *Computers & Industrial Engineering*, vol. 68, pp. 35–41, 2014.
[21] P. Brucker, *Scheduling Algorithms*, Springer, Berlin, Germany, 3rd edition, 2001.
[22] L. Sun, A. J. Yu, and B. Wu, "Single machine common flow allowance group scheduling with learning effect and resource allocation," *Computers and Industrial Engineering*, vol. 139, Article ID 106126, 2020.
[23] R. L. Graham, E. L. Lawler, J. K. Lenstra, and A. H. G. R. Kan, "Optimization and approximation in deterministic sequencing and scheduling: a survey," *Annals of Discrete Mathematics*, vol. 5, pp. 287–326, 1979.
[24] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge University Press, Cambridge, UK, 2nd edition, 1967.

Hindawi
Complexity
Volume 2021, Article ID 6615824, 13 pages
https://doi.org/10.1155/2021/6615824

Research Article

Solution Algorithms for Single-Machine Group Scheduling with Learning Effect and Convex Resource Allocation

Wanlei Wang,[1] **Jian-Jun Wang**,[2] **and Ji-Bo Wang**[3]

[1]*College of Mechanical and Electronic Engineering, Dalian Minzu University, Dalian 116600, China*
[2]*School of Economics and Management, Dalian University of Technology, Dalian 116024, China*
[3]*School of Science, Shenyang Aerospace University, Shenyang 110136, China*

Correspondence should be addressed to Wanlei Wang; wwl@dlnu.edu.cn

Received 22 December 2020; Revised 19 January 2021; Accepted 10 February 2021; Published 3 March 2021

Academic Editor: Lei Xie

Copyright © 2021 Wanlei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper deals with a single-machine resource allocation scheduling problem with learning effect and group technology. Under slack due-date assignment, our objective is to determine the optimal sequence of jobs and groups, optimal due-date assignment, and optimal resource allocation such that the weighted sum of earliness and tardiness penalties, common flow allowances, and resource consumption cost is minimized. For three special cases, it is proved that the problem can be solved in polynomial time. To solve the general case of problem, the heuristic, tabu search, and branch-and-bound algorithms are proposed.

1. Introduction

In the conventional scheduling models and problems, it is generally assumed that the job processing times are constants, but in practice, examples can be found to illustrate that the job processing times are not necessarily constants (Shabtay and Steiner [1], Biskup [2], and Azzouz et al. [3]). More recently, Zhu et al. [4] considered resource allocation single-machine scheduling problems with learning effects and group technology. For the linear and convex resource allocation models, they proved that problem of minimizing the weighted sum of makespan and total resource cost can be solved in polynomial time. Lu et al. [5] revisited the same model with Zhu et al. [4], but they considered the case of resource availability is limited. For the makespan minimization subject to limited resource availability, they proposed heuristic and branch-and-bound algorithms. Sun et al. [6] and Lv et al. [7] worked on single-machine scheduling group problems with learning effects and resource allocation. Under the slack (SLK) due-date assignment, for the linear weighted sum of scheduling cost and resource consumption cost minimization, Sun et al. [6] proved that the problem can be solved in polynomial time. However, Lv et al. [7] showed that the results of Sun et al. [6] are incorrect by two counter-examples, and Lv et al. [7] also provided the corrected results under a special case. In this paper, we will consider the same model with Sun et al. [6] and Lv et al. [7], i.e., three popular features in the recent years: group technology, resource allocation, and learning effect. The contributions of this study are given as follows: (1) we study the SLK assignment single-machine group scheduling problem along with learning effect and convex resource allocation; (2) the optimal properties are provided for the total cost (including earliness, tardiness, common flow allowances, and resource consumption cost) minimization; (3) we propose the heuristic, tabu search, and branch-and-bound algorithms to solve the problem.

The reminder of this paper is organized as follows. In Section 2, the relevant literature review is presented. The problem statement is presented in Section 3. Section 4 gives some properties of the problem. In Section 5, some special cases are discussed. In Section 6, for the general case, solution algorithms are proposed. Finally, the conclusions are given in Section 7.

2. Literature Review

In this section, we restrict our literature review to papers that study scheduling problems with learning effects, resource allocation, and/or group technology.

In manufacturing environments, after learning, the time required for workers (machines) to process some jobs is decreasing, which causes scheduling problems with learning effects (Biskup [2]). Wang et al. [8] considered the flow shop problem with a learning effect. Under two-machine and release dates, the goal is to minimize the weighted sum of makespan and total completion time. They proposed a branch-and-bound algorithm and a multiobjective memetic algorithm to solve the problem. Wang et al. [9] considered flow shop problems with truncated learning effects. For the makespan and total weighted completion time minimizations, they proposed heuristics and branch-and-bound algorithms. Yan et al. [10] and Wang et al. [11] studied single-machine scheduling problems with learning effects and release times. Sun et al. [12] investigated flow shop problem with general position weighted learning effects. For the total weighted completion time minimization, they proposed some heuristics and a branch-and-bound algorithm to solve the problem.

In addition, scheduling problems with resource allocation (controllable processing times) have also attracted considerable interest from researchers (Shabtay and Steiner [1]), that is, the scheduler can control processing times of jobs by allocating a common continuously nonrenewable resource. Kayvanfar et al. [13, 14] considered single-machine scheduling with controllable processing times. For the total tardiness and earliness minimization, Kayvanfar et al. [13] proposed a mathematical model and three heuristic techniques; Kayvanfar et al. [14] proposed a drastic hybrid heuristic algorithm. Lu and Liu [15] considered single machine scheduling problems with resource allocation and position-dependent workloads. They proposed a bicriteria analysis for the scheduling cost and total resource consumption cost. Tsao et al. [16] considered energy-efficient single-machine scheduling problem with controllable processing time. Under differential electricity pricing, they proposed a mixed-integer programming model and a fuzzy control approach. Mor et al. [17] considered single-machine scheduling problems with resource-dependent processing times. For a large set of that the scheduling criterion can be represented as one that includes positional penalties; they proposed heuristic algorithms to solve the problems. Kayvanfar et al. [18] considered unrelated parallel machines scheduling problem with controllable processing times. For the linear weighted sum of tardiness, earliness, jobs compressing and expanding costs, and makespan minimization, they proposed a mixed integer programming model and some heuristics. Zarandi and Kayvanfar [19] considered a biobjective identical parallel machines scheduling problem with controllable processing times. The goal is to simultaneously minimize total cost of tardiness and earliness as well as compression and expansion costs of job processing times and makespan. They proposed two multiobjective evolutionary algorithms to solve the biobjective problem. Kayvanfar et al. [20] studied identical parallel machine scheduling problem with controllable processing times. For the linear weighted sum of tardiness, earliness, and job compressions/expansion cost minimization, they proposed a mixed integer linear programming model and some heuristic algorithms to solve the problem.

A third possible aspect is that scheduling problems with group technology (GT, see studies by Mosier et al. [21] and Webster and Baker [22]), that is, GT is an approach to manufacturing that seeks to improve efficiency in high-volume production by exploiting the similarities of different products and activities in their production (see studies by Yang and Yang [23] and Ji et al. [24]). Xu et al. [25] and Liu et al. [26] considered single-machine group scheduling with deteriorating jobs and ready times. For the makespan minimization, Xu et al. [25] proved that some special cases can be solved in polynomial time; for the general case, Liu et al. [26] proposed a branch-and-bound algorithm. Li and Zhao [27] considered single machine scheduling problem with group technology. Under multiple due windows assignment, they proved that the total cost (including earliness, tardiness, and due windows) minimization can be solved in polynomial time. Zhang and Xie [28] studied the single machine scheduling problem with position dependent processing times. Under group technology and ready times, for the makespan minimization, they proved that a special case can be solved in polynomial time. Ji and Li [29] studied single-machine group scheduling with variable job processing times (including resource allocation, learning effects, and deteriorating jobs). They proved that two versions of problem can be solved in polynomial time. Zhang et al. [30] considered single-machine group scheduling problems with position-dependent processing times. They proved that the makespan and the total completion time minimization can be solved in polynomial-time algorithm, respectively. Muştu, and Eren [31]) studied the single-machine scheduling problem with sequence-independent setup times and time-dependent learning and forgetting effects. They proved that the makespan minimization is ordinary NP-hard, and they also proposed an integer nonlinear programming and a dynamic programming to solve the problem. Extensive surveys of different scheduling models and problems with group technology can be found in studies by Allahverdi [32] and Neufeld et al. [33].

More recently, Wang et al. [34] and Lu et al. [35] delved into single-machine resource allocation scheduling problems with learning effects. He et al. [36] considered the single-machine resource allocation scheduling with truncated job-dependent learning effect. Under linear and convex resource allocations, polynomial time algorithms are developed to solve the problem. Li et al. [37] considered single-machine scheduling with general job-dependent learning curves and controllable processing times. They proved that some regular and nonregular objective minimizations can be solved in polynomial time. Wang et al. [38] considered single-machine scheduling with truncated learning effects and resource allocation. For total weighted flow time cost and total resource consumption cost, they provided a bicriteria analysis. Geng et al. [39] and Sun et al. [40] investigated two-machine no-wait flow shop scheduling with resource allocation and learning effect. For the common due date assignment, Geng et al. [39] proved that two versions of the scheduling criteria and resource consumption cost can be solved in polynomial time. For the slack due-date assignment, Sun et al. [40] proved that three versions of the scheduling criteria and resource consumption cost can be solved in polynomial time. Liu and Jiang [41] explored due-date assignment scheduling problems with job-

dependent learning effects and convex resource allocation. Zhang et al. [30] considered single-machine group scheduling problems with position-dependent learning effects. They proved that the makespan and total completion time minimizations can be solved in polynomial time algorithm. Wang and Liang [42] and Liang et al. [43] investigated single-machine resource allocation scheduling with deteriorating jobs and group technology. Liao et al. [44] studied a two-competing group parallel machines scheduling problem with truncated job-dependent learning effects. Under serial-batching machines, for the makespan minimization, they proposed a greedy algorithm.

To the best of our knowledge, apart from the recent papers of Sun et al. [6] and Lv et al. [7], the single-machine slack due-date assignment scheduling problem with resource allocation, group technology, and learning effects has not been investigated. In this paper, we consider the same model as in Sun et al. [6] and Lv et al. [7], but with the tabu search and branch-and-bound algorithms to solve the problem.

3. Problem Formulation

We have n jobs grouped into f groups (i.e., $(G_1, G_2, \ldots, G_f)$) to be processed by a single machine, where the number of jobs in the group G_i $(i = 1, 2, \ldots, f)$ is m_i, i.e., $\sum_{i=1}^{f} m_i = n$. Each group G_i $(i = 1, 2, \ldots, f)$ has an independent setup time s_i and contains m_i jobs which are processed consecutively. Let $J_{i,j}$ be the j th job in group G_i, $i = 1, 2, \ldots, f$, $j = 1, 2, \ldots, m_i$, i.e., group G_i has jobs $J_{i,1}, J_{i,2}, \ldots, J_{i,m_i}$. Let $J_{[i],[j]}$ be the job in the i th group position and j th internal job position. As in the study by Liang et al. [43], if the job $J_{i,j}$ is scheduled in r th position in group G_i, the actual processing time of job $J_{i,j}$ is

$$p_{i,j}^{A} = \left(\frac{p_{i,j} r^{a_{i,j}}}{u_{i,j}}\right)^{\eta}, \tag{1}$$

where η is a constant positive parameter, $p_{i,j}$ is the normal processing time of job $J_{i,j}$, $a_{i,j} \le 0$ is the learning rate (see the study by Biskup [2]) of job $J_{i,j}$, $u_{i,j} \ge u_i' > 0$, u_i' is the minimal resource allocation to the jobs of group G_i (if $u_{i,j}$ is close to zero, $p_{i,j}^{A}$ will be close to infinity, which is not realistic; hence, we set $u_{i,j} \ge u_i' > 0$). Let $C_{i,j}$ $(d_{i,j})$ be the completion time (due-date) of job $J_{i,j}$. For the slack (SLK) due-date assignment, we assume that the due-date of job $J_{i,j}$ is $d_{i,j} = p_{i,j}^{A} + q_i$, where q_i is the common flow allowance for group G_i and q_i is a decision variable. Let $E_{i,j} = \max\{d_{i,j} - C_{i,j}, 0\}$ $(T_{i,j} = \max\{C_{i,j} - d_{i,j}, 0\})$ be the earliness (tardiness) of job $J_{i,j}$; our goal is to find the optimal group sequence π_G^*, job sequence π_i^* $(i = 1, 2, \ldots, f)$ within group G_i, and resource allocation such that the following cost is minimized:

$$\sum_{i=1}^{f}\sum_{j=1}^{m_i}\left(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i\right) + \sum_{i=1}^{f}\sum_{j=1}^{m_i} \nu_{i,j} u_{i,j}, \tag{2}$$

where α_i, β_i, and γ_i are the nonnegative parameters of group G_i and $\nu_{i,j}$ represents the per unit cost of the resource $u_{i,j}$ allocated to job $J_{i,j}$. Using the three-field classification (see the studies by Shabtay and Steiner [1], Biskup [2], and Azzouz et al. [3]), the problem can be denoted as

$$1\left|\mathrm{GT}, s_i, \mathrm{CRA}, \mathrm{SLK}\right| \sum_{i=1}^{f}\sum_{j=1}^{m_i}\left(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i\right) + \sum_{i=1}^{f}\sum_{j=1}^{m_i} \nu_{i,j} u_{i,j}, \tag{3}$$

where GT denotes the group technology and CRA represents the convex function of resource allocation (1).

4. Some Properties

Obviously, an optimal sequence exists that starts at time zero and without any machine idle time between all the jobs. Similar to the study by Adamopoulos and Pappis [45], we have the following.

Lemma 1. *If* $C_{[i],[j]} \le d_{[i],[j]} \Longrightarrow C_{[i],[j-1]} \le d_{[i],[j-1]}$, $(i = 1, 2, \ldots, f; j = 1, 2, \ldots, m_i)$.

If $C_{[i],[j]} \ge d_{[i],[j]} \Longrightarrow C_{[i],[j+1]} \ge d_{[i],[j+1]}$, $(i = 1, 2, \ldots, f; j = 1, 2, \ldots, m_i)$.

Lemma 2. *For the problem* $1|GT, s_i, CRA, SLK| \sum_{i=1}^{f}\sum_{j=1}^{m_i} (\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i} \nu_{i,j} u_{i,j}$, *there exists an optimal schedule such that the optimal value of* $q_{[i]}$ *coincides with the job completion time of the group* $G_{[i]}$, *i.e.,* $q_{[i]} = C_{[i],[k_i-1]} = S_{[i]} + s_{[i]} + \sum_{j=1}^{k_i-1} p_{[i],[j]}^{A}$, *where*

$$k_{[i]} = \min\left\{m_{[i]}, \max\left\{0, \left\lceil \frac{m_{[i]}\left(\beta_{[i]} - \gamma_{[i]}\right)}{\alpha_{[i]} + \beta_{[i]}} \right\rceil\right\}\right\}, \tag{4}$$

and $S_{[i]}$ *is the starting time of group* $G_{[i]}$.

Lemma 3 (see [43]). *For a given sequence* π, *the optimal resource allocation is*

$$u_{[i],[j]}^{*} = \max\left[u_{[i],[j]}, u_{[i]}'\right], \tag{5}$$

where

$$u_{[i],[j]} = \begin{cases} \left(\dfrac{\eta\left(\alpha_{[i]} j + \sum_{h=i}^{f} m_{[h]}\gamma_{[h]}\right)}{\nu_{[i],[j]}}\right)^{(1/(\eta+1))} \left(p_{[i],[j]} j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))}, & i = 1, \ldots, f; j = 1, \ldots, k_{[i]-1}, \\ \left(\dfrac{\eta\beta_{[i]}\left(m_{[i]} - j\right)}{\nu_{[i],[j]}}\right)^{(1/(\eta+1))} \left(p_{[i],[j]} j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))}, & i = 1, \ldots, f; j = k_{[i]}, \ldots, m_{[i]}. \end{cases} \tag{6}$$

From the study by Liang et al. [43], we have

$$\begin{aligned}
&\sum_{i=1}^{f}\sum_{j=1}^{m_i}\left(\alpha_i E_{i,j}+\beta_i T_{i,j}+\gamma_i q_i\right)+\sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}\\
&\quad=\sum_{i=1}^{f}\left(\sum_{h=i}^{f}m_{[h]}\gamma_{[h]}\right)s_{[i]}\\
&\quad+\left(\eta^{(1/(\eta+1))}+\eta^{(-\eta/(\eta+1))}\right)\sum_{i=1}^{f}\sum_{j=1}^{k_{[i]}-1}\left(\alpha_{[i]}j+\sum_{h=i}^{f}m_{[h]}\gamma_{[h]}\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))}\\
&\quad+\left(\eta^{(1/(\eta+1))}+\eta^{(-\eta/(\eta+1))}\right)\sum_{i=1}^{f}\sum_{j=k_{[i]}}^{m_{[i]}}\left(\beta_{[i]}\left(m_{[i]}-j\right)\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))}.
\end{aligned}\tag{7}$$

Lemma 4 (see [43]). *For the problem* $1|GT, s_i, CRA, SLK|\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,}\ \ j+\beta_i T_{i,j}+\gamma_i q_i)+\ \sum_{i=1}^{f}\ \sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$, *if the sequence of groups is given by* $\pi_G=[G_{[1]},G_{[2]},\ldots,G_{[f]}]$, *the optimal job sequence in the group* $G_{[i]}$ $(i=1,2,\ldots,f)$ *can be obtained by the following assignment problem (**AP-i**):*

$$(\mathbf{AP}-\mathbf{i})\mathrm{Min}\sum_{j=1}^{m_{[i]}}\sum_{h=1}^{m_{[i]}}\vartheta_{[i],j,h}x_{[i],j,h},\tag{8}$$

$$s.t.\ \sum_{h=1}^{m_{[i]}}x_{[i],j,h}=1,\quad i=1,2,\ldots,f;\ j=1,2,\ldots,m_{[i]},\tag{9}$$

$$\sum_{j=1}^{m_{[i]}}x_{[i],j,h}=1,\quad i=1,2,\ldots,f;\ h=1,2,\ldots,m_{[i]},\tag{10}$$

$$x_{[i],j,h}=0\text{ or }1,\tag{11}$$

where

$$\vartheta_{[i],j,h}=\begin{cases}\left(\eta^{(1/(\eta+1))}+\eta^{(-\eta/(\eta+1))}\right)\left(\alpha_{[i]}h+\sum_{l=i}^{f}m_{[l]}\gamma_{[l]}\right)^{(1/(\eta+1))}\left(\nu_{[i],j}p_{[i],j}h^{a_{[i],j}}\right)^{(\eta/(\eta+1))}, & i=1,2,\ldots,f;\ j=1,2,\ldots,m_{[i]};\ h=1,2,\ldots,k_{[i]}-1\\ \left(\eta^{(1/(\eta+1))}+\eta^{(-\eta/(\eta+1))}\right)\left(\beta_{[i]}\left(m_{[i]}-h\right)\right)^{(1/(\eta+1))}\left(\nu_{[i],j}p_{[i],j}h^{a_{[i],j}}\right)^{(\eta/(\eta+1))}, & i=1,2,\ldots,f;\ j=1,2,\ldots,m_{[i]};\ h=k_{[i]},k_{[i]}+1,\ldots,m_{[i]}.\end{cases}\tag{12}$$

Lemma 5. *The term* $\sum_{i=1}^{f}(\sum_{h=i}^{f}m_{[h]}\gamma_{[h]})s_{[i]}$ *is minimized if* $s_{[1]}\le s_{[2]}\le\cdots\le s_{[f]}$.

Proof. It is similar to the proof of Liang et al. [43]. □

Lemma 6. *If the optimal job sequence within each group is given, the term,*

$$\begin{aligned}
&\left(\eta^{(1/(\eta+1))}+\eta^{(-\eta/(\eta+1))}\right)\sum_{i=1}^{f}\sum_{j=1}^{k_{[i]}-1}\left(\alpha_{[i]}j+\sum_{h=i}^{f}m_{[h]}\gamma_{[h]}\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))}\\
&+\left(\eta^{(1/(\eta+1))}+\eta^{(-\eta/(\eta+1))}\right)\sum_{i=1}^{f}\sum_{j=k_{[i]}}^{m_{[i]}}\left(\beta_{[i]}\left(m_{[i]}-j\right)\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))},
\end{aligned}\tag{13}$$

is minimized if $m_{[1]}\gamma_{[1]} \geq m_{[2]}\gamma_{[2]} \geq \cdots \geq m_{[f]}\gamma_{[f]}$.

Proof. By using simple group interchanging technique, the result can be easily obtained. □

5. Polynomial Time Solvable Cases

5.1. Case 1. As in the study by Liao et al. [44], if the groups have agreeable conditions, i.e., if $s_i \leq s_h$ implies $m_i\gamma_i \geq m_h\gamma_h$ for all groups G_i and G_h, the problem $1|\mathrm{GT}, s_i, \mathrm{CRA}, \mathrm{SLK}, (s_i \leq s_h) \Longrightarrow (m_i\gamma_i \geq m_h\gamma_h)| \sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ can be solved in polynomial time.

Lemma 7 (see [44]). *For the problem* $1|GT, s_i, CRA, SLK, (s_i \leq s_h) \Longrightarrow (m_i\gamma_i \geq m_h\gamma_h)| \sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$, *if the groups have agreeable conditions, i.e., if* $s_i \leq s_h$ *implies* $m_i\gamma_i \geq m_h\gamma_h$ *for all groups* G_i *and* G_h, *then the optimal group sequence* π_G^* *can be obtained by sequencing groups in nondecreasing order of* s_i, *or equivalently, the optimal group sequence* π_G^* *can be obtained by sequencing groups in nonincreasing order of* $m_i\gamma_i$.

For the problem $1|\mathrm{GT}, s_i, \mathrm{Con}, (s_i \leq s_h) \Longrightarrow (m_i\gamma_i \geq m_h\gamma_h)| \sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$, the optimal solution algorithm is given as follows:

Theorem 1 (see [44]). *The problem,*

$$1\left|\mathrm{GT}, s_i, \mathrm{CRA}, \mathrm{SLK}, \left(s_i \leq s_h\right) \Longrightarrow \left(m_i\gamma_i \geq m_h\gamma_h\right)\right| \sum_{i=1}^{f}\sum_{j=1}^{m_i}\left(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i\right) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}, \tag{14}$$

can be solved by Algorithm 1 in $O(n^3)$ *time.*

5.2. Case 2. In this subsection, a special case will be considered, i.e., if $s_i = s$, $m_i\gamma_i = \overline{m\gamma}$, for $i = 1, 2, \ldots f$.

Lemma 8. *For the problem* $1|GT, s_i, CRA, SLK, s_i = s, m_i\gamma_i = \overline{m\gamma}| \sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$, *the optimal group sequence* π_G^* *can be obtained by an assignment problem.*

Proof. From Lemma 4, the sequence π_i^* $(i = 1, 2, \ldots, f)$ within the group G_i can be obtained. For the group G_i, if $s_i = s$, and $m_i\gamma_i = \overline{m\gamma}$ $(i = 1, 2, \ldots, f)$, from (7), we have

$$\begin{aligned}
&\sum_{i=1}^{f}\sum_{j=1}^{m_i}\left(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i\right) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j} \\
&= \frac{f(f+1)s\overline{m\gamma}}{2} + \left(\eta^{(1/(\eta+1))} + \eta^{(-\eta/(\eta+1))}\right)\sum_{i=1}^{f}\sum_{j=1}^{k_{[i]}-1}\left(\alpha_{[i]}j + (f-i+1)\overline{m\gamma}\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))} \\
&+ \left(\eta^{(1/(\eta+1))} + \eta^{(-\eta/(\eta+1))}\right)\sum_{i=1}^{f}\sum_{j=k_{[i]}}^{m_{[i]}}\left(\beta_{[i]}\left(m_{[i]} - j\right)\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))}.
\end{aligned} \tag{15}$$

Let $x_{i,r} = 1$, if group G_i is assigned to position r, and $x_{i,r} = 0$ otherwise. For $((f(f+1)s\overline{m\gamma})/2)$ is a constant, hence the optimal group sequence π_G^* can be obtained by an assignment problem (**AP-G**):

$$\begin{aligned}
&(\mathrm{AP-G})\ \mathrm{Min}\sum_{i=1}^{f}\sum_{r=1}^{f}\vartheta_{i,r}x_{i,r} \\
&s.t.\ \sum_{r=1}^{f}x_{i,r} = 1, \quad i = 1, 2, \ldots, f, \\
&\sum_{i=1}^{f}x_{i,r} = 1, \quad r = 1, 2, \ldots, f, \\
&x_{i,r} = 0 \text{ or } 1,
\end{aligned} \tag{16}$$

where the optimal sequence of group G_i is $\pi_i^* = (J_{i,(1)}, J_{i,(2)}, \ldots, J_{i,(m_i)})$ (by Lemma 4), and

$$\begin{aligned}
\vartheta_{i,r} = &\left(\eta^{(1/(\eta+1))} + \eta^{(-\eta/(\eta+1))}\right)\sum_{j=1}^{k_i-1}\left(\alpha_i j + (f-r+1)\overline{m\gamma}\right)^{(1/(\eta+1))} \\
&\cdot\left(\nu_{i,(j)}p_{i,(j)}j^{a_{i,(j)}}\right)^{(\eta/(\eta+1))} \\
&+ \left(\eta^{(1/(\eta+1))} + \eta^{(-\eta/(\eta+1))}\right)\sum_{j=k_i}^{m_i}\left(\beta_i\left(m_i - j\right)\right)^{(1/(\eta+1))} \\
&\cdot\left(\nu_{i,(j)}p_{i,(j)}j^{a_{i,(j)}}\right)^{(\eta/(\eta+1))}.
\end{aligned} \tag{17}$$

For the problem $1|\mathrm{GT}, s_i, \mathrm{CRA}, \mathrm{SLK}, s_i = s, m_i\gamma_i = \overline{m\gamma}| \sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$, the optimal solution algorithm is given as follows: □

Step 1. Calculate $k_{[i]}$ by using (4), $i = 1, 2, \ldots, f$.
Step 2. The optimal sequence between groups is arranged in nondecreasing order of $s_{[i]}$.
Step 3. The jobs in each group are arranged according to the assignment problem **AP** (Lemma 4).
Step 4. Calculate the optimal resource allocation $u^*_{[i][j]}$ according to (5).
Step 5. Calculate the optimal common flow allowance $q_{[i]} = C_{[i],[k_{[i]}]-1}$ and the corresponding optimal objective function $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ by using (7).

Algorithm 1: Optimal solution for Case 1.

Theorem 2. *The problem,*

$$1\left|\mathrm{GT}, s_i, \mathrm{CRA}, \mathrm{SLK}, s_i = s, m_i\gamma_i = \overline{m\gamma}\right| \sum_{i=1}^{f}\sum_{j=1}^{m_i} \cdot \left(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i\right) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}, \tag{18}$$

can be solved by Algorithm 2 in $O(n^3)$ time.

Proof. Time complexity of Step 1 is $O(n)$; time complexity of Step 2 is $O(f \log f)$ time. Step 3 needs $\sum_{i=1}^{f} O(m_i^3) \le O(n^3)$ time. Steps 4-5 need $O(n)$ time, respectively. Thus, the total time complexity of Algorithm 2 is $O(n^3)$ time. □

5.3. Case 3. In this subsection, it is assumed that the number of groups f is a given constant.

Theorem 3. *For the $1|GT, s_i, CRA, SLK|\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ problem, an optimal schedule can be solved in $O(f!n^3)$ time, i.e., the problem $1|GT, s_i, CRA, SLK|\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ can be solved in polynomial time if f is a given constant.*

Proof. From Lemma 4, if the schedule of groups is given, then an optimal schedule can be obtained in $\sum_{i=1}^{f} O(n_i^3) \le O(n^3)$ time. Obviously, there are $f!$ possible group schedules, hence an optimal schedule can be solved in $O(f!n^3)$ time.

Based on Theorem 2, an algorithm can be proposed to solve the problem $1|\mathrm{GT}, s_i, \mathrm{CRA}, \mathrm{SLK}|\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$. □

6. General Case

For the general case of the problem $1|\mathrm{GT}, s_i, \mathrm{CRA}, \mathrm{SLK}|\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,\quad j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$, the complexity is an open question. Hence, the heuristic algorithm and branch-and-bound (B&B) algorithm might be a good way to solve the problem.

6.1. Heuristic Algorithm. From Lemma 4, the optimal job sequence within the same group can be obtained, and the optimal resource allocation of a given schedule can be obtained by Lemma 3. In this subsection, we apply the well-known heuristic procedure from the study by Nawaz et al. [46], and the following heuristic algorithm can be proposed.

6.2. Tabu Search Algorithm. Tabu search (TS) algorithm is a metaheuristic algorithm first proposed by Glover [47]. In this subsection, tabu search (TS) is used to find a near-optimal solution (Xu et al. [48]). The initial sequence used in the TS algorithm is chosen by the nondecreasing order of s_i, and the maximum number of iterations for the TS algorithm is set at $100f$, where f is the number of groups. As in the study by Wu et al. [49], the implementation of the TS algorithm is given as follows:

6.3. A Lower Bound. Let $\pi_G = [\pi_{GS}, \pi_{GU}]$ be a sequence of groups in which π_{GS} is the scheduled part, and π_{GU} is a unscheduled part. Assume that there are g groups in π_{GS}; from (7), it is noticed that the terms,

$$\begin{aligned}
&\sum_{i=g+1}^{f}\left(\sum_{h=i}^{f} m_{[h]}\gamma_{[h]}\right)s_{[i]}, \\
&\cdot\left(\eta^{(1/(\eta+1))} + \eta^{(-\eta/(\eta+1))}\right)\sum_{i=g+1}^{f}\sum_{j=1}^{k_{[i]}-1}\left(\alpha_{[i]}j + \sum_{h=i}^{f} m_{[h]}\gamma_{[h]}\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))} \\
&+\left(\eta^{(1/(\eta+1))} + \eta^{(-\eta/(\eta+1))}\right)\sum_{i=g+1}^{f}\sum_{j=k_{[i]}}^{m_{[i]}}\left(\beta_{[i]}\left(m_{[i]} - j\right)\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))},
\end{aligned} \tag{19}$$

Step 1. Calculate $k_{[i]}$ by using (4), $i = 1, 2, \ldots, f$.
Step 2. The optimal sequence between groups can be obtained by Lemma 8.
Step 3. The jobs in each group are arranged according to the assignment problem **AP** (Lemma 4).
Step 4.Calculate the optimal resource allocation $u^*_{[i],[j]}$ according to (5).
Step 5. Calculate the optimal common flow allowance $q_{[i]} = C_{[i][k_{[i]}]-1}$ and the corresponding optimal objective function $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ by using (7).

Algorithm 2: Optimal solution for Case 2.

can be minimized by Lemmas 5–6. Hence, the lower bound can be obtained by the following formula:

$$\begin{aligned} LB = {} & \sum_{i=1}^{g}\left(\sum_{h=i}^{g} m_{[h]}\gamma_{[h]} + \sum_{h=g+1}^{f} m_{\langle h\rangle}\gamma_{\langle h\rangle}\right)s_{[i]} + \sum_{i=g+1}^{f}\left(\sum_{h=i}^{f} m_{\langle h\rangle}\gamma_{\langle h\rangle}\right)s_{(i)} \\ & + \left(\eta^{(1/(\eta+1))} + \eta^{(-\eta/(\eta+1))}\right)\left(\begin{array}{l} \sum_{i=1}^{g}\sum_{j=1}^{k_{[i]}-1}\left(\alpha_{[i]}j + \sum_{h=i}^{g} m_{[h]}\gamma_{[h]} + \sum_{h=g+1}^{f} m_{\langle h\rangle}\gamma_{\langle h\rangle}\right)^{(1/(\eta+1))} \\ \qquad \times\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))} \\ + \sum_{i=g+1}^{f}\sum_{j=1}^{k_{[i]}-1}\left(\alpha_{[i]}j + \sum_{h=i}^{g} m_{[h]}\gamma_{[h]} + \sum_{h=g+1}^{f} m_{\langle h\rangle}\gamma_{\langle h\rangle}\right)^{(1/(\eta+1))} \\ \qquad \times\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))} \end{array}\right) \\ & + \left(\eta^{(1/(\eta+1))} + \eta^{(-\eta/(\eta+1))}\right)\left(\begin{array}{l} \sum_{i=1}^{g}\sum_{j=k_{[i]}}^{m_{[i]}}\left(\beta_{[i]}\left(m_{[i]} - j\right)\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))} \\ + \sum_{i=g+1}^{f}\sum_{j=k_{[i]}}^{m_{[i]}}\left(\beta_{[i]}\left(m_{[i]} - j\right)\right)^{(1/(\eta+1))}\left(\nu_{[i],[j]}p_{[i],[j]}j^{a_{[i],[j]}}\right)^{(\eta/(\eta+1))} \end{array}\right), \end{aligned} \tag{20}$$

where $s_{(g+1)} \le s_{(g+2)} \le \cdots \le s_{(f)}$ and $m_{\langle g+1\rangle}\gamma_{\langle g+1\rangle} \ge m_{\langle g+2\rangle}\gamma_{\langle g+2\rangle} \ge \cdots \ge m_{\langle f\rangle}\gamma_{\langle f\rangle}$ (note that $s_{(i)}$ and $m_{\langle i\rangle}\gamma_{\langle i\rangle}$ $((i = g+1, r+2, \ldots, f))$ do not necessarily correspond to the same group).

6.4. Branch-and-Bound (B&B) Algorithm. The branch-and-bound (B&B) algorithm search follows a depth-first strategy; this algorithm assigns groups in a forward manner starting from the first position (assign a group to a node).

6.5. An Example for B&B Algorithm. Consider an example in which there are 13 jobs belonging to 5 groups $G_1 = \{J_{11}, J_{12}\}$, $G_2 = \{J_{21}, J_{22}, J_{23}\}$, $G_3 = \{J_{31}, J_{32}, J_{33}\}$, $G_4 = \{J_{41}, J_{42}\}$, and $G_5 = \{J_{51}, J_{52}, J_{53}\}$. The processing times of each job, learning rate, and setup time of each group and other parameters are shown in Tables 1 and 2.

From Algorithm 3 (HA), the initial sequence is $[G_1, G_2, G_4, G_5, G_3]$, and the objective function value (upper bound) is $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j} = 295.8274$. According to Algorithm 4 (B&B algorithm), the following search tree can be obtained, which is represented by Figure 1. The numbers in Figure 1 represent the lower bound values, and G_0 is defined as the level 0.

At level 1, i.e., $g = 1$, for group G_1, from formula (8), the lower bound is

Table 1: Numerical parameters.

	s_i	α_i	β_i	γ_i	u_i'
G_1	3	1	2	1	1
G_2	5	1	2	1	1
G_3	7	1	2	1	1
G_4	4	1	2	1	1
G_5	6	1	2	1	1

Table 2: Numerical parameters.

	$\eta = 1$												
	J_{11}	J_{12}	J_{21}	J_{22}	J_{23}	J_{31}	J_{32}	J_{33}	J_{41}	J_{42}	J_{51}	J_{52}	J_{53}
$p_{i,j}$	4	6	5	3	5	4	7	7	6	3	2	8	5
$a_{i,j}$	−0.16	−0.21	−0.22	−0.12	−0.24	−0.23	−0.32	−0.31	−0.15	−0.17	−0.19	−0.2	−0.26
$\nu_{i,j}$	5	4	7	6	8	6	3	4	2	5	6	4	5

(1) Input: $k_{[i]} \longleftarrow$ using equation (4) for $i = 1, 2, \ldots, f$; the internal job sequence $\pi_i^* \longleftarrow$ using Lemma 4 for each group G_i, $i = 1, 2, \ldots, f$.
(2) Output: the suboptimal resource allocation $u_{[i],[j]}^*$, suboptimal common flow allowance $q_{[i]} =, C_{[i][k_{[i]}]-1}$, and corresponding objective function value $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) +, \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$.
3) For each π_i^*, groups are scheduled by the nondecreasing order of s_i, $(i = 1, 2, \ldots, f)$;
(4) For each π_i^*, groups are scheduled by the nonincreasing order of $m_i\gamma_i$, $(i = 1, 2, \ldots, f)$;
(5) From Step 3 and Step 4, the smallest objective function value $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i), + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ is selected as the original group sequence π_G;
(6) Pick the two groups from the first and second position of the list of Step 5 and find the best sequence for these two groups by calculating $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ for the two possible sequences. Do not change the relative positions of these two groups with respect to each other in the remaining steps of the algorithm. Set $i = 3$;
(7) Pick the group in the ith position of the list generated in Step 5 and find the best sequence by placing it at all possible i positions in the partial sequence found in the previous step, without changing the relative positions to each other of the already assigned groups. The number of enumerations at this step equals i;
(8) **If** $i > f$, **then** stop
(9) **Otherwise,**
(10) $i \longleftarrow i + 1$, and return to Step 7;
(11) Calculate the suboptimal resource allocation $u_{[i],[j]}^*$ according to (5). Calculate the suboptimal common flow allowance $q_{[i]} = C_{[i][k_{[i]}]-1}$ and the corresponding optimal objective function $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ by using (7).

Algorithm 3: HA.

Step 1. Finding the upper bound: use Algorithm 3 to obtain an initial solution for the problem.
Step 2. Bounding: calculate the lower bound (see (8)) for the node. If the lower bound for an unfathomed partial schedule of groups is larger than or equal to the value of the objective function of the initial solution, eliminate the node and all the nodes following it in the branch. Calculate the objective function value of the completed schedule; if it is less than the initial solution, replace it as the new solution; otherwise, eliminate it.
Step 3. Termination: continue until all nodes have been explored.

Algorithm 4: B&B algorithm.

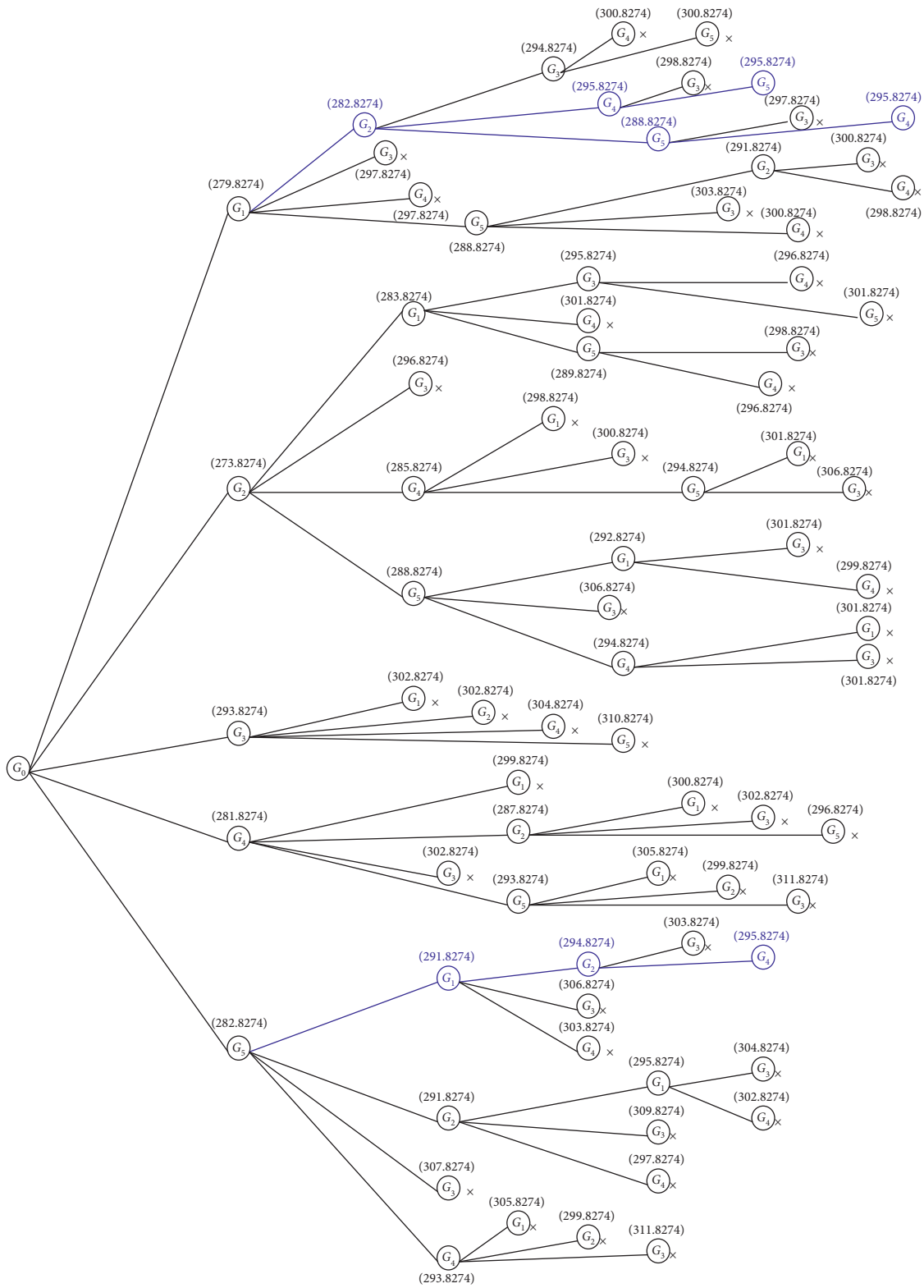


Figure 1: Search tree of the B&B algorithm for the example (× denotes pruning).

Step 1. Calculate $k_{[i]}$ by using (4), $i = 1, 2, \ldots, f$.
Step 2. The jobs in each group are arranged according to the assignment problem **AP** (Lemma 4).
Step 3. List all the group schedules.
Step 4. For each group schedule, calculate the corresponding objective value $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$.
Step 5. Comparing all the objective values $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$, the minimum one is optimal, and its corresponding schedule is the optimal sequence of the problem $1|\mathrm{GT}, s_i, \mathrm{CRA}, \mathrm{SLK}|\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$.

Algorithm 5: Optimal solution for Case 3.

Step 1. Let the tabu list be empty and the iteration number be zero.
Step 2. Set the initial sequence of the TS algorithm, calculate its objective function, and set the current sequence as the best solution π^*.
Step 3. Search the associated neighborhood of the current sequence and resolve if there is a sequence π^{**} with the smallest objective function in associated neighborhood and it is not in the tabu list.
Step 4. If π^{**} is better than π^*, then let $\pi^* = \pi^{**}$. Update the tabu list and the iteration number.
Step 5. If there is not a sequence in associated neighborhood but it is not in the tabu list or the maximum number of iterations is reached, then output the final sequence. Otherwise, update tabu list and go to Step 3.

Algorithm 6: TS.

$$\begin{aligned}
LB(G_1) &= (2\times 1 + 3\times 1 + 3\times 1 + 3\times 1 + 2\times 1)\times 3 \\
&\quad + (3\times 1 + 3\times 1 + 3\times 1 + 2\times 1)\times 4 + (3\times 1 + 3\times 1 + 2\times 1)\times 5 \\
&\quad + (3\times 1 + 2\times 1)\times 6 + 2\times 1\times 7 \\
&\quad + 2\times\begin{pmatrix} (2\times(2-1))^{(1/2)}\times(4\times 5)^{(1/2)} \\ +(2\times(3-1))^{(1/2)}\times(3\times 6)^{(1/2)} + (2\times(3-2))^{(1/2)}\times\left(5\times 7\times 2^{-0.22}\right)^{(1/2)} \\ +(2\times(3-1))^{(1/2)}\times(3\times 7)^{(1/2)} + (2\times(3-2))^{(1/2)}\times\left(6\times 4\times 2^{-0.23}\right)^{(1/2)} \\ +(2\times(3-1))^{(1/2)}\times(6\times 2)^{(1/2)} + (2\times(3-2))^{(1/2)}\times\left(5\times 5\times 2^{-0.26}\right)^{(1/2)} \\ +(2\times(2-1))^{(1/2)}\times(2\times 6)^{(1/2)} \end{pmatrix} \\
&= 279.8274.
\end{aligned} \tag{21}$$

The calculation process of lower bounds of the remaining node is similar to that of this node. From Figure 1, the optimal sequences are $[G_1, G_2, G_4, G_5, G_3]$, $[G_1, G_2, G_5, G_4, G_3]$, and $[G_5, G_1, G_2, G_4, G_3]$, and the optimal value of objective function is $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j} = 295.8274$.

6.6. Computational Experiments. An enumeration algorithm (i.e., Algorithm 5), heuristic algorithm (i.e., Algorithm 3), TS algorithm (i.e., Algorithm 6), and B&B algorithm (i.e., Algorithm 4) were programmed in C++ and carried out on a CPU Intel Core i5-8250U 1.4 GHz PC with 8.00 GB RAM. The number of jobs and groups $n = 50, 100, 150, 200$ and $f = 8, 9, 10, 11, 12$ were tested, and each group must contain at least one job. The parameters setting can be obtained as follows: s_i, $v_{i,j}$, α_i, β_i, and γ_i were drawn from a discrete uniform distribution in [1, 10]; $p_{i,j}$ were drawn from a discrete uniform distribution in [1, 100]; $a_{i,j}$ were drawn from a uniform distribution in [−0.1, −1]; and $\eta = 2$. To avoid the contingency, each problem instance was conducted 20 times, setting the maximum CPU time per instance at 3600 seconds. The percentage relative error of the solution produced by Algorithms 3 and 6 is calculated as

$$\frac{Z(Ai) - Z^*}{Z^*}\times 100\%, \tag{22}$$

where $i = 4, 5$, $Z(Ai)$, and Z^* are the objective function values $\sum_{i=1}^{f}\sum_{j=1}^{m_i}(\alpha_i E_{i,j} + \beta_i T_{i,j} + \gamma_i q_i) + \sum_{i=1}^{f}\sum_{j=1}^{m_i}\nu_{i,j}u_{i,j}$ generated by Algorithm 1 and Algorithm 4, respectively.

On the other hand, "A3-CPU time (s)" ("A4-CPU time (s)," "A5-CPU time (s)," and "A6-(B&B) CPU time (s)") as the running time of Algorithm 5 (Algorithms 3–4) is defined. The results are summarized in Table 3. From Table 3, it is easy to find that the maximum relative error percentage of Algorithm 3 is less than 6.1% for $n \le 200$, and performance of Algorithm 3 (i.e., the heuristic algorithm) performs very well in terms of the

Table 3: Results of Algorithms.

		A3-CPU time (s)		A4 (HA)-CPU time (s)		A5 (TS)-CPU time (s)		A6 (B&B)-CPU time (s)		Error percentage of Algorithm A4 (%)		Error percentage of Algorithm A5 (%)	
	f	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max
n	8	3.971	5.582	0.032	0.091	0.022	0.052	27.382	32.422	2.561	5.071	9.972	10.063
	9	28.362	34.652	0.053	0.123	0.043	0.194	39.029	57.875	1.361	2.622	5.962	21.044
	10	104.129	130.994	0.048	0.121	0.052	0.354	42.344	98.673	0.961	1.142	7.651	20.212
	11	1000.221	1089.481	0.054	0.282	0.115	0.352	91.898	583.131	1.272	1.831	4.033	14.321
50	12	3600	3600	0.061	0.823	0.133	0.491	173.384	875.419	4.112	6.032	5.841	28.402
	8	4.229	8.229	0.034	0.092	0.033	0.191	32.125	39.009	1.371	2.632	7.284	9.792
	9	43.345	50.812	0.032	0.112	0.042	0.194	39.245	45.673	1.854	1.983	4.082	30.125
	10	264.826	406.541	0.044	0.214	0.651	0.423	36.304	69.695	3.114	3.835	5.641	23.164
	11	1997.841	3600	0.052	0.292	0.691	0.784	66.647	300.213	2.812	3.051	4.914	23.815
100	12	3600	3600	0.076	0.651	0.146	0.802	195.528	893.523	1.343	1.692	6.961	21.573
	8	7.851	19.331	0.033	0.112	0.034	0.415	40.554	49.069	2.781	3.742	6.462	20.744
	9	57.825	89.071	0.043	0.255	0.047	0.324	52.417	82.521	1.561	1.952	10.184	39.163
	10	329.781	497.362	0.045	0.291	0.066	0.694	109.697	209.871	2.052	2.931	7.322	19.972
	11	3193.322	3600	0.057	0.331	0.095	0.911	159.225	323.745	3.131	4.052	6.913	27.331
150	12	3600	3600	0.071	0.992	0.139	0.684	243.252	904.032	2.663	3.054	9.771	28.324
	8	11.861	32.881	0.042	0.212	0.047	0.309	51.859	69.672	0.781	1.516	4.421	18.714
	9	98.651	122.325	0.047	0.215	0.059	0.575	74.957	194.634	1.181	1.844	5.521	22.605
	10	475.841	779.563	0.051	0.421	0.067	0.688	193.577	488.581	2.724	3.114	8.821	16.345
200	11	3600	3600	0.061	0.985	0.101	0.691	296.335	723.316	1.641	2.954	6.601	26.125
	12	3600	3600	0.079	1.431	0.139	0.704	435.229	1012.532	1.773	2.265	8.223	28.108

relative error percentages. Moreover, Table 3 shows that the mean CPU time (s) for the enumeration algorithm (i.e., Algorithm 5) is larger than the B&B algorithm (i.e., Algorithm 4).

7. Conclusions

In this paper, we studied the single-machine resource allocation scheduling problem with learning effect and group technology. The goal is to determine the optimal sequence of jobs and groups, the optimal common flow allowances, and the optimal resource allocation such that the weighted sum of the scheduling cost and the resource allocation cost is minimized. For some special cases (i.e., cases $(s_i \le s_h) \Longrightarrow (m_i\gamma_i \ge m_h\gamma_h)$, $s_i = s, m_i\gamma_i = \overline{m\gamma}$, and f is a given constant), it was shown that the problem can be solved in polynomial time. For the general case of the problem, the heuristic, tabu search, and B&B algorithms were proposed. The results show that the maximum relative percentage error of the proposed heuristic algorithm (i.e., Algorithm 3) from optimal solutions is less than 6.1% for all sizes of instances.

Further research may focus on the extensions of this model to more complicated machine setting (such as flow shop and/or parallel machines) or study other nonregular objective functions (such as due-window assignment scheduling problems with position-dependent weights, see the study by Wang et al. [50]).

Data Availability

The data used to 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.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (71903020 and 71672019) and the Natural Science Foundation of Liaoning Province, China (20180550743). Ji-Bo Wang was also supported by the Natural Science Foundation of Liaoning Province, China (2020-MS-233).

References

[1] D. Shabtay and G. Steiner, "A survey of scheduling with controllable processing times," *Discrete Applied Mathematics*, vol. 155, no. 13, pp. 1643–1666, 2007.

[2] D. Biskup, "A state-of-the-art review on scheduling with learning effects," *European Journal of Operational Research*, vol. 188, no. 2, pp. 315–329, 2008.

[3] A. Azzouz, M. Ennigrou, and L. Ben Said, "Scheduling problems under learning effects: classification and cartography," *International Journal of Production Research*, vol. 56, no. 4, pp. 1642–1661, 2018.

[4] Z. Zhu, L. Sun, F. Chu, and M. Liu, "Single-machine group scheduling with resource allocation and learning effect," *Computers & Industrial Engineering*, vol. 60, no. 1, pp. 148–157, 2011.

[5] Y.-Y. Lu, J.-B. Wang, P. Ji, and H. He, "A note on resource allocation scheduling with group technology and learning effects on a single machine," *Engineering Optimization*, vol. 49, no. 9, pp. 1621–1632, 2017.

[6] L. Sun, A. J. F. Yu, and B. Wu, "Single machine common flow allowance group scheduling with learning effect and resource allocation," *Computers and Industrial Engineering*, vol. 139, Article ID 106126, 2020.

[7] D.-Y. Lv, S.-W. Luo, J. Xue, J.-X. Xu, and J.-B. Wang, "A note on single machine common flow allowance group scheduling with learning effect and resource allocation," *Computers & Industrial Engineering*, vol. 151, Article ID 106941, 2021.

[8] J.-B. Wang, J. Xu, and J. Yang, "Bi-criterion optimization for flow shop with a learning effect subject to release dates," *Complexity*, vol. 2018, Article ID 9149510, 12 pages, 2018.

[9] J. B. Wang, F. Liu, and J. J. Wang, "Research on m -machine flow shop scheduling with truncated learning effects *m*-machine flow shop scheduling with truncated learning effects," *International Transactions in Operational Research*, vol. 26, no. 3, pp. 1135–1151, 2019.

[10] P. Yan, J.-B. Wang, and L.-Q. Zhao, "Single-machine bi-criterion scheduling with release times and exponentially time-dependent learning effects," *Journal of Industrial and Management Optimization*, vol. 15, no. 3, pp. 1117–1131, 2019.

[11] J.-B. Wang, M. Gao, J.-J. Wang, L. Liu, and H. He, "Scheduling with a position-weighted learning effect and job release dates," *Engineering Optimization*, vol. 52, no. 9, pp. 1475–1493, 2020.

[12] X. Sun, X.-N. Geng, and F. Liu, "Flow shop scheduling with general position weighted learning effects to minimise total weighted completion time," *Journal of the Operational Research Society*, 2020.

[13] V. Kayvanfar, I. Mahdavi, and G. M. Komaki, "Single machine scheduling with controllable processing times to minimize total tardiness and earliness," *Computers & Industrial Engineering*, vol. 65, no. 1, pp. 166–175, 2013.

[14] V. Kayvanfar, I. Mahdavi, and G. H. M. Komaki, "A drastic hybrid heuristic algorithm to approach to JIT policy considering controllable processing times," *International Journal of Advanced Manufacturing Technology*, vol. 69, no. 1–4, pp. 257–267, 2013.

[15] Y.-Y. Lu and J.-Y. Liu, "A note on resource allocation scheduling with position-dependent workloads," *Engineering Optimization*, vol. 50, no. 10, pp. 1810–1827, 2018.

[16] Y.-C. Tsao, V.-V. Thanh, and F.-J. Hwang, "Energy-efficient single-machine scheduling problem with controllable job processing times under differential electricity pricing," *Resources*," *Conservation and Recycling*, vol. 161, Article ID 104902, 2020.

[17] B. Mor, D. Shabtay, and L. Yedidsion, "Heuristic algorithms for solving a set of NP-hard single-machine scheduling problems with resource-dependent processing times," *Computers & Industrial Engineering*, vol. 153, Article ID 107024, 2021.

[18] V. Kayvanfar, G. M. Komaki, A. Aalaei, and M. Zandieh, "Minimizing total tardiness and earliness on unrelated parallel machines with controllable processing times," *Computers & Operations Research*, vol. 41, pp. 31–43, 2014.

[19] M. H. F. Zarandi and V. Kayvanfar, "A bi-objective identical parallel machine scheduling problem with controllable processing times: a just-in-time approach," *International Journal of Advanced Manufacturing Technology*, vol. 77, no. 1–4, pp. 545–563, 2015.

[20] V. Kayvanfar, M. Zandieh, and E. Teymourian, "An intelligent water drop algorithm to identical parallel machine scheduling with controllable processing times: a just-in-time approach," *Computational and Applied Mathematics*, vol. 36, no. 1, pp. 159–184, 2017.

[21] C. T. Mosier, D. A. Elvers, and D. Kelly, "Analysis of group technology scheduling heuristics," *International Journal of Production Research*, vol. 22, no. 5, pp. 857–875, 1984.

[22] S. Webster and K. R. Baker, "Scheduling groups of jobs on a single machine," *Operations Research*, vol. 43, no. 4, pp. 692–703, 1995.

[23] S.-J. Yang and D.-L. Yang, "Single-machine scheduling simultaneous with position-based and sum-of-processing-times-based learning considerations under group technology assumption," *Applied Mathematical Modelling*, vol. 35, no. 5, pp. 2068–2074, 2011.

[24] M. Ji, K. Chen, J. Ge, and T. C. E. Cheng, "Group scheduling and job-dependent due window assignment based on a common flow allowance," *Computers & Industrial Engineering*, vol. 68, pp. 35–41, 2014.

[25] Y.-T. Xu, Y. Zhang, and X. Huang, "Single-machine ready times scheduling with group technology and proportional linear deterioration," *Applied Mathematical Modelling*, vol. 38, no. 1, pp. 384–391, 2014.

[26] F. Liu, J. Yang, and Y.-Y. Lu, "Solution algorithms for single-machine group scheduling with ready times and deteriorating jobs," *Engineering Optimization*, vol. 51, no. 5, pp. 862–874, 2019.

[27] W.-X. Li and C.-L. Zhao, "Single machine scheduling problem with multiple due windows assignment in a group technology," *Journal of Applied Mathematics and Computing*, vol. 48, no. 1-2, pp. 477–494, 2015.

[28] X. Zhang and Q. Xie, "Single machine group scheduling with position dependent processing times and ready times," *Mathematical Problems in Engineering*, vol. 2015, Article ID 206230, 9 pages, 2015.

[29] P. Ji and L. Li, "Single-machine group scheduling problems with variable job processing times," *Mathematical Problems in Engineering*, vol. 2015, Article ID 758919, 9 pages, 2015.

[30] X. Zhang, L. Liao, W. Zhang, T. C. E. Cheng, Y. Tan, and M. Ji, "Single-machine group scheduling with new models of position-dependent processing times," *Computers & Industrial Engineering*, vol. 117, pp. 1–5, 2018.

[31] S. Muştu and T. Eren, "The single machine scheduling problem with setup times under an extension of the general learning and forgetting effects," *Optimization Letters*, 2020.

[32] A. Allahverdi, "The third comprehensive survey on scheduling problems with setup times/costs," *European Journal of Operational Research*, vol. 246, no. 2, pp. 345–378, 2015.

[33] J. S. Neufeld, J. N. D. Gupta, and U. Buscher, "A comprehensive review of flowshop group scheduling literature," *Computers & Operations Research*, vol. 70, pp. 56–74, 2016.

[34] D. Wang, M.-Z. Wang, and J.-B. Wang, "Single-machine scheduling with learning effect and resource-dependent processing times," *Computers & Industrial Engineering*, vol. 59, no. 3, pp. 458–462, 2010.

[35] Y.-Y. Lu, G. Li, Y.-B. Wu, and P. Ji, "Optimal due-date assignment problem with learning effect and resource-dependent processing times," *Optimization Letters*, vol. 8, no. 1, pp. 113–127, 2014.

[36] H. He, M. Liu, and J.-B. Wang, "Resource constrained scheduling with general truncated job-dependent learning effect," *Journal of Combinatorial Optimization*, vol. 33, no. 2, pp. 626–644, 2017.

[37] L. Li, P. Yan, P. Ji, and J.-B. Wang, "Scheduling jobs with simultaneous considerations of controllable processing times and learning effect," *Neural Computing and Applications*, vol. 29, no. 11, pp. 1155–1162, 2018.

[38] J. B. Wang, D. Y. Lv, J. Xu, P. Ji, and F. Li, "Bicriterion scheduling with truncated learning effects and convex

controllable processing times," *International Transactions in Operational Research*, vol. 28, no. 3, pp. 1573–1593, 2021.

[39] X.-N. Geng, J.-B. Wang, and D. Bai, "Common due date assignment scheduling for a no-wait flowshop with convex resource allocation and learning effect," *Engineering Optimization*, vol. 51, no. 8, pp. 1301–1323, 2019.

[40] X. Sun, X.-N. Geng, J.-B. Wang, and F. Liu, "Convex resource allocation scheduling in the no-wait flowshop with common flow allowance and learning effect," *International Journal of Production Research*, vol. 57, no. 6, pp. 1873–1891, 2019.

[41] W. Liu and C. Jiang, "Due-date assignment scheduling involving job-dependent learning effects and convex resource allocation," *Engineering Optimization*, vol. 52, no. 1, pp. 74–89, 2020.

[42] J.-B. Wang and X.-X. Liang, "Group scheduling with deteriorating jobs and allotted resource under limited resource availability constraint," *Engineering Optimization*, vol. 51, no. 2, pp. 231–246, 2019.

[43] X.-X. Liang, M. Liu, Y.-B. Feng, J.-B. Wang, and L.-S. Wen, "Solution algorithms for single-machine resource allocation scheduling with deteriorating jobs and group technology," *Engineering Optimization*, vol. 52, no. 7, pp. 1184–1197, 2020.

[44] B. Liao, X. Wang, X. Zhu, S. Yang, and P. M. Pardalos, "Less is more approach for competing groups scheduling with different learning effects," *Journal of Combinatorial Optimization*, vol. 39, no. 1, pp. 33–54, 2020.

[45] G. I. Adamopoulos and C. P. Pappis, "Single machine scheduling with flow allowances," *Journal of the Operational Research Society*, vol. 47, no. 10, pp. 1280–1285, 1996.

[46] M. Nawaz, E. E. Enscore, and I. Ham, "A heuristic algorithm for the m-machine, n-job flow-shop sequencing problem*mn*-machine,-job flow-shop sequencing problem," *Omega*, vol. 11, no. 1, pp. 91–95, 1983.

[47] F. Glover, "Tabu search-Part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.

[48] K. Xu, Z. Feng, and K. Jun, "A tabu-search algorithm for scheduling jobs with controllable processing times on a single machine to meet due-dates," *Computers & Operations Research*, vol. 37, no. 11, pp. 1924–1938, 2010.

[49] C.-C. Wu, W.-H. Wu, W.-H. Wu, P.-H. Hsu, Y. Yin, and J. Xu, "A single-machine scheduling with a truncated linear deterioration and ready times," *Information Sciences*, vol. 256, pp. 109–125, 2014.

[50] J.-B. Wang, B. Zhang, L. Li, D. Bai, and Y.-B. Feng, "Due-window assignment scheduling problems with position-dependent weights on a single machine," *Engineering Optimization*, vol. 52, no. 2, pp. 185–193, 2020.

Hindawi
Complexity
Volume 2021, Article ID 5539195, 14 pages
https://doi.org/10.1155/2021/5539195

Research Article

Green Credit, Financial Ecological Environment, and Investment Efficiency

Meng Qi

College of Marxism, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Meng Qi; qimeng@sdust.edu.cn

Received 12 January 2021; Revised 10 February 2021; Accepted 20 February 2021; Published 2 March 2021

Academic Editor: Lei Xie

Copyright © 2021 Meng Qi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article uses the "Green Credit Guidelines" issued in 2012 as a quasi-natural experiment, using the statistics of A-share listed companies from 2008 to 2017, using the PSM-DID model to examine the effect and mechanism of green credit policies on the investment efficiency of heavily polluting companies, and taking into consideration the heterogeneous influence of the financial ecological environment on the relationship between the two. The research indicates that, after the Green Credit Guidelines were promulgated, the investment efficiency of heavy-polluting companies has been slightly improved compared with non-heavy-polluting companies and that the impact is more obvious in regions with better financial ecological environment. The research conclusions confirm the beneficial effects of the Green Credit Guidelines policy on the prudent investment of companies that cause serious pollution to the environment and improve investment efficiency, a provision of empirical evidence for financial leverage to drive the green economy transformation.

1. Introduction

In June 2015, the Environmental Protection Department of Liaoning Province, in conjunction with financial institutions, imposed green credit restrictions on 37 companies that violated environmental laws. These companies are mainly involved in heating, chemical, financial, smelting, papermaking, electroplating, and other industries that seriously pollute the environment. The environmental protection department will follow up the supervision of these enterprises, and the restrictions can only be lifted after the rectification of their problems is in place. Contrary to the restrictions on loans in the first quarter of 2013, the Agricultural Bank of China Zhejiang Branch issued 7.2 billion yuan in loans to 47 green environmental protection projects, 3.5 billion yuan increased over the same period of the previous year; the bank's "green loan" balance reached 30.5 billion yuan. Its loans are mainly for infrastructure projects such as clean energy, sewage treatment, garbage treatment, and energy-saving services. The above two scenarios are real cases of economic transformation driven by green credit. Especially, since the promulgation of the "Green Credit Guidelines" in 2012 (hereinafter referred to as the "Guidelines"), China's green credit scale has gradually expanded. As of the end of June 2020, China's green credit balance has exceeded 11 trillion yuan, taking a leading position in the world; China's stock of green bonds is 1.2 trillion yuan, which takes the runner-up position in the world.

From the perspective of the implementation path of green credit, on the one hand, by raising the loan threshold for heavily polluting industries, companies are forced to undergo green transformation and upgrading. On the other hand, it focuses on supporting the financing of environmental protection industry, clean energy, sewage treatment, and other livelihood projects, alleviating their financing constraints, and promoting the rapid development of green industries. The promulgation of the Green Credit Guidelines is a significant measure for financial services in the real economy, aiming to enhance the efficiency of fund spending in heavily polluting industries and environmentally friendly industries. Therefore, after the reform and exploration in recent years, what is the effect of the policy? Can the inefficient investment of heavily polluting enterprises be truly

improved? This series of questions urgently need to be tested by theoretical research.

Looking back to existing research, green credit policies mainly play a part in the transmission of financial entities and how the effect of the policy depends on the response of microenterprises. Aizawa and Chaofei [1] and Naveiro and Aoussat [2] pointed out that the core purpose of green finance is to promote the coordinated and sustainable development of economic and ecological benefits. However, sustainable development is inseparable from the efficient investment efficiency of enterprises; information asymmetry and agency problems often make enterprise investment deviate from the optimal level, resulting in low investment efficiency of enterprises [3, 4]. The issuance of green loans does not improve public expectations of enterprises in the green industry [5]. The promulgation of the green finance policy requires polluting companies to disclose their environmental information and reduce the information asymmetry between banks and enterprises; by adjusting financial resources, the financing costs and investment risks of polluting companies can be increased [6, 7]. Therefore, through the promulgation of the green credit policy, the amount of financing that heavily polluting companies can obtain from financial institutions has been reduced. Enterprise management must reassess the future capital operation status, carefully select investment projects, and change the direction of investment to reduce inefficient investment and improve the investment efficiency of heavily polluting enterprises. Green credit policy can encourage enterprises to pay attention to early prevention and control measures rather than late mitigation measures [8]. In addition, the implementation effect of green credit policies between regions is closely related to the differences in the financial ecological environment. The financial ecological environment will have a certain degree of impact on corporate debt financing costs, financing structure, debt maturity, etc. [9–12]. Green credit policy is effective in suppressing the investments of energy-intensive industries. Liu et al. and Michael [13, 14] showed that the quality of the financial ecological environment has a certain degree of influence on the credit financing capacity and credit term structure of enterprises. Dosi [15] found that the financial ecological environment's constraints and incentives for the operation of the financial market are affected by many factors such as social economy, systems, law, people's living standards, and education which ultimately lead to different behaviors of financial entities. In summary, the impact of green credit policies on corporate investment efficiency will vary due to differences in the financial ecological environment. Therefore, under the strategic background of financial services in the real economy, how the green credit policy guides the investment behavior of companies that cause serious pollution to the environment is a practical issue that needs to be immediately handled in the theoretical and academic circles. Studying the relationship between the two has important theoretical and practical significance for improving green finance to serve the real economy and driving the green transformation of heavily polluting enterprises.

Based on this, this article selects my country's A-share listed companies from 2008 to 2017 as the research sample, distinguishing heavy-polluting companies from nonheavy-polluting companies; this paper uses the PSM-DID model to explore the effect and mechanism of the "Guidelines" on corporate investment efficiency and examine the heterogeneous effects of the financial ecological environment on the relationship between the two to test the effectiveness and region of the "Guidelines" difference. The research contributions of this paper are mainly reflected in two aspects: (1) from the perspective of capital demand and capital utilization, it explains the impact of green credit on the investment efficiency of heavily polluting enterprises and further explores its internal mechanisms. (2) On the basis of the study of the relationship between the two, we further explored the impact of the heterogeneity of the financial ecological environment and provided an empirical reference for promoting the smooth implementation of green credit.

2. Theoretical Analysis and Research Hypothesis

2.1. Green Credit and Corporate Investment Efficiency. The purpose of the CBRC's "Guidelines" is to use the "green credit" of banking financial institutions to change the existing unreasonable credit structure and effectively stop both environmental and social risks to support the real economy with a higher standard and boost the adjustment of economic structure as well as the transformation of economic development mode. The "Guidelines" clearly stated that special credit guidelines should be formulated for restricted categories and industries with major environmental and social risks that are subject to national key regulation, implementation of differentiated and dynamic credit policies, and risk exposure management systems. After the "Guidelines" are promulgated, heavy-polluting companies may have the following two changes: (1) heavy-polluting companies focus on the influence of the "Guidelines" on them. Therefore, according to the assessment standards of environmental and social risks, they reestimate the future capital status and policy changes of the company and invest more cautiously. (2) Compared with non-heavy-polluting enterprises, under the limited capital level, heavy-polluting enterprises can reduce ineffective investment and improve investment efficiency; this is a necessary condition for heavy-polluting enterprises to survive. Thus, how do these two aspects play a role in reality?

Concerning the first change, in the context of the country's increased environmental control, if heavily polluting companies continue their original investment plans, they will indeed face large fines or even suspend business for rectification, such as high environmental taxes, administrative fines, and and taking off the market, etc. Therefore, if heavy-polluting companies want to survive, they must reevaluate the future financial situation and policy changes, invest prudently, and allocate more resources to effective and efficient investment. In response to the second change, heavy-polluting companies are listed as restricted credit by financial institutions. Compared with non-heavy-polluting

companies, it is more difficult to obtain financial support from financial institutions, and the availability of financing is significantly reduced. Affected by the "Guidelines," the amount of new loans that heavily polluting companies can obtain from banks has decreased. The scale of new loans is often difficult to cover the funding gap, and there is a risk of rupture of the capital chain. The phenomenon of excessive investment as mentioned above has improved. Mengze and Wei [16] confirmed that green credit can improve microeconomic efficiency from three aspects: reducing transaction costs, diversifying or reducing enterprise innovation risks, and supervising invested enterprises or projects. Therefore, this article believes that the green credit policy can form an "elimination mechanism" to improve the efficiency of internal capital allocation of heavily polluting enterprises and increase effective investment through the dynamic game between banks and enterprises. When the green credit policy is implemented, companies must disclose environmental information. The higher the quality of nonfinancial information is, the closer the amount of external financing obtained is to the optimal financing amount, which not only alleviates underinvestment but also avoids overinvestment [17, 18].

From the analysis mentioned above, this article puts forth the hypotheses listed as follows.

Hypothesis 1. After the "Guidelines" are issued, the investment efficiency of heavily polluting enterprises can be significantly improved compared to non-heavy-polluting enterprises.

How does green credit affect the investment efficiency of polluting companies? Peeters [19] believed that when the external macroeconomic uncertainty is high, the investment income of enterprises will become unstable. In this way, the capital needs of enterprises for projects under construction, investment in fixed assets, equipment renewal and transformation, scientific and technological development fees, and trial production of new products will be reduced; at the same time, under the condition of financing constraints, the manager's capital discretionary power will also be reduced along with the debt repayment responsibility so that the company will significantly reduce excessive investment behavior [20]. After the "Guidelines" were issued, the environmental and social risks faced by heavily polluting companies restricted the financing channels. Reduced planned investments, especially investment plans that originally had a negative impact on the ecological environment, are more likely to be forced to suspend or terminate, resulting in a substantial reduction in capital requirements and excessive investment behavior. On the contrary, enterprises will improve the efficiency of the use of existing funds, realize the improvement of microeconomic efficiency, increase effective investment, and accelerate the transformation and upgrading of enterprises. From the analysis mentioned above, this article puts forth the hypotheses listed as follows.

Hypothesis 2. The promulgation of the "Guidelines" has reduced the capital needs of heavily polluting companies and has improved the efficiency of the use of existing funds by heavily polluting companies, thereby affecting the investment efficiency of companies.

2.2. The Impact of Financial Eco-Environment Heterogeneity. The so-called "financial ecological environment" refers to a series of external environments and basic conditions for financial operation. According to its components, a good financial eco-environment is represented by high-speed economic development, a sound legal environment, high-level social credibility, favorable financial sector independence, and sophisticated intermediary services and social security [21]. The financial eco-environment differs observably from region to region, but in terms of China, the overall financial eco-environment in the eastern part and eastern seaside region is good to some degree, while the eco-environment in the middle and western regions is comparatively poor. The conditions of financial eco-environment will significantly affect the allocative efficiency of financial resources in a region. For example, when the financial eco-environment in a certain area is better, the financial market, legal system, and integrity system in that area are significantly better than those in other areas, and the government of this area will lessen the interference in its market. It shows that the ability of banks to recognize the environmental information risks of borrowing companies and the ability to transform risk compensation are deeply affected by the financial eco-environment [22]. According to the research of previous scholars, the financial eco-environment can indeed significantly affect the allocative efficiency of credit capital. For example, Fazzari et al. [23] found that the financial ecology is positively related to the allocative efficiency of credit capital. Wang et al. [24] found that the effectiveness of financial development in promoting economic growth is affected by the external financial eco-environment in which it is located.

Narrow down to this paper, the impact of green credit on corporate investment efficiency is also surely affected by the financial eco-environment. On the one hand, with the effective implementation of the green credit policy, in areas with better financial eco-environment, the financial market, legal system, and integrity system are relatively complete, and heavily polluting enterprises are subject to stricter supervision and more rigid qualifications from the banks. At this time, under the circumstances of high financing constraints, the green credit policy can play a dynamic game with the banks and be more beneficial in promoting the improvement of the microeconomic efficiency in heavily polluting enterprises which will increase and investment efficiency with the amount of current funds. Meanwhile, when a bank evaluates the credit rating of an enterprise, its external environmental advantages are bound to be included in the assessment, which will increase the possibility of heavily polluting enterprises of obtaining bank loans, and the corresponding loan interest rates and guarantee fees will decrease together [25]. On the other hand, according to the previous analysis, in the areas with poor financial ecological environment, the absence of effective implementation of the

green credit policy is due to the relatively low marketization, poor financial and legal system, weak resource allocation efficiency, etc., which may lead to the possibility of poor implementation; hence, its influence towards heavily polluting enterprises may lower. That is, in areas with poor financial ecological environment, the impact of the green credit policy on heavily polluting enterprises is very low; in addition, the bank's profit-driven motive will not restrict loans to heavily polluting enterprises so that the investments of heavily polluting enterprises will not be greatly affected. From the analysis mentioned above, this article puts forth the hypotheses listed as follows.

Hypothesis 3. The impact of green credit on the investment efficiency of enterprises is more obvious in areas with good financial and ecological environment.

3. Research Design

3.1. Sample Selection and Data Sources. This paper takes the A-share listed companies from 2008 to 2017 as the research object, drawing on the handling methods from Li and Feng [26] as well as Su and Lian [27] to ascertain the heavy-polluting enterprises by calculating the pollution emission intensity in various industries, which is also the experimental group identified in this paper. The detailed steps are as follows.

Firstly, the industrial sulfur dioxide, smoke (powder), and liquid and solid waste emissions published by the National Bureau of Statistics every year through the China Statistical Yearbook are determined as pollutant emissions, and the pollutant emissions per unit of the output value of various industries are calculated, which can be expressed as $UE_{ij} = E_{ij}/O_i\,(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$, where UE_{ij} is the emission per unit output value of pollutant emission j of industry i, E_{ij} is the total emission of pollutant emission j of industry i, and O_i is the total output value of industry i.

Secondly, standardize the discharge amount of pollutants per unit of the output value of various industries to make it within the range of [0, 1]:

$$UE_{ij}^{s} = \left[UE_{ij} - \min\left(UE_j\right)\right] / \left[\max\left(UE_j\right) - \min\left(UE_j\right)\right], \tag{1}$$

where UE_{ij}^{s} is the emission per unit output value of pollutant emission j of the normalized industry i, $\min(UE_j)$ is the minimum emission of pollutant emission j in all industries, and $\max(UE_j)$ is the maximum emission of pollutant emission j in all industries.

Thirdly, the emission intensity γ_i of industry i is calculated, and the heavily polluting industry and non-heavily polluting industry are distinguished according to the median of γ_i:

$$\gamma_i = \sum_{j=1}^{n} UE_{ij}^{s}. \tag{2}$$

Specifically, this paper has calculated the pollution emission intensity of various industries in 2011, the year before the publication of the Guidelines, based on the 2012 China Statistical Yearbook, and identified 20 heavily polluting industries such as power industry, thermal production and supply industries, and paper making and paper products' industry. Based on this, the paper eliminates the ST and *ST enterprises, financial enterprises, and major variables with serious missing and abnormal data and finally ascertains 18,349 observations, including 2852 listed companies which are divided into 1002 experimental groups of listed companies and a control group of 1850 listed companies. The financial eco-environment data in this paper come from China Regional Financial Eco-environmental Evaluation (2013-2014), and other financial and microsurvey data are from the database of Tai'an (CSMAR). To reduce the influence of extreme values on the research conclusion, the main continuous variables are processed with winsorization up and down to 1%.

3.2. Variable Definition

3.2.1. Investment Efficiency. This paper draws on the research results from Richardson [28], AGcB et al. [29], and Wang et al. [30] to use the regression residuals of the following models to represent the investment efficiency of enterprises:

$$\text{Invest}_{i,t} = \alpha_0 + \alpha_1 \text{Invest}_{i,t-1} + \alpha_2 \text{Size}_{i,t-1} + \alpha_3 \text{Cash}_{i,t-1} + \alpha_4 \text{Lev}_{i,t-1} + \alpha_5 \text{Growth}_{i,t-1} + \alpha_6 \text{Return}_{i,t-1} + \alpha_7 \text{Age}_{i,t-1} + \varepsilon_{i,t}. \tag{3}$$

In the formula, Invest represents a new investment, which is equal to (cash spent in the purchase and construction of fixed assets, intangible assets, and other long-term assets – net cash received from the fixed assets, intangible assets, and other assets' disposal) divided by total assets; Size stands for the size of the company, which is equal to the logarithm of the company's overall assets; Cash represents cash and cash equivalents, which is equal to (monetary capital plus tradable financial assets) divided by total assets; Lev represents the asset-liability ratio, which is equal to the company's total liabilities divided by total assets; Growth stands for investment opportunities,

which is equal to the enterprise's increase rate of main business revenue; Return represents the enterprise's annual stock yield, which is equal to the annual stock return rate considering the cash bonus; and Age represents the enterprise's age, which is equal to the logarithm of the enterprise's listed years. Formula (3) regresses by year and industry to obtain the residual and takes the absolute value of the residual and records it as Absinvest, which is the investment efficiency. The larger the value is, the lower the investment efficiency and the higher the nonefficient investment of the enterprise owns.

3.2.2. Financial Ecology Environment. This paper makes conclusions on the research of Xie [31], Wei [32], etc., and uses all references of the financial eco-environment of different areas in China covered by the research group of the "Evaluation of Financial Ecology Environment of China" of the Institute of Finance, Chinese Academy of Social Sciences, to determine the level of financial ecological environment in each region. Since there are no data of the financial ecological environment comprehensive index in 2009, 2011, and 2012, this paper uses the practice of Deng Jianping and Zeng Yong for reference and uses data of 2008 to replace 2009, data of 2010 to replace the data of 2011, data of 2013 to replace the data of 2012, and data of 2014 to replace the data of 2015 to investigate the heterogeneous influence of financial ecological environment from 2008 to 2015. The specific practice is to find the average value of all references of financial eco-environment. If the comprehensive index of the financial eco-environment in a certain area is greater than the average value, then it is assigned to 1, which demonstrates that the financial eco-environment in the region is relatively not bad. If the index is less than the average value, it is assigned to 0, which shows that the financial and ecological environment in the region is relatively bad.

3.2.3. Control Variable. To reduce the impact of other factors, this paper draws on relevant research and sets the following control variables from the aspects of profitability, development capacity, operating capacity, and administration of an enterprise: corporate size, total asset return rate, sales growth rate, total asset turnover, the proportion of independent directors, and the firm's age; in addition to this, this paper further controls the year fixed effect and the industry fixed effect. The definitions of the main variables in this paper are shown in Table 1.

3.3. Model Selection. The double difference (DID) model is a policy evaluation model that is widely used in academia. The basic principle is to divide the sample into an experimental group affected by the policy and a control group not affected by the policy and then compare the experimental group and the control group to get the policy effect. The application of this model needs to satisfy that the experimental group and the control group have similar changing trends (common trends) before the implementation of the policy. However, due to the large differences between the heavy-polluting enterprises and non-heavy-polluting enterprises in this article in terms of different industry norms and management methods, there are certain self-selection problems in the research samples, which make the common trend assumption difficult to meet. In response to this problem, this paper draws on the propensity score matching proposed by Heckman et al. [33]. PSM can better solve the problem of sample selection bias, but it often ignores the endogeneity problem between variables, while DID can solve the endogeneity problem through difference and obtain the policy processing effect, but it has certain defects in solving the sample bias problem. Therefore, a combination of the two methods is chosen to evaluate the impact of green credit on the investment efficiency of enterprises.

The basic idea of the model is as follows: in the control group not affected by the policy, find company j (matching) that is very similar to company i in the experimental group so that $i \approx j$; repeated operations can be matched to a set with a common value range Sp; the experimental group and the control group of this set can better meet the common trend assumption in the double difference. Propensity score matching is to match individuals with the same propensity score together. Typical matching methods include nearest neighbor matching, radius matching, and kernel matching. This paper uses the kernel matching method to determine the weight, and its expression is shown in the following:

$$\omega(i, j) = \frac{K\left[\left(x_j - x_i\right)/h\right]}{\sum_{k:D_{k=0}} K\left[\left(x_k - x_i\right)/h\right]}, \tag{4}$$

where h is the designated bandwidth and $K(\cdot)$ is the kernel function.

The specific PSM-DID process can be divided into the following steps:

(1) Using the logit model to estimate the propensity score:

$$P\left(Z_i\right) = P\left(D_i = 1 \mid Z_i\right) = \frac{\Lambda\left(Z_i'\beta\right)}{\left(1 + \exp\left(Z_i'\beta\right)\right)}. \tag{5}$$

(2) Calculating the changes in investment efficiency of heavily polluting companies and non-heavy-polluting companies before and after the issuance of the Guidelines.

(3) Subtracting the change in investment efficiency of heavy-polluting enterprises before and after the issuance of the Guidelines minus the change in investment efficiency of matching non-heavy-polluting enterprises before and after the issuance of the Guidelines to obtain the average treatment effect after the issuance of the Guidelines, as shown in formula (3.6).

Table 1: Main definitions of variables.

Variable	Symbol	Definition
Investment efficiency	Absinvest	According to the absolute value of the residual of the regression model, the larger the value gets, the lower the investment efficiency and the higher the nonefficient investment are
Is it heavily polluting enterprise?	Pollution	Dummy variable, Pollution = 1, which indicates heavily polluting enterprise; Pollution = 0, which indicates non-heavily polluting enterprise
Policy implementation time	After	Dummy variable, After = 1, which indicates 2012 and subsequent years; After = 0 which indicates 2011 and previous years
Difference-in-differences variable	Pollution × After	The product of Pollution and After indicates the net effect of policy implementation
Financial ecology environment	Fe	Dummy variable, Findex = 1, which indicates the regional financial ecology environment is relatively good; Findex = 0, which indicates the regional financial ecology environment is relatively bad
Corporate size	Size	The logarithm of enterprise's total assets
Total asset return rate	Roa	Net profit/total assets
Sales growth rate	Growth	(The current amount of business revenue of the current year – the synchronous amount of business revenue of last year)/the synchronous amount of business revenue of last year
Total assets turnover	Tat	Business revenue/(total asset + final balance of initial balance)/2
Proportion of independent directors	Ids	Number of independent directors/number of directors
Firm's age	Age	The logarithm of the enterprise's listed years

$$A\widehat{T}T = \frac{1}{N^*}\sum_{i:i\in I_1\cap S_p}\left[\left(Y_T t_1 i - Y_T t_0 i\right) - \sum_{j:j\in I_0\cap S_p} w(i,j)\left(Y_C t_1 i - Y_C t_0 i\right)\right]. \tag{6}$$

In the formula, T represents a heavily polluting company in the experimental group, C represents a non-heavy-polluting company in the control group, t_0 represents before the release of the Guidelines, t_1 represents after the release of the Guidelines, i represents a company in the experimental group, j represents a company in the control group, $w(i, j)$ represents the weight after core matching, I_1 represents the experimental group company before matching, I_0 represents the control group company before matching, and N^* represents the number of enterprises in the experimental group in the set $I\ I_1 \cap S_p$.

Based on the above analysis, this paper sets the regression model based on the PSM-DID method which is as follows:

$$\text{Absinvest}_{it} = \beta_0 + \beta_1\text{Pollution}_{it} + \beta_2\text{After}_{it} + \beta_3\text{Pollution}_{it}\times\text{After}_{it} + \beta_4\text{Control}_{it} + \varepsilon_{it}. \tag{7}$$

In the formula, Control represents the control variable, which includes the control of the fixed effect of the year and the industry; the coefficient β_3 represents the net effect of the policy implementation on the experimental group. If $\beta_3 < 0$, it means that the promulgation of the "Guidelines" has reduced the inefficient investment of heavily polluting enterprises, that is, improved the investment efficiency of heavily polluting enterprises.

In addition, to further explore the dynamic effects of policy implementation, we study the expected effects and lag effects before and after the issuance of the "Guidelines." This paper draws on the practice of Fan et al. [34], and the following model is constructed:

$$\begin{aligned}\text{Absinvest}_{it} = {} & \eta_0 + \eta_1\text{Pollution}_{it} + \eta_2\text{After}_{it} + \sum_{t=2008}^{t=2011}\eta_t\text{Pollution}_{it}\times\text{Preyear}_t \\ & + \eta_{2012}\text{Pollution}_{it}\times\text{Current}_{2012} + \sum_{t=2013}^{t=2017}\eta_t\text{Pollution}_{it}\times\text{Postyear}_t \\ & + \eta_3\text{Control}_{it} + \varepsilon_{it}.\end{aligned} \tag{8}$$

4. Outcomes of Practice

4.1. Descriptive Statistics of the Main Variables. Tables 2 and 3 list the descriptive statistics of the main variables in this article. It can be seen from Table 2 that the average investment efficiency of the full sample of companies is 0.0238, indicating that the overall investment efficiency of listed companies in my country is relatively good, and the inefficient investment is not serious. For the heavy-polluting companies in the experimental group, before the "Guidelines" were issued, the average investment efficiency was 0.0316. After the promulgation of the Guidelines, the average investment efficiency was 0.0244, a decrease of 23%. To a certain extent, the promulgation of the Guidelines has greatly reduced the inefficient investment of heavily polluting enterprises and improved the investment efficiency. The mean value of the investment efficiency of non-heavy-polluting companies in the control group before the issuance of the Guidelines was 0.0259. The average value after the issuance of the Guidelines is 0.0210, a decrease of 19%, which shows that the inefficiency investment of all types of enterprises has decreased after the issuance of the Guidelines. Compared with non-heavy-polluting enterprises, the impact of inefficient investment by heavy-polluting enterprises is greater. The specific causality and internal mechanism need to be empirically tested.

4.2. Model Checking

4.2.1. Impact of Green Credit on Investment Efficiency

(1) Average Treatment Effect. We evaluated the influence of green credit policies on the inefficient investment of companies that cause serious pollution through a panel regression model, with the results listed in Table 4. Result (1) includes only the key explanatory variable, namely, inefficient investment, and the coefficient is significantly negative, indicating that the promulgation of the Guidelines has indeed significantly reduced the inefficient investment of heavily polluting enterprises, improved the efficiency of capital allocation, and thereby increased investment efficiency; so, Hypothesis 1 is not rejected. On the basis of model 1, control variables are gradually added to the model to form results (2)–(7). Clearly, the coefficient of Pollution × After is still negative, which further supports the conclusion of Hypothesis 1.

(2) Dynamic Action. To study the dynamic effects of green credit on investment efficiency, this paper further uses a hybrid panel model, with the results shown in Table 5. The dynamic panel regression model shows that, in the year the Guidelines were issued, the implementation of the policy did not have a significant impact on the investment efficiency of heavily polluting enterprises. The main reason is that financial institutions have a time lag in formulating their own credit policies in accordance with the Guidelines. The investment efficiency of heavily polluting enterprises is mainly calculated by the investment projects, and it takes a certain amount of time to change the investment plan. Since 2013, the net effect of policy implementation has been significantly negative, at least at the level of 1%, indicating that the Guidelines have shown microeffects since the second year of policy promulgation, and the inefficient investment of heavily polluting enterprises has been significantly reduced. The regression coefficient in 2013 was −0.0071, and it was significant at the 1% level. That is to say, after the "Guidelines" were issued, the effect of policy implementation in 2013 was the most obvious. Although the absolute value of the regression coefficient decreased slightly in 2014 compared with 2013, the absolute value of the regression coefficient has increased year by year since 2014, which also means that the implementation process of the Guidelines has stabilized, and the degree of influence has gradually increased.

(3) Test Based on the PSM-DID Model. To reduce the systematic differences between the heavily polluting enterprises in the experimental group and the non-heavily polluting enterprises in the common trend assumption and obtain more robust research conclusions, this paper further tests the above conclusions by using PSM-DID. If there is no important dissimilarity between the experimental group and the control group after the matching, the propensity score matching needs to meet the balance between the nature hypothesis and the common support hypothesis. The hypothesis of equilibrium property means that there is no important dissimilarity between the experimental group and the control group after matching, while the common support hypothesis means that both the experimental group and the control group have enough overlapping areas in the value range.

The matching of the propensity score can be completed in the following two steps. First, the appropriate matching variables are selected, the conditional distribution of variables is estimated, that is, green credit, and then the propensity score is calculated, where the propensity score refers to the probability that the sample enterprise will be affected by the guidance after controlling the observable factors. Second, according to the calculated tendency score and the variable green credit in the first step, the conditional distribution of the investment efficiency of the resulting variable is estimated. Due to the propensity score covering the "synthesis" of the influence of all matching variables on investment efficiency, the consistency of the control propensity score can effectively guarantee the independence of green credit and investment efficiency, which greatly reduces the estimation bias caused by the difference of control variables between different enterprises.

Referring to the related study and the selection of this paper, finally, the enterprise size (Size), the asset-liability ratio (Lev), the growth rate of business income (Growth), the concentration of equity (Shrcr), and the total asset turnover rate (Tat) are determined as matching variables; in addition, the year is further controlled. Table 6 shows the balance property test results of each matching variable after kernel matching.

Table 2: Description of the main variables.

Variable	N	Mean	SD	Min	Max
Absinvest	18,349	0.0238	0.0247	0.0002	0.1361
Size	18,349	22.0830	1.2516	19.7200	25.8306
Roa	18,349	0.0418	0.0531	−0.1429	0.2107
Lev	18,349	0.4444	0.2077	0.0530	0.8790
Growth	18,349	0.1934	0.4504	−0.5657	2.8958
Tat	18,349	0.6717	0.4683	0.0651	2.6503
Ids	18,287	0.3720	0.0530	0.3077	0.5714
Age	18,349	2.1714	0.7286	0.6931	3.2189

Table 3: Descriptive statistical analysis of the main variables before and after the policy.

Variable	Experimental group						Control group					
	Before			After			Before			After		
	Number of samples	Mean value	Standard deviation	Number of samples	Mean value	Standard deviation	Number of samples	Mean value	Standard deviation	Number of samples	Mean value	Standard deviation
Absinvest	2126	0.0316	0.0283	4749	0.0244	0.0240	2880	0.0259	0.0264	8594	0.0210	0.0230
Size	2126	21.9186	1.2182	4749	22.2319	1.2445	2880	21.7705	1.1741	8594	22.1461	1.2667
Roa	2126	0.0454	0.0649	4749	0.0419	0.0580	2880	0.0474	0.0490	8594	0.0391	0.0480
Lev	2126	0.4812	0.1970	4749	0.4207	0.2084	2880	0.4773	0.1989	8594	0.4374	0.2102
Growth	2126	0.1959	0.3507	4749	0.1470	0.3993	2880	0.2294	0.4798	8594	0.2063	0.4856
Tat	2126	0.7949	0.4749	4749	0.6809	0.4396	2880	0.7461	0.5374	8594	0.6113	0.4468
Ids	2097	0.3649	0.0513	4749	0.3709	0.0517	2849	0.3666	0.0508	8592	0.3761	0.0545
Age	2126	2.1416	0.6388	4749	2.2627	0.7043	2880	2.1319	0.7259	8594	2.1416	0.7588

Table 4: The impact of green credit on investment efficiency.

	(1) $M1$	(2) $M2$	(3) $M3$	(4) $M4$	(5) $M5$	(6) $M6$	(7) $M7$
Pollution	0.0165***	0.0196***	0.0196***	0.0199***	0.0190***	0.0191***	0.0194***
	(0.0020)	(0.0022)	(0.0021)	(0.0021)	(0.0021)	(0.0022)	(0.0020)
After	0.0085***	−0.0067***	−0.0068***	−0.0069***	−0.0079***	−0.0082***	−0.0079***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Pollution × After	0.0029***	−0.0030***	−0.0030***	−0.0030***	−0.0029***	−0.0028***	−0.0026***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Size		−0.0020***	−0.0021***	−0.0021***	−0.0020***	−0.0020***	−0.0015***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Roa			0.0147***	0.0109***	0.0158***	0.0158***	0.0069*
			(0.0037)	(0.0038)	(0.0038)	(0.0038)	(0.0039)
Growth				0.0020***	0.0023***	0.0023***	0.0021***
				(0.0005)	(0.0005)	(0.0005)	(0.0005)
Tat					−0.0037***	−0.0036***	−0.0032***
					(0.0005)	(0.0005)	(0.0005)
Ids						0.0063***	0.0046
						(0.0034)	(0.0034)
Age							−0.0033***
							(0.0003)
cons	0.0164***	0.0591***	0.0599***	0.0603***	0.0598***	0.0580***	0.0546***
	(0.0018)	(0.0038)	(0.0038)	(0.0038)	(0.0038)	(0.0040)	(0.0039)
R^2 adjusted	0.0588	0.0669	0.0678	0.0690	0.0722	0.0726	0.0795
F	26.7022	29.9979	29.7287	29.8250	30.0947	29.6057	31.0591
N	18349	18349	18349	18349	18349	18287	18287

Notes: standard deviations are in parentheses; *, **, and *** indicate significant differences at $p < 0.01$, $p < 0.05$, and $p < 0.001$, respectively.

Therefore, there is no exact method to determine the matching effect, but the smaller the absolute value of the standard deviation after matching, the better the matching effect. Rosenbaum and Rubin [35] pointed out that if the absolute value of the standard deviation between the matched variables is greater than 20%, the matching effect is not good. Chang et al. [36] pointed out that if the absolute value of standard deviation between variables after matching

Table 5: Dynamic effect of green credit on investment efficiency.

Variable	Panel regression model
Pollution × Preyear 2008	−0.0019 (0.0024)
Pollution × Preyear 2009	0.0000 (.)
Pollution × Preyear 2010	−0.0090*** (0.0022)
Pollution × Preyear 2011	−0.0032 (0.0022)
Pollution × Current 2012	−0.0035* (0.0021)
Pollution × Postyear 2013	−0.0071*** (0.0020)
Pollution × Postyear 2014	−0.0055*** (0.0020)
Pollution × Postyear 2015	−0.0062*** (0.0019)
Pollution × Postyear 2016	−0.0066*** (0.0019)
Pollution × Postyear 2017	−0.0079*** (0.0019)
_cons	0.0539*** (0.0039)
R^2 adjusted	0.0807
F	25.7534
N	18287.0000

is less than 5%, then the matching effect is reliable and acceptable. To sum up, the absolute value of the standard deviation of the matching variable, enterprise scale, asset-liability ratio, operating income growth rate, equity concentration, and total asset turnover rate after matching is less than 5%, so the matching variable and matching method selected in this paper can be considered appropriate, and the matching estimation result is also reliable. On the contrary, from the point of view of the mean difference between the experimental group and the control group before and after matching, there is a significant difference in the mean value of matching variables before matching and no important dissimilarity after matching, which also shows that the matching estimation results in this paper are reliable.

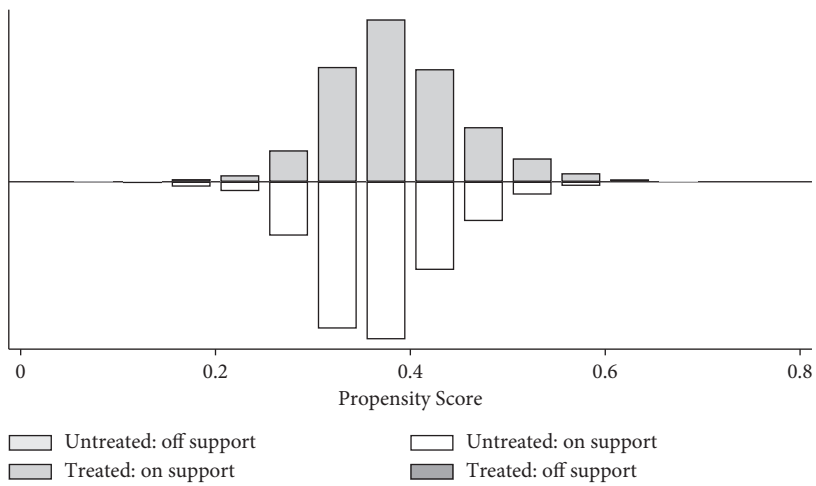
The above test shows that if the propensity score matching the balance property hypothesis is satisfied, then the common support hypothesis requires that the propensity scores of the experimental group and the control group should have enough common value range; that is to say, the propensity scores of the experimental group and the control group should have similar distribution characteristics. Figure 1 is the result of the common support hypothesis test. After the kernel matching, the observation values are in the common value range, so the observation values after matching and in the common support range will not affect the samples.

In summary, after matching the propensity scores, the data of the heavy-polluting enterprises in the experimental group and the non-heavy-polluting enterprises in the control group have been well balanced. At this time, the heavy-polluting enterprises and the non-heavy-polluting enterprises have similar characteristics, thus ensuring the common trend assumption in the DID model is satisfied. The table lists the test results of the PSM-DID model. It is not difficult to see from Table 7 that the double difference term is −0.001, which is significant at the 10% level, making it clear that the promulgation of the "Guidelines" has reduced the inefficiency of enterprises' investment; compared with non-heavy-polluting companies, the impact of heavy-polluting companies is greater.

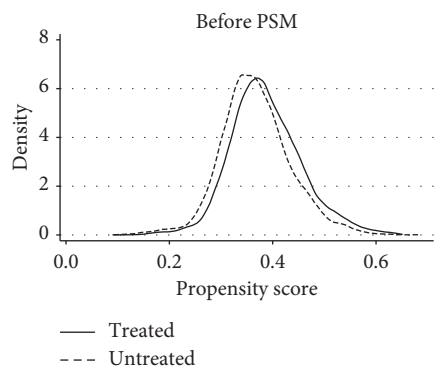
4.2.2. Mechanism Test of Green Credit Affecting Investment Efficiency. The previous research shows that, after the Guidelines were proclaimed, heavy-polluting companies have reduced inefficient investment and improved investment efficiency compared with non-heavy-polluting enterprises. We will further examine the mechanism by which the Guidelines affect the investment efficiency of heavily polluting enterprises. After the "Guidelines" were issued, heavy-polluting companies lacked sources of funds. Under the influence of national policies, the planned investment of heavy-polluting companies may decrease. In particular, investment plans that originally had a negative impact on the ecological environment are more likely to be forced to suspend or terminate, resulting in a substantial reduction in capital requirements. Instead, more attention is paid to the use of existing funds, which improves the efficiency of the use of existing funds, thereby reducing inefficient investment and improving investment efficiency. To verify the accuracy of this hypothesis, this article draws on the research methods by Baron and Kenny [37] and Satterthwaite [38]. It is checked through the following three steps: in the first step, the difference item is used to regress the capital demand; if the regression coefficient of the difference item is significant, it indicates that the promulgation of the "Guidelines" has significantly affected the capital needs of heavily polluting enterprises. The second step is to use the difference term to regress the investment efficiency; assuming that the regression coefficient of the difference term is significant, it means that the promulgation of the "Guidelines" has affected the investment efficiency of heavily polluting enterprises; since this step has been completed before, it will be omitted. The third step is to use the difference term and capital demand to regress the investment efficiency; if the difference term coefficient is no longer significant or the significance is reduced or significant, but the absolute value of the coefficient is reduced, it means that the promulgation of the "Guidelines" will indeed affect the enterprise through the capital demand investment efficiency. According to the above steps, this article designs the following model:

Table 6: Test of the equilibrium property of matching variables.

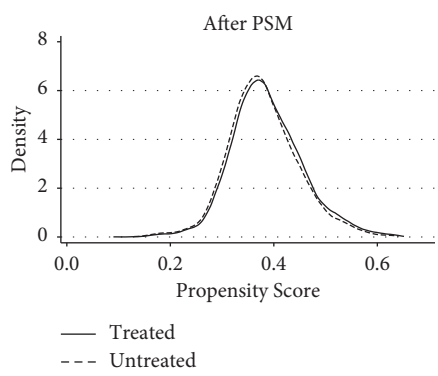
Variable		Mean		Standard deviation (%)	Standard deviation reduction margin (%)	T-statistics
		Experimental group	Control group			
Size	Before	22.135	22.052	6.7	70.6	4.36***
	After	22.135	22.111	2.0		1.13
Lev	Before	0.4394	0.4475	−3.9	94.0	−2.54**
	After	0.4394	0.4398	−0.2		−0.14
Growth	Before	0.1621	0.2121	−11.4	94.0	−7.29***
	After	0.1617	0.1587	0.7		0.45
Shrcr	Before	0.3599	0.3452	9.8	75.1	6.44***
	After	0.3599	0.3562	2.4		1.42
Tat	Before	0.7162	0.6451	15.3	81.3	9.97***
	After	0.7162	0.7030	2.9		1.59



(a)

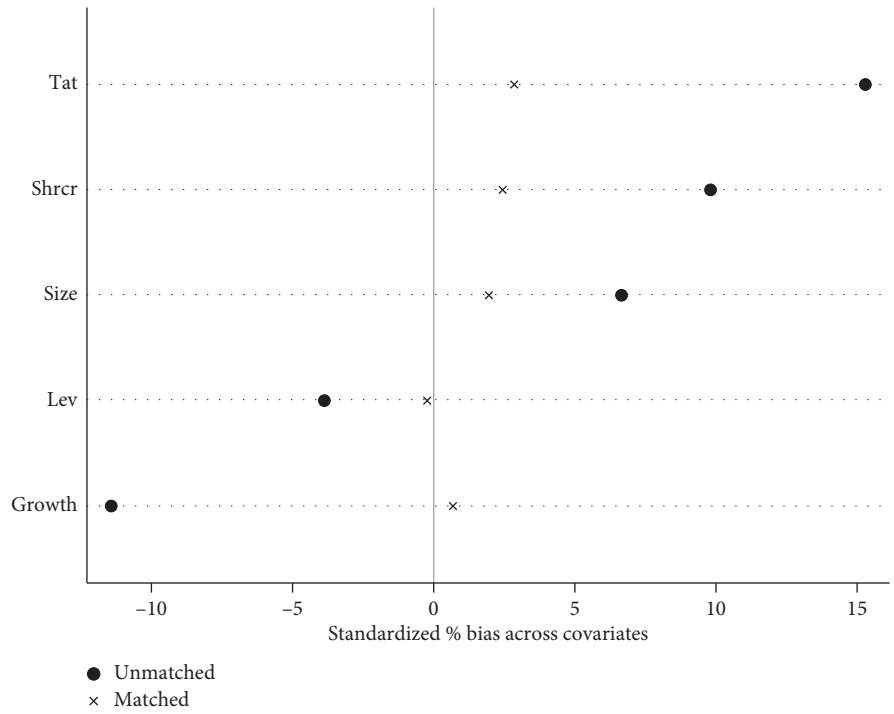


(b)



(c)

Figure 1: Continued.



(d)

Figure 1: Common support hypothesis testing.

Table 7: Test results of PSM-DID.

Variable	Before the promulgation of the Guidelines			After the promulgation of the Guidelines			Difference-in-difference
	Experimental group	Control group	Difference	Experimental group	Control group	Difference	
Absinvest	0.032	0.026	0.006	0.024	0.019	0.005	−0.001
SE			0.001			0.001	0.001
$\lvert t \rvert$			8.88			9.04	1.71
$p > \lvert t \rvert$			≤0.001***			≤0.001***	0.087*

Standard deviations are in parentheses; *, **, and *** indicate significant differences at $p < 0.01$, $p < 0.05$, and $p < 0.001$, respectively.

$$D_{it} = \delta_0 + \delta_1 \text{Pollution}_{it} + \delta_2 \text{After}_{it} + \delta_3 \text{Pollution}_{it} \times \text{After}_{it} + \delta_4 \text{Control}_{it} + \varepsilon_{it}, \tag{9}$$

$$\begin{aligned} \text{Absinvest}_{it} = {} & \mu_0 + \mu_1 \text{Pollution}_{it} + \mu_2 \text{After}_{it} + \mu_3 \text{Pollution}_{it} \times \text{After}_{it} \\ & + \mu_4 D_{it} + \mu_5 \text{Control}_{it} + \varepsilon_{it}. \end{aligned} \tag{10}$$

Equation (4.1) is step one, and equation (4.2) is step three, where D_{it} represents the capital requirement. Drawing on the research of Wang and Song (2014), the funding requirements in this article include liquidity funding requirements (Dld) and long-term funding requirements (Dfd). Liquidity capital requirements refer to the short-term or daily business needs of the company; the uncertainty of the company's liquidity capital quota will affect the company's future financing behavior and then affect the company's investment behavior. Liquidity requirements = short-term borrowing + short-term bonds payable + additional issuance and allotment based on short-term demand-short-term borrowing based on long-term demand. Long-term capital demand refers to the long-term capital needs of the company or project construction and investment. The company's project investment generally comes from long-term funds, and the increase in long-term capital uncertainty will inevitably reduce the company's investment behavior. Long-term funding needs = long-term borrowing + additional issuance and allotment based on long-term needs-long-term borrowing based on short-term needs. The capital requirements in this article are all measured by changes in capital requirements, which are changes in capital requirements = (current capital requirements−preliminary capital requirements)/preliminary capital requirements.

Table 8: Mechanism of green credit affecting investment efficiency.

	(1) Liquidity demand	(2) Long-term funding needs	(3) Investment efficiency
Pollution	0.2564*	−0.0713	0.0070***
	(0.1472)	(0.1977)	(0.0017)
After	−0.1225	−1.1532***	−0.0063***
	(0.1280)	(0.2045)	(0.0017)
Pollution × After	−0.1257*	−0.0634	−0.0017
	(0.0700)	(0.1050)	(0.0010)
Dld			0.0003***
			(0.0001)
_cons	−7.6161***	−12.8218***	0.0730***
	(1.1241)	(1.6411)	(0.0132)
R^2 adjusted	0.0227	0.0162	0.0441
F	11.1497	8.7302	30.5698
N	18287	18287	18287

Standard deviations are in parentheses; *, **, and *** indicate significant differences at $p < 0.01$, $p < 0.05$, and $p < 0.001$, respectively.

Table 9: Test of the heterogeneity of the financial ecological environment.

Variable	Mixed panel model	
	Fe = 1 (1)	Fe = 0 (2)
Pollution	0.0128***	0.0135***
	(0.0036)	(0.0034)
After	−0.0017	−0.0054***
	(0.0013)	(0.0020)
Pollution × After	−0.0035**	−0.0014
	(0.0012)	(0.0015)
Control	Control	Control
Year	Control	Control
Industry_cons	Control	Control
	0.0505***	0.0623***
R^2 adjusted	(0.0052)	(0.0069)
R^2 adjusted	0.0844	0.0669
F	22.4277	9.9204

Standard deviations are in parentheses; *, **, and *** indicate significant differences at $p < 0.01$, $p < 0.05$, and $p < 0.001$, respectively.

Table 8 shows the test results of the mechanism for green credit affecting investment efficiency. Judging from the coefficients of the difference terms of models (1) and (2), the promulgation of the Guidelines has reduced the liquidity capital requirements and long-term capital requirements of enterprises, but only the demand for liquidity funds is significant. The explained variable of model (3) is the investment efficiency, that is, the test result of formula (4.1); the difference term of model (3) is no longer significant. Therefore, it can be shown that the influence of the Guidelines on corporate investment efficiency is indeed affected by the demand for funds, especially the demand for liquidity funds, with Hypothesis 2 not rejected.

4.2.3. Impact of Financial Ecological Environment Heterogeneity. It can come to light from Table 9 that the absolute value of the regression coefficient of the differential term with a better regional financial eco-environment is greater than the regression coefficient of the differential term with a poor regional financial eco-environment, and the regional differential term with a good financial eco-environment is significant at the level of 5%. We can find a better financial ecological environment can indeed increase the impact of the Guidelines on investment efficiency, and Assumption 3 cannot be rejected. It needs to be further pointed out that the quality of the financial eco-environment can change the effect of the implementation of green credit policies; areas with a better financial eco-environment are conducive to the implementation and transmission of policies. In this study, the areas with better financial eco-environment include Shanghai, Beijing, Zhejiang, Guangdong, Jiangsu, Fujian, Tianjin, Shandong, Chongqing, Liaoning, and Anhui.

4.3. Robustness Test. For making the research conclusions more robust, this article first uses a placebo test and moves the policy time forward by 1 year. It is assumed that the policy is implemented in 2011 (fictitious policy), and the inspection period is set to 2007–2016. If the regression result of the difference term under the fictitious policy approach is still significant, it means that the original estimation result is likely to be biased. On the contrary, if the difference term is no longer significant, it can be explained to a certain extent that the virtual policy will not have an impact on the investment efficiency of the company, and the result of regression is robust. This article uses a placebo test to verify the impact of green credit on investment efficiency. The results show that the difference terms are no longer significant, as shown in Table 10. It can be seen that the above conclusions of this article are robust. This article also constructed a new experimental group and a control group based on the heavily polluting companies disclosed in the List of Listed Companies' Environmental Inspection Industry Classification Management as the experimental group of this article. The above process is verified again, and the result shows that the difference term is still significantly negative, as shown in Table 11. It can be seen that the classification standards of

Table 10: Robust test (1).

Variable	Before the promulgation of the Guidelines			After the promulgation of the Guidelines			Difference-in-difference
	Experimental group	Control group	Difference	Experimental group	Control group	Difference	
Absinvest	0.031	0.025	0.006	0.025	0.020	0.005	−0.001
SE			0.001			0.001	0.001
$\|t\|$			7.8			8.53	0.73
$p > \|t\|$			$\leq 0.001^{***}$			$\leq 0.001^{***}$	$p = 0.464$

Standard deviations are in parentheses; *, **, and *** indicate significant differences at $p < 0.01$, $p < 0.05$, and $p < 0.001$, respectively.

Table 11: Robust test (2).

Variable	Before the promulgation of the Guidelines			After the promulgation of the Guidelines			Difference-in-difference
	Experimental group	Control group	Difference	Experimental group	Control group	Difference	
Absinvest	0.032	0.025	0.007	0.023	0.020	0.004	−0.003
SE			0.001			0.001	0.001
$\|t\|$			10.24			7.38	3.74
$p > \|t\|$			$\leq 0.001^{***}$			$\leq 0.001^{***}$	$\leq 0.001^{***}$

Standard deviations are in parentheses; *, **, and *** indicate significant differences at $p < 0.01$, $p < 0.05$, and $p < 0.001$, respectively.

heavily polluting enterprises will not affect the research hypothesis of this article, so the above conclusions still sound.

5. Research Conclusions and Inspiration

This article takes all listed companies in both Shanghai and Shenzhen stock exchange market from 2008 to 2017 as an object, takes the promulgation of the 2012 Green Credit Guidelines as a quasi-natural experiment, and uses the PSM-DID model to empirically examine the impact of green credit policies on the investment efficiency of companies that cause heavy pollution. The conclusions are as follows: first, compared to the enterprises with minor pollution issues, the promulgation of the "Guidelines" has indeed significantly reduced the inefficient investment of heavy-polluting enterprises, thereby improving investment efficiency, and the policy effect has stabilized. Second, the promulgation of the "Guidelines" restricts the sources of funds for heavily polluting companies and improves the efficiency of companies' use of existing funds by affecting their capital needs. Third, regional policies with a better financial eco-environment are easier to implement and transmit; therefore, the promulgation of the Guidelines in these regions has a more obvious impact on the investment efficiency of heavily polluting enterprises.

The research conclusions of this paper have important guiding significance for supporting the construction of national ecological civilization and the development of green finance. The policy implications of the conclusions of this article are as follows: first, the promulgation of the green credit policy has imposed financing constraints on companies which can pollute the environment, using green to force heavily polluting companies to transform and upgrade, reduce the inefficient investment of heavily polluting companies, and improve the investment efficiency of heavily polluting companies. This also shows that the effect of the national macropolicy has been reflected in the micromarket to a certain extent. The country should further deepen the green finance policy, establish a more complete green finance evaluation mechanism, and achieve high-quality economic and social development. Second, credit financing is a booster for industrial restructuring. Green credit policy can adjust the allocation of financial resources and then affect the investment and financing decisions of heavy-polluting enterprises. Financial institutions should vigorously develop green finance to help the country's supply-side structural reforms. Third, the financial eco-environment can give an impetus to the effective implementation of green credit policies. Local governments should strengthen the construction of the financial ecological environment, while promoting economic development and financial stability, they must create a good financial ecological atmosphere so that national policies can be transmitted smoothly and local enterprises can be effectively guaranteed.

Data Availability

The raw data supporting the conclusions of this article are available upon request to the author, without undue reservation, to any qualified researcher.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] M. Aizawa and Y. Chaofei, "Green credit, green stimulus, green revolution? China's mobilization of banks for environmental cleanup," *The Journal of Environment & Development*, vol. 19, no. 2, pp. 119–144, 2010.

[2] R. M. Naveiro and A. Aoussat, "Systematic literature review of eco-innovation models: opportunities and recommendations for future research," *Journal of Cleaner Production*, vol. 149, pp. 1278–1302, 2017.

[3] S. C. Myers, "Determinants of corporate borrowing," *Journal of Financial Economics*, vol. 5, no. 2, pp. 147–175, 1977.

[4] M. P. Narayanan, "Managerial incentives for short-term results," *The Journal of Finance*, vol. 40, no. 5, p. 1469, 1985.

[5] C. Luo, Q. S. Fan, and Q. Zhang, "Investigating the influence of green credit on operational efficiency and financial performance based on hybrid econometric models," *International Journal of Financial Studies*, vol. 5, no. 4, p. 27, 2017.

[6] M. Lemmon and M. R. Roberts, "The response of corporate financing and investment to changes in the supply of credit," *Journal of Financial and Quantitative Analysis*, vol. 45, no. 3, pp. 555–587, 2010.

[7] M. Allet and M. Hudon, "Green microfinance: characteristics of microfinance institutions involved in environmental management," *Journal of Business Ethics*, vol. 126, no. 3, pp. 395–414, 2015.

[8] J. Sun, H. F. Wang, and B. Zhang, "Money talks: the environmental impact of China's green credit policy," *Journal of Policy Analysis and Management*, vol. 38, no. 3, p. 653, 2019.

[9] R. Levine, "Financial development and economic growth: views and agenda," *Journal of Economic Literature*, vol. 6, pp. 688–726, 1997.

[10] S. Labatt, "Environmental finance: a guide to environmental risk assessment and financial products," *Transplantation*, vol. 66, no. 8, pp. 405–409, 2002.

[11] P. Guo, "Financial policy innovation for social change: a case study of China's green credit policy," *International Review of Sociology*, vol. 24, no. 1, pp. 69–76, 2014.

[12] C. X Jing, "Financial ecological environment, ownership and debt-financing structure," *Journal of Shandong University(Philosophy and Social Sciences*, vol. 03, pp. 100–106, 2011.

[13] J. Y. Liu, Y. Xia, Y. Fan et al., "Assessment of a green credit policy aimed at energy-intensive industries in China based on a financial CGE model," *Journal of Cleaner Production*, vol. 163, p. 293, 2017.

[14] F. M. Michael, "Does the source of capital affect capital structure?" *Review of Financial Studies*, vol. 19, no. 1, pp. 45–79, 2005.

[15] G. Dosi, "Finance, innovation and industrial change," *Journal of Economic Behavior & Organization*, vol. 13, no. 3, pp. 299–319, 1990.

[16] H. Mengze and L. Wei, "A comparative study on environment credit risk management of commercial banks in the asia-pacific region," *Business Strategy and the Environment*, vol. 24, no. 3, pp. 159–174, 2013.

[17] M. Faulkender and R. Wang, "Corporate financial policy and the value of cash," *The Journal of Finance*, vol. 61, no. 4, pp. 1957–1990, 2006.

[18] G. C. Biddle and G. Hilary, "Accounting quality and firm-level capital investment," *The Accounting Review*, vol. 81, no. 5, pp. 963–982, 2006.

[19] M. Peeters, "Does demand and price uncertainty affect Belgian and Spanish corporate investment?. discussion papers," *REL-Recherches Economiques de Louvain*, 2001.

[20] R. D'Mello and M. Miranda, "Long-term debt and overinvestment agency problem," *Journal of Banking & Finance*, vol. 34, no. 2, pp. 324–335, 2010.

[21] L. Yang, *Evaluation of Urban Financial Ecological Environment in China*, People's Publishing House, New Delhi, India, 2005.

[22] N. Bloom, S. Bond, and J. Van Reenen, "Uncertainty and investment dynamics," *Review of Economic Studies*, vol. 74, no. 2, pp. 391–415, 2007.

[23] S. M. Fazzari, R. G. Hubbard, and B. C. Petersen, "Financing constraints and corporate investment," *Brookings Papers on Economic Activity*, vol. 1, pp. 141–195, 1988.

[24] E. Wang, X. Liu, J. Wu et al., "Green credit, debt maturity, and corporate investment-evidence from China," *Sustainability*, vol. 11, no. 3, p. 583, 2019.

[25] R. J. Indjejikian, "Discussion of accounting information, disclosure, and the cost of capital," *Journal of Accounting Research*, vol. 45, no. 2, pp. 421–426, 2007.

[26] L. Li and F. Tao, "Selection of optimal environmental regulation intensity for Chinese manufacturing industry-based on the green TFP perspective," *China Industrial Economics*, vol. 05, pp. 70–82, 2012.

[27] D. Su and L. Lian, "Does green credit policy affect corporate financing and investment? evidence from publicly listed firms in pollution-intensive industries," *Financial Research*, vol. 12, pp. 123–137, 2018.

[28] S. Richardson, "Over-investment of free cash flow," *Review of Accounting Studies*, vol. 11, no. 2-3, pp. 159–189, 2006.

[29] AGcB, BGH, and CRSV, "How does financial reporting quality relate to investment efficiency?" *Journal of Accounting and Economics*, vol. 48, no. 2–3, pp. 112–131, 2009.

[30] X. Wang, F. Chen, O. K. Hope et al., "Financial reporting quality and investment efficiency of private firms in emerging markets," *Social Science Electronic Publishing*, vol. 86, no. 4, pp. 1255–1288, 2011.

[31] D. Xie, "Financial ecological environment, property right and governance effect of debt," *Economic Research*, vol. 44, no. 05, pp. 118–129, 2009.

[32] Z. Wei, "Financial ecological environment and corporate financial constraints-evidence from Chinese listed firms," *Accounting Research*, vol. 05, pp. 73–80, 2014.

[33] J. J. Heckman, H. Ichimura, and P. E. Todd, "Matching as an econometric evaluation estimator: evidence from evaluating a job training programme," *The Review of Economic Studies*, vol. 64, no. 4, pp. 605–654, 1997.

[34] Z. Fan, F. Peng, and C. Liu, "Political connections and economic growth: evidence from the DMSP/OLS satellite data," *Economic Research*, vol. 51, no. 1, pp. 114–126, 2016.

[35] P. R. Rosenbaum and D. B. Rubin, "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score," *The American Statistician*, vol. 39, no. 1, pp. 33–38, 1985.

[36] X. Chang, S. Dasgupta, and G. Hilary, "The effect of auditor quality on financing decisions," *The Accounting Review*, vol. 84, no. 4, pp. 1085–1117, 2009.

[37] R. M. Baron and D. A. Kenny, "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations," *Journal of Personality and Social Psychology*, vol. 51, no. 6, p. 1173, 1986.

[38] D. Satterthwaite, "Environmental transformations in cities as they get larger, wealthier and better managed," *The Geographical Journal*, vol. 163, no. 2, pp. 216–224, 1997.

Hindawi
Complexity
Volume 2021, Article ID 6670873, 13 pages
https://doi.org/10.1155/2021/6670873

Research Article

Multiview Graph Learning for Small- and Medium-Sized Enterprises' Credit Risk Assessment in Supply Chain Finance

Cong Wang,[1] Fangyue Yu,[1] Zaixu Zhang,[1] and Jian Zhang [2]

[1]*School of Economics and Management, China University of Petroleum, Qingdao, China*
[2]*School of Government, Central University of Finance and Economics, Beijing, China*

Correspondence should be addressed to Jian Zhang; zjpolicy@163.com

Received 25 December 2020; Revised 28 January 2021; Accepted 1 February 2021; Published 15 February 2021

Academic Editor: Baogui Xin

Copyright © 2021 Cong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, supply chain finance (SCF) is exploited to solve the financing difficulties of small- and medium-sized enterprises (SMEs). SME credit risk assessment is a critical part in the SCF system. The diffusion of SME credit risk may cause serious consequences, leading the whole supply chain finance system unstable and insecure. Compared with traditional credit risk assessment models, the supply chain relationship, credit condition of SME, and core enterprises should all be considered to rate SME credit risk in SCF. Traditional methods mix all indicators from different index systems. They cannot give a quantitative result on how these index systems work. Furthermore, traditional credit risk assessment models are heavily dependent on the number of annotated SME data. However, it is implausible to accumulate enough credit risky SMEs in advance. In this paper, we propose an adaptive heterogenous multiview graph learning method to tackle the small sample size problem for SMEs' credit risk forecasting. Three graphs are constructed by using indicators from supply chain operation, SME financial indicator, and nonfinancial indicator individually. All the graphs are integrated in an adaptive manner, providing a quantitative explanation on how the three parts cooperate. The experimental analysis shows that the proposed method has good performance for determining whether SME is risky or nonrisky in SCF. From the perspective of SCF, SME financing ability is still the main factor to determine the credit risk of SME.

1. Introduction

As a result of the COVID-19 pandemic, global industrial chain and supply chain have suffered a severe setback. Industrial chains across the world are experiencing an unprecedented crisis, which has rarely been encountered nearly the century. The accompanying adjustment of supply chain and industrial chain has a great impact on small- and medium-sized enterprises (SMEs). The survival predicament faced by SMEs also arises in China. According to the National Bureau of Statistics, the SMEs in China played an essential role in the national economy, accounting for more than 99% of the total number of enterprises, more than 80% employed people in industrial enterprises, more than 70% of the national total technological innovation results, more than 60% of the GDP, and more than 50% of the national total tax revenue. However, the financing difficulties faced by SMEs have always been the key issues for their development and survival.

Supply chain finance (SCF) is a series of financing modes designed to solve the SMEs financing problem [1], which integrates capital into supply chain management. With the core enterprise as the center and the real trade as the background, SCF transforms the uncontrollable risk of a single enterprise into controllable risk of the whole supply chain enterprise through effective control of capital flow, information flow, and logistics. It effectively builds a benign industrial ecology of banks, core enterprises, and SMEs, promoting the interactive development of capital and industry. In SCF, credit risk is regarded as one of the core issues, which must be considered seriously, especially that of SMEs in the upstream or downstream of core enterprises. Compared with banks and core enterprises, SMEs are more prone to encounter credit risk. In the SCF system, once the

credit risk occurs, the credit status of enterprises in the chain will be magnified and even spread to the whole supply chain due to the connectivity of the supply chain [2–4]. However, credit risk is inevitable. Therefore, it is necessary to establish a credit risk assessment model for SMEs to effectively control risks, forming a stable SCF system, both in the construction and operation process.

The SMEs' credit risk assessment system under the SCF environment has always been the focus of the academic world and the financial world [5]. The deep integration between core enterprises and SMEs in SCF can reflect the future capital capacity and cash flow of SMEs. This characteristic provides a solid data base for credit risk prediction in SCF. The academic world constructs the credit risk assessment index system from different perspectives. Constructing the credit risk assessment system based on different factors is one of the branches. Lekkakos et al. considered core enterprise qualification as one of the main factors for SME credit evaluation [6]. Wuttke found that the buyer has an impact on supply chain finance [7]. These literatures investigated the correlation between SCF and SME credit risk theoretically. However, some indicators are neither difficult to obtain in practice nor to analyze quantitatively [8].

Another branch established the credit risk assessment system for SMEs in SCF from different perspectives. Altman and Sabato [5] designed an SME credit risk evaluation system, including five financial indicators such as liquidity, earnings, leverage, coverage ratio, and business activity ratio. However, only financial indicators are selected as the main features for SME credit risk assessment and other nonfinancial indicators are neglected. Yurdakul [9] combined both finance and nonfinance indicators to evaluate the SME credit risk in Turkey, including the revenue generation capacity, cost control, operational efficiency and profitability, short-term liquidity, capital structure, and other solvency indicators. In order to consider the factors synthetically on SME credit risk assessment in SCF, many other indicators were also involved in the system. From the perspective of supply chain, Rostamzadeh et al. [10] proposed to take operational and major policy risks into considerations. Mou et al. [11] constructed the evaluation system from four aspects: industrial status, operation status, asset status, and credit history. Among these systems, there may be overlapping or contradictory phenomena among the credit risk assessment indicators selected from different views. It should be emphasized that this issue is difficult to solve for most of the conventional prediction model. Furthermore, most of existing credit risk assessment systems focus on selecting single discriminative indicator (e.g., three-class indicator), without considering the overall discrimination of the whole perspectives (e.g., one or two-class indicator). They cannot give quantitative results about the relationships of the different views.

Recently, machine learning approaches have been widely applied as a substitute to traditional qualitative and statistical analysis methods on credit risk assessment field [12, 13]. Logistic regression (LR) is the basic method to assess credit risk for its simplicity and usability [14]. However, it is insufficient to describe the complex situation of credit risk. The prediction accuracy needs further enhancement. Support vector machine [15] (SVM) is another approach to tackle credit risk problem. It has been proved that the SVM-based credit risk assessment model is effective and advantageous than the LR model based on principal component analysis (PCA) [16]. Danenas and Garsva identified risky enterprises in imbalanced datasets based on linear SVM and particle swarm optimization [17]. The authors of [18] applied a nonlinear SVM classifier with genetic optimization for credit risk assessment. The artificial neural network (ANN) has also attracted a wide range of attention in credit risk forecasting. Adnan made a comprehensive investigation of different supervised neural models and learning schemes for credit risk evaluation [19]. The deep belief network (DBN), as a special paradigm of ANN, has been found to yield the best performance on the CDS dataset [20]. In addition, tree-structured [21] and graph-structured [22] methods have also been shown to be effective on credit risk assessment.

Various approaches have shown their advantageous on credit risk evaluation. However, to assess the credit risks of the SMEs in SCF environment, we need to construct the model from a perspective of the supply chain, rather than only assessing repayment ability [23]. The funds and credit condition of SME, core enterprises, and supply chain relationship are to be involved in the machine learning algorithms. These information from different views need to get organized properly. Most of approaches concatenate the indicators together to evaluate SME credit risk in supply chain finance. By concatenating the indicators, Zhang et al. [23] and Xu and He [24] applied SVM and restricted Boltzmann machine (RBM), respectively, for assessment. From the perspective of feature fusion, these early fusion approaches treat all features in the same way, which weakens the internal connections among different views. Other literatures ensemble multiple machine learning models to make a comprehensive decision. Zhang et al. [25] considered the leading enterprise's credit status and the relationships developed in the supply chain by assembling SVM and BP. Zhu et al. [26] proposed a two-stage hybrid model by integrating the results of LR and ANN to boost the accuracy of the single model. Nevertheless, all methods mentioned above are heavily dependent on the annotated SME data tightly. Once the number of annotated samples decreases, the algorithm performance will decline rapidly. However, it seems implausible to accumulate enough credit risky SMEs in reality. Compared with the method above, graph-based learning can work with small number of annotated data by modeling the relationships of samples, which has been widely used in many other fields [27–29]. It is difficult to construct a graph to cover local manifolds of different view features. Therefore, how to integrate multiview heterogenous features into graph learning is worth pondering.

In this paper, we propose a heterogenous multiview graph learning method for SMEs' credit risk forecasting. The credit risky or nonrisky SME is recognized by diffusion process on the fused multiview manifolds. Considering different views having different impacts on credit risk recognition, we learn the view weights and credit risk scores

simultaneously. The main contributions of our work are summarized as follows:

(1) Tackling the problem of lacking enough annotated data, we propose a multiview graph-based semi-supervised learning approach by modeling the local manifold between samples for SME credit risk assessment in SCF.

(2) The proposed multiview graph-based learning approach integrates multiview features in a comprehensive way. The SME credit risk scores are achieved by learning the complementariness of different views automatically, which improves interpretability.

(3) Comprehensive experiments are conducted to empirically analyze the proposed SME credit risk in the SCF method. The experimental results on the collected dataset demonstrate the effectiveness of the proposed method.

The remainder of this paper is organized as follows. In Section 2 and Section 3, we discuss the single graph learning and multiview graph learning. Section 4 describes the data and variables. Section 5 demonstrates the experimental results. The last section is the conclusion.

2. Single Graph-Based Learning for Classification

Graph-based learning [30, 31] has attracted great interests in classification tasks due to its effectiveness and flexibility to various areas. A graph describes the pairwise relationships based on the given data, where vertices are labeled and unlabeled samples and edges indicate the relationships of vertices. Generally, single graph-based learning can be formulated as follows.

Assume that there are N data points $X = \{x_1, x_2, \ldots, x_N\}$ in the dataset. Each sample $x_i \in \mathfrak{R}^d$ can be represented by a d-dimensional feature vector. Without loss of generality, we assume the first N_l ($N_l \ll N$) samples are labeled from dataset $X_L = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{N_l}, y_{N_l})\}$ by experts in advance, where $y_i = \{-1, +1\}$ is the class label. Our goal is to assign an optimal label y_i for each of the rest $N - N_l$ unlabeled data points through graph learning.

A graph is defined as $G = (V, E, W)$. V is the set of vertices, where each sample x_i is a vertex v_i in G, including both labeled and unlabeled samples. $E = e_{ij}$, $(i = 1, 2, \ldots, N; j = 1, 2, \ldots, N)$ is the set of edges, where e_{ij} links the adjacent vertices v_i and v_j. $W = w_{ij}$, $(i = 1, 2, \ldots, N; j = 1, 2, \ldots, N)$ is termed as affinity matrix, where w_{ij} is the weight of e_{ij}, indicating the strength of vertices v_i and v_j. The value of w_{ij} can be either discrete (e.g., $\{0, 1\}$) or continuous (e.g., 0.54).

Generally, graph-based classification task is formulated in a regularization framework:

$$\arg\min_f \left\{\Omega(f) + \lambda R_{\text{emp}}(f)\right\}, \tag{1}$$

where f is the to-be-learned relevance score vector defined on domain $(-1, 1)$, $R_{\text{emp}}(f)$ is an empirical loss function, $\Omega(f)$ is a regularizer term on f, and λ is a nonnegative parameter.

$\Omega(f)$ implements the smoothness assumption on the graph by considering data points on the same manifold are likely to share the same label.

The regularizer on the graph is defined as follows:

$$\Omega(f) = \frac{1}{2}\sum_{i,j=1}^{N} W_{ij}\left(\frac{f_i}{\sqrt{D_{ii}}} - \frac{f_j}{\sqrt{D_{jj}}}\right)^2, \tag{2}$$

where D is a diagonal matrix given by $D_{ii} = \sum_j W_{ij}$, that is, D_{ii} is the sum of the i-th row of W.

Let $\Theta = D^{-(1/2)} W D^{-(1/2)}$, and the normalized graph Laplacian is denoted as follows:

$$L = I - \Theta, \tag{3}$$

here I is a unit diagonal matrix. Finally, the regularizer can be rewritten as follows:

$$\Omega(f) = f^T L f. \tag{4}$$

The empirical loss function $R_{\text{emp}}(f)$ conducts the consistency of labeled vertices by forcing the assigned labels close to the initial labels.

$$R_{\text{emp}}(f) = \sum_{i=1}^{N} (f_i - y_i)^2, \tag{5}$$

where y_i is the initial label vector. Three possible values $\{1, 0, -1\}$ can be assigned to y_i: 1 if vertex i is the positive sample, -1 if it is the negative sample, and 0 if it is unlabeled. The closed-form solution for the minimization is found to be

$$f = \left(I + \frac{1}{\lambda} L\right)^{-1} y. \tag{6}$$

After obtaining f, data point x_i can be classified according to its sign, i.e., positive if $f_i > 0$ and negative otherwise. In addition, the relative value of data points in X can also be ranked according to the learned f.

3. Multiview Graph-Based Learning for Classification

Single graph learning has improved to be an effective way for classification task, especially when labeled samples are limited. However, to make a comprehensive evaluation on SME credit risk, many different aspect information should be taken into consideration, such as the funds and credit condition of SME, core enterprises, and supply chain operation. These complementary knowledges contained in multiple views to comprehensively represent the credit status of SME. In this situation, more than one view features can be used to measure the affinity between vertices.

Suppose we have M sets of features. Accordingly, M graphs can be constructed, denoted as $G_1, G_2, \ldots, G_M$. The regularization framework in equation (2) can be extended to handle multiview features by combining each graph with a set of weighting coefficient. Then, the framework is written as follows:

$$\Omega_{\text{fix}}(f) = \frac{1}{2}\sum_{g=1}^{M}\sum_{i,j=1}^{N}\alpha_g W_{g,ij}\left(\frac{f_i}{\sqrt{D_{g,ii}}} - \frac{f_j}{\sqrt{D_{g,jj}}}\right)^2. \quad (7)$$

The coefficient is correlative with the importance of the g-th view, constrained by $0 < \alpha_g < 1$ and $\sum_{g=1}^{M}\alpha_g = 1$. W_g and D_g are the affinity matrix and the diagonal matrix for graph G_g, respectively. The value of α_g can be set as prior knowledge in advance by domain experts according to the importance of different views. Then, the solution of equation (1) can be derived as follows:

$$f = \left(I + \frac{\sum_{g=1}^{M}\alpha_g L_g}{\lambda}\right)^{-1} y, \quad (8)$$

where L_g is the normalized graph Laplacian of G_g. Equation (8) amounts to learning on a fused graph, where several normalized graph Laplacians are combined through the coefficient α_g. However, there has a lot of confusion about which view is the more important one for classification task. Subsequently, in most cases, it is not an easy way to set value for α_g in advance. It is desired to incorporate the influence of α_g into the above learning framework. To avoid trivial solutions, a relaxation on weight coefficient by changing α_g to α_g^r is performed on equation (7).

$$\Omega_{a\,dp}(f) = \frac{1}{2}\sum_{g=1}^{M}\sum_{i,j=1}^{N}\alpha_g^r W_{g,ij}\left(\frac{f_i}{\sqrt{D_{g,ii}}} - \frac{f_j}{\sqrt{D_{g,jj}}}\right)^2, \quad (9)$$

where r is a hyperparameter larger than 1 and r is used to avoid only the smoothest graph play the key role ($\alpha_g = 1$), while graphs from the other views are invalid ($\alpha_g = 0$). Finally, the cost function of multiview graph-based learning is reformulated as follows:

$$\begin{aligned} H(f,\alpha) &= \left\{\underbrace{\frac{1}{2}\sum_{g=1}^{M}\sum_{i,j=1}^{N}\alpha_g^r W_{g,ij}\left(\frac{f_i}{\sqrt{D_{g,ii}}} - \frac{f_j}{\sqrt{D_{g,jj}}}\right)^2}_{\Omega_{\text{adp}}(f)} + \underbrace{\lambda\sum_{i=1}^{N}(f_i - y_i)^2}_{R_{\text{emp}}(f)}\right\} \\ &= \sum_{g=1}^{M}\alpha_g^r f^T L_g f + \lambda\|f - y\|^2. \end{aligned} \quad (10)$$

Equation (10) holds that data are in local manifold spaces for different views, respectively. Combination of multiple features can be done through a weighted union of graphs generated by different views. Each graph resembles a weak classifier generated from a single cue, and together they form a stronger classifier.

The relevance value \$f\$ and weight coefficient α_g are achieved by minimizing equation (10).

$$\begin{aligned} &\text{argmin}_{f,\alpha_1,\alpha_2,\ldots,\alpha_M} H(f,\alpha), \\ &s.t. \quad \alpha_g > 0, \quad \sum_{g=1}^{M}\alpha_g = 1. \end{aligned} \quad (11)$$

The optimization strategy is to learn f and α_g simultaneously by fixing one and optimizing another for each iteration. We first fix α and optimize f.

The partial derivative of the objective function with respect to f is

$$\frac{\partial H(f,\alpha)}{\partial f} = \sum_{g=1}^{M}\alpha_g^r L_g f + \lambda(f - y). \quad (12)$$

By setting $((\partial H(f,\alpha))/\partial f) = 0$, we have

$$f = \left(I + \frac{\sum_{g=1}^{M}\alpha_g^r L_g}{\lambda}\right)^{-1} y. \quad (13)$$

Then, we fix f to optimize α_g. The partial derivative of the objective function with respect to α is

$$\frac{\partial H(f,\alpha)}{\partial \alpha_g} = r\alpha_g^{r-1} f^T L_g f. \quad (14)$$

Similarly, we can obtain α_g with Lagrange multipliers because $\sum_{g=1}^{M}\alpha_g = 1$.

$$\alpha_g = \frac{\left(1/\left(f^T L_g f\right)\right)^{(1/(r-1))}}{\sum_{g=1}^{M}\left(1/\left(f^T L_g f\right)\right)^{(1/(r-1))}}. \quad (15)$$

4. Data and Variable

4.1. Data. To validate the effectiveness of the multiview graph-based learning method in forecasting SMEs' credit risk in SCF, we construct a proper database of Chinese SME. SCF is a new and developing branch of financing. Because of imperfection of the theoretical study, development of SCF in China is less than satisfactory. Until recent years, SCF business is gradually on the right track, making the supporting businesses and data disclosure relatively perfect. However, it is still difficult to gather a complete dataset of SCF, especially unlisted companies. Therefore, the listed companies are selected as the main data source.

Due to the huge difference in factors and characteristics of different industries, it may lead to the decline of the prediction model by using the whole industry data. Manufacturing industry is the key field of SCF, which attracted widely attention of researchers. However, traditional manufactured products (e.g., auto industry) were so complex, which has long chains from suppliers to distributors, that thorough collection of SCF data was impossible. Compared with traditional manufacturing, the mobile manufacturing supply chain data scale is moderate (e.g., a mobile phone has more than 80 parts, while a car has tens of thousands of parts). The division of labor among upstream suppliers, midstream manufacturers, and downstream distributors in mobile phone manufacturing industry is relatively clear. Furthermore, mobile phones are time sensitive, which requires enterprises on the chain to cooperate closely. Taking all these into account, we select 104 quoted SMEs data in Wind dataset. After deleting invalid data, our final dataset consists of 924 samples over the period (2011–2019).

Some literatures [26, 32] identified credit risk by considering whether the enterprise is special treatment (ST). They held that the ST and ∗ST companies are in financial crisis [33]. These companies may have higher credit risks than the general ones. ST company has suffered losses for two consecutive years. However, it cannot reflect the credit risk when a company suffered losses in the first year. The Z-score model [34] is adopted to define the credit risk, which is widely used in enterprise financial health measurement. Finally, the samples, whose Z-score < 1.8, are considered as risk enterprises in the experiments, while the remaining samples are nonrisk type. In total, the dataset consists of 265 risky enterprises and 659 nonrisky enterprises. In the experimental settings, 40 samples are treated as known risky enterprises denoted by −1 and 40 are known as nonrisky enterprises denoted by +1. The rest unknown samples (denoted by 0) need to be validated by the multiview graph-based learning. The sample distribution is given in Table 1.

4.2. Independent Variables. In compliance with the widely accepted 5C principle, the credit risk assessment system is constructed by three views, such as supply chain operation, SME financial indicator, and nonfinancial indicator. To facilitate the observation, the descriptions of the independent variables for supply chain operation, SME financial indicator, and nonfinancial indicator are shown in Tables 2–4, respectively.

4.3. Evaluations. To evaluate the effectiveness of proposed methods, average accuracy, Type I error, and Type II error are adopted in the experiments. These evaluation criteria are defined as follows:

Table 1: Sample distribution of dataset.

	Dataset	Labeled (training)	Unlabeled (testing)
Risky	265	40 (−1)	225 (0)
Nonrisky	659	40 (+1)	619 (0)

Table 2: List of independent variables for supply chain operations (V_1).

Indexes	Variables	Mean	Std. Deviation
C_1	Customer concentration	27.78245	28.70553
C_2	Supplier concentration	19.9285	23.77576
C_3	CE current ratio	0.71919	0.73108
C_4	CE quick ratio	0.51692	0.52582
C_5	CE return on net assets	1.90677	2.52005
C_6	CE net profit rate of sales	2.05564	2.28251
C_7	CE asset liability ratio	26.04776	17.08954
C_8	CE inventory turnover days	45.55526	37.55354
C_9	CE accounts receivable turnover	3.68589	3.02582

∗CE represents the core enterprise in supply chain

Table 3: List of independent variables for supply chain operations (V_2).

Indexes	Variables	Mean	Std. Deviation
C_{10}	EBIT	44186.84275	195438.89078
C_{11}	Gross profit	116784.07683	489589.42188
C_{12}	Net interest rate of total assets	6.14171	11.5233
C_{13}	Retained earnings	109346.28328	476524.70955
C_{14}	Cash turnover	953.97242	2781.79251
C_{15}	Inventory turnover days	85.66379	140.64208
C_{16}	Accounts receivable turnover	13.29238	98.98973
C_{17}	Turnover rate of current assets	1.26009	0.88023
C_{18}	Turnover of fixed assets	28.04207	146.34457
C_{19}	Asset liability ratio	38.03026	25.63465
C_{20}	Quick ratio	1.836	2.5333
C_{21}	Interest cover ratio	170.74379	2559.79827
C_{22}	Current ratio	2.22639	2.70785

Table 4: List of independent variables for supply chain operations (V_3).

Indexes	Variables	Mean	Std. Deviation
C_{23}	Total R&D expenditure	22777.54545	82601.12925
C_{24}	Market value of the company	88.86682	242.42438
C_{25}	Enterprise value	9.64595	58.71563
C_{26}	Staff quality	15.19087	21.52928
C_{27}	GDP	10.505	2.51509
C_{28}	CPI	2.589	1.12239
C_{29}	Consumer spending	25.60751	7.66701

$$\text{Average Accuracy} = \frac{N}{M},$$
$$\text{type I error} = \frac{E_p}{M_p}, \tag{16}$$
$$\text{type II error} = \frac{E_n}{M_n},$$

here M is the total number of samples in the testing set, M_p and M_n are the number of positive and negative samples, respectively ($M = M_p + M_n$), N is the number of all correctly predicted samples, and E_p and E_n are the incorrectly predicted positive and negative samples.

Average accuracy is used to evaluate the overall performance of the model. The higher, the better. Sometimes, the imbalance between positive and negative samples may lead an inaccurate evaluation. Type I and Type II errors are also adopted as a supplement to evaluate the performance of the model. The lower, the better. Type I error describes the accuracy of positive samples. In SME credit risk assessment, Type II error with a high ratio will result in a loss of potential customers, which incorrectly classify nonrisky enterprises into risky. Type II error presents the accuracy of negative samples. Incorrectly classifying risky enterprises into nonrisky will lead a high ratio of Type II error, which may expose great risk on bank, core enterprises, and SMEs in SCF. In order to build a stable SCF system, Type II error is more vital than the other two criteria.

4.4. Experimental Settings. The experiments are performed on a computer which has Intel Xeon (R) 2.13 GHz 8 processors 8 GB RAM and the 64 bit Windows10 system. The experiments are conducted based on Python, NumPy, and Scikit-learn. The data are firstly normalized by the following:

$$X_{\text{norm}} = \frac{X - \mu}{\sigma}, \tag{17}$$

where X_{norm} is the normalized data, X is the source data, and μ and σ are the mean value and standard deviation of X. Then, principal component analysis (PCA) is performed on X_{norm} for dimension reduction. By preserving 90% energy, 18-dimensional features are prepared for LR, SVM, and GL. Particularly for M-GL, PCA is performed on each view separately. For GL and M-GL, a vertex x_i is linked with its nearest 5 neighbors and the Euclidean distance is used for measurement. The parameter λ in equation (6) and equation (8) is set to 0.01, and r in equation (8) is set to 1.9. The initial weights α_g for the three views are set equally.

5. Results and Discussion

5.1. Experimental Settings. In this part, we compare the proposed multiview graph learning for SME credit risk assessment in SCF with several related works.

(1) *LR*. Logistic regression [35] is performed on the processed features from all the three views.

(2) *SVM*. Support vector machine is performed on the processed features from all the three views.

(3) *GL*. Single graph learning is performed by using the same features as LR and SVM. Only one graph is constructed by using the processed features.

(4) *M-GL*. The proposed multiview graph-based learning approach integrates multiview features individually, which can adaptively assign different weights for the three views. Three graphs are constructed by using the processed features from the three views correspondingly.

Figures 1–3 describe the obtained results for the four models when the number of negative and positive training samples is 40, respectively. Upon inspecting the results in Figures 1–3, we notice that the proposed M-GL performs best on all the three evaluation criteria for SME credit risk assessment in SCF. It demonstrates the superiority of the proposed M-GL. From the perspective of feature fusion, all methods listed above integrate the features from the three views. However, M-GL treats these features as heterogeneous elements, which assigns different weights of each view for credit risk assessment. The other works treat the three views as a whole. They cannot distinguish which view is more valuable for SME credit risk assessment in SCF. Though some works (e.g., LR) can obtain the most important independent variables, the models depend too much on these features, leading a high data sensibility. On the contrary, graph-based methods (GL and M-GL) exploit the smoothness between samples. The loss or inaccuracy of one feature has little influence on the assessment model. During data collection, not all indicators can be easily obtained for every SME.

The noise resistance is an important characteristic in practical financial applications. In the next subsection, we will investigate the impact of different financial views for SME credit risk assessment in SCF.

5.2. The Impact of Different Financial Views. For traditional SME credit risk assessment methods, it is difficult to make a quantitative investigation on which one is the most informative view on risk assessment. They either concentrate the features as a whole or treat each indicator individually. Few works [32] give discriminative analysis on each indicator. However, the effectiveness of single indicator cannot provide evidences to guarantee the overall effectiveness of the corresponding view. The proposed M-GL constructs three individual subgraphs by using the features from supply chain operation, SME financial indicator, and nonfinancial indicator. By means of novel graph construction and optimization method, M-GL can adaptively distinguish the importance of different views. The upside of this kind of SME credit risk assessment is suitable for building the secure and stable SCF system. For example, the SME can improve their credit artificially by tampering a certain financial data to deceive the bank or core enterprises. It is of great danger for the whole SCF system. Tampering with a small amount of data will not affect the risk assessment results obviously for

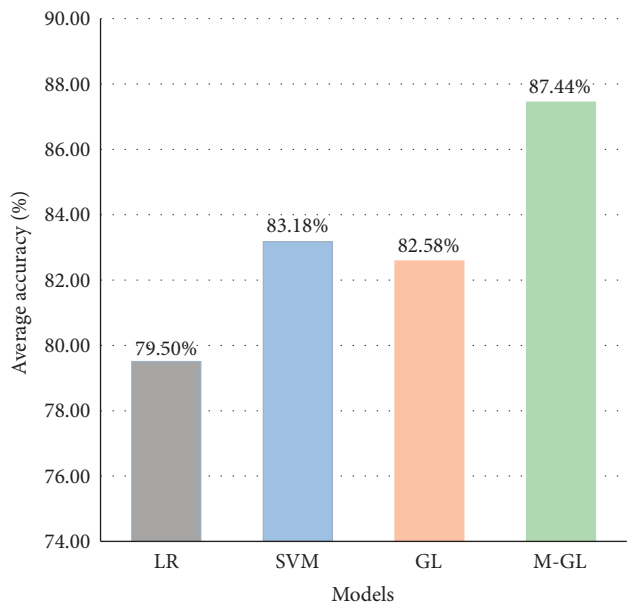


Figure 1: Average accuracy for different methods.

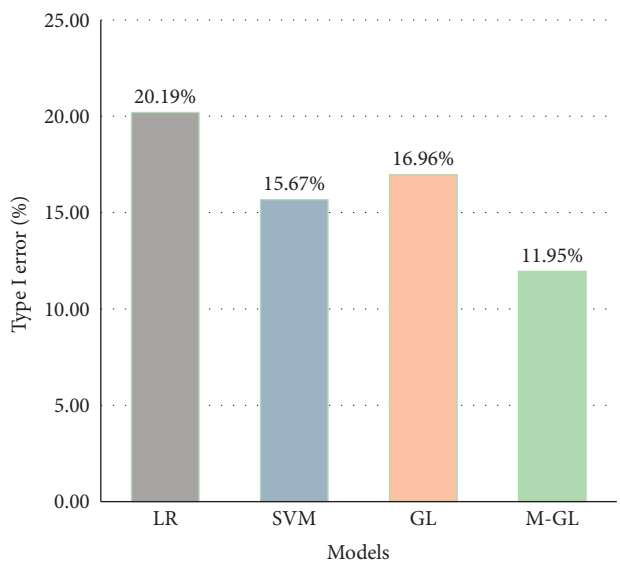


Figure 2: Type I error for different methods.

the proposed M-GL, which is one of the strengths of our approach. Table 5 gives the learned weight coefficient α_g for each view. α_1, α_2, and α_3 are the learned weight coefficients for supply chain operation, SME financial indicator, and nonfinancial indicator, respectively. From Table 5, we can see that SME financial indicator is the main factor for SME credit risk assessment in SCF. However, this does not mean that supply chain relationship is useless. The view of supply chain relationship is also an effective supplement, which is also informative to assess the SME credit risk assessment in SCF. In contrast, nonfinancial indicator has the lowest value among the three views. Partly because we investigate the mobile manufacturing supply chain data, the difference of influencing factors between different industries have been weakened during data collection phase.

5.3. The Impact on the Number of Labeled Samples. The proposed M-GL is a graph-based transductive classification. An advantage of transductive methods is that it may be able to make better predictions with fewer labeled samples. In this part, we investigate the performance for different

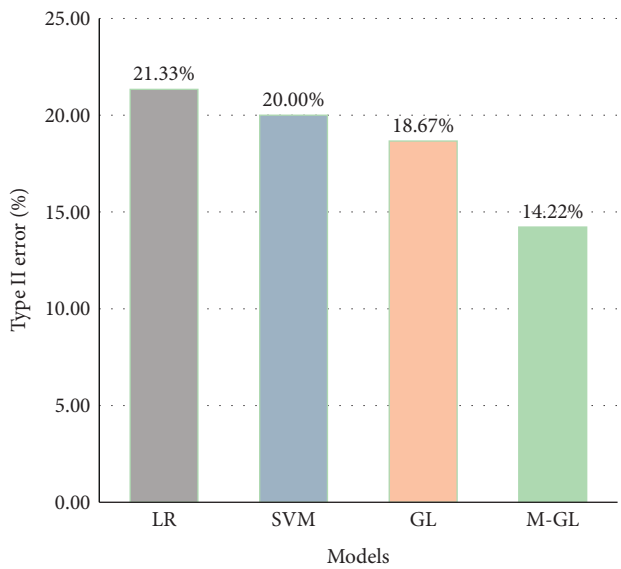


Figure 3: Type II error for different methods.

Table 5: The impact of different financial views.

α_1	α_2	α_3
0.284	0.598	0.118

amounts of initial labeled samples. Figure 4 reports the average accuracy, and Figures 5 and 6 report Type I and Type II errors of SVM and M-GL with varying amounts of positive and negative samples. M-GL can be regarded as label diffusion processes through three subgraphs. A good graph method needs to propagate labels to unlabeled samples. It is observed that M-GL outperforms SVM throughout the experiments. Especially, when the number of labeled samples is 10, the advantage of M-GL is more apparent than SVM on the three evaluation criteria. In the practical financial forecasting, labeling samples are always inconvenient, especially the risky SME. It has a great influence on traditional methods when the labeled samples are insufficient. Nevertheless, we can obtain lots of unlabeled SME information. M-GL models the relationship of these few labeled samples and large amounts of unlabeled samples through label diffusion processes on the smooth manifold for SME credit risk forecasting. Exploiting the large amounts of unlabeled samples to improve the accuracy of credit risk forecasting is also an advantage of M-GL. In the next subsection, we will investigate the influence of smoothness on the classifier's performances.

5.4. The Impact on the Number of Nearest Neighbors. In our experiment, the number of nearest neighbors k is set to 5 in graph construction. During label diffusion, we need a smooth graph. A proper value of k will lead a stable and excellent performance. A large or a small k will degrade the performance. When k is small, it fails to capture the relationship of samples. On the contrary, when k is large, mismatched vertices may be brought into the graph. This explains the trends of curves shown in Figures 7–9. After k reaches an appropriate value, the performances begin to decline.

Because of the smoothness of the graph, we can obtain acceptable results through label diffusion. When we need to select partners from some SMEs, M-GL can give an appropriate credit ranking among the candidate enterprises.

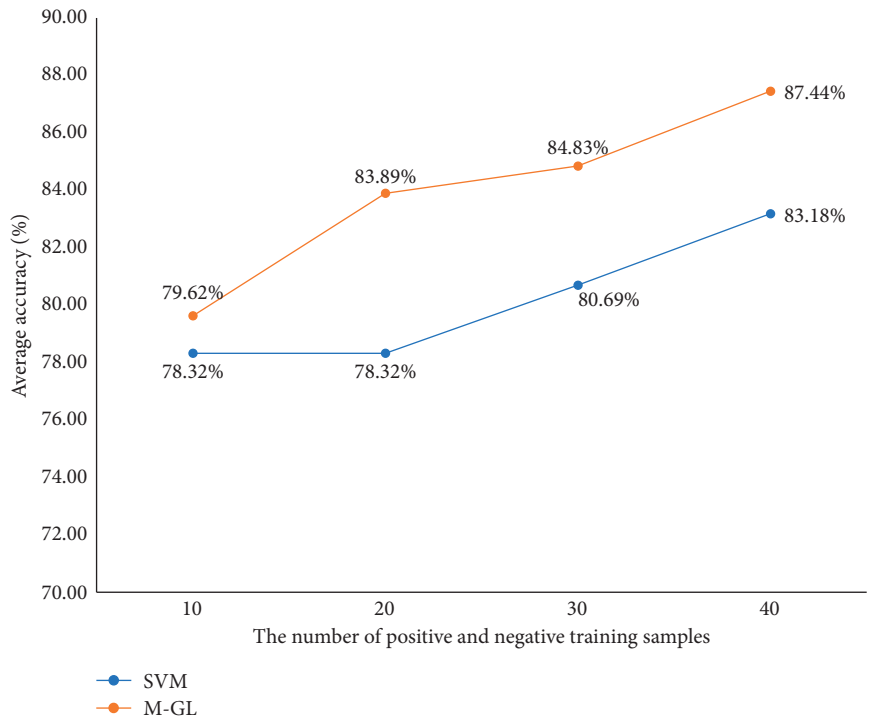


Figure 4: Variations of average accuracy with the increased number of labeled samples for M-GL.

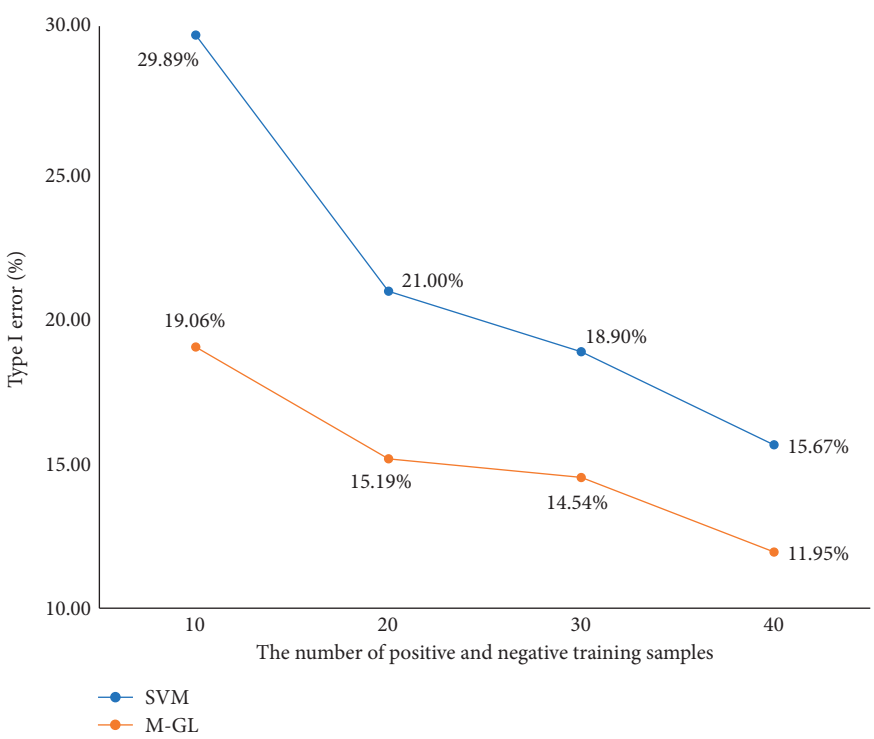


Figure 5: Variations of Type I error with the increased number of labeled samples for M-GL.

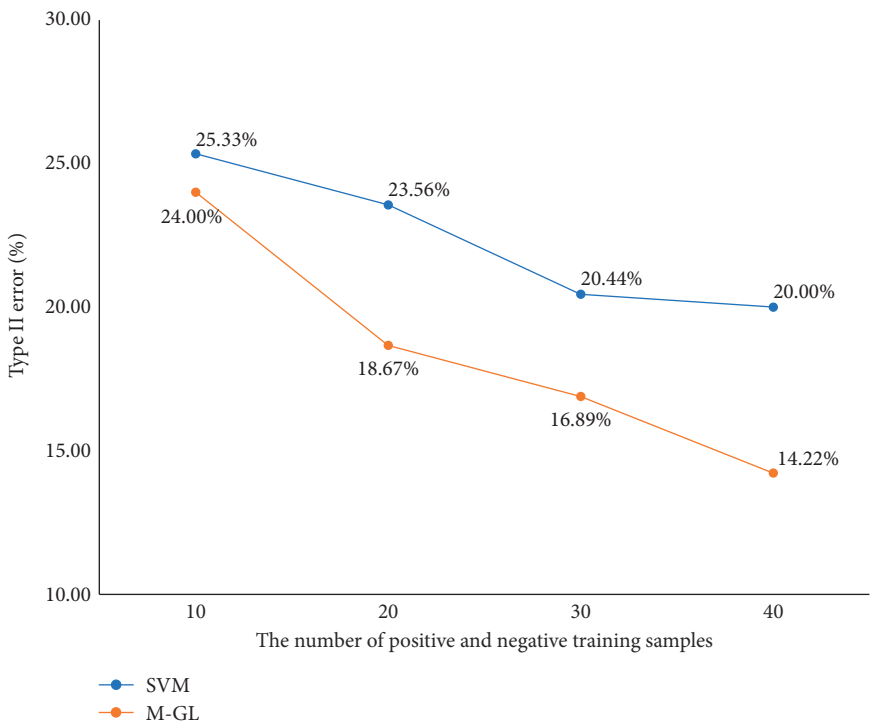


Figure 6: Variations of Type II error with the increased number of labeled samples for M-GL.

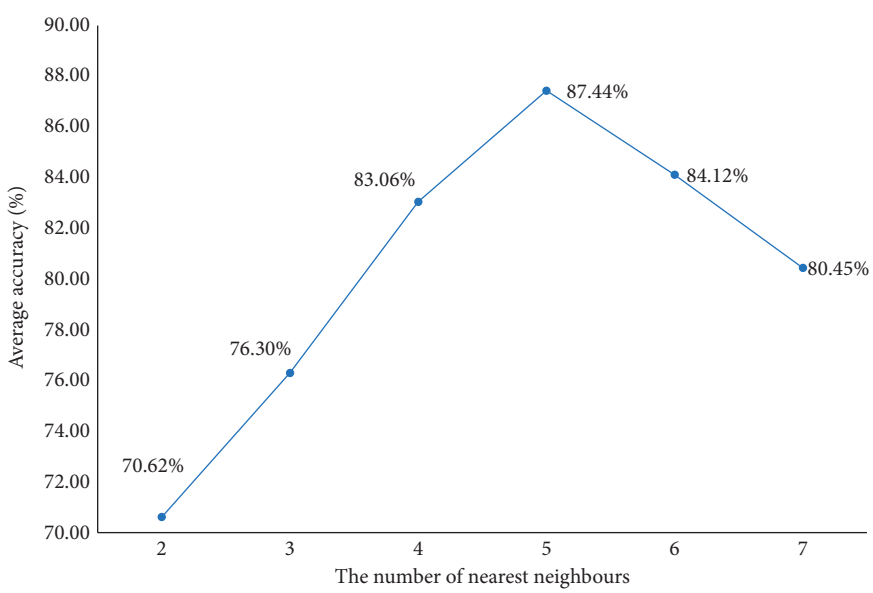


Figure 7: Variations of average accuracy with the increased number of nearest neighbors for M-GL.

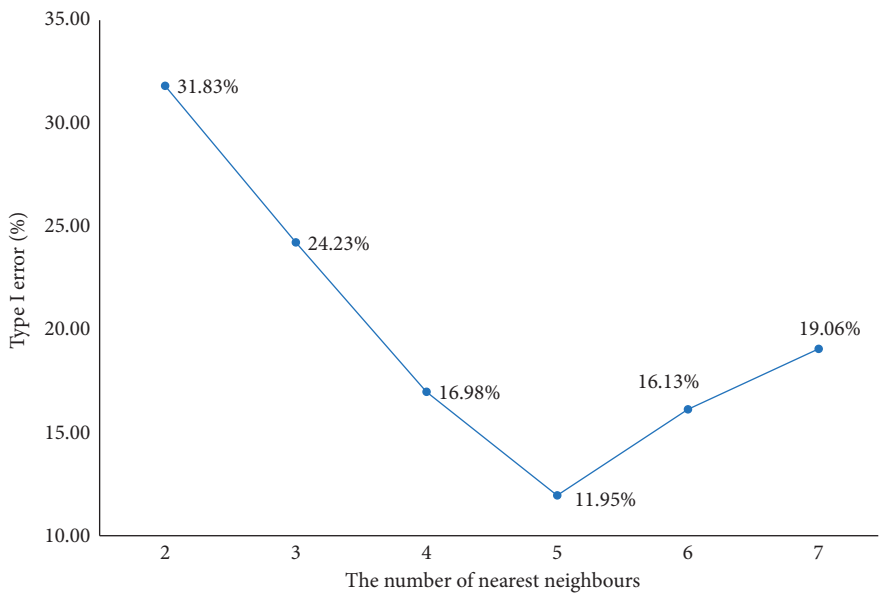


Figure 8: Variations of Type I error with the increased number of nearest neighbors for M-GL.

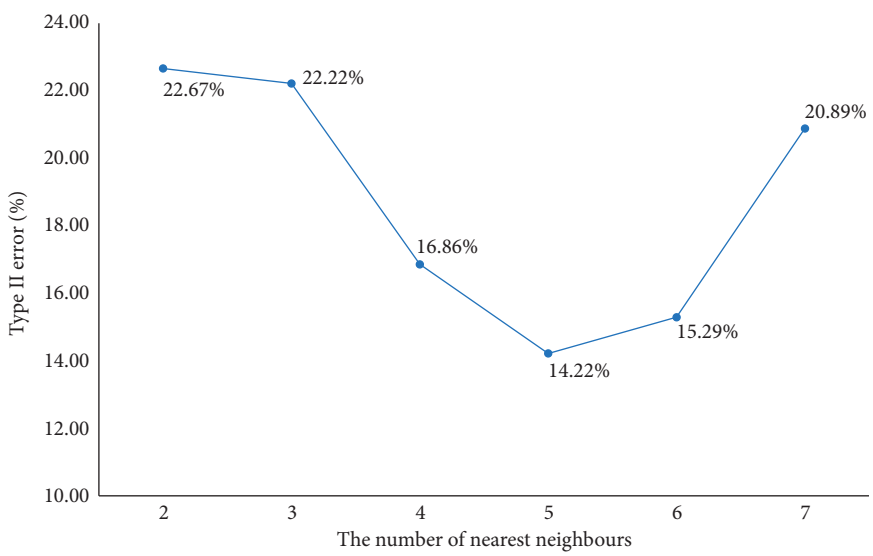


Figure 9: Variations of Type II error with the increased number of nearest neighbors for M-GL.

6. Conclusions

The primary purpose of the paper is to assess SME credit risks in SCF by means of multiview graph-based learning. The credit risk evaluation system is constructed from three aspects: supply chain operation, SME financial indicator, and nonfinancial indicator. Different from traditional methods, the proposed methods treat the three views individually, aiming to give a quantitative comparison for which view is more important for credit risk assessment. The paper also provides a detailed analysis on the impact of different financial views for SME credit risk assessment in SCF. To sum up, M-GL shows its effectiveness to assess SME credit risks in SCF. Especially, M-GL has good performance in dealing with few labeled samples. In addition, SME financial indicator is the most critical view for SME credit risk assessment in SCF. Supply chain operation is a supplementary view, which is also a beneficial view to enhance the performance of assessment. Furthermore, M-GL not only uses the features from the three views but also exploits the relationships among samples. Therefore, the proposed M-GL for SME credit risk assessment is robust, which is applicable to prevent credit fraud by tampering with few data. When the bank or core enterprise needs to pick a partner, M-GL could give a relative sort order from the candidate enterprises.

Machine learning methods have shown their effectiveness in credit risk assessment. However, assessing SME credit risk in SCF is still a real challenge. Data

collection and annotation is a natural barrier to apply the new machine learning method for SME credit risk assessment in SCF, such as deep learning approaches. Moreover, mobile manufacturing supply chain data are utilized in this paper, where differences between industries are neglected. In the future, we will investigate SME credit risk assessment in SCF with a large dataset for comprehensive analysis.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (no. 18cx04007B), Qingdao Social Science Planning Project (no. QDSKL2001044), and Shandong Social Science Planning Research Project (no. 18CJJJ13).

References

[1] W. A. Abbasi, Z. Wang, and A. Alsakarneh, "Overcoming smes financing and supply chain obstacles by introducing supply chain finance," *HOLISTICA-Journal of Business and Public Administration*, vol. 9, no. 1, pp. 7–22, 2018.

[2] E. Letizia and F. Lillo, "Corporate payments networks and credit risk rating," *EPJ Data Science*, vol. 8, no. 1, p. 21, 2019.

[3] R. C. Basole, M. A. Bellamy, H. Park, and J. Putrevu, "Computational analysis and visualization of global supply network risks," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1206–1213, 2016.

[4] Z. Zhao, D. Chen, L. Wang, and C. Han, "Credit risk diffusion in supply chain finance: a complex networks perspective," *Sustainability*, vol. 10, no. 12, p. 4608, 2018.

[5] E. I. Altman and G. Sabato, "Modelling credit risk for SMEs: evidence from the U.S. Market," *Abacus*, vol. 43, no. 3, pp. 332–357, 2007.

[6] S. D. Lekkakos, A. Serrano, and A. E. Ellinger, "Supply chain finance for small and medium sized enterprises: the case of reverse factoring," *International Journal of Physical Distribution & Logistics Management*, vol. 46, 2016.

[7] D. A. Wuttke, C. Blome, H. Sebastian Heese, and M. Protopappa-Sieke, "Supply chain finance: optimal introduction and adoption decisions," *International Journal of Production Economics*, vol. 178, pp. 72–81, 2016.

[8] Q. An and Y. Zhao, "Construction of credit risk evaluation system for small and medium-sized enterprises: based on principal component analysis and logistic model," in *Proceedings of the 5th International Conference on Economics, Management, Law and Education (EMLE 2019)*, pp. 244–250, Atlantis Press, Krasnodar, Russia, October 2020.

[9] M. Yurdakul, "AHP as a strategic decision-making tool to justify machine tool selection," *Journal of Materials Processing Technology*, vol. 146, no. 3, pp. 365–376, 2004.

[10] R. Rostamzadeh, M. K. Ghorabaee, K. Govindan, A. Esmaeili, and H. B. K. Nobar, "Evaluation of sustainable supply chain risk management using an integrated fuzzy TOPSIS- CRITIC approach," *Journal of Cleaner Production*, vol. 175, pp. 651–669, 2018.

[11] W. Mou, W.-K. Wong, and M. McAleer, "Financial credit risk evaluation based on core enterprise supply chains," *Sustainability*, vol. 10, no. 10, p. 3699, 2018.

[12] Y. Liu and L. Huang, "Supply chain finance credit risk assessment using support vector machine–based ensemble improved with noise elimination," *International Journal of Distributed Sensor Networks*, vol. 16, no. 1, Article ID 1550147720903631, 2020.

[13] H. Jiang, W.-K. Ching, K. F. C. Yiu, and Y. Qiu, "Stationary mahalanobis kernel svm for credit risk evaluation," *Applied Soft Computing*, vol. 71, pp. 407–417, 2018.

[14] D. Fantazzini and S. Figini, "Default forecasting for small-medium enterprises: does heterogeneity matter?" *International Journal of Risk Assessment and Management*, vol. 11, no. 1-2, pp. 138–163, 2009.

[15] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[16] H.-q. Hu, L. Zhang, D.-h. Zhang, and L. CHEN, "Research on finance credit risk assessment of supply chain based on svm," *Soft Science*, vol. 5, pp. 1–7, 2011.

[17] P. Danenas and G. Garsva, "Selection of support vector machines based classifiers for credit risk domain," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3194–3204, 2015.

[18] J. Zhou and T. Bai, "Credit risk assessment using rough set theory and ga-based SVM," in *Proceedings of the 2008 The 3rd International Conference on Grid and Pervasive Computing-Workshops*, pp. 320–325, IEEE, Kunming, China, May 2008.

[19] A. Khashman, "Neural networks for credit risk evaluation: investigation of different neural models and learning schemes," *Expert Systems with Applications*, vol. 37, no. 9, pp. 6233–6239, 2010.

[20] C. Luo, D. Wu, and D. Wu, "A deep learning approach for credit scoring using credit default swaps," *Engineering Applications of Artificial Intelligence*, vol. 65, pp. 465–470, 2017.

[21] N. Ghatasheh, "Business analytics using random forest trees for credit risk prediction: a comparison study," *International Journal of Advanced Science and Technology*, vol. 72, pp. 19–30, 2014.

[22] P. Giudici, B. Hadji-Misheva, and A. Spelta, "Network based credit risk models," *Quality Engineering*, vol. 32, no. 2, pp. 199–211, 2020.

[23] L. Zhang, H. Hu, and D. Zhang, "A credit risk assessment model based on svm for small and medium enterprises in supply chain finance," *Financial Innovation*, vol. 1, no. 1, p. 14, 2015.

[24] R.-Z. Xu and M.-K. He, "Application of deep learning neural network in online supply chain financial credit risk assessment," in *Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA)*, pp. 224–232, IEEE, Guiyang, China, April 2020.

[25] H. Haiqing, Z. Lang, and Z. Daohong, "Research on smes credit risk assessment from the perspective of supply chain finance-a comparative study on the svm model and bp model," *Management Review*, vol. 11, 2016.

[26] Y. Zhu, C. Xie, G.-J. Wang, and X.-G. Yan, "Predicting China's SME credit risk in supply chain finance based on machine learning methods," *Entropy*, vol. 18, no. 5, p. 195, 2016.

[27] W. Liu, J. Wang, and S.-F. Chang, "Robust and scalable graph-based semisupervised learning," *Proceedings of the IEEE*, vol. 100, no. 9, pp. 2624–2638, 2012.

[28] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, "Image annotation via graph learning," *Pattern Recognition*, vol. 42, no. 2, pp. 218–228, 2009.
[29] Z. Kang, H. Pan, S. C. Hoi, and Z. Xu, "Robust graph learning from noisy data," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1833–1843, 2019.
[30] M. Culp and G. Michailidis, "Graph-based semisupervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 1, pp. 174–179, 2007.
[31] L. Wang, W. Sun, Z. Zhao, and F. Su, "Modeling intra- and inter-pair correlation via heterogeneous high-order preserving for cross-modal retrieval," *Signal Processing*, vol. 131, pp. 249–260, 2017.
[32] Y. Zhu, L. Zhou, C. Xie, G.-J. Wang, and T. V. Nguyen, "Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach," *International Journal of Production Economics*, vol. 211, pp. 22–33, 2019.
[33] X. Chen, X. Wang, and D. D. Wu, "Credit risk measurement and early warning of smes: an empirical study of listed smes in China," *Decision Support Systems*, vol. 49, no. 3, pp. 301–310, 2010.
[34] E. I. Altman, "Predicting financial distress of companies: revisiting the z-score and zeta models," in *Handbook of Research Methods and Applications in Empirical Finance*Edward Elgar Publishing, Cheltenham, EN, UK, 2013.
[35] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, Vol. 398, John Wiley & Sons, Hoboken, NJ, USA, 2013.

Hindawi
Complexity
Volume 2021, Article ID 6640781, 8 pages
https://doi.org/10.1155/2021/6640781

Research Article

An Infeasible Incremental Bundle Method for Nonsmooth Optimization Problem Based on CVaR Portfolio

Jia-Tong Li,[1] Jie Shen ,[2] and Na Xu[2]

[1]*College of Science, Northeast Forestry University, Harbin 150040, China*
[2]*School of Mathematics, Liaoning Normal University, Dalian 116029, China*

Correspondence should be addressed to Jie Shen; tt010725@163.com

Received 12 November 2020; Revised 30 November 2020; Accepted 28 December 2020; Published 18 January 2021

Academic Editor: Abdelalim Elsadany

Copyright © 2021 Jia-Tong Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For CVaR (conditional value-at-risk) portfolio nonsmooth optimization problem, we propose an infeasible incremental bundle method on the basis of the improvement function and the main idea of incremental method for solving convex finite min-max problems. The presented algorithm only employs the information of the objective function and one component function of constraint functions to form the approximate model for improvement function. By introducing the aggregate technique, we keep the information of previous iterate points that may be deleted from bundle to overcome the difficulty of numerical computation and storage. Our algorithm does not enforce the feasibility of iterate points and the monotonicity of objective function, and the global convergence of the algorithm is established under mild conditions. Compared with the available results, our method loosens the requirements of computing the whole constraint function, which makes the algorithm easier to implement.

1. Introduction

Optimization problems arise in the wide range of practical applications, and they have been successfully solved by utilizing various methods, especially by state of the art approaches [1–3]. For an actual engineering optimization problem [2], a new optimal mutation strategy based on the complementary advantages of five mutation strategies is designed to develop a novel improved differential evolution algorithm with the wavelet basis function; the proposed method can improve the search quality, accelerate convergence and avoid fall into local optimum and stagnation. Parametric analysis and optimization are conducted for a novel geothermal combined cooling and power system [3], and not only the combined system performs better than the separate system but also the n-nonane brings the lowest total product unit cost to the proposed system. Nonsmooth optimization (NSO) problems are in general difficult to solve. Lots of approaches are proposed to solve these problems [4–8]. Among others, bundle methods are considered as one of the most efficient and promising methods. Infeasible bundle methods [9,10] can be viewed as the unconstrained proximal-like bundle methods applied to improvement functions, and the main advantage superior to other methods is that it does not require the feasibility of the iterate points and the monotonicity of the objective function. Conditional value-at-risk (CVaR) is currently the main tool to measure financial risk when we face portfolio for selected risky assets, and the study of CVaR model usually brings about the following nonsmooth optimization problem:

$$\begin{aligned} &\min_{(x,\alpha)\in R^n\times R} \quad F_\theta(x,\alpha) = \alpha + \frac{1}{(1-\theta)J}\sum_{j=1}^{J}\left(-x^T y_j - \alpha\right)^+, \\ &\text{s.t,} \quad -x^T m \le -M, \sum_{i=1}^{n} x_i = 1, x \ge 0, \end{aligned} \tag{1}$$

where $F_\theta(x,\alpha)$ is the approximate performance function which is convex and $\theta \in (0,1)$ is the probability level, $Y_{J\times n} = (y_{ji})$ is the scenario matrix, J is the number of scenarios, and its element y_{ji} denotes the yield rate of risky asset under scenario j. The vector $y_j = (y_{j1}, y_{j2}, \ldots, y_{jn})^T$ is the vector of yield rate of scenario, j. $(\cdot)^+ = \max\{0,\cdot\}$. The vector $y \in R^n$

denotes the yield rate at the end of investment which is uncertain, and $m = E(y)$, the constant M is the given rate of expected return. We rewrite problem (1) in more general form:

$$\begin{aligned} &\min_{(x,\alpha)\in R^n\times R} \quad F_\theta(x,\alpha), \\ &\text{s.t} \quad c(x,\alpha)\le 0, \end{aligned} \tag{2}$$

where $c(x,\alpha) = \max\left\{c_j(x,\alpha)|, tjn \in q\widetilde{J}h =_{\{1,2,\ldots,n+3\}}\right\}$ is the pointwise maximum of finite many convex functions, $c_1(x,\alpha) = -x^Te_1$, $c_2(x,\alpha) = -x^Te_2, \ldots,$ $c_n(x,\alpha) = -x^Te_n$, $c_{n+1}(x,\alpha) = -x^Tm + M$, and $c_{n+2}(x,\alpha) = x^Te - 1$, $c_{n+3}(x,\alpha) = -x^Te + 1$ are called the component functions of c, and $e_i = (0,\ldots,1^{\text{ith}},\ldots,0)^T \in R^n, e = (1,1,\ldots,1)^T \in R^n$. Obviously, function $c\colon R^{n+1} \longrightarrow R$ is convex and nondifferentiable.

Problem (2) involves a special constraint function, namely, the max-valued function. Qing-Ye Zhang and Yan Gao [11] once transform the objective function in the CVaR model into a piecewise smooth function and present an algorithm by employing common proximal bundle methods, where the values of every component function need to be evaluated at each iteration. An incremental method for solving convex finite min-max problem provides us a new approach to deal with max-valued function, which extends the philosophy of the incremental approaches [12], and it does not need to evaluate the actual value of max-valued function. The idea has already been applied to linearly constrained min-max problems [13] and to inequality constraint min-max problems [14]. The algorithms provided by the above two papers are feasible and descent methods. But in some cases, the feasibility of iterate points is difficult to realize since we have to make a tradeoff between the search for feasibility and for the reduction of the objective function.

Motivated by the work [14, 15], we provide an iterate method which is strongly connected with the incremental technique [12] and the infeasible idea [9]. The algorithm we design does not require the evaluation of each component function of the constraint function. In other words, we employ the incomplete knowledge of the constraint function to construct a lower approximate model for improvement function associated with the original problem, which reduces the number of elements in bundle and decreases the difficulty of implementation of the algorithm. Based on the approximate model, a descent test is presented. Even though the model is a more rough approximation to improvement function, the proposed algorithm still possesses good convergence properties under mild conditions.

This paper is organized as follows: In Section 2, we introduce basic notations, concepts, and existing results which are the basis for the construction of approximate model and the design of overall algorithm. A new cutting-planes framework and the way to update the elements in bundle are also given in this section. Section 3 presents an infeasible incremental bundle algorithm for solving problem (2). The convergence analysis of the presented algorithm is discussed in Section 4, and the optimal solution to problem (2) is obtained under mild conditions.

2. Preliminaries and the Construction of Subproblem

We denote by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ the usual inner product and norm in R^n, respectively. The subdifferential of a convex function $f\colon R^n \longrightarrow R$ at x is defined by $\partial f(x) = \{p \in R^n | f(y) \ge f(x) + \langle p, y - x\rangle, \forall y \in R^n\}$. Let

$$D = \left\{(x,\alpha) \in R^{n+1} | c(x,\alpha) \le 0\right\} \tag{3}$$

be the feasible set of problem (2), and for sake of the notation simplicity, we leave out the confidence parameter θ in (2) and indicate $F_\theta(x,\alpha)$ by $F(x,\alpha)$.

Lemma 1. *Suppose that $f_i(x)$ $(i \in I)$ is the convex differentiable function and the index set I is finite. Then, the subdifferential of function $f(x) = \max\{f_i(x)|it \in nI\}$ is $\partial f(x) = co\{\nabla f_i(x)|it \in nIq(x)\}$, where "co" denotes the convex hull, $I(x) = \{i \in I|f_it(x)n = qfh(x)\}$.*

Definition 1. For given $(y,\beta) \in R^{n+1}$, the following function,

$$H_{(y,\beta)}(x,\alpha) = \max\{F(x,\alpha) - F(y,\beta), c(x,\alpha)\}, \quad (x,\alpha) \in R^{n+1}, \tag{4}$$

is called the improvement function associated with problem (2).

The following lemmas present some useful properties of improvement function which play an important role in designing our algorithm.

Lemma 2. *[9] The subdifferential of improvement function $H_{(y,\beta)}(\cdot,\cdot)$ at point (x,α) is given by*

$$\partial H_{(y,\beta)}(x,\alpha) = \begin{cases} \partial F(x,\alpha), & \text{if } F(x,\alpha) - F(y,\beta) > c(x,\alpha), \\ \operatorname{conv}\{\partial F(x,\alpha) \cup \partial c(x,\alpha)\}, & \text{if } F(x,\alpha) - F(y,\beta) = c(x,\alpha), \\ \partial c(x,\alpha), & \text{if } F(x,\alpha) - F(y,\beta) < c(x,\alpha). \end{cases} \tag{5}$$

Lemma 3. *[16] Suppose that the Slater constraint qualification is satisfied for problem (2), the following statements are equivalent:*

(i) $(\overline{x}, t\overline{\alpha})$ is a solution to problem (2)

(ii) $\min\left\{H_{(\overline{x},t\overline{\alpha})}(y,\beta)|(y,\beta) \in R^{n+1}\right\} = H_{(\overline{x},t\overline{\alpha})}(\overline{x}, t\overline{\alpha})$

(iii) $0 \in \partial H_{(\overline{x},t\overline{\alpha})}(\overline{x},t\overline{\alpha})$, *i.e.,* $0 \in \partial\phi(\overline{x},t\overline{\alpha})$, *where* $\phi(\cdot,\cdot) = H_{(\overline{x},t\overline{\alpha})}(\cdot,\cdot)$

Suppose that (x^k,α^k) is the current stability center (the current last serious step) and (y^i,β^i) is the trial point generated from previous iterations. We indicate $H_{(x^k,\alpha^k)}(\cdot,\cdot)$ by $H_k(\cdot,\cdot)$. Choose arbitrarily one component function $c_j$ofc and define the values of the linearization functions of $F$and c_j at point (x^k,α^k):

$$\begin{aligned} s_{F_i}^k &= F\left(y^i,\beta^i\right) + \langle g_F^i, \left(x^k,\alpha^k\right) - \left(y^i,\beta^i\right)\rangle, \\ t_{c_{ji}}^k &= c_j\left(y^i,\beta^i\right) + \langle g_{c_j}^i, \left(x^k,\alpha^k\right) - \left(y^i,\beta^i\right)\rangle, \end{aligned} \tag{6}$$

where $g_F^i \in \partial F(y^i,\beta^i)$, $g_{c_j}^i \in \partial c_j(y^i,\beta^i)$. Bundle B_l keeps memory of the information of previous iterations:

$$\begin{aligned} &B_l = B_l^1 \cup B_l^2 \cup \left\{\left(x^k,\alpha^k\right), F\left(x^k,\alpha^k\right), c\left(x^k,\alpha^k\right)\right\}, k = k(l), \text{ with} \\ &B_l^1 \subseteq \cup_{i<l}\left\{F\left(y^i,\beta^i\right), c_j\left(y^i,\beta^i\right), \quad s_{F_i}^k, t_{c_{ji}}^k, g_F^i \in \partial F\left(y^i,\beta^i\right), g_{c_j}^i \in \partial c_j\left(y^i,\beta^i\right)\right\}, \\ &B_l^2 \subseteq \cup_{i<l}\left\{\left(\widehat{\varepsilon}_i^k, \widehat{g}^i \in \partial_{\widehat{\varepsilon}_i^k} H_k\left(x^k,\alpha^k\right)\right)\right\}, \text{ (if there is compression)}, \end{aligned} \tag{7}$$

where $k(l)$ denotes the index of the last serious step preceding the iteration l, when it is clear from the context, we do not specify the dependence of k on the current iteration index l. The element $(\widehat{\varepsilon}_i^k, \widehat{g}^i)$ in B_l^2 is the aggregate couple that is introduced when compression of bundle is implemented for controlling the size of bundle, and we will discuss aggregate technique later in detail.

Lemma 4. *Define*

$$\begin{aligned} e_i^k &= \begin{cases} F\left(x^k,\alpha^k\right) - s_{F_i}^k + c^+\left(x^k,\alpha^k\right), & \text{if } F\left(y^i,\beta^i\right) - F\left(x^k,\alpha^k\right) \geq c\left(y^i,\beta^i\right), \\ -t_{c_{ji}}^k + c^+\left(x^k,\alpha^k\right), & \text{if } F\left(y^i,\beta^i\right) - F\left(x^k,\alpha^k\right) < c\left(y^i,\beta^i\right), \end{cases} \\ g_{H_k}^i &= \begin{cases} g_F^i, & \text{if } F\left(y^i,\beta^i\right) - F\left(x^k,\alpha^k\right) \geq c\left(y^i,\beta^i\right), \\ g_{c_j}^i, & \text{if } F\left(y^i,\beta^i\right) - F\left(x^k,\alpha^k\right) < c\left(y^i,\beta^i\right), \end{cases} \end{aligned} \tag{8}$$

where notations $s_{F_i}^k, t_{c_{ji}}^k, g_F^i, g_{c_j}^i, (y^i,\beta^i)$, and (x^k,α^k) are from (6) and (7). Then, we have $e_i^k \geq 0$ and $g_{H_k}^i \in \partial_{e_i^k} H_k(x^k,\alpha^k)$.

Proof. By definitions of $s_{F_i}^k, t_{c_{ji}}^k$ and $g_F^i \in F(y^i,\beta^i), g_{c_j}^i \in \partial c_j(y^i,\beta^i)$, we have $F(x^k,\alpha^k) - s_{F_i}^k \geq 0$ and $c_j(x^k,\alpha^k) - t_{c_{ji}}^k \geq 0$. Since $c^+(x^k,\alpha^k) \geq 0, c^+(x^k,\alpha^k) \geq c_j(x^k,\alpha^k)$, it follows from (8) that $e_i^k \geq 0$. According to the definition of improvement function $H_k(\cdot,\cdot)$, we have

$$\begin{aligned} H_k(y,\beta) &\geq \max\begin{cases} F(y,\beta) - F\left(x^k,\alpha^k\right) \\ c(y,\beta) \end{cases} \geq \max\begin{cases} F(y,\beta) - F\left(x^k,\alpha^k\right) \\ c_j(y,\beta) \end{cases} \\ &= \max\begin{cases} -F\left(x^k,\alpha^k\right) + s_{F_i}^k + \langle g_F^i, (y,\beta) - \left(x^k,\alpha^k\right)\rangle \\ t_{c_{ji}}^k + \langle g_{c_j}^i, (y,\beta) - \left(x^k,\alpha^k\right)\rangle \end{cases} \\ &= c^+\left(x^k,\alpha^k\right) + \langle g_{H_k}^i, (y,\beta) - \left(x^k,\alpha^k\right)\rangle - \\ &\quad \begin{cases} F\left(x^k,\alpha^k\right) - s_{F_i}^k + c^+\left(x^k,\alpha^k\right), & \text{if } F\left(y^i,\beta^i\right) - F\left(x^k,\alpha^k\right) \geq c\left(y^i,\beta^i\right), \\ -t_{c_{ji}}^k + c^+\left(x^k,\alpha^k\right), & \text{if } F\left(y^i,\beta^i\right) - F\left(x^k,\alpha^k\right) < c\left(y^i,\beta^i\right). \end{cases} \end{aligned} \tag{9}$$

Notice that $H_k(x^k,\alpha^k) = c^+(x^k,\alpha^k)$ and the above inequality implies $g_{H_k}^i \in \partial_{e_i^k} H_k(x^k,\alpha^k)$.

Now, we are in position to define the cutting-planes model for $H_k(\cdot,\cdot)$:

$$\psi_l(y,\beta) = c^+\left(x^k,\alpha^k\right) + \max\left\{\max_{i\in B_l^1}\left\{-e_i^k + \langle g_{H_k}^i, (y,\beta) - \left(x^k,\alpha^k\right)\rangle\right\},\right.$$
$$\left.\max_{i\in B_l^2}\left\{-\hat{\varepsilon}_i^k + \langle \hat{g}^i, (y,\beta) - \left(x^k,\alpha^k\right)\rangle\right\}\right\}. \tag{10}$$

By Lemma 3 and $\hat{g}^i \in \partial_{\hat{\varepsilon}_i^k} H_k(x^k,\alpha^k)$, ψ_l is a lower approximate model for H_k, that is,

$$\psi_l(y,\beta) \le H_k(y,\beta), \quad \forall (y,\beta) \in R^{n+1}. \tag{11}$$

Given μ_l, a positive proximal parameter, the next trial point (y^l,β^l) is generated by solving the following quadratic programming:

$$\min_{(y,\beta)\in R^{n+1}} \psi_l(y,\beta) + \frac{\mu_l}{2}\left\|(y,\beta) - \left(x^k,\alpha^k\right)\right\|^2. \tag{12}$$

Obviously, (y^l,β^l) is unique and

$$\left(y^l,\beta^l\right) = \left(x^k,\alpha^k\right) - \frac{1}{\mu_l}\hat{g}^l, \quad \text{where } \hat{g}^l \in \partial\psi_l\left(y^l,\beta^l\right). \tag{13}$$

In order to measure whether the new trial point (y^l,β^l) provides sufficient decrease of $H_k(\cdot,\cdot)$, we define the nominal decrease δ_l by

$$\delta_l = H_k\left(x^k,\alpha^k\right) - \psi_l\left(y^l,\beta^l\right) - \frac{\mu_l}{2}\left\|\left(y^l,\beta^l\right) - \left(x^k,\alpha^k\right)\right\|^2$$
$$= \hat{\varepsilon}_l^k + \frac{1}{2\mu_l}\left\|\hat{g}^l\right\|^2, \tag{14}$$

where

$$\hat{\varepsilon}_l^k = H_k\left(x^k,\alpha^k\right) - \psi_l\left(y^l,\beta^l\right) - \frac{1}{\mu_l}\left\|\hat{g}^l\right\|^2 \ (\ge 0). \tag{15}$$

Since (y^l,β^l) is the solution to problem (12) and ψ_l is a lower approximation to H_k, we have

$$H_k\left(x^k,\alpha^k\right) \ge \psi_l\left(x^k,\alpha^k\right) \ge \psi_l\left(y^l,\beta^l\right) + \frac{\mu_l}{2}\left\|\left(y^l,\beta^l\right) - \left(x^k,\alpha^k\right)\right\|^2. \tag{16}$$

It follows that $\delta_l \ge 0$. Now, we present the descent test. Let $m \in (0,1)$ be a given parameter, when (y^l,β^l) satisfies:

$$H_k\left(y^l,\beta^l\right) \le c_j^+\left(x^k,\alpha^k\right) - m\delta_l, \tag{17}$$

then we declare a serious step: $(x^{k+1},\alpha^{k+1}) = (y^l,\beta^l)$; otherwise, a null step is declared: $(x^{k+1},\alpha^{k+1}) = (x^k,\alpha^k)$. Notice that here we only employ the information of one component function c_j of c to measure the "property" of the new generated trial point (y^l,β^l), unlike [9]; therein, they use the whole information of c.

Obviously, for any $(y,\beta) \in R^{n+1}$,

$$H_k(y,\beta) \ge \psi_l(y,\beta) \ge \psi_l\left(y^l,\beta^l\right) + \langle \hat{g}^l, (y,\beta) - \left(y^l,\beta^l\right)\rangle$$
$$= H_k\left(x^k,\alpha^k\right) - \langle \hat{g}^l, (y,\beta) - \left(x^k,\alpha^k\right)\rangle - \hat{\varepsilon}_l^k, \tag{18}$$

Hence, $\hat{g}^l \in \partial_{\hat{\varepsilon}_l^k} H(x^k,\alpha^k)$. If $\hat{\varepsilon}_l^k$ and $\|\hat{g}^l\|$ are very small, it follows that the approximate optimality condition is satisfied and the algorithm stops (see Step 3 in Algorithm 1).

As iterations go along, the number of elements in bundle B_l increases and the burden of computation and storage increases simultaneously. When the size of bundle becomes too big, it is necessary to compress and to clean the model. We have to discard some elements from bundle B_l^1 and append the aggregate couple $(\hat{\varepsilon}_l^k, \hat{g}^l)$ into bundle B_l^2 since the aggregate couple summarizes the information of previous iterate points [17]. For this purpose, we introduce the aggregate function:

$$l_{k,l}(y,\beta) = H_k\left(x^k,\alpha^k\right) + \langle \hat{g}^l, (y,\beta) - \left(x^k,\alpha^k\right)\rangle - \hat{\varepsilon}_l^k, \quad (y,\beta) \in R^{n+1}. \tag{19}$$

It has the following equivalent expression:

$$l_{k,l}(y,\beta) = \psi_l\left(y^l,\beta^l\right) + \langle \hat{g}^l, (y,\beta) - \left(y^l,\beta^l\right)\rangle, \quad (y,\beta) \in R^{n+1}. \tag{20}$$

For each k, the following conditions guarantee the bundle technique applied to function $H_k$either produces a serious step after a finite number of null steps, or the current serious step (x^k,α^k) is the minimum of H_k[18]:

$$\begin{cases} (a) & \psi_l(y,\beta) \le H_k(y,\beta), & \text{for all } l \ge 1; \\ (b) & l_{k(l),l}(y,\beta) \le \psi_{l+1}(y,\beta), & \left(y^l,\beta^l\right) \text{ is a null step;} \\ (c) & H_k\left(y^l,\beta^l\right) + \langle g_{H_k}^l, (y,\beta) - \left(y^l,\beta^l\right)\rangle, & \left(y^l,\beta^l\right) \text{ is a null step.} \end{cases} \tag{21}$$

Conditions (b) and (c) mean that the bundle B_l should contain both the aggregate information and the last generated information.

For model function ψ_l defined by (10) to satisfy conditions (a)-(c) when passing to the next iteration $l+1$, we should consider two cases of the lth iteration being a null step or a serious step. When there is a null step, the updates of the elements in bundle and the model do not present any problems since the improvement function is fixed between consecutive serious steps. Now suppose that the lth iteration

Input: $\varepsilon_1 > 0, \varepsilon_2 > 0, \gamma > 0$, $\mu_{\max} > \mu_{\min} > 0, \overline{m} \in (0,1)$, and $|B_{\max}| \geq 2$.
Step 1: given an initial proximal parameter $\mu_0 > 0$. Choose $(x^0, \alpha^0) \in R^{n+1}$ and one component function c_j of c, set $(y^0, \beta^0) = (x^0, \alpha^0)$, and compute $(F(y^0, \beta^0), c_j(y^0, \beta^0), g_F^0 \in \partial F(y^0, \beta^0), g_{c_j}^0 \in \partial c_j(y^0, \beta^0))$. Set $k = 0, l = 1$, $s_{F_0}^0 = F(y^0, \beta^0), t_{c_{j0}}^0 = c_j(y^0, \beta^0)$.
Define $B_1^1 = \left\{(s_{F_0}^0, t_{c_{j0}}^0, F(y^0, \beta^0), c_j(y^0, \beta^0), g_F^0, g_{c_j}^0)\right\}$, $B_1^2 = \varnothing$.
Step 2: find the solution (y^l, β^l) to problem (12). Compute $\widehat{g}^l = \mu_l((x^k, \alpha^k) - (y^l, \beta^l)), \widehat{\varepsilon}_l^k = H_k(x^k, \alpha^k) - \psi_l(y^l, \beta^l) - (1/\mu_l)\|\widehat{g}^l\|^2, \delta_l = \widehat{\varepsilon}_l^k + (1/2\mu_l)\|\widehat{g}^l\|^2$, $(F(y^l, \beta^l), c_j(y^l, \beta^l), g_F^l, g_{c_j}^l)$, and $(s_{F_l}^k, t_{c_{jl}}^k)$ using (6) written i with l.
Step 3: if $\widehat{\varepsilon}_l^k \leq \varepsilon_1$ and $\|\widehat{g}^l\|^2 \leq \varepsilon_2$, then stop, (x^k, α^k) is an approximate solution to problem (2).
Step 4: compute $H_k(y^l, \beta^l)$. If $H_k(y^l, \beta^l) \leq c_j^+(x^k, \alpha^k) - \overline{m}\delta_l$, then set $(x^{k+1}, \alpha^{k+1}) = (y^l, \beta^l)$ (serious step) and $\mu_{l+1} = \max\{(\mu_l/\gamma), \mu_{\min}\}$. Otherwise, set $(x^{k+1}, \alpha^{k+1}) = (x^k, \alpha^k)$ (null step) and $\mu_{l+1} = \min\{\gamma\mu_l, \mu_{\max}\}$.
Step 5: set $B_{l+1}^1 = B_l^1, B_{l+1}^2 = B_l^2$, if $|B_{l+1}| = |B_{\max}|$, then delete at least two elements from B_{l+1}, and append $(\widehat{\varepsilon}_l^k, \widehat{g}^l)$ to B_{l+1}^2. Append $(F(y^l, \beta^l), c_j(y^l, \beta^l), g_F^l, g_{c_j}^l, s_{F_l}^k, t_{c_{jl}}^k)$ to B_{l+1}^1.
Step 6: if a serious step is taken, choose $\overline{j} \in \widetilde{J}(x^{k+1}, \alpha^{k+1}) = \left\{j \in \widetilde{J} | c_j t(x^{k+1}, \alpha^{k+1}) n = qch(x^{k+1}, \alpha^{k+1})\right\}$ such that the function $c_{\overline{j}}$ is the component function of c we choose for constructing the cutting-planes model for $H_{k+1}(\cdot,\cdot)$. Replace component function c_j by $c_{\overline{j}}$ and compute $t_{c_{ji}}^k$ again. Update $s_{F_i}^{k+1}, t_{c_{ji}}^{k+1}$ for $i \in B_{l+1}^1$ by (22) and update $\widehat{\varepsilon}_i^{k+1}$ for $i \in B_{l+1}^2$ by (23).
Set $k = k + 1$.
Step 7: set $l = l + 1$, go to Step 2.
End of Algorithm 1.

Algorithm 1: Infeasible incremental bundle method for CVaR portfolio.

produces a serious step, i.e., $(x^{k+1}, \alpha^{k+1}) = (y^l, \beta^l)$, it means that for the next iteration we have to work with the new function $H_{k+1}(\cdot,\cdot) = H_{(x^{k+1}, \alpha^{k+1})}(\cdot,\cdot)$. Since serious step can be infeasible and the monotonicity of the objective function is not enforced, we can only make sure $H_k(x^{k+1}, \alpha^{k+1}) < H_k(x^k, \alpha^k)$, it is possible that $F(x^{k+1}, \alpha^{k+1}) > F(x^k, \alpha^k)$, and in that case, we have $H_{k+1}(\cdot,\cdot) \leq H_k(\cdot,\cdot)$, and, as a consequence, the cutting-planes model for $H_k(\cdot,\cdot)$ may not be a lower approximation for $H_{k+1}(\cdot,\cdot)$. Thus, the old model must be revised and adjusted to ensure that the conditions $(a - c)$ are satisfied for the new function $H_{k+1}(\cdot,\cdot)$.

Next lemma provides one approach to make sure that all elements in bundle correspond to approximate subgradients of new improvement function H_{k+1} at (x^{k+1}, α^{k+1}); hence, the model ψ_{l+1} is still a lower approximation function for H_{k+1} after adjusting the bundle. For elements in B_l^1, we only need to update the values of the linearization functions of F and c_j at (x^{k+1}, α^{k+1}). For elements in B_l^2, we have to invoke the information of F and c in order to make sure the aggregate subgradient $\widehat{g}^i$ also corresponds to some approximate subgradient of H_{k+1} at (x^{k+1}, α^{k+1}).

Lemma 5. *Suppose that the trial point (y^l, β^l) generated by (12) is a serious step, i.e., $(x^{k+1}, \alpha^{k+1}) = (y^l, \beta^l)$. Then, the following statements hold:*

(i) For each $i \in B_l^1$,

$$\begin{aligned} s_{F_i}^{k+1} &= s_{F_i}^k + \langle g_F^i, \left(x^{k+1}, \alpha^{k+1}\right) - \left(x^k, \alpha^k\right)\rangle, \\ t_{c_{ji}}^{k+1} &= t_{c_{ji}}^k + \langle g_{c_j}^i, \left(x^{k+1}, \alpha^{k+1}\right) - \left(x^k, \alpha^k\right)\rangle. \end{aligned} \tag{22}$$

Furthermore, $g_{H_{k+1}}^i \in \partial_{e_i^{k+1}} H_{k+1}(x^{k+1}, \alpha^{k+1})$, where $e_i^{k+1} \geq 0$ and $g_{H_{k+1}}^i$ are defined in (8) written with k replaced by $k + 1$.

(ii) For each $i \in B_l^2$, define

$$\begin{aligned} \widehat{\varepsilon}_i^{k+1} &= \widehat{\varepsilon}_i^k + c^+\left(x^{k+1}, \alpha^{k+1}\right) - c^+\left(x^k, \alpha^k\right) \\ &\quad + \left(F\left(x^{k+1}, \alpha^{k+1}\right) - F\left(x^k, \alpha^k\right)\right)^+ \\ &\quad + \langle \widehat{g}^i, \left(x^k, \alpha^k\right) - \left(x^{k+1}, \alpha^{k+1}\right)\rangle. \end{aligned} \tag{23}$$

Then, $\widehat{\varepsilon}_i^{k+1} \geq 0$ and $\widehat{g}^i \in \partial_{\widehat{\varepsilon}_i^{k+1}} H_{k+1}(x^{k+1}, \alpha^{k+1})$.

Proof. For each $i \in B_l^1$, we have, for all $(y, \beta) \in R^{n+1}$,

$$\begin{aligned} F(y, \beta) &\geq F\left(y^i, \beta^i\right) + \langle g_F^i, (y, \beta) - \left(y^i, \beta^i\right)\rangle \\ &= F\left(x^k, \alpha^k\right) + \langle g_F^i, (y, \beta) - \left(x^k, \alpha^k\right)\rangle - \left(F\left(x^k, \alpha^k\right) - s_{F_i}^k\right) \\ &\geq F\left(x^{k+1}, \alpha^{k+1}\right) + \langle g_F^i, (y, \beta) - \left(x^{k+1}, \alpha^{k+1}\right)\rangle \\ &\quad - \left(F\left(x^{k+1}, \alpha^{k+1}\right) - \left(s_{F_i}^k + \langle g_F^i, \left(x^{k+1}, \alpha^{k+1}\right) - \left(x^k, \alpha^k\right)\rangle\right)\right). \end{aligned} \tag{24}$$

It follows from (22) and $g_F^i \in_{F(x^k,\alpha^k)-s_{F_i}^k} F(x^{k+1},\alpha^{k+1})$ that

$$\begin{aligned} F\left(x^{k+1},\alpha^{k+1}\right) &\geq F\left(x^k,\alpha^k\right) + \langle g_F^i, \left(x^{k+1},\alpha^{k+1}\right) - \left(x^k,\alpha^k\right)\rangle - \left(F\left(x^k,\alpha^k\right) - s_{F_i}^k\right) \\ &= s_{F_i}^k + \langle g_F^i, \left(x^{k+1},\alpha^{k+1}\right) - \left(x^k,\alpha^k\right)\rangle = s_{F_i}^{k+1}. \end{aligned} \tag{25}$$

Therefore, $F(x^{k+1},\alpha^{k+1}) - s_{F_i}^{k+1} \geq 0$ and $g_F^i \in_{F(x^{k+1},\alpha^{k+1})-s_{F_i}^{k+1}} F(x^{k+1},\alpha^{k+1})$. Similarly,

we have $g_{c_j}^i \in_{c_j(x^{k+1},\alpha^{k+1})-t_{c_{ji}}^{k+1}} c_j(x^{k+1},\alpha^{k+1})$. The remaining proof can be finished by imitating the proof of Lemma 2.3 in [9], where the quantities $(l,k,s_{F_i}^k,t_{c_{ji}}^k)$ are replaced by $(l+1,k+1,s_{F_i}^{k+1},t_{c_{ji}}^{k+1})$, respectively.

For each $i \in B_l^2$, the update pattern is just the one in [9], so we omit the proof.

To sum up, no matter whether the l th iteration produces a null step or a serious step, as long as,

$$\begin{aligned} &B_{l+1}^1 \subseteq \cup_{i<l+1}\left\{\left(F\left(y^i,\beta^i\right), c_j\left(y^i,\beta^i\right), s_{F_i}^{k+1}, t_{c_{ji}}^{k+1}, g_F^i, g_{c_j}^i\right)\right\}, \\ &B_{l+2}^2 \subseteq \cup_{i<l+1}\left\{\left(\widehat{\varepsilon}_i^{k+1}, \widehat{g}^i\right)\right\}, \end{aligned} \tag{26}$$

the model

$$\begin{aligned} \psi_{l+1}(y,\beta) = c^+\left(x^{k+1},\alpha^{k+1}\right) + \max\Big\{&\max_{i\in B_{l+1}^1}\left\{-e_i^{k+1} + \right. \\ &\left.\langle g_{H_{k+1}}^i, (y,\beta) - \left(x^{k+1},\alpha^{k+1}\right)\rangle\right\}, \\ &\max_{i\in B_{l+1}^2}\left\{-\widehat{\varepsilon}_i^{k+1} + \langle \widehat{g}^i, (y,\beta) - \left(x^{k+1},\alpha^{k+1}\right)\rangle\right\}\Big\} \end{aligned} \tag{27}$$

satisfies condition (a) in (21) written with l replaced by $l+1$ and k replaced by $k+1$, and the point (x^{k+1},α^{k+1}) indicates the $(k+1)$th stability center which may be (x^k,α^k) if a null step is executed. Furthermore, if $(\widehat{\varepsilon}_l^k, \widehat{g}^l) \subseteq B_{l+1}^2$, then ψ_{l+1} satisfies condition (b) in (21), and if $(F(y^l,\beta^l), c_j(y^l,\beta^l), s_{F_l}^{k+1}, t_{c_{jl}}^{k+1}, g_F^l, g_{c_j}^l) \subseteq B_{l+1}^1$, then ψ_{l+1} satisfies condition (c) in (21).

3. Infeasible Incremental Bundle Algorithm for Problem (2)

Remark 1. If a serious step is declared, we have

$$F\left(x^{k+1},\alpha^{k+1}\right) - F\left(x^k,\alpha^k\right) \leq c_j^+\left(x^k,\alpha^k\right) - \overline{m}\delta_l, \tag{28}$$

$$c\left(x^{k+1},\alpha^{k+1}\right) \leq c_j^+\left(x^k,\alpha^k\right) - \overline{m}\delta_l. \tag{29}$$

4. Convergence Analysis

In this part, we discuss the convergence of Algorithm 1 by borrowing the main idea from [9]. From now on we assume that Algorithm 1 produces an infinite sequence of iterate points. As usual in the convergence analysis of bundle methods, we consider two cases: the number of serious steps is infinite and the number of serious steps is finite, and the last serious step is followed by infinitely many null steps. Let $L_s = \left\{l \mid (y^l,\beta^l)t \text{ is a serious step}\right\}$ be the set which collects the indices of serious steps in the sequence $\left\{(y^l,\beta^l)\right\}$.

Proposition 1. *For any iteration index $k_0 \geq 0$ of serious step, it holds that*

$$\left(x^k,\alpha^k\right) \in \left\{(x,\alpha) \in R^{n+1} \mid c(x,\alpha) \leq c^+\left(x^{k_0},\alpha^{k_0}\right)\right\}, \quad \forall k \geq k_0. \tag{30}$$

In particular, if $(x^{k_1},\alpha^{k_1}) \in D$ for some $k_1 \geq 0$, then $(x^k,\alpha^k) \in D$, for all $k \geq k_1$.

Proof. In Algorithm 1, the descent test is designed as follows: $H_k(y^l,\beta^l) \leq c_j^+(x^k,\alpha^k) - \overline{m}\delta_l$. Since function c is the pointwise maximum of finite convex functions $c_j\,(j \in \widetilde{J})$, it follows that $c^+(x^k,\alpha^k) \geq c_j^+(x^k,\alpha^k)$. If a serious step is declared, we have $H_k(y^l,\beta^l) \leq c^+(x^k,\alpha^k) - \overline{m}\delta_l$. Next, we adopt the techniques similar to [9], and the desired conclusions can be obtained. □

Proposition 2. *Suppose F is bounded below on D and Algorithm 1 generates an infinite number of serious steps. Then $\left\{\widehat{\varepsilon}_l^k\right\}_{l\in L_s} \longrightarrow 0$ and $\left\{\widehat{g}^l\right\}_{l\in L_s} \longrightarrow 0$.*

Proof. It follows from (29) and $c(x^k,\alpha^k) = \max\left\{c_j(x^k,\alpha^k) \mid j \in \widetilde{J}\right\}$ that if a serious step is declared,

$$c\left(x^{k+1},\alpha^{k+1}\right) \leq c_j^+\left(x^k,\alpha^k\right) - \overline{m}\delta_{l(k+1)} \leq c\left(x^k,\alpha^k\right) - \overline{m}\delta_{l(k+1)}. \tag{31}$$

Hence, the sequence $\{c(x^k,\alpha^k)\}$ is decreasing. Notice the update pattern of proximal parameter sequence $\{\mu_l\}$ in Algorithm 1, we have $\mu_l \leq \mu_{\max}$ for all l, i.e., the condition $\mu_l \leq \overline{\mu}$ for some $\overline{\mu} > 0$ holds. Following the thinking of [6], we can prove $\left\{\widehat{\varepsilon}_l^k\right\}_{l\in L_s} \longrightarrow 0$ and $\left\{\widehat{g}^l\right\}_{l\in L_s} \longrightarrow 0$.

Next proposition discusses the conditions which ensure the boundedness of sequence of serious steps $\{(x^k,\alpha^k)\}$.

Proposition 3. *Suppose that problem (2) has a solution $(\overline{x}, t\overline{\alpha})$ and Algorithm 1 produces an infinite number of serious steps. If the feasible set D is bounded or there exists some iteration index k_1 such that $F(\overline{x}, t\overline{\alpha}) \leq F(x^k,\alpha^k) + c_j^+(x^k,\alpha^k)$ for all $k \geq k_1$ and for some $j \in \widetilde{J}$, then the sequence $\{(x^k,\alpha^k)\}$ is bounded.*

Proof. For the first case that the feasible set D is bounded, since D is the level set of function c, the convexity of function c implies that all the level sets of function c are bounded. Then, the boundedness of $\{(x^k, \alpha^k)\}$ follows from the conclusion of Proposition 1. For the second case, since $\widehat{g}^l \in \partial_{\varepsilon_l^k} H_k(x^k, \alpha^k)$ and $\delta_l = \widehat{\varepsilon}_l^k + (1/2\mu_l)\|\widehat{g}^l\|^2$, we have

$$\begin{aligned}\left\|\left(x^{k+1}, \alpha^{k+1}\right) - (\overline{x}, \overline{\alpha})\right\|^2 &= \left\|\left(x^k, \alpha^k\right) - (\overline{x}, \overline{\alpha})\right\|^2 - \frac{2}{\mu_l}\langle \widehat{g}^l, \left(x^k, \alpha^k\right) - (\overline{x}, \overline{\alpha})\rangle + \frac{1}{\mu_l^2}\left\|\widehat{g}^l\right\|^2 \\ &\le \left\|\left(x^k, \alpha^k\right) - (\overline{x}, \overline{\alpha})\right\|^2 + \frac{2}{\mu_l}\left(H_k(\overline{x}, \overline{\alpha}) - H_k\left(x^k, \alpha^k\right) + \delta_l\right).\end{aligned} \tag{32}$$

Observe that $F(\overline{x}, \overline{\alpha}) \le F(x^k, \alpha^k) + c_j^+(x^k, \alpha^k) \le F(x^k, \alpha^k) + c^+(x^k, \alpha^k)$ for all $k \ge k_1$ and $c_j(\overline{x}, \overline{\alpha}) \le 0$, $c^+(x^k, \alpha^k) \ge 0$, we have

$$H_k(\overline{x}, \overline{\alpha}) - H_k\left(x^k, \alpha^k\right) = \max\left\{F(\overline{x}, \overline{\alpha}) - F\left(x^k, \alpha^k\right), c(\overline{x}, \overline{\alpha})\right\} - c^+\left(x^k, \alpha^k\right) \le 0. \tag{33}$$

By combining (32) and (33), it follows that

$$\left\|\left(x^{k+1}, \alpha^{k+1}\right) - (\overline{x}, \overline{\alpha})\right\|^2 \le \left\|\left(x^k, \alpha^k\right) - (\overline{x}, \overline{\alpha})\right\|^2 + \frac{2}{\mu_l}\delta_l, \quad \forall k \ge k_l. \tag{34}$$

By [19], $\sum_{l \in L_s} \delta_l < +\infty$, and the conclusion we have just proved in Proposition 2, the sequence $\{\|(x^{k+1}, \alpha^{k+1}) - (\overline{x}, \overline{\alpha})\|\}$ is convergent, and hence the sequence $\{(x^k, \alpha^k)\}$ is bounded. □

Theorem 1. *Suppose that problem (2) satisfies the Slater constraint qualification and its solution set is nonempty. Assume that Algorithm 1 generates an infinite number of serious steps, which is bounded. Then, all the accumulation points of the sequence $\{(x^k, \alpha^k)\}$are solutions to problem (2). And, if there exists some iteration index k_1 such that $F(\overline{x}, \overline{\alpha}) \le F(x^k, \alpha^k) + c_j^+(x^k, \alpha^k)$for all $k \ge k_1$ and for some $j \in \widetilde{J}$, the whole sequence $\{(x^k, \alpha^k)\}$ converges to a solution to problem (2).*

Proof. Since the proof is very similar to the one in [9], we omit it.

Now we consider the second case that the Algorithm 1 generates finite serious steps followed by infinitely many null steps, i.e., there exists an index last $= \max\{l | lt \in nL_s\}$, and the corresponding last serious step index is denoted by k_{last}. Notice that the improvement function is fixed for $k \ge k_{\text{last}}$, that is, $H_k(\cdot, \cdot) = H_{k_{\text{last}}}(\cdot, \cdot)$, $\forall k \ge k_{\text{last}}$.

Theorem 2. *Suppose that problem (2) satisfies the Slater constraint qualification and Algorithm 1 produces a finite number of serious steps, then $(x^{k_{last}}, \alpha^{k_{last}})$ is a solution to problem (2).*

Proof. We consider the case $l \ge l_{\text{last}}$ and denote $H(\cdot, \cdot) = H_{k_{\text{last}}}(\cdot, \cdot)$. By imitating the proof of Theorem 4.5 in [9], we obtain

$$H(\overline{y}, \overline{\beta}) + \frac{\mu_{\max}}{2}\left\|(\overline{y}, \overline{\beta}) - \left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right)\right\|^2 \le H\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right), \tag{35}$$

$$\psi_{l_i}\left(y^{l_i}, \beta^{l_i}\right) \longrightarrow H(\overline{y}, \overline{\beta}), \quad \text{i} \longrightarrow +\infty, \tag{36}$$

where $(\overline{y}, \overline{\beta})$ is an accumulation point of (y^l, β^l), i.e., $(y^{l_i}, \beta^{l_i}) \longrightarrow (\overline{y}, \overline{\beta})$ as $i \longrightarrow +\infty$. Since we assume that k_{last} is the last index for serious step, the descent test is not satisfied for $l \ge$ last:

$$\begin{aligned}H\left(y^l, \beta^l\right) - c_j^+\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right) &> -\overline{m}\delta_l \\ &\ge -\overline{m}\left(H\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right) - \psi_l\left(y^l, \beta^l\right)\right).\end{aligned} \tag{37}$$

From Step 6 in Algorithm 1, we know if a serious step is declared, we choose $\overline{j}$ from $\widetilde{J}(x^{k+1}, \alpha^{k+1}) = \left\{j \in \overline{J} | c_j(x^{k+1}, \alpha^{k+1}) = c(x^{k+1}, \alpha^{k+1})\right\}$ such that the function $c_{\overline{j}}$ is the component function of c we chose for constructing the cutting-planes model for $H_{k+1}(\cdot, \cdot)$. Therefore, according to (37),

$$H\left(y^l, \beta^l\right) - H\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right) \ge -\overline{m}\left(H\left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right) - \psi_l\left(y^l, \beta^l\right)\right). \tag{38}$$

Taking the limits along the specified subsequence as $i \longrightarrow +\infty$, by using (36), we obtain $0 \le (1 - \overline{m})(H(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}) - H(y^l, \beta^l))$. Since $\overline{m} \in (0, 1)$, we have $H(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}) \le H(y^l, \beta^l)$. Taking into account (36), we have $(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}) = (\overline{y}, \overline{\beta})$. It follows from (35) that $(\overline{y}, \overline{\beta})$ is the optimal solution to the following problem:

$$\min_{(y,\beta)\in R^{n+1}} H(y,\beta) + \frac{\mu_{\max}}{2}\left\|(y,\beta) - \left(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}}\right)\right\|^2. \quad (39)$$

According to the optimality condition, $0 \in \partial H(y^l, \beta^l) = \partial H(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}})$, where $H(\cdot,\cdot) = H_{k_{\text{last}}}(\cdot,\cdot)$. By Lemma 3, $(x^{k_{\text{last}}}, \alpha^{k_{\text{last}}})$ is a solution to problem (2). □

5. Conclusions

We present an infeasible incremental bundle method for the CVaR portfolio nonsmooth optimization problem; the algorithm is easier to implement since it only employs the information of the objective function and one component function of constraint functions. The algorithm does not enforce the feasibility of iterate points and the monotonicity of objective function, but the global convergence is established under mild conditions.

Data Availability

No data were used to support this study.

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Nature Science Foundation of China (grant nos. 61877032 and 11601061), the Foundation of Educational Committee of Liaoning Province (LQ2019019), and the PhD Research Startup Foundation of Liaoning Normal University (203070091909).

References

[1] Y. Song, D. Wu, W. Deng et al., "Multi-population parallel co-evolutionary differential evolution for parameter optimization," in *Energy Conversion and Management*, Elsevier, Amsterdam, Netherlands, 2020.

[2] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Applied Soft Computing*, vol. 23, Article ID 106724, 2020.

[3] C. Wu, X. Xu, Q. Li et al., "Performance assessment and optimization of a novel geothermal combined cooling and power system integrating an organic flash cycle with an ammonia-water absorption refrigeration cycle," *Energy Conversion and Management*, vol. 227, no. 2021, Article ID 113562, 2020.

[4] W. Hare, C. Sagastizábal, and M. Solodov, "A proximal bundle method for nonsmooth nonconvex functions with inexact information," *Computational Optimization and Applications*, vol. 63, no. 1, pp. 1–28, 2016.

[5] J. Shen, Z.-Q. Xia, and L.-P. Pang, "A proximal bundle method with inexact data for convex nondifferentiable minimization," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 66, no. 9, pp. 2016–2027, 2007.

[6] J. Shen, X.-Q. Liu, F.-F. Guo, and S.-X. Wang, "An approximate redistributed proximal bundle method with inexact data for minimizing nonsmooth nonconvex functions," *Mathematical Problems in Engineering*, vol. 2015, Article ID 215310, 9 pages, 2015.

[7] K. C. Kiwiel, "Proximity control in bundle methods for convex nondifferentiable minimization," *Mathematical Programming*, vol. 46, no. 1–3, pp. 105–122, 1990.

[8] J. Shen, J. Lv, J. Lv, F.-F. Gao, and R. Zhao, "A new proximal Chebychev center cutting plane algorithm for nonsmooth optimization and its convergence," *Journal of Industrial & Management Optimization*, vol. 14, no. 3, pp. 1143–1155, 2018.

[9] S. Claudia and S. Mikhail, "An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter," *SIAM. Journal on Optimization*, vol. 16, pp. 146–169, 2005.

[10] C. Lemaréchal, A. Nemirovskii, and Y. Nesterov, "New variants of bundle methods," *Mathematical Programming*, vol. 69, no. 1–3, pp. 111–147, 1995.

[11] Q.-Y. Zhang and Y. Gao, "A nonsmooth optimization method based on CVaR portfolio optimization problem," *Chinese Journal of Management Science*, vol. 25, no. 10, pp. 11–19, 2017.

[12] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.

[13] C. Tang, H. Chen, and J. Jian, "An improved partial bundle method for linearly constrained minimax problems," *Statistics, Optimization & Information Computing*, vol. 4, pp. 84–98, 2016.

[14] J. B. Jian, C. M. Tang, and F. Tang, "A feasible descent bundle method for inequality constrained Minimax problems (in Chinese)," *Science China Mathematics*, vol. 45, pp. 2001–2024, 2015.

[15] M. Gaudioso, G. Giallombardo, and G. Miglionico, "An incremental method for solving convex finite Min-Max problems," *Mathematics of Operations Research*, vol. 31, no. 1, pp. 173–187, 2006.

[16] K. C. Kiwiel, "Methods of descent for nondifferentiable optimization," in *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, Germany, 1985.

[17] J. Frédéric Bonnans, J. C. Gilbert, C. Lemaréchal et al., *Bundle Methods, the Quest for Descent, Numerical Optimization*, Springer, Berlin Germany, 2006.

[18] R. Correa and C. Lemaréchal, "Convergence of some algorithms for convex minimization," *Mathematical Programming*, vol. 62, no. 1–3, pp. 261–275, 1993.

[19] B. T. Polyak, *Introduction to Optimization*, Optimization Software, Inc., Publications Division, New York, NY, USA, 1987.

Hindawi
Complexity
Volume 2021, Article ID 6672131, 14 pages
https://doi.org/10.1155/2021/6672131

Research Article

Constrained Multiobjective Equilibrium Optimizer Algorithm for Solving Combined Economic Emission Dispatch Problem

M. A. El-Shorbagy [1,2] and A. A. Mousa [3]

[1]*Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia*
[2]*Department of Basic Engineering Science, Faculty of Engineering, Menoufia University, Shebin El-Kom 32511, Egypt*
[3]*Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*

Correspondence should be addressed to M. A. El-Shorbagy; mohammed_shorbagy@yahoo.com

Received 14 November 2020; Revised 22 December 2020; Accepted 29 December 2020; Published 15 January 2021

Academic Editor: Baogui Xin

Copyright © 2021 M. A. El-Shorbagy and A. A. Mousa. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This research implements a recent evolutionary-based algorithm of equilibrium optimizer to resolve the constrained combined economic emission dispatch problem. This problem has two objective functions that represent the minimizing of generation costs and minimizing the emission of environmental pollution caused by generators. The proposed algorithm integrates the dominant criteria for multiobjective functions that allow the decision-maker to detect all the Pareto boundaries of constrained combined economic emission dispatch problem. In order to save the effort for the decision-maker to select the best compromise alternative, a cluster study was carried out to minimize the size of the Pareto boundary to an acceptable size, representing all the characteristics of the main Pareto frontier. On the other hand, in order to deal with the infringement of constraints, a repair algorithm was used to preserve the viability of the particles. The proposed algorithm is applied to solve the standard 30-bus IEEE system with 6 generators to validate its robustness and efficiency to produce a well-distributed Pareto frontier for constrained combined economic emission dispatch problem. Compared with other studies, good results in solving constrained combined economic emission dispatch problem are obtained and a reasonable reduced Pareto set is found.

1. Introduction

The robust and efficient economic planning, operation, and distribution of power systems have always presided a vital role in the power system industry. Saving a small percent in the operation of the power systems produces a reasonable reduction in the operating cost and in the amount of fuel consumed [1]. Finding the optimum operating cost is the key purpose of the classic constrained combined economic emission dispatch problem (CEEDP) [2]. Recently, for large scale electric power system, modern system optimization theory methods are applied with the cost savings [3]. There are three directions to solve CEEDP [4].

Traditionally, the first direction is to simplify the multiobjective optimization problem (MOP) to a single objective problem. Traditional methods used to convert MOP into a single objective problem are either the aggregating of objective functions as in the weighted sum method or the optimization of the most important objective and the treatment of others as constraints as in the ε-constraint method or the penalty factor approach [5]. Then, various numerical optimization methods have been employed to handle this single objective problem such as the augmented Lagrangian method and gradient method, for example, in weighted sum [6], the ε-constraint method [7]. The most important weaknesses of these methods are that it cannot deal with nonconvex function and tends to find weak set of nondominated solutions. On the other hand, goal programming is also implemented to deal with CEEDP [8]. In this approach, a specified target is assigned for each objective to be achieved and then aims to minimize the deviation from the desired targets to the objective functions.

The second direction is to solve the single objective CEEDP by any single objective meta-heuristics algorithm such as artificial bee colony (ABC) [9], gravitational search algorithm (GSA) [10], and Gaussian particle swarm optimization (GPSO) [11] or by any hybrid single objective meta-heuristics algorithms such as hybrid particle swarm optimization (PSO) algorithm and firefly algorithm (FA) [12], hybrid ABC algorithm and simulated annealing algorithm (SA) [13], and PSO-GSA algorithm [14].

The third direction is to handle both objectives of CEEDP simultaneously, by using meta-heuristics-based MOO techniques, as competing objective functions instead of transforming the MOP formulation to a single objective problem, as dynamic random neighborhood PSO (DRN-PSO) [15], nondominated sorting genetic algorithm (NSGA) [16], niched Pareto genetic algorithm (NPGA) [17], fuzzy clustering-based particle swarm (FCPSO) [18], modified shuffled frog leaping algorithm (MSFLA) [19], real coded genetic algorithm (RCGA) [20], and strength Pareto evolutionary algorithm (SPEA) [21]. The implementation and development of evolutionary-based multiobjective algorithms have significantly grown since they employ a population of individuals in their search. On the other hand, multiple Pareto optimal frontiers can be detected in one single run instead of running the algorithm many times to get multiple Pareto optimal frontiers. These algorithms can be robustly implemented to overcome most of the disadvantages of traditional methods [22].

Due to the importance of CEEDP, researchers have recently proposed new methods for solving CEEDP. In [23], the authors seek to find the best compromise alternative using the quantum genetic algorithm. In [24], Mellal and Williams proposed a hybrid cuckoo optimization algorithm with the binary approach and penalty function to solve CEEDP. In [25], Kim et al. presented a neural network approach to find the optimal solution of CEEDP. In [26], Sundaram proposed a chaotic-based approach to explore the search space to deal with CEEDP. In [27], a novel approach, based on a hybrid algorithm combining a genetic algorithm and a modified Hooke and Jeeves method, was presented to solve the CEEDP with equality constraints. In [28], a novel MOP was proposed to address the problems of economic dispatch and power shedding at the same time. In [29], CEEDP was formulated as a MOP in which the two goals, fuel cost, and emissions of pollutants were optimized at the same time as meeting constraints and addressed by evolutionary multiobjective algorithms. In addition, an empirical study of constraint management has been discussed. In [30], an ameliorated grey wolf optimization algorithm is proposed to solve the CEEDP that coordinates the behavior of grey wolves, random exploration, opposition learning, and local random search. Furthermore, a new paradigm of the CEEDP-based MOP is proposed in [31], where CEEDP has been solved by relatively recent multiobjective algorithms. In [32], Sundaram and Erdogmus proposed a hybrid evolutionary multiobjective optimization system using nondominated sorting genetic algorithm II (NSGA II) and multiobjective PSO to solve the CEEDP. The hybrid approach with a constraint management system is capable of balancing the tasks of exploration and exploitation. In [33], an interesting algorithm, biogeography-based learning particle swarm optimization (BLPSO), was applied to solve the CEEDP with different types of constraints. Simulation results of BLPSO overcome local trap and improve the convergence of the solution. In [34], the authors presented a deterministic optimization based on the augmented Lagrangian to solve CEEDP. In [35], the authors reformulated centralized CEEDP into a decentralized model. In [36], Srivastava and Das presented a new evolutionary-based meta-heuristic method, which implemented to solve CEEDP. The performance of the suggested algorithm was evaluated by using twenty-nine benchmark functions and CEEDP. Finally, Zhang et al. [37] presented a decentralized power system for the combined economic emission dispatch problem (CEEDP) model.

One of the key challenges in the solving of constrained MOPs is the handling of constraints. The constraints management strategies are different for the meta-heuristic algorithms [38]. There are four common methods for constraint handling:

(i) Discarding unfeasible options

(ii) The inclusion of a penalty function to reduce the fitness of infeasible solutions

(iii) The development of algorithm mechanism to always generate feasible solutions

(iv) Repairing the unfeasible solutions to be feasible

Handling strategies of constraints i, iii, and iv are specifically applicable to the case of multiobjective. On the other hand, constraint-handling strategy (ii) is not applied explicitly to multiobjective algorithms, due to the fact that the fitness assignment relies on the nondominance rank of the solution, not on the objective function values of the solution [39].

There are several new methods for constraint management that have recently been suggested, such as a modern methodology of constraint management that consistently takes into account proximity, diversity, and feasibility [40], dynamic constraint-handling mechanism [41], adaptive repair method for constraint handling [42], etc. Also, in [43], three typical constraint-handling approaches were implemented into a multiobjective PSO algorithm: (1) dominance of feasible solutions; (2) penalty function; and (3) multiobjective optimization strategy, where the constraints are treated as additional objectives to be optimized.

The Pareto front of MOPs consists of a very large set of Pareto solutions. Selecting a single alternative from this huge set is potentially intractable for any decision-maker (DM) [44]. In order to allow the DM to classify and coordinate solutions, some means of reducing/organizing the set of nondominated solutions are implemented to shrink the size of the Pareto optimal set, which facilitates finding the optimal operating alternative [45]. Several studies have concerned this issue by implementing filtering and cluster analysis to minimize the optimum size of the Pareto to a rational size, enabling the DM to choose the best compromise solution [46]. Algorithms of clustering can be divided into three categories: density-based clustering,

hierarchical clustering, and centroid-based clustering (partitioning clustering methods) [47]. One of the most common partitioning clustering methods is the K-means, where it is distinguished by its efficiency and robustness in clustering data analysis [48].

There are many meta-heuristics algorithms that have been proposed to solve optimization problems such as genetic algorithms (GA) [49], artificial immune system [50], particle swarm optimization (PSO) [51], ant colony optimization (ACO) [52], artificial bee colony (ABC) [53], bacterial foraging algorithm (BFA) [54], cat swarm optimization (CSO) [55], glowworm swarm optimization algorithm (GSOA) [56], firefly optimization algorithm (FOA) [57], monkey algorithm (MA) [58], krill herd algorithm (KHA) [59], cuckoo search algorithm (CSA) [60], whale optimization algorithm (WOA) [61], sine cosine algorithm (SCA) [62], grasshopper optimization algorithm (GOA) [63], salp swarm algorithm [64], equilibrium optimizer algorithm (EOA) [65], gradient-based optimizer (GBO) [66], slime mould algorithm (SMA) [67], and Harris Hawks optimization (HHO) [68].

As a new algorithm, the equilibrium optimizer (EO) mimics physics-based dynamic source and sink models used to estimate and consider equilibrium states. In this paper, a constrained multiple objective equilibrium optimizer algorithm (EOA) is proposed to solve the combined economic emission dispatch problem (CEEDP) in the power system by modifying the mechanism of the EOA. The proposed methodology is tested by the standard IEEE 30-bus system with six generator units to validate the robustness of the proposed algorithm. In addition, a clustering strategy is used to choose the most compromised set of alternatives based on DM preferences. This paper's major contributions are as follows:

(1) A new approach based on an EOA to solve CEEDP is introduced and tested

(2) Clustering was performed in order to minimize the Pareto border to an appropriate size

(3) A repair algorithm was used to tackle the constraints and unfeasible solution

(4) The robustness and reliability of the proposed algorithm to generate a well-distributed Pareto frontier is tested by the IEEE standard 30-bus system with 6 generators

(5) Numerical analysis results validate the high performance of the proposed algorithm

This paper is organized as follows: MOP formulation is presented in Section 2. The formulation of the combined economic emission dispatch problem is described in Section 3. A constrained multiobjective EOA is presented in Section 4. The simulation analysis is analyzed in Section 5. Finally, the conclusion is presented in Section 6.

2. Multiobjective Optimization Problem

A broad range of applications in architecture, computer science, and many other fields include optimizing several objectives at the same time [69]. In multiobjective optimization problem (MOP), there is no way to decide the best solution where a set of alternatives with different trade-offs among the multiple objectives is retained, which is known as Pareto optimal frontier, instead of getting a single alternative. In the following, the MOP is presented with several general notes and concepts [70].

Definition 1 (formulation of multiobjective optimization problem). Formally, a MOP is stated as follows:

$$\begin{aligned} &\text{Min } f(x) = \left[f_1(x), f_2(x), \ldots, f_m(x)\right]^T, \\ &\text{s.t.} x \in X. \end{aligned} \tag{1}$$

The vector $x \in \mathbb{R}^n$ is defined as a vector of n decision variables in the optimization problem formulation. The feasible set $X \subseteq \mathbb{R}^n$ is implicitly determined by a set of constraints. The vector function $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is defined by m scalar objective functions.

Definition 2 (Pareto dominance). Vector z^1 Pareto-dominates vector z^2, which is denoted by $z^1 \succ_{\text{pareto}} z^2$, if and only if

$$\begin{aligned} &\forall_i \in \{1, \ldots, m\}\colon z_i^1 \le z_i^2, \\ &\exists_i \in \{1, \ldots, m\}\colon z_i^1 < z_i^2, \end{aligned} \tag{2}$$

i.e., the vector z^1 Pareto-dominates vector z^2 if the previous two conditions are satisfied; thus, in order to solve MOP, the set of solutions $x \in X$ should be found whose images $z = f(x)$ are not dominated (nondominated set) by any other vector in the feasible space.

Definition 3 (Pareto optimality concept). A feasible solution $x^* \in X$ is a Pareto optimal solution if there does not exist any feasible solution $x \in X$ such that $f(x) \prec_{\text{pareto}} f(x^*)$.

Definition 4 (the set of Pareto optimal solutions). The Pareto optimal set, PO*, is the set, which is defined as

$$\text{PO}^* = \{x \in X | \text{óy} \in \text{X}\colon \text{f}(\text{y}) \prec_{\text{Pareto}} \text{f}(\text{x})\}. \tag{3}$$

Definition 5 (Pareto optimal frontier). For a Pareto optimal set, PO*, the Pareto optimal frontier, PF*, is stated as

$$\text{PF}^* = \{f(x) = (f_1(x), \ldots, f_m(x)) | x \in \text{PO}^*\}. \tag{4}$$

Definition 6 (ideal objective vector). This vector z_i^I is a vector whose components are the optimal/minimal values of each objective. The component of $z^I \in R^m$ can be got by minimizing every objective separately subject to the set of constraints, that is,

$$\begin{aligned} z_i^I &= \text{Min} f_i(x), \\ &\text{s.t. } x \in X \text{ for } i = 1, \ldots, m. \end{aligned} \tag{5}$$

By using the ideal vector, the lower limits of the Pareto optimal frontier for each objective function are determined.

Definition 7 (nadir objective vector) [71]. Nadir vector is the upper bound of the optimal Pareto frontier. The nadir objective vector components z^{nad} can be estimated from the payoff table as depicted in Figure 1.

3. The Formulation of Combined Economic Emission Dispatch Problem

The CEEDP involves the optimization of two multiple conflicting objectives, generation operation cost, and pollutant emission, which must be addressed at the same time. The problem formulation is as follows.

3.1. Objective Functions. The formulation of the objective functions of the problem, which are objective for the economy and environmental objective, is seen in this subsection.

3.1.1. Generation Operation Cost (Economy Objective). The traditional CEEDP problem is to decide the optimum active power produced by the power system with minimizing the total cost of the generation operation and satisfying the overall load requirement. The economy objective can be formulated mathematically as [72]

$$f_1(P_{Gi}) = C_t = \sum_{i=1}^{n} C_i(P_{Gi}) = \sum_{i=1}^{n}\left(a_i + b_i P_{Gi} + c_i P_{Gi}^2\right) + e_i \sin\left(f_i\left(P_{Gi} - P_{Gi,\min}\right)\right)\$/h, \tag{6}$$

where C_t is the overall cost of fuel that is used for generation operations ($\$/h$), the parameters a_i, b_i, c_i, e_i, f_i are the coefficients of the cost for each generator i, P_{Gi} is the generated active power (p.u) by each generator i, $P_{Gi,\min}$ is the minimal generation limit of unit i, and n is the generator numbers. The valve point loading effect is overlooked in this study.

3.1.2. Pollutant Emission (Environmental Objective). The other environmental objective is the pollutant emission, which can be defined as the amount of all types of pollutant emission, such as nitrogen oxides (NO_x) and sulfur dioxide 2 (SO_2). In this study, only "nitrogen oxide" pollutant emissions are taken into account without a lack of generality. As a quadratic and exponential function, the amount of emission pollutants is indicated as

$$f_2(P_{Gi}) = E_{NO_x} = \sum_{i=1}^{n}\left[\alpha_i + \beta_i P_{Gi} + \gamma_i P_{Gi}^2 + \xi_i e^{(\lambda_i P_{Gi})}\right]\left(\frac{\text{ton}}{h}\right), \tag{7}$$

where the parameters $\alpha_i, \beta_i, \gamma_i, \xi_i, \lambda_i$ are the coefficients of each generator's NO_x emission.

3.2. Constraints. The formulation of CEEDP is restricted by a set of constraints as follows.

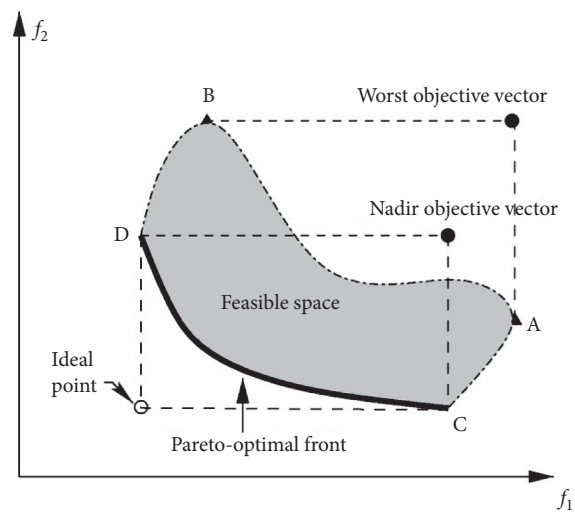


Figure 1: Ideal, nadir, and worst objective vectors.

3.2.1. The Set of Active Power Constraints. The total active power constraints delivered by the generator must provide the total load demand P_D and the transmission line losses P_{loss}:

$$\sum_{i=1}^{n} P_{Gi} - P_{\text{loss}} = P_D, \tag{8}$$

where P_D is the total load demand (p.u.) and P_{loss} is the total transmission line losses (p.u.).

The losses of the transmission line are given as follows [73]:

$$\begin{aligned} P_{\text{loss}} &= \sum_{i=1}^{n}\sum_{i=1}^{n}\left[A_{ij}\left(P_i P_j + Q_i Q_j\right) + B_{ij}\left(Q_i P_j - P_i Q_j\right)\right], \\ P_i &= P_{Gi} - P_{Di}, \\ Q_i &= Q_{Gi} - Q_{Di}, \\ A_{ij} &= \frac{R_{ij}}{V_i V_j}\cos\left(\delta_i - \delta_j\right), \\ B_{ij} &= \frac{R_{ij}}{V_i V_j}\sin\left(\delta_i - \delta_j\right), \end{aligned} \tag{9}$$

where the system parameters are defined as follows:

(i) n is the number of buses,

(ii) R_{ij} is the series resistance linking buses i and j,

(iii) V_i is the voltage magnitude at bus i,

(iv) δ_i is the voltage angle at bus i,

(v) P_i is the real active power injection at bus i, and

(vi) Q_i is the reactive power injection bus i.

3.2.2. Upper Bound and Lower Bound of Active Power Generation. The active power P_{Gi} delivered by each generator in the power system is constrained by its upper bound and lower bound, i.e.,

$$\begin{aligned} &P_{Gi\min} \le P_{Gi} \le P_{Gi\max}, \\ &Q_{Gi\min} \le Q_{Gi} \le Q_{Gi\max}, \\ &V_{i\min} \le V_i \le V_{i\max}, \quad i = 1, \ldots, n, \end{aligned} \tag{10}$$

where

(i) $P_{Gi\min}$ is the lower bound active power generated, and

(ii) $P_{Gi\max}$ is the upper bound active power generated.

3.2.3. The System Security Constraints. The CEEDP formulation shall take into account the limited proportion of system lines in violation or in near violation of security limits; these lines are known as the critical lines. The system critical lines that connect in the optimal solution are only taken into account. The experiences of the DM help to detect the critical lines. The protection limit can be maintained by minimizing the following equation:

$$S = f(P_{Gi}) = \sum_{j=1}^{k} \frac{\left|T_j(P_G)\right|}{T_j^{\max}}, \tag{11}$$

where

(i) $T_j(P_G)$ is the real active power flow,

(ii) $T_j^{\max}$ is the maximum limit of the active power flow through the jth line, and

(iii) k is the number of critical lines.

The line active power flow through the jth line is defined in terms of the control variables P_{Gs}, by implementing the generalized generation distribution factors (GGDFs) [74], and is defined as follows:

$$T_j(P_G) = \sum_{i=1}^{m} \left(D_{ij} P_{Gi}\right), \tag{12}$$

where the parameter D_{ij} is the GGDF for line j, due to generator i. For protection, the loading of the transmission line S_l is limited by its maximum, which is given as $S_l \le S_{l\max}$, $l = 1, \ldots, m_l$, where m_l is the number of transmission line in the power system.

4. Constrained Multiobjective Equilibrium Optimizer Algorithm

In this section, we introduce a novel evolutionary-based algorithm depending on the EOA [65] which simulates the equilibrium and dynamic m states related to the mass balance models.

4.1. Brief Introduction to Equilibrium Optimizer Algorithm. The equilibrium optimizer algorithm (EOA) is an optimizer that was firstly presented by Farmarzi [65] which first appeared in 2020. It simulates the balance and dynamic states associated to models for mass balance, in which each position (concentration of particles) is randomly updated in order to reach equilibrium state (fitness). The equilibrium optimizer is easy to use. Also, it has an adaptive dynamic control parameter. EOA is started with the initial particle positions (initial population C_i, $i = 1, 2, \ldots,$ no. of particles) and problem's dimensions (dim) as in the following equation:

$$C_{\text{initial}} = \text{rand}\,(\text{search particles\_no}, \text{dim}) \times (ub - lb) + lb, \tag{13}$$

where C_{initial} locates the initial positions of the particles; the bounds $lb$and $ub$are the specified lower and upper bounds respectively of the decision optimization variables.

4.1.1. Equilibrium Pool and Candidates (Ceq). The equilibrium/balance state is the EOA's final state of convergence. At the beginning of the algorithm, equilibrium candidates are allocated to support a particle search pattern. During the optimization process, the four best so-far particles are determined in addition to another particle whose concentration is the arithmetic mean of the four particles. EOA has an exploration scheme using four candidates and an exploitation scheme using the average mean. For constructing a vector called the balance pool, these five particles are called candidates of equilibrium:

$$C_{\text{eq,pool}} = \left(C_{\text{eq},1}, C_{\text{eq},2}, C_{\text{eq},3}, C_{\text{eq},4}, C_{\text{eq,av}}\right). \tag{14}$$

The position of each particle in each iteration, by the same probability, is updated using the random selection among the chosen candidates. Then, the particle positions are regularly updated. The updating procedure of the EOA is as in the following equation:

$$\begin{aligned} C_{\text{new}} &= C_{\text{eq}} + \frac{G}{\lambda}(1 - F) + \left(C_{\text{old}} - C_{\text{eq}}\right) \times F, \\ F &= a_1 \,\text{sign}\left(r_1 - 0.5\right)\left(e^{-\lambda t} - 1\right), \\ G &= \begin{cases} 0.5 r_1, & \text{if } r_2 \ge GP, \\ 0, & \text{if } r_2 < GP, \end{cases} \\ t &= \left(1 - \frac{T}{T_{\max}}\right)^{a_2 \left(T/T_{\max}\right)}, \end{aligned} \tag{15}$$

where C_{old} is the current position vector and C_{new} is the new position vectors of the particle. From the equilibrium pool, we pick one concentration vector randomly which nominated by C_{eq}. λ is a random vector between 0 and 1; a_1 and a_2 are constants ($a_1 = 2$ and $a_1 = 1$), r, r_1, r_2 are random numbers between 0 and 1, T is the current iteration counter, and $T_{\max}$ is the number of iterations maximum. In each iteration, for each particle's position, the considered objective function is measured to estimate its status. The

equilibrium pool is also updated to include the four best particles to date in every iteration.

4.2. Dominance Criteria. To improve the EOA to deal with MOPs, we incorporate the dominance criteria in the proposed algorithm, which are described as follows:

For any problem having multiple objective functions (say f_j, $j = 1, \ldots, k$, $k > 1$), any two particles x^1 and x^2 can have one of two cases, "one dominates the other" or "none dominates the other."

The particle x^1 dominates the particle x^2, if the following two conditions are satisfied [75]. Note that the operator $\prec$ denotes worse and the operator $\succ$ denotes better.

(1) The particle x^1 is no worse than x^2 in all objectives, or $f_j(x^1) \nprec f_j(x^2)$, $\forall j = 1, \ldots, p$.

(2) The particle x^1 is strictly better than x^2 in at least one objective, or $f_j(x^1) \succ f_j(x^2)$ for at least one $j \in \{1, 2, \ldots, k\}$.

If any of the above two conditions is infringed, the particle x^1 dose not dominate the particle x^2. This algorithm is repeatedly for all populations, all particles that are not dominated by any other individuals are from "the non-dominated set".

4.3. Archive/Selection Strategy. The vector function F is represented by m-dimensional, where each coordinate is one of the objectives (Figure 2). In each iteration counter t, the aim of the archive/selection technique is to create a new set of solutions. This technique uses the old archive set $A^{(t-1)}$ to update the archive content $A^{(t)}$. Overall, the purpose of this archive is to collect useful data about optimization problems during the run and update the content of the stored data.

4.4. Identifying the K-Centroids of Pareto Front. The solution of MOP by using the concept of dominance produces a large size of Pareto optimal solutions, not a single elite optimum solution [76]. The selection of a solution to achieve various objectives is the most critical task in practical engineering applications. Also, decreasing the size of the Pareto set, that is available to the DM to select, saves effort and time. Cluster analysis has been implemented to shrink the volume of the Pareto frontier to a reasonable volume using a predetermined size. K-means is an important partitioning clustering which is used to classify n observations into K clusters. Consider the Pareto front which is N points with m dimensions, where m is the number of objectives. The following steps are used to identify the K-centroids of the Pareto front:

(i) Step 1: the initial K cluster centers $Ce_1, Ce_2, \ldots, Ce_K$ are randomly selected from the Pareto frontier $Pf^* = \{Pf_1^*, Pf_2^*, \ldots, Pf_N^*\}$.

(ii) Step 2: each agent form the Pareto set Pf_i^*, $i = 1, 2, \ldots, N$ is allocated to cluster C_j, $j \in \{1, 2, \ldots, k\}$ if and only if

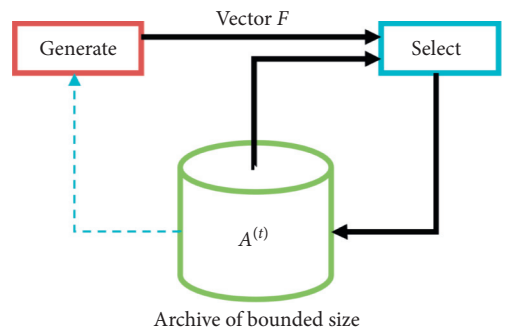


Figure 2: Block structure of archive-based selection algorithm.

$$\left\| Pf_i^* - Ce_j \right\| < \left\| Pf_i^* - Ce_p \right\|, \quad p = 1, 2, \ldots, K \& j \neq p. \tag{16}$$

(iii) Step 3: centers of new cluster $Ce_1, Ce_2, \ldots, Ce_K$ are determined as follows:

$$Ce_i^* = \frac{1}{ne_i} \sum_{Pf_j^* \in C_i} Pf_j^*, \quad \forall i = 1, 2, \ldots, K, \tag{17}$$

where ne_i is the elements number of the cluster C_j.

(iv) Step 4: if $Ce_i^* = Ce_i$, $\forall i = 1, 2, \ldots, K$, then the algorithm stops; otherwise, go to Step 2.

4.5. Basic Algorithm. The structure of the proposed algorithm is shown in Figure 3. In order to guarantee the obtaining for the true Pareto optimal set, Figure 3 shows how the Pareto set is updated and how it is applied in the procedure of proposed algorithm. In addition, an archiving-based selection algorithm was used to ensure retaining the Pareto optimal frontier and monitoring of the whole domain of the nondominated set. On the other hand, the violation in the constraint will be handled using a repair method [77], where it separates and repairs any infeasible particle in population at each generation. Firstly, an initial reference point R (feasible reference point) is defined for the repair phase. Infeasible particles in the population are then repaired according to the repair process until they are feasible. The repair process produces a new viable particle (z) instead of an unfeasible one (q) on a segment identified by the two points (R and q). This segment can be expanded equally by a specified parameter $\mu \in [0, 1]$ on both sides. Therefore, the new feasible particle z is generated as

$$z = \begin{cases} \gamma \cdot q + (1 - \gamma) \cdot R, \\ \gamma \cdot R + (1 - \gamma) \cdot q, \end{cases} \tag{18}$$

where $\gamma = (1 + 2\mu)\delta - \mu$ and $\delta \in [0, 1]$ is a random number. Figure 4 shows a schematic view of the constraint-handling method for the generated particles, finally, defining the K-centroids of Pareto front using the K-means clustering algorithm.

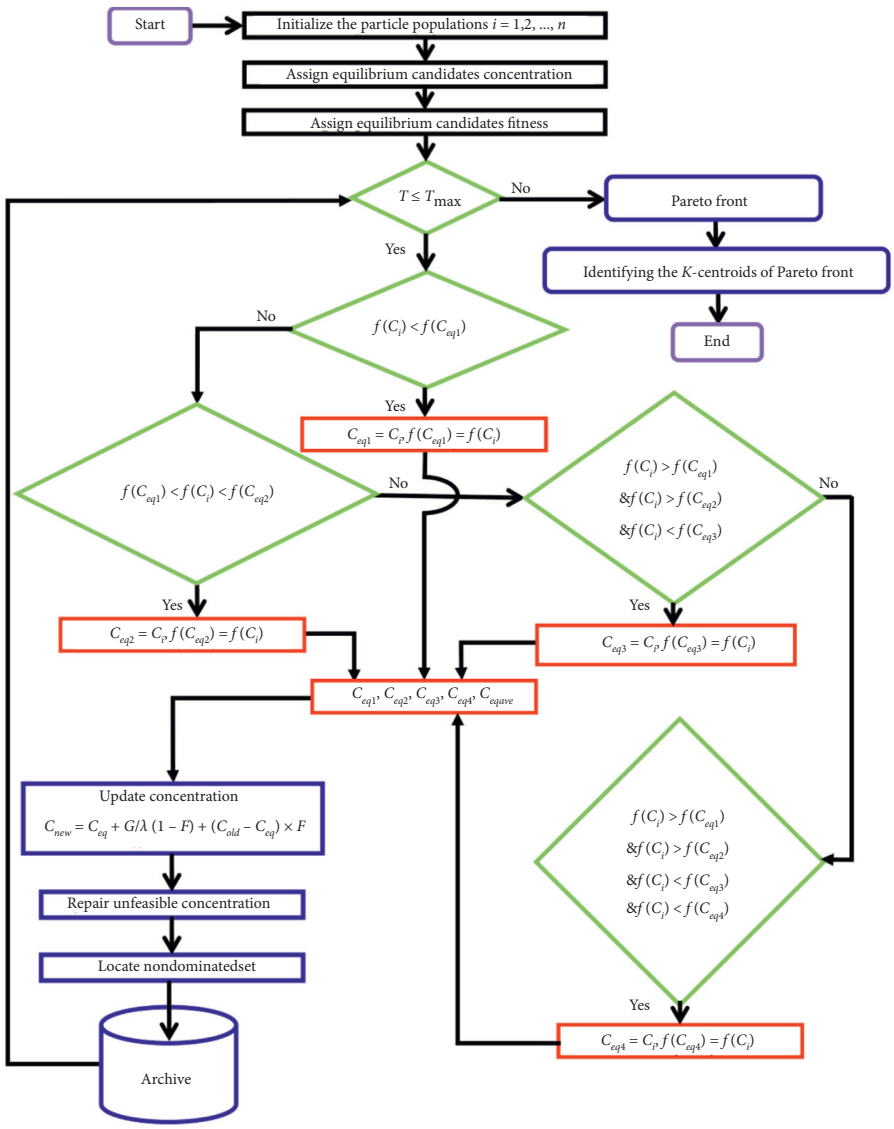


Figure 3: The structure of the proposed algorithm.

5. Simulation Analysis

The standard IEEE system having a 30-bus system and six generator units is elected to confirm the robustness and the proposed algorithm efficiency. The data of the test system consists of data for running the generators located at buses 25–30 and 41 lines. Figure 5 illustrates the diagram of this system. Fuel cost with the generator's limits and pollutant emission NO_x parameters for this system are shown in Table 1, where the down ramp rate is considered as 10%. The transmission B-loss coefficient matrix is determined by using a load flow software as seen in Table 2, as used in [78]. The overall demand for power is 2.834 p.u. in the system at the base of 100 MVA. All information for this system is given in the MATLAB power system simulation package [79].

The proposed algorithm is coded in MATLAB (R2016b) and implemented on the computer with Intel Core i5, 1.80 GHz and 4 GB RAM. Like any meta-heuristic algorithm, the proposed algorithm requires a set of parameters that influence its performance. The controlled parameters of the proposed algorithm are shown in Table 3.

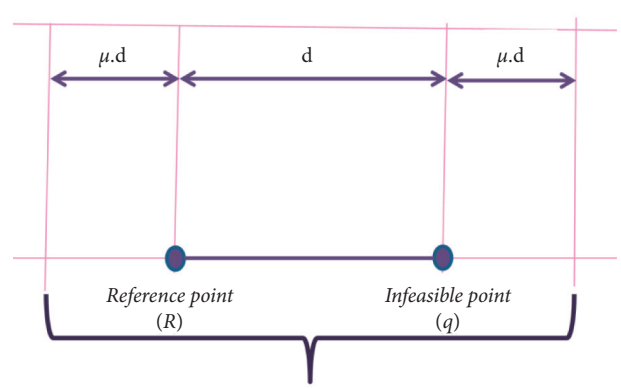


Figure 4: Constraint-handling method.

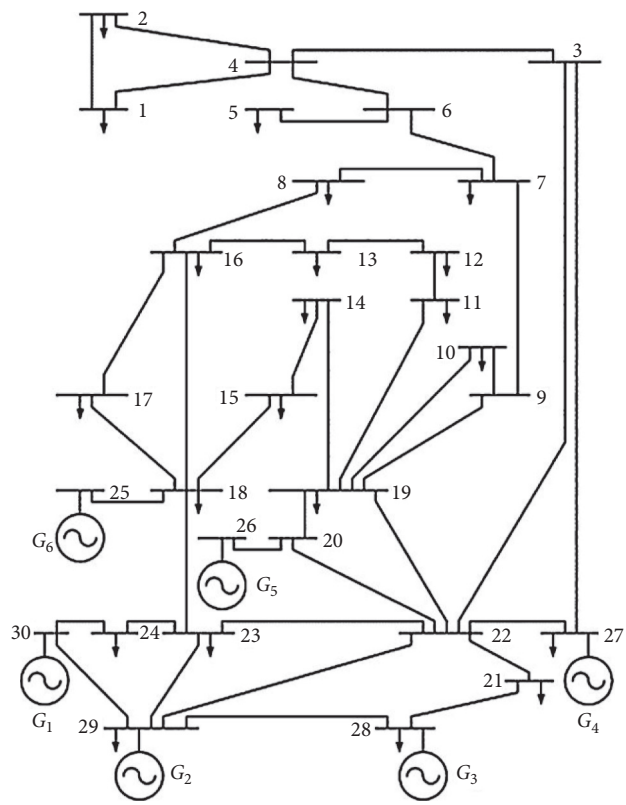


Figure 5: IEEE system line diagram of 30-bus 6 generators.

Table 1: Generator limits, fuel cost coefficients, and emission coefficients of the IEEE 30-bus system.

Generator (i)	Generation limits		Fuel cost coefficients			Emission coefficients				
	$P_{Gi\min}$	$P_{Gi\max}$	a_i	b_i	c_i	α_i	β_i	γ_i	$\xi_i \times 10^{-6}$	λ_i
1	0.05	0.50	10	200	100	4.091	−5.554	6.490	200	2.857
2	0.05	0.60	10	150	120	2.543	−6.047	5.638	500	3.333
3	0.05	1.00	20	180	40	4.258	−5.094	4.586	1.0	8.000
4	0.05	1.20	10	100	60	5.326	−3.550	3.380	2000	2.000
5	0.05	1.00	20	180	40	4.258	−5.094	4.586	1.0	8.000
6	0.05	0.60	10	150	100	6.131	−5.555	5.151	10	6.667

Table 2: Coefficient matrix of transmission B-loss for the test system: IEEE 30 bus.

B_{ij} =	0.02180	0.01070	−0.00036	−0.00110	0.00055	0.00330
	0.01070	0.01704	−0.00010	−0.00179	0.00026	0.00280
	−0.00040	−0.00020	0.02459	−0.01328	−0.01180	−0.00790
	−0.0011	−0.00179	−0.01328	0.0265	0.0098	0.0045
	0.00055	0.00026	−0.0118	0.0098	0.0216	−0.0001
	0.0033	0.0028	−0.00792	0.0045	−0.00012	0.02978
B_{0i} =	1.0731e−05	0.0017704	−0.0040645	0.0038453	0.0013832	0.0055503
B_{00} =	0.0014					

Table 3: The proposed algorithm parameters.

Parameter	Value
No. of particles	50
The iterations maximum number ($T_{\max}$)	100
Probability of generation (GP)	0.5
a_1	2
a_2	1
Number of clusters (K)	5–100

To examine the proposed algorithm, it is implemented to solve CEEDP with the following 3 cases of IEEE system:

(i) Case 1: single objective function, where the fuel cost was minimized as constrained nonlinear programming

(ii) Case 2: single objective function, where the generation emission was minimized as constrained nonlinear programming

(iii) Case 3: minimizing the two objective functions (fuel cost and generation emission) as a MOP using domination criteria

For the first two cases (case 1 and case 2), the obtained results by applying the proposed algorithm were compared to other optimization literature algorithms such as DRN-PSO [15], NSGA [16], NPGA [17], FCPSO [18], MSFLA [19], RCGA [20], and SPEA [21] as illustrated in Table 4. From Table 4, it was concluded that the proposed method provided a better minimum fuel cost for case 1 and found minimum generation emissions for case 2 than the other algorithms. For case 3, Figure 6 represents the Pareto optimal set for the CEEDP obtained by the proposed algorithm. From the figure, it can be concluded that the proposed algorithm is capable of getting the Pareto frontier for the CEEDP.

Figure 7 represents the convergence analysis of the proposed method to obtain the optimal solution. Figure 7(a) represents the cost function convergence versus the generation; it declares that the cost function has converged to its optimal value in the first 50 generations. On the other hand, Figure 7(b) represents the emission function convergence versus the generation; it declares that the emission function has converged to its optimal value in the first 150 generations. Figure 7(c) represents the locus of the ideal points which is starting at point $(631.7933, 0.2161)$ and ending at point $(599.0424, 0.1928)$. Finally, Figure 7(d) represents the locus of the nadir points which starts from point $(600.0432, 0.1932)$ and ends at point $(0.2204, 639.1063)$.

Implementation of dominance criteria provides not a single optimum solution, but a huge set of solutions as in Figure 1. For practical engineering applications, the selection of the best solution or limited set that will satisfy different objectives to some extent is the most important task. Also, minimizing the size of the Pareto set is available to the DM to select and save effort and time. Cluster analysis was implemented to shrink the size of the Pareto optimal set to a small set with a predetermined size K as in Table 5. Figures 8–11 represent some reduced Pareto sets depending on the chosen K parameter which may save effort to choose the best compromise alternative from the Pareto optimal set.

5.1. Discussion. It is clear from Table 4 that the suggested algorithm has done better than all state-of-the-art approaches in case 1 and case 2. While Figure 6 indicates that the proposed algorithm was able to solve CEEDP as a multiobjective optimization problem and gave a uniform distribution of the Pareto optimal curve and obtained very good solutions at both ends of the curve. Furthermore, Figure 7 indicates the accelerated convergence of the proposed algorithm, where the cost function has converged to its optimum value over the first 50 generations and the emission function has converged to its optimum value over the first 150 generations. In addition, Figures 8–11 indicate the potential of the proposed algorithm to help DM find a compromise solution on the whole Pareto front by splitting it into a variety of regions using the K-means clustering algorithm.

Table 4: Optimal solutions by different methods: case 1 and case 2.

Algorithm	Cases			
	Case 1 (minimizing fuel costs)		Case 2 (minimizing generation emissions)	
DRN-PSO [15]	591.1517	0.2150	643.8616	0.1949
NSGA [16]	600.3100	0.2238	633.8300	0.1946
NPGA [17]	600.2200	0.2206	636.0400	0.1943
FCPSO [18]	600.1300	0.2223	638.3577	0.1942
MSFLA [19]	600.1114	0.2225	638.2425	0.1942
RCGA [20]	611.6935	0.2285	648.5301	0.1932
SPEA [21]	600.3400	0.2241	640.4200	0.1942
Proposed approach	**599.0424**	0.2204	639.1063	**0.1928**

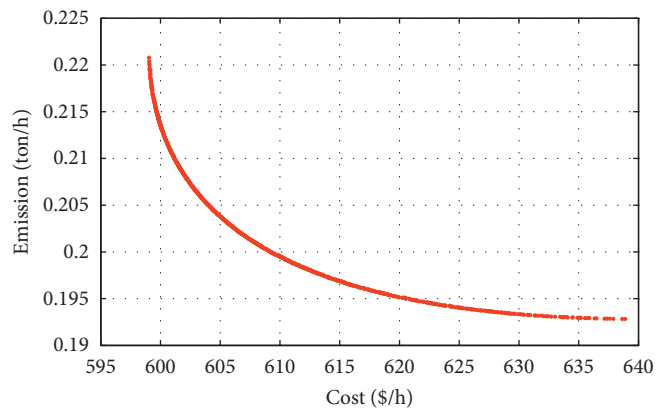


Figure 6: CEEDP Pareto optimal set obtained by the proposed algorithm.

A comparative study has been investigated in this subsection to examine the proposed method concerning the assumptions of convexity, the smoothness of the Pareto front, the number of the set of nondominated frontier, the handling constraint, and the closeness to the true Pareto optimal set.

Firstly, unfortunately, the CEEDP is a nonlinear optimization problem and having multimodal functions. Also, it needs many mathematical assumptions such as differential functions, analytics, and convexity. Therefore, traditional optimization techniques implement derivatives and gradients of a function, in general, not valid to detect the optimal solution. On the other hand, the proposed method does not need any assumption regarding derivatives, gradients, and convexity.

Other evolutionary-based algorithms lack the shortage in numbers of nondominated solutions in the Pareto frontier

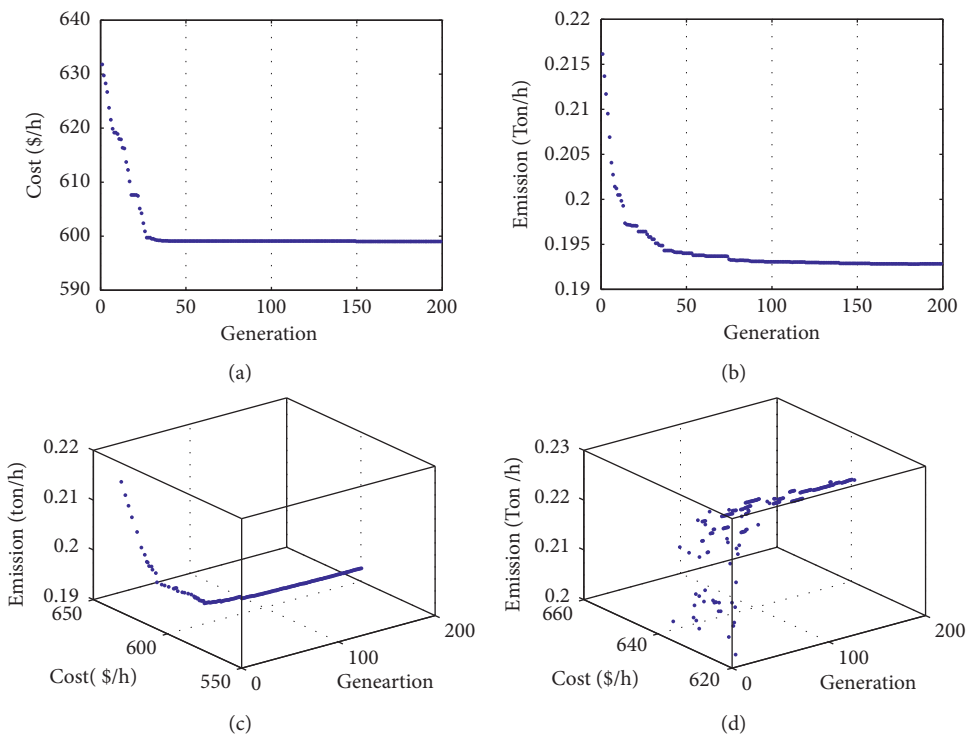


Figure 7: Pareto optimal convergence. (a) Cost ($/h) versus generation. (b) Emission (Ton/h) versus generation. (c) Ideal points versus generation. (d) Nadir points versus generation.

Table 5: Implementation of K-means method to reduce the shrink Pareto optimal frontier.

K	Size of reduced Pareto set	Ideal point
100	100	(0.1933, 601)
50	50	(0.1930, 599.8)
10	10	(0.1929, 599.1)
5	5	(0.1929, 599.1)

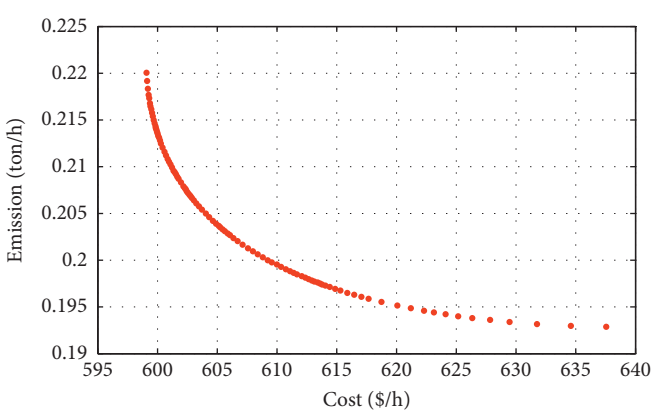


Figure 8: Compromising 100 alternatives from the Pareto set.

(does not cover the whole Pareto frontier). On the other side, the proposed method has been effectively exploring the whole search space using dominance criteria and to handle unfeasibility by implementing the repair algorithm which enables the algorithm to deal with constrained optimization. Also, the proposed algorithm is capable of efficiently discovering a large set of the Pareto front, spanning the whole Pareto optimal frontier.

Secondly, the proposed method keeps track of all nondominated solutions detected during the process of optimization (i.e., the algorithm has no restriction on the number of Pareto optimal solutions), by using an archiving-based selection technique that ensures the convergence towards the Pareto optimal set. On the other hand, the proposed algorithm is a meta-heuristic-based multiple-objective optimization method where it uses a population of particles in its search. So, multiple solutions (Pareto optimal) can be found in one single run. On the contrary to traditional techniques, which require a lot of runs to get the Pareto optimal solutions.

Furthermore, it is observed from the simulation analysis that the proposed method is absolutely better than other algorithms, where its behavior is completely good and it is capable of retaining the Pareto frontier for the CEEDP problem. Finally, unlike other methods, the proposed method does not leave the DM confused in front of choosing a suitable solution from among all the

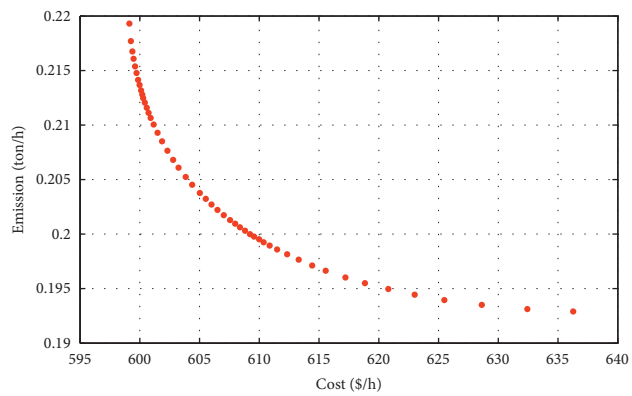


Figure 9: Compromising 50 alternatives from the Pareto set.

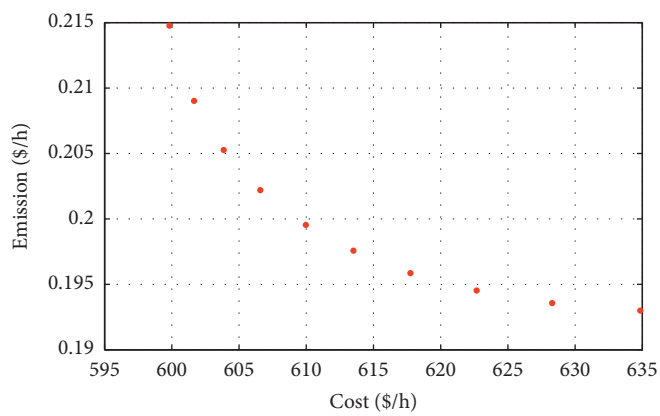


Figure 10: Compromising 10 alternatives from the Pareto set.

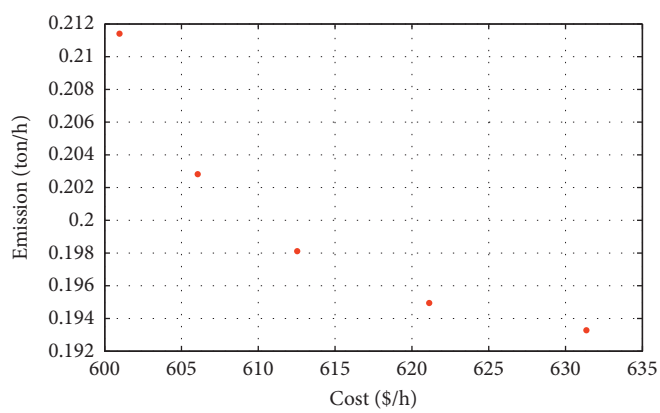


Figure 11: Compromising 5 alternatives from the Pareto set.

nondominated sets obtained but divides the Pareto front into a set of clusters with a center for each cluster by one of the most famous methods of cluster analysis, namely, *K*-means.

6. Conclusion

This paper investigates a new optimization system called constrained multiobjective equilibrium optimizer for dealing with the constrained combined economic emission dispatch problem (CEEDP) with two objectives that reflect minimizing the generation cost and minimizing the environmental pollution emission. The proposed algorithm incorporates the dominance criteria to handle the multiobjective functions that enable the DM to detect all the Pareto frontiers. Cluster analysis was implemented to decrease the size of the Pareto frontier to a reasonable size that reflects all the characteristics of the original Pareto frontier. On the other hand, the repair method was applied to handle the constraints and feasibility of particles. Solving the

standard 30-bus IEEE system demonstrates the superiority of the proposed algorithm to generate a well-distributed Pareto frontier.

The following points are the major contribution points of this research:

(a) The proposed algorithm was applied efficiently to deal with constrained multiobjective CEEDP effectively, without limiting the treatment of more than two objectives

(b) The obtained Pareto optimal frontier is well distributed

(c) By means of cluster analysis, the proposed algorithm according to DM needs help to track the resolution of the Pareto frontier

(d) Incorporating the repair method enables the algorithm to retain the particles feasibility

This paper can be developed as a future research study as follows:

(i)Constrained multiobjective equilibrium optimizer algorithm used in this analysis also needs further testing and development and can also be paired with other approaches to further improve its efficiency

(ii) Economic dispatch issues in this analysis may be expanded to cover losses in repairing engine and operating costs

Data Availability

All data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that this article content has no conflicts of interest.

Acknowledgments

The authors extend their appreciation for the great support of Deanship of Scientific Research, Taif University, for funding Taif University Researcher Supporting Project Number Tursp-2020/48, Taif University, Taif, Saudi Arabia.

References

[1] S. F. Brodesky and R. W. Hahn, "Assessing the influence of power pools on emission constrained economic dispatch," *IEEE Transactions on Power*, vol. 1, no. 1, pp. 57–62, 1986.

[2] M. S. Osman, M. A. Abo-Sinna, and A. A. Mousa, "An ε-dominance-based multiobjective genetic algorithm for economic emission load dispatch optimization problem," *Electric Power Systems Research*, vol. 79, no. 11, pp. 1561–1567, 2009.

[3] M. A. Hussein, A. A. EL-Sawy, E. L.-S. M. Zaki, and A. A. Mousa, "Genetic algorithm and rough sets based hybrid approach for economic environmental dispatch of power systems," *British Journal of Mathematics & Computer Science*, vol. 4, no. 20, pp. 2978–2999, 2014.

[4] A. Farag, S. Al-Baiyat, and T. C. Cheng, "Economic load dispatch multiobjective optimization procedures using linear programming techniques," *IEEE Transactions on Power Systems*, vol. 10, no. 2, pp. 731–738, 1995.

[5] C. S. Chang, K. P. Wong, and B. Fan, "Security-constrained multiobjective generation dispatch using bicriterion global optimisation," *IEEE Proceedings-Generation, Transmission and Distribution*, vol. 142, no. 4, pp. 406–414, 1995.

[6] J. X. Xu, C. S. Chang, and X. W. Wang, "Constrained multiobjective global optimisation of longitudinal interconnected power system by genetic algorithm," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 143, no. 5, pp. 435–446, 1996.

[7] Y. T. Hsiao, H. D. Chiang, C. C. Liu, and Y. L. Chen, "A computer package for optimal multi-objective VAR planning in large scale power systems," *IEEE Transactions on Power Systems*, vol. 9, no. 2, pp. 668–676, 1994.

[8] B. S. Kermanshahi, Y. Wu, K. Yasuda, and R. Yokoyama, "Environmental marginal cost evaluation by non-inferiority surface (power systems)," *IEEE Transactions on Power Systems*, vol. 5, no. 4, pp. 1151–1159, 1990.

[9] S. Mouassa and T. Bouktir, "Artificial bee colony algorithm for solving economic dispatch problems with non-convex cost functions," *International Journal of Power and Energy Conversion*, vol. 8, no. 2, pp. 146–165, 2017.

[10] R. K. Swain, N. C. Sahu, and P. K. Hota, "Gravitational search algorithm for optimal economic dispatch," *Procedia Technology*, vol. 6, pp. 411–419, 2012.

[11] S. Komsiyah, "Computational methods of Gaussian particle swarm optimization (GPSO) and lagrange multiplier on economic dispatch issues (case study on electrical system of Java-Bali IV area)," *The European Physical Journal*, EDP Sciences, vol. 68, , 2014.

[12] S. Arunachalam, T. AgnesBhomila, and M. Ramesh Babu, "Hybrid particle swarm optimization algorithm and firefly algorithm based combined economic and emission dispatch including valve point effect," in *Proceedings of the 5th International Conference on Swarm, Evolutionary, and Memetic Computing*, December 2014.

[13] S. Arunachalam, R. Saranya, and N. Sangeetha, "Hybrid artificial bee colony algorithm and simulated annealing algorithm for combined economic and emission dispatch including valve point effect," in *Proceedings of the 4th International Conference on Swarm, Evolutionary, and Memetic Computing*, pp. 354–365, Springer, Chennai, India, December 2013.

[14] M. Ahmad, W. Ali, H. Farooq, M. Jamil, M. Ali, and A. U. Rehman, "Solving the problem of economic load dispatch for a small scale power system using a novel hybrid PSO-GSA algorithm," in *Proceedings of the 2018 International Symposium on Recent Advances in Electrical Engineering (RAEE)*, pp. 1–6, IEEE, October 2018.

[15] M. A. El-hosseini, R. A. El-Sehiemy, and Y. H. Amira, "A multiobjective dynamic particle swarm optimizer for environmental/economic dispatch problem," *Engineering Research Journal*, vol. 37, no. 4, pp. 379–388, 2014.

[16] P. K. Hota, A. K. Barisal, and R. Chakrabarti, "Economic emission load dispatch through fuzzy based bacterial foraging algorithm," *International Journal of Electrical Power and Energy Systems*, vol. 32, no. 7, pp. 794–803, 2010.

[17] M. A. Abido, "A niched Pareto genetic algorithm for multiobjective environmental/economic dispatch," *International*

Journal of Electrical Power & Energy Systems, vol. 25, no. 2, pp. 97–105, 2003.

[18] S. Agrawal, B. K. Panigrahi, and M. K. Tiwari, "Multiobjective particle swarm algorithm with fuzzy clustering for electrical power dispatch," *IEEE Trans. on Evolutionary Computation.*vol. 12, no. 5, pp. 529–541, 2008.

[19] S. Reddy and K. Vaisakh, "Economic emission load dispatch by modified shuffled frog leaping algorithm," *International Journal of Computer Applications*, vol. 31, pp. 35–42, 2011.

[20] R. El-Sehiemy, M. A. El-Hosseini, and A. E. Hassanien, "Multiobjective real coded genetic algorithm for economic/environmental dispatch problem," *Studies in Informatics and Control*, vol. 22, no. 2, pp. 113–122, 2013.

[21] K. Deb, *Multi-objective Optimization Using Evolutionary Algorithms*, Wiley, New York,NY, USA, 2001.

[22] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," *Evolutionary Computation*, vol. 3, no. 1, pp. 1–16, 1995.

[23] A. A. Mousa and E. E. Elattar, "Best compromise alternative to EELD problem using hybrid multiobjective quantum genetic algorithm," *International Journal of Applied Mathematics & Information Sciences*, vol. 8, no. 6, pp. 2889–2902, 2014.

[24] M. A. Mellal and E. J. Williams, "Cuckoo optimization algorithm with penalty function and binary approach for combined heat and power economic dispatch problem," *Energy Reports*, vol. 6, pp. 2720–2723, 2020.

[25] M. J. Kim, T. S. Kim, R. J. Flores, and J. Brouwer, "Neural-network-based optimization for economic dispatch of combined heat and power systems," *Applied Energy*, vol. 265, Article ID 114785, 2020.

[26] A. Sundaram, "Multiobjective multi-verse optimization algorithm to solve combined economic, heat and power emission dispatch problems," *Applied Soft Computing*, vol. 91, Article ID 106195, 2020.

[27] M. F. Tabassum, M. Saeed, N. Ahmad Chaudhry et al., "Evolutionary simplex adaptive Hooke-Jeeves algorithm for economicload dispatch problem considering valve point loading effects," *Ain Shams Engineering Journal*, 2020.

[28] I. R. Fitri and J.-S. Kim, "Economic dispatch problem using load shedding: centralized solution," *IFAC-Papers OnLine*, vol. 52, no. 4, pp. 40–44, 2019.

[29] J. N. Kuk, R. A. Gonçalve, L. M. Pavelski, S. M. Guse Scós Venske, C. P. de Almeida, and A. T. Ramirez Pozo, "An empirical analysis of constraint handling on evolutionary multi-objective algorithms for the environmental/economic load dispatch problem," *Expert Systems with Applications*, vol. 165, Article ID 113774, 2021.

[30] D. Singh and J. S. Dhillon, "Ameliorated grey wolf optimization for economic load dispatch problem," *Energy*, vol. 169, pp. 398–419, 2019.

[31] O. T. Altinoz, "The distributed many-objective economic/emission load dispatch benchmark problem," *Swarm and Evolutionary Computation*, vol. 49, pp. 102–113, 2019.

[32] A. Sundaram and P. Erdogmus, "Solution of combined economic emission dispatch problem with valve-point effect using hybrid NSGA II-MOPSO," in *Particle Swarm Optimization with Applications*InTech, London, UK, 2017.

[33] X. Chen, K. Li, B. Xu, and Z. Yang, "Biogeography-based learning particle swarm optimization for combined heat and power economic dispatch problem," *Knowledge-Based Systems*, vol. 208, Article ID 106463, 2020.

[34] E. H. Talbi, L. Abaali, R. Skouri, and M. E. Moudden, "Solution of economic and environmental power dispatch problem of an electrical power system using BFGS-AL algorithm," *Procedia Computer Science*, vol. 170, pp. 857–862, 2020.

[35] S. Zhou, Z. Hu, W. Gu et al., "Combined heat and power system intelligent economic dispatch: a deep reinforcement learning approach," *International Journal of Electrical Power & Energy Systems*, vol. 120, Article ID 106016, 2020.

[36] A. Srivastava and D. K. Das, "A new Kho-Kho optimization Algorithm: An application to solve combined emission economic dispatch and combined heat and power economic dispatch problem," *Engineering Applications of Artificial Intelligence*, vol. 94, Article ID 103763, 2020.

[37] R. Zhang, K. Yan, G. Li, T. Jiang, X. Li, and H. Chen, "Privacy-preserving decentralized power system economic dispatch considering carbon capture power plants and carbon emission trading scheme via over-relaxed ADMM," *International Journal of Electrical Power & Energy Systems*, vol. 121, Article ID 106094, 2020.

[38] C. A. C. Coello and A. Carlos, "A survey of constraint handling techniques used with evolutionary algorithms," Technical Report Lania-RI-99-04, Laboratorio Nacional de Informática Avanzada, Xalapa, Mexico, 1999.

[39] Z. Michalewicz, "A survey of constraint handling techniques in evolutionary computation methods," in *Proceedings of the 4th Annual Conference on Evolutionary Programming*, vol. 4, pp. 135–155, San Diego, CA, USA, February 1995.

[40] Q. Long, "A constraint handling technique for constrained multi-objective genetic algorithm," *Swarm and Evolutionary Computation*, vol. 15, pp. 66–79, 2014.

[41] Y. Yang, J. Liu, and S. Tan, "A constrained multi-objective evolutionary algorithm based on decomposition and dynamic constraint-handling mechanism," *Applied Soft Computing*, vol. 89, Article ID 106104, 2020.

[42] F. Samanipour and J. Jelovica, "Adaptive repair method for constraint handling in multi-objective genetic algorithm based on relationship between constraints and variables," *Applied Soft Computing*, vol. 90, Article ID 106143, 2020.

[43] J. Xu, P. Yang, G. Liu, Z. Bai, and W. Li, "Constraint handling in constrained optimization of a storage ring multi-bend-achromat lattice," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 988, Article ID 164890, 2020.

[44] A. A. Mousa and M. A. El-Shorbagy, "Identifying a satisfactory operation point for fuzzy multiobjective environmental/economic dispatch problem," *American Journal of Mathematical and Computer Modelling*, vol. 1, no. 1, pp. 1–14, 2016.

[45] A. A. Mousa, M. A. El-Shorbagy, and M. A. Farag, "K-means-Clustering based evolutionary algorithm for multi-objective resource allocation problems," *Applied Mathematics & Information Sciences*, vol. 11, no. 6, pp. 1–12, 2017.

[46] M. A. El-Shorbagy, A. Y. Ayoub, I. M. El-Desoky, and A. A. Mousa, "A novel genetic algorithm based k-means algorithm for cluster analysis," in *International Conference on Advanced Machine Learning Technologies and Applications*, pp. 92–101, Springer, Cham, Switzerland, 2018.

[47] H. P. Benson and S. Sayin, "Towards finding global representations of the efficient set in multiple objective mathematical programming," *Naval Research Logistics*, vol. 44, no. 1, pp. 47–67, 1997.

[48] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.

[49] M. A. E. Shorbagy, A. A. Mousa, and M. Farag, "Solving nonlinear single-unit commitment problem by genetic

algorithm based clustering technique," *Review of Computer Engineering Research*, vol. 4, no. 1, pp. 11–29, 2017.

[50] P. Saurabh and B. Verma, "An efficient proactive artificial immune system based anomaly detection and prevention system," *Expert Systems with Applications*, vol. 60, pp. 311–320, 2016.

[51] A. A. A. Mousa and M. A. El-Shorbagy, "Enhanced particle swarm optimization based local search for reactive power compensation problem," *Applied Mathematics*, vol. 3, no. 10, pp. 1276–1284, 2012.

[52] A. A. Mousa and I. M. El_Desoky, "Stability of Pareto optimal allocation of land reclamation by multistage decision-based multipheromone ant colony optimization," *Swarm and Evolutionary Computation*, vol. 13, pp. 13–21, 2013.

[53] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Technical Report-TR06, Erciyes University, Melikgazi, Turkey, 2005.

[54] W. Zhao and L. Wang, "An effective bacterial foraging optimizer for global optimization," *Information Sciences*, vol. 329, pp. 719–735, 2016.

[55] L. Guo, Z. Meng, Y. Sun, and L. Wang, "Parameter identification and sensitivity analysis of solar cell models with cat swarm optimization algorithm," *Energy Conversion and Management*, vol. 108, no. 15, pp. 520–528, 2016.

[56] M. Marinaki and Y. Marinakis, "A glowworm swarm optimization algorithm for the vehicle routing problem with stochastic demands," *Expert Systems with Applications*, vol. 46, pp. 145–163, 2016.

[57] S. Verma and V. Mukherjee, "Firefly algorithm for congestion management in deregulated environment," *Engineering Science and Technology, An International Journal*, vol. 19, no. 3, pp. 1254–1265, 2016.

[58] Y. Zhou, X. Chen, and G. Zhou, "An improved monkey algorithm for a 0-1 knapsack problem," *Applied Soft Computing*, vol. 38, pp. 817–830, 2016.

[59] A. L. A. Bolaji, M. A. Al-Betar, M. A. Awadallah, A. T. Khader, and L. M. Abualigah, "A comprehensive review: krill Herd algorithm (KH) and its applications," *Applied Soft Computing*, vol. 49, pp. 437–446, (2016).

[60] M. Shehab, A. T. Khader, M. Laouchedi, and O. A. Alomari, "Hybridizing cuckoo search algorithm with bat algorithm for global numerical optimization," *The Journal of Supercomputing*, vol. 75, no. 5, pp. 2395–2422, 2019.

[61] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.

[62] Y. Abo-elnaga and M. A. El-Shorbagy, "Multi-sine cosine algorithm for solving nonlinear bilevel programming problems," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 421–432, 2020.

[63] M. A. El-Shorbagy and A. M. El-Refaey, "Hybridization of grasshopper optimization algorithm with genetic algorithm for solving system of non-linear equations," *IEEE Access*, vol. 8, pp. 220944–220961, 2020.

[64] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: a bio-inspired optimizer for engineering design problems," *Advances in Engineering Software*, vol. 114, pp. 163–191, 2017.

[65] A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, "Equilibrium optimizer: a novel optimization algorithm," *Knowledge-Based Systems*, vol. 191, Article ID 105190, 2020.

[66] I. Ahmadianfar, O. Bozorg-Haddad, and X. Chu, "Gradient-based optimizer: a new Metaheuristic optimization algorithm," *Information Sciences*, Information Sciences, vol. 540, , pp. 131–159, 2020.

[67] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: a new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020.

[68] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.

[69] C. A. Coello Coello, S. González Brambila, J. Figueroa Gamboa et al., "Evolutionary multiobjective optimization: open research areas and some challenges lying ahead," *Complex & Intelligent Systems*, vol. 6, pp. 221–236, 2020.

[70] A. A. Mousa, M. A. El-Shorbagy, and W. F. Abd El-Wahed, "Local search based hybrid particle swarm optimization for multiobjective optimization," *International Journal of Swarm and Evolutionary Computation*, vol. 3, pp. 1–14, 2012.

[71] K. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, MA, USA, 1999.

[72] J. Zahavi and L. Eisenberg, "Economic–environmental power dispatch," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 5, no. 5, pp. 485–489, 1985.

[73] D. Hazarika and P. K. Bordoloi, "Modified loss coefficients in the determination of optimum generation scheduling," *IEE Proceedings C Generation, Transmission and Distribution*, vol. 138, no. 2, pp. 166–172, 1991.

[74] W. Ng, "Generalized generation distribution factors for power system security evaluations," *IEEE Transactions on Power Apparatus and Systems*, vol. 100, no. 3, pp. 1001–1005, 1981.

[75] M. A. Farag, M. A. El-Shorbagy, A. A. Mousa, and I. M. El-Desoky, "A new hybrid metaheuristic algorithm for multi-objective optimization problems," *International Journal of Computational Intelligence Systems–Atlantis Press Publisher*, vol. 13, no. 1, pp. 920–940, 2020.

[76] M. A. El-Shorbagy, A. Y. Ayoub, A. A. Mousa, and I. M. El-Desoky, "An enhanced genetic algorithm with new mutation for cluster analysis," *Computational Statistics*, vol. 34, pp. 1355–1392, 2019.

[77] M. A. E. Shorbagy and A. A. Mousa, "Chaotic particle swarm optimization for imprecise combined economic and emission dispatch problem," *Review of Information Engineering and Applications*, vol. 4, no. 1, pp. 20–35, 2017.

[78] S. Hemamalini and P. S. Sishaj, "Emission constrained economic dispatch with valve point effect using particle swarm optimization," in *Proceedings of the 2008 TENCON*, Hyderabad, India, November 2008.

[79] R. Zimmerman and D. Gan, "MATPOWER: a matlab power system simulation package," 1997, http://www.pserc.cornell.edu/matpower/.

Hindawi
Complexity
Volume 2021, Article ID 6660568, 24 pages
https://doi.org/10.1155/2021/6660568

Research Article

Mechanism of User Participation in Co-creation Community: A Network Evolutionary Game Method

Fanshun Zhang,[1] **Congdong Li**,[1] **and Cejun Cao**[2]

[1]*School of Management, Jinan University, Guangzhou 510632, China*
[2]*School of Management Science and Engineering, Chongqing Technology and Business University, Chongqing 400067, China*

Correspondence should be addressed to Congdong Li; licd@jnu.edu.cn and Cejun Cao; caocejun0601@tju.edu.cn

Received 15 November 2020; Revised 17 December 2020; Accepted 21 December 2020; Published 12 January 2021

Academic Editor: Baogui Xin

Copyright © 2021 Fanshun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Active participation closely associates with the sustainable operation of co-creation communities. Different from recent studies on the promotion of sustainable operation by identifying the internal and external motivations of user participation, this paper aims to analyze the mechanism regarding how different motivations affect the decision of user participation from group-level perspective. To better understand the mechanism, internal and external motivations are, respectively, captured by return-cost analysis and user interactive network. Afterwards, a network evolutionary game model was formulated to analyze the dynamic strategy selection (e.g., active participation and passive participation) of all users. In addition, the stable equilibrium and evolutionary path of strategies are analyzed through computational experiments. Results indicate the following: (a) Rewards have an influence on the promotion of active participation. However, with the continued growth of rewards, this promotion does not make sense sustainably. (b) The promotional effect of information noise on the selection of active participation can be found when passive participation is the dominant strategy. However, the inhibitory effect can be seen in populations that mainly adopt active participation. (c) The scale-free feature of user interactive network inhibits the selection of active participation when active participation is the dominant strategy in populations. Results found here is beneficial for managers to implement the specified policies and thus to achieve the sustainability of co-creation community.

1. Introduction

Co-creation community has significantly changed the role of users by breaking up the difference between producers and customers [1] and brought greater opportunities for promoting revenues and profits of companies [2]. About 60% of the world's biggest companies use online communities, and they allow users to share experiences and contribute creativity for product/service improvement [3]. Since every user acts as a content creator in co-creation communities, user-generated contents play a critical role in product/service innovation [4]. However, many co-creation communities fail to survive because of the insufficient user participation [5]. Consequently, the key challenge for managers is to encourage user participation and to create a thriving community [4, 5].

User participation, which is defined as the extent to which users actively engage in the community's activities and interact with other users [6], is an emerging research topic in the field of marketing science [7] and information management systems [8]. Based on the definition of [6], users in co-creation communities can be divided into two types: passive users and active users. The former receives useful knowledge from communities without shared behavior. The latter is glad to share their experiences, ideas, and intelligences in a voluntary way, which plays an indispensable role in performance improvement of product innovation and profit making. However, Nielsen [9] portrayed that participation inequality was concluded as the result of the 90-9-1 rule. 90% of users are lurkers (e.g., they read or observe, but they do not contribute). 9% of users contribute occasionally. 1% of users actively participate and account for

most contributions. Thus, extensive studies on motivational identification of user participation were discussed from both internal and external perspectives, to promote active participation and thus to achieve the sustainability of co-creation community.

Based on the theory of personal behavior (e.g., commitment theory) and social behavior (e.g., social interaction theory), studies on internal aspects explained the motivation of user participation according to some views from psychological science and social science [4]. For example, Xu and Li [8] examined the influence of extrinsically oriented motivation (e.g., reputation and reciprocity) and intrinsically oriented motivation (e.g., altruism and enjoyment) on user participation. Zhou [10] tested the effect of social interaction (e.g., compliance and identification) on user participation. Tsai et al. [6] defined community participation as the interaction among members and their activity involvement, and several influential factors were analyzed from individual aspect (e.g., extraversion and need for affiliation), group aspect (e.g., identification and perceived critical mass), and relationship aspect (e.g., relationship satisfaction and relational trust). Roberts et al. [11] divided motivations of user participation into three aspects: motivations of innovation (e.g., interesting, passion, and skill development), motivations of contributing to community innovation activities (e.g., recognition, feedback, and ties to community) and motivations of directly collaborating with companies (e.g., desire for better products, career opportunities and economic goals). Chen et al. [12] discussed the intention of future participation from a perspective of knowledge sharing. They verified that knowledge sharing behavior had a great influence on user participation via customer learning value, socially integrative value, and hedonic value. On the other hand, a few studies concentrate on external motivations that relate to the external environment of communities. For example, Fang and Zhang [13] found that the network size which closely associated with number of users had a great influence on lurkers' participation. Based on the extensive studies on motivation of user participation, Malinen [14] concluded that group processes also affected online participation as well. For example, individuals (especially those with similar interests) associating with each other have a great influence on their decision.

Although many empirical studies were conducted to identity motivation of user participation, valuable topics should be further discussed on three aspects:

Firstly, motivations identified by researchers are based on questionnaire investigation for individual users, which ignores user interaction in co-creation communities. Basically, users build their interactive networks by publishing and responding to posts, and users who directly interact with each other over this network can be motivated to participate [15]. For example, user interaction was treated as group-level variables to analyze the influence of the interactive network on user participation [16]. *Secondly,* the initial motivations were identified by lots of researchers based on the assumption that the process of user participation is monocyclic and static [17]. However, user participation is a dynamic process in which users need to continuously decide whether to actively participate and when to actively participate next time. Users may choose active participation at present, adopt a passive strategy later, and use a combined strategy in the end. In this process, the decision is influenced by diverse motivations, and the influence of different motivations on user participation changes over time [14, 18]. For example, free knowledge acquisition motives active participation at first, while rewards attract users to select the enduring active participation. *Thirdly*, despite the significant contributions made by existing studies on motivation identification, analyses of a mechanism regarding how different motivations influence user participation are still scarce in these studies. In other words, a mechanism regarding whether to actively participate, when to actively participate, and how such a strategy can perform well has not been established. As [17] argued, an analytical and mathematical method is needed to study the mechanism by using controlled variables and so as to understand the impact of various motivations on user participation. This work will bring opportunities for prediction of user participation behavior and creation of personalized services for different users [19]. A similar viewpoint of mechanism design based on different motivations was also suggested by [20].

Based on statements previously mentioned, the research question of this paper is how to formulate an adaptive model involving user interaction, dynamic decision-making processes, and motivations of users to analyze the mechanism regarding how different motivations affect the decision of user participation from group-level perspective. Fortunately, the network evolutionary game approach provides a powerful tool to explain the emergency of the given strategies (e.g., active participation and passive participation) in structural populations.

Three main tasks are implemented. *Firstly,* internal motivations (e.g. rewards, reputations, time costs and knowledge spillovers) are treated as returns and costs of users in the decision-making processes of user participation. A traditional evolutionary game model (TEGM) based on return-cost analysis is developed to analyze the stably evolutionary strategies (e.g. active strategy, passive strategy, or mixed strategy) between two individual users who are randomly selected from co-creation communities. *Secondly,* to integrate user interaction into the dynamic decision-making processes, a network evolutionary game model (NEGM) is applied to analyze the strategy selection of all users who are connected by a user interactive network. In this model, not only the effect of internal motivations but also the influence of external motivations (e.g., network size, network structure, and information noise) can be well discussed. *Thirdly,* since the evolution of user participation behavior has non-linear features, computational experiments are used to model the dynamic decision-making processes. The strategy evolution of user participation is presented by computational experiments.

The rest of this paper is organized as follows. Section 2 presents a critical literature review. Section 3 introduces the TEGM and the NEGM. The computational experiments are implemented in Section 4. Finally, conclusions and future directions are discussed in Section 5.

2. Literature Review

The relevant studies are reviewed from three angles: (a) user participation in co-creation community; (b) modeling user participation through an evolutionary game model in networks; and (c) discovering the dynamic processes of strategy selection via computational experiments.

2.1. User Participation in Co-Creation Community. Since user participation closely associates with the sustainability of co-creation community, it has received a particularly growing attention in recent years. For example, Kohler et al. [21] highlighted the importance of experience in encouraging active participation, and they found that active participation was enhanced by enjoyable experiences (e.g., inspiring, intrinsically motivating, involving, and funny). Akman et al. [22] divided co-creation activities into four types, and they examined a mechanism regarding why users contributed their creativities in multiple activities. Based on different styles of employee communication, Lee and van Dolen [23] tested the influence of individual user sentiment on user behavior (e.g., creative behavior and community participation). To explore the reasons for volunteer participation in value co-creation activities, Lee and Kim [24] employed expectancy-value theory and empirically identified that three customer benefits (e.g., cognition, social integration, and hedonism) significantly affected their intentions of constant participation. According to the Stimulus-Organism paradigm, Kamboj and Rahman [4] developed a theoretical model to examine antecedents of branding co-creation from the perspective of user participation. Different from many studies focusing on motivational identification, Chepurna and Criado [25] tried to identify the barriers of value co-creation. Based on motivational aspects and deterrent identification, they developed a theoretical model to understand user participation behavior. Shen et al. [26] found that behavior of value co-creation was achieved by the interactivity of online community and online trust. Chen et al. [27] proposed a knowledge-sharing model to illustrate the reason why users choose to share their experiences for co-creation activities.

In summary, the literature mentioned onwards discussed user participation or user behavior in co-creation community from different perspectives. Table 1 presents the summary of these relevant studies. Five parts, including research question, motivational types, decision-making process, consideration of user interaction, and research method, are detailed to conclude the research focus of recent studies. In particularly, motivated by [4], motivations for user participation can be divided into two categories. On the one hand, internal motivation closely associates with the cognition of individual users (e.g., self-development, interest, social communication, and a sense of belonging to a social group). On the other hand, external motivation refers to external environments (e.g., network size, network structure, and information noise). As Jiang et al. [18] portrayed, recent studies assumed that the influence of initial motivation sustained for a long time, and users did not change their decisions over a longer time span. However, users' motivations often change over time, and users will change their strategies all the time. Therefore, the decision-making process is divided into two types: static process and dynamic process. Shen et al. [26] argued that online interaction was the critical antecedent of value co-creation behavior. The perceived interactivity, such as perceived quality and variety of information sharing, had a positive influence on active participation in co-creation activities via online trust. Therefore, user interaction is given extra consideration in the decision-making process of user participation in co-creation community. Research method is a suitable artifact (e.g., theoretical model, empirical research, and mathematical model) to contributes their main ideas.

From Table 1, user (continued) participation is a hot topic which is widely discussed in many papers. Based on the assumption of a static decision-making process, lots of researchers employed empirical methods to examine the influence of internal motivations on user participation from the investigation of individual users. However, few studies have investigated a mechanism of user participation (which also can be understood as a strategy selection of active participation or passive participation). In other words, there is a dearth of research directed towards analyzing the question concerning how user interaction influences the decision of user participation. Besides, a mathematical model for analyzing the dynamic decision-making processes is absent [14]. Thus, how to employ a mathematical model characterizing user participation as a dynamic decision-making problem is the special contribution of this paper.

2.2. Modeling User Participation through an Evolutionary Game Model in Networks. User participation behavior is a multiperiod decision-making process which users may continuously change their strategy based on different internal and external motivations from time to time. Vassileva [20] suggested that the mathematical model based on game-theoretical analysis was a good method for modeling this process. Particularly, as a popular method for modeling social interaction, TEGM provides an effective decision-making guidance for players with multiperiod games [28, 29]. TEGM assumes that players have bounded rationality (e.g., players are reluctant to acknowledge complete information of others). For the first game with new players, they will make decisions based on the evaluation of their returns and costs. Next, they will adjust their strategies based on their previous actions (especially the successful strategy). Replicator dynamics, which is denoted by differential equation, describes the dynamic frequency change of the discrete decision in evolutionary games.

Based on different motivations, users may choose active participation at present, adopt a passive strategy later, and use a combined strategy in the end. Different motivations can be transformed in accordance with the returns and costs of users in multiperiod games. All users will adjust their strategies to maximize their returns (e.g., knowledge acquisition, reputation improvement, and satisfaction from

Table 1: Summary of recent studies on user participation in co-creation community.

Study (year)	RQ or RO	Types of motivations		Decision-making process		User interaction			Research method	
		Internal	External	Static	Dynamic	Yes	No	TM	ER	MM
Reference [21] (2011)	Attracting sustainable engagement	✓		✓			✓		✓	
Reference [22] (2019)	Sustainable participation	✓		✓			✓		✓	
Reference [23] (2015)	Creativity or participation	✓			✓		✓		✓	
Reference [24] (2018)	Continued user participation	✓		✓			✓		✓	
Reference [4] (2017)	Intention to active participation	✓		✓			✓		✓	
Reference [25] (2018)	Barriers identification of co-creation	✓	✓	✓		✓		✓		
Reference [26] (2020)	Intention for co-creation activities	✓	✓	✓			✓		✓	
Reference [27] (2018)	Sustainable contribution	✓		✓			✓		✓	
Reference [6] (2012)	Promotion of participation	✓		✓			✓		✓	
Reference [11] (2014)	Engagement in innovation	✓		✓		✓		✓		
Reference [13] (2019)	User continued participation	✓	✓	✓			✓		✓	
This paper	Mechanism of user participation	✓	✓		✓	✓				✓

social exchange) in each game. Based on statements mentioned previously, TEGM provides a useful tool for modeling return-cost-based game between two players who are randomly selected from co-creation communities. However, user interaction, which is formulated by user-interactive behaviors (e.g., publish a post and respond to a post), is hard to be well analyzed in this model. Besides, it is also difficult to describe the influence of external environments on user participation. Therefore, researchers attempt to find another appropriate method to model user participation behaviors.

Fortunately, as an extensive method of TEGM, NEGM characterizes user interaction and dynamic decision-making process as a multiagent game model. This method was initially employed to analyze human cooperation in the field of physical science. NEGM fills up the gaps in which some factors (e.g., user interaction and the influence of external environment) are ignored in TEGM [30]. Particularly, user interaction is captured by a user interactive network where the connected edges indicate the presence of user interaction. Two-player game in NEGM generally occurs in two players who are connected by the given user interactive network. Besides, the consideration of user interaction is beneficial to examine the influence of external motivations (e.g., group preference, number of users and information noise caused by external environments) on user participation behaviors. Specially, group preference can be interpreted as the scale-free characteristic in complex networks. The number of users can be captured by the network size.

In recent years, NEGM has been widely used in the application of multiagent systems. For example, Esmaeilyfard et al. [31] developed a game-theoretical model to analyze user participation in ephemeral social vehicular networks. Al-Dhanhani et al. [17] employed a multiagent game model to discuss human cooperation in social applications under four circumstances (e.g., traditional Tit for Tat with fixed history, the generous Tit for Tat, reputation-based Tit for Tat, and group reputation-based Tit for Tat). Xu et al. [32] developed a multiagent game model for exploring user participation in collaborative filtering-based recommendation systems, and they found that the satisfactory recommendation depended on a high expectation of recommendation quality. Liu et al. [33] characterized patent cooperation as a multiagent system, and they constructed an industry and university & research model to analyze the strategy selection of all users. Jiang et al. [18] investigated the evolution of knowledge-sharing behavior via a multigame model. However, there is a dearth of research directed towards investigating user participation in co-creation community [20].

Accordingly, the focus of this paper (e.g., internal motivations, external motivations, dynamic decision-making process, and consideration of user interaction) can be well modeled by the mathematical method of NEGM.

2.3. Discovering the Dynamic Processes of Strategy Selection via Computational Experiments. In TEGM, the dynamic processes of strategy selection are easily addressed by listing average returns of each player, analyses of replicator dynamics, calculation of stable evolutionary equilibrium, and presentation of strategy selection path. However, user participation in co-creation community is not determined by a simple accumulation of all results held in a single period of time. Analyses of NEGM involve the formulation of user interactive network, the establishment of an evolutionary game model, and the setting of evolutionary rules (e.g., how to change user's strategy based on their previous actions) in user interactive network [30]. In other words, this process modeled by NEGM has a nonlinear characteristic, which is difficult to be solved by an analysis tool in TEGM.

Fortunately, the computational experiment is a useful tool to understand the complexity and dynamics of the user participation system. The dynamic decision-making process can be well analyzed by setting environments, defining a theoretical model, designing an algorithm to simulate a multiagent system, and identifying the influence of different parameters on the decision made by players. Therefore, this method is widely applied to studying user interaction and user behaviors in social communication [18].

Motivated by [18], a user participation system can be divided into four parts. First, the user interactive network is set as the environment which influences the strategy

selection of all users. Second, TEGM is set as the theoretical model based on return-cost analysis to describe the game between two players who are randomly selected from populations. Third, the multi-agent system of user participation is developed to present the dynamic decision-making process including network setting, evolutionary rules, and TEGM. Fourth, the influence of different parameters (e.g., network structure and network size) on user participation can be analyzed based on the result of computational experiments.

3. NEGM of User Participation in Co-Creation Community

User participation in co-creation community is a decision-making process which many users combined by user interactive network make decision on whether to active-participate or passive-participate in communities. According to literature review, this process can be captured by a multiagent model which integrates micro individual behavior (e.g., game between two users) and macroscopic phenomenon (e.g., individual behavior influenced by user interaction) into a systematic framework. Next, the micro individual behavior is described by TEGM in Section 3.1. The macroscopic phenomenon is elaborated by NEGM in Section 3.2. The framework of this paper is presented in Figure 1.

3.1. TEGM of User Participation and Its Analysis. In this section, the individual behavior of all users is formulated as a two-players game. Problem description, model formulation, and model analysis are detailed as follows.

3.1.1. Problem Description. Individual behavior of user participation is a decision-making process which user needs to decide whether to active participate or passive participate in a cocreation community. Since user participates in communities via online communicating technology, users fail to know what strategies are chosen by other users and how many returns are obtained by other users. Thus, users usually make their decisions based on the return-cost analysis which is abstracted from internal motivations. Table 2 shows the measurement of returns and costs in this decision-making process.

On one hand, users can obtain returns from their participations. Specially, internal motivations (e.g., interest, making friends, and reputations) account for active participation, while passive participation is driven by free way of knowledge acquisition and social communications. Besides, rewards paid by communities motivate active participation in communities. On the other hand, users should pay some costs, such as time, relationship maintenance, and a little money paid for knowledge acquisition.

3.1.2. Model Formulation. To develop a suitable model for better understanding problem description, the basic assumptions are elaborated as follows:

Assumption 1. Given that there is no difference among all users, two players in TEGM are the same stakeholders. In other words, a symmetric TEGM is proposed here.

Assumption 2. When a user decides to participate in a community, two strategies regarding active participation and passive participation can be chosen. They will make decisions based on their return-cost analysis. To present a comprehensive overview of various factors, parameter setting is summarized in Table 3.

Assumption 3. Denote x as the probability that user B chooses active strategy, and 1-x indicates the probability that user B chooses passive strategy. For user A, y and 1-y respectively represent the probability of choosing active participation or choosing passive participation. Three interactive modes (e.g. {active, passive}, {passive, passive}, and {active, active}) between two players are obtained from game matrix (see Table 4). The best situation is that both two players simultaneously choose active strategy, which brings opportunities for sustainable operation of communities.

3.1.3. Model Analysis. Based on the literature review, replicator dynamics method [34] is employed to analyze the TEGM.

The expected returns of user B that adopts active strategy are

$$U_{\mathrm{BA}} = y(R_1 - C_1) + (1 - y)(R_1 - C_1 + W). \tag{1}$$

The expected returns of user B that adopts passive strategy are

$$U_{\mathrm{BP}} = y(R_2 - C_2). \tag{2}$$

The average expected returns of user B under the mixed strategies are

$$\overline{U_B} = xU_{\mathrm{BA}} + (1 - x)U_{\mathrm{BP}} = (C_2 - W - R_2)xy + (R_1 - C_1 + W)x + (R_2 - C_2)y. \tag{3}$$

Similarly, the average expected returns of user A under the mixed strategies are

$$\overline{U_A} = (C_2 - W - R_2)xy + (R_1 - C_1 + W)y + (R_2 - C_2)x. \tag{4}$$

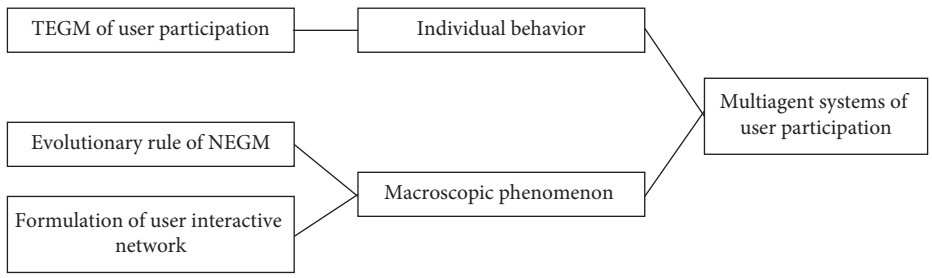


Figure 1: Framework of this paper.

Table 2: Measurement of returns and costs.

Items	Performance measurements
Returns	Self-development, enjoyment, knowledge learning, recognition, making friends, empowerment, need for affiliation, identification, relationship satisfaction, ties to community, reciprocity, reputation, passion, and other internal motivations
Costs	Lots of time, relationship maintenance, their own knowledge, a few moneys paid for knowledge acquisition
Rewards	Rewards from community, including tangible rewards and intangible rewards

Table 3: Parameter setting of TEGM.

Parameter	Description
For users who passively participate in co-creation community	
R_2	Returns obtained by passive users, such as knowledge spillover.
C_2	Costs paid by passive users are their time, and few moneys paid for knowledge acquisition.
For users who actively participate in co-creation community	
R_1	Returns obtained by active users depend on feature of users. For example, narcissistic person prefers reputations from other users. Social enthusiasts concentrate on making friends, social communications, and relationship satisfaction.
C_1	Costs paid by active users include their time, knowledge, and relationship maintenance.
W	Rewards paid by co-creation community and free to acknowledge knowledge.

Table 4: Game matrix of user participation in co-creation community.

		User A	
		Active (y)	Passive ($1-y$)
User B	Active (x)	$R_1 - C_1$; $R_1 - C_1$	$R_1 - C_1 + W$; $R_2 - C_2$
	Passive ($1-x$)	$R_2 - C_2$; $R_1 - C_1 + W$	0; 0

According to the theory of TEGM, if the expected returns of a strategy selected by one user are greater than the average expected returns of populations, this strategy will spread in the population [28]. The frequency of such strategies can be described as the replicator dynamics which is denoted by the differential equation. Specially, the replicator dynamics system of user B is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x\left(U_{\mathrm{BA}} - \overline{U}\right) = x(1-x)\left[\left(R_1 - C_1 + W\right) + \left(C_2 - W - R_2\right)y\right]. \tag{5}$$

Similarly, the replicator dynamics system of user A is

$$\frac{\mathrm{d}y}{\mathrm{d}t} = y\left(U_{\mathrm{AA}} - \overline{U}\right) = y(1-y)\left[\left(R_1 - C_1 + W\right) + \left(C_2 - W - R_2\right)x\right]. \tag{6}$$

Based on equations (5) and (6), five equilibrium points, namely, $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$ and (x^*, y^*), are obtained, where

$$x^* = y^* = \frac{C_1 - W - R_1}{C_2 - W - R_2}. \tag{7}$$

The stability of equilibrium points is obtained by analysis of the Jacobian matrix of replicator dynamics system. Based on [34], the Jacobian matrix is defined by equation (8):

$$J = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \tag{8}$$

where

$$\begin{aligned} a_{11} &= (1-2x)\left[\left(R_1 - C_1 + W\right) + \left(C_2 - W - R_2\right)y\right], \\ a_{12} &= x(1-x)\left(C_2 - W - R_2\right), \\ a_{21} &= y(1-y)\left(C_2 - W - R_2\right), \\ a_{22} &= (1-2y)\left[\left(R_1 - C_1 + W\right) + \left(C_2 - W - R_2\right)x\right]. \end{aligned} \tag{9}$$

The determinant and the trace of Jacobian matrix are, respectively, denoted by $\det J$ and $\mathrm{tr} J$, where

$$\det J = a_{11}a_{22} - a_{12}a_{21},$$
$$\operatorname{tr} J = a_{11} + a_{22}. \tag{10}$$

When $\det J > 0$, $\operatorname{tr} J < 0$, the equilibrium point is locally stable. In other words, the equilibrium point is the evolutionary stable strategy (ESS) of TEGM.

Based on the analysis mentioned onwards, three ESSs are discussed as follows:

Situation 1. If $R_1 - C_1 < R_2 - C_2$ and $W < C_1 - R_1$, ESS of group will reach "passive participation."

Situation 2. If $R_1 - C_1 < R_2 - C_2$ and $W > C_1 - R_1$, ESS of group will reach "active participation."

Situation 3. If $R_1 - C_1 < R_2 - C_2$ and $W > \max\{C_1 - R_1, C_2 - R_2\}$, ESS of group will reach the mixed strategies of $(1, 0)$ and $(0, 1)$. It indicates that active and passive strategy are interchangeably chosen by users.

TEGM based on replication dynamics method is beneficial to analyze the multiperiod games of two players who are randomly selected from population and to design a mechanism of user participation in a microenvironment. However, evolutionary results of user participation are not a simple accumulation of all game states held in a single time period. Such evolutionary results, which are significantly influenced by the decision of other users, community's environment (e.g., group preference and number of users), and information noise, need to be considered for modeling user participation behavior in a macroscopic environment.

In sum, since the evolution has nonlinear features and evolutionary equilibrium points are hard to be calculated by TEGM, NEGM including game theory and complex network is considered next.

3.2. NEGM of User Participation in Co-Creation Community. Once a user participates in a community, a user interactive network will be continuously enhanced via the created relationship with other users. The decision of user participation behavior is significantly influenced by user interactive network (e.g., group preference and number of users). For modeling user interaction and their game in networks, the step of NEGM formulation is detailed.

Step 1. Developing a network to present user interaction in a co-creation community. Specially, users who directly interact with the given user are called as the neighbors of the given user.

Step 2. At each time, all users calculate their returns based on the TEGM and compare with the returns obtained by their neighbors. Then, based on the evolutionary rules, they decide on whether to imitate neighbors' strategies.

Model assumptions, definition of network, and evolutionary rules are elaborated as follows.

3.2.1. Model Assumptions. Based on the properties of network observed by researchers focusing on co-creation community and the application of evolutionary game model, several assumptions are given.

Assumption 1. Since the focus of this paper is the connection among all users, the undirected network is represented by $G(V, E)$. $V = \{v_i | i = 1, 2, \ldots, n\}$ is a set of nodes (e.g., users in a co-creation community). $E = \{e_{ij} | i, j = 1, 2, \ldots, n\}$ is a set of edges (e.g., user interactions in a co-creation community). G is expressed by an adjacency matrix $A = (a_{ij})$. $a_{ij} = 1$ indicates the presence of an undirected line, while $a_{ij} = 0$ indicates the absence of an undirected line. Due to the undirected characteristic, there is no difference between a_{ij} and a_{ji}.

Assumption 2. All networks used in this paper are based on the simulated network. Community managers can adjust the parameter of the model to fit the real-world network and thus to comprehensively describe different situations in practice.

The characteristic of user interaction in co-creation community can be concluded from two aspects. On one hand, preferential attachment is the main principle of user interactive behavior. It indicates that users prefer to contact with persons with a high influence (e.g., leader users in co-creation community). Due to this feature, user interactive network generally emerges as a scale-free network [35]. On the other hand, users usually publish or respond posts within the given groups which persons have the same interests (e.g., improvement of products/services and feedback of using experiences). Such user behaviors lead to user interactive network consisting of many small groups. Specially, connections are dense within the small groups, while connections are sparser between different small groups [36].

To further verify the scale-free feature and the occurrence of small groups, the authors did an empirical study via real data which is collected from the famous co-creation community of Pollen Club. Pollen Club is a social networking platform for customers who use the electronic equipment (e.g., mobile phone) produced by Huawei. Users can publish posts to share their experiences, to suggest for product improvement, to ask for guidance, and to provide a feedback of phone using. For emphasizing the importance of co-creation, posts discussing suggestions for product improvement are considered. Huawei P40, launched at 21:00 p.m., March 26, 2020, is the newest mobile phone which receives a growing attention. To explore the evolution of users' behaviors from the beginning, the time period for this case is from March 26, 2020 to September 5, 2020. The data extracted from Pollen Club includes user name, time, and topic. Finally, 64302 posts, 8197 users' names as well as posting time are collected by the tool of Octopus Data Collector. Based on this data, user interactive network of $8197 * 8197$ is formulated by Python tool. According to [37], scale-free features and the

characteristic of small group can be respectively proved by cumulative degree distribution and hierarchical organization (see Figure 2).

From Figure 2, empirical results provide a powerful verification that user interactive network in co-creation community has scale-free features and the occurrence of small groups.

To formulate the interactive network mentioned onwards, network generation model concerning the growth of users and user interactions is chosen. More details of the proposed model are discussed in the next section.

Assumption 3. Given that user participation behavior is significantly influenced by return-cost analysis of other users, the evolutionary rule (e.g. replicator dynamics) is chosen to update users' decisions. The strategy selection depends on the last game result, and the memory length is set to 1.

Next sections introduce formulation of user interactive network and evolutionary rules.

3.2.2. Definition of User Interactive Network. Based on Assumption 2, the evolving network model proposed by [36] is employed to generate networks similar as the desired networks including scale-free features and the occurrence of small groups. The processes of network generation are elaborated.

Step 1 (Initialization). The initial network includes M small groups where m_0 nodes are connected as the complete graph. M nodes respectively selected from M small groups are connected as the complete graph, which indicates that both two small groups are connected by the initial edge at least.

Step 2 (Growth). At each time t, a new node participates in a network and joins in a small group j. Then, the new node connects $m(m < m_0)$ nodes within group j through inner-group links. Meanwhile, the new node connects $n(n < m)$ nodes in other $M-1$ groups through inter-group links with a probability of $\alpha(0 < \alpha < 1)$. For taking account preferential principle into the network generation model, the interactive attachments are concluded:

(a) *Inner-group preferential attachment.* The probability $(P(s_{ij}))$ that a new node connects node i in small group j depends on the degree s_{ij}. $P(s_{ij})$ is defined as

$$P\left(s_{ij}\right) = \frac{s_{ij}}{\sum_k s_{jk}}, \tag{11}$$

where

s_{ij} is the degree of node i in group j. k is the total number of nodes in group j. $\sum_k s_{jk}$ is the total degree of nodes in group j.

(b) *Inter-group preferential attachment.* The probability $(P(l_{ih}))$ that a new node connects node i in a small group $h(h \neq j)$ depends on the degree l_{ih}. $P(l_{ih})$ is defined as

$$P\left(l_{ih}\right) = \frac{l_{ih}}{\sum_{m,n,n\neq j} l_{m,n}}, \tag{12}$$

where l_{ih} is the degree of node i in the externa group $h(h \neq j)$. n is the sum of external group (apart from group j). m is the number of nodes which connects external node in group $n(n \neq j)$. $\sum_{m,n,n\neq j} l_{m,n}$ is the sum of nodes which connect the external node.

Step 3 (Formulation). To generate the desired network mentioned onwards, the algorithm is given in Algorithm 1.

3.2.3. Evolutionary Rules of NEGM. Based on Assumption 3, replicator dynamics method is chosen.

$J_i = \left\{j_1, j_2, \ldots, j_{k_i}\right\}$ (k_i is the degree of node i) is a set of nodes which are connected by node i. In other words, J_i is all neighbors of node i. At time t, $p_i(t)$ is the probability that a user actively participates in a community. The expected returns of node i and node j is calculated as

$$U_i = \sum_{i\in I_i}\left(U^i_{\mathrm{BA}} * p_i(t) + U^i_{\mathrm{BP}} * \left(1 - p_i(t)\right)\right) = \sum_{i\in I_i}\left(y * \left(R_1 - C_1\right) * p_i(t) + (1-y) * \left(R_1 - C_1 + W\right) * p_i(t) + y * \left(R_2 - C_2\right) * \left(1 - p_i(t)\right)\right), \tag{13}$$

$$U_j = \sum_{j\in J_i}\left(U^j_{\mathrm{BA}} * p_j(t) + U^j_{\mathrm{BP}} * \left(1 - p_j(t)\right)\right) = \sum_{j\in J_i}\left(y * \left(R_1 - C_1\right) * p_j(t) + (1-y) * \left(R_1 - C_1 + W\right) * p_j(t) + y * \left(R_2 - C_2\right) * \left(1 - p_j(t)\right)\right). \tag{14}$$

At time $t+1$, if $U_i > U_j$, the initial strategy is still chosen by node i. Otherwise, the decision of node i evolves as the neighbor j's strategy. The probability $p_i(t+1)$ that a user actively takes part in a community is

$$p_i(t+1) = p_i(t)\left(1 - P_{s_i - s_j}\right) + p_j(t) P_{s_i - s_j}, \tag{15}$$

where $P_{s_i \longrightarrow s_j}$ is the probability that node i imitate the strategy of user j, and it is defined as

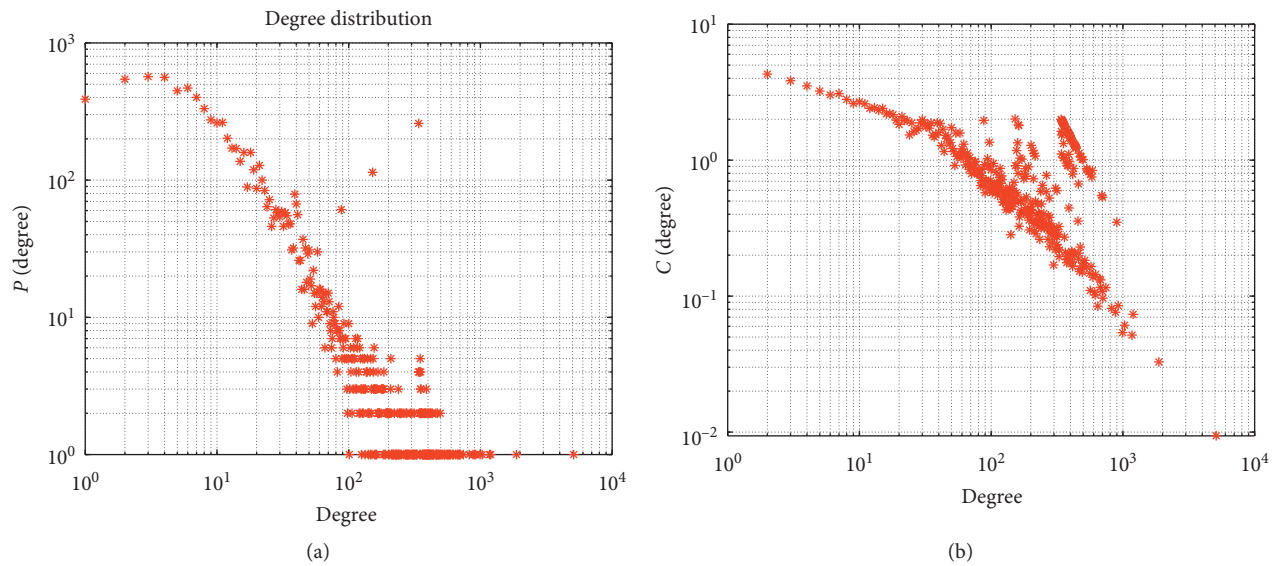


Figure 2: Result of cumulative degree distribution and hierarchical organization. (a) Cumulative degree distribution. (b) Hierarchical organization.

Input: The initial number of nodes m_0, the desired edges N and the iteration number $t = N - m_0$, the number of additive edges m within the inner group at each time t, the number of additive edges n in external groups at each time t, the probability α that a new node connects the node in external groups, number of small groups M, and size of each small group Group_Size.
Initially: For $t = 0$, $M * m_0$ number of nodes within the inner group are connected as a complete graph. M number of nodes which are, respectively, selected from different groups are connected as a complete graph.
For $1 \le t \le N$, do
Step 1: Formulate the network G_t based on equations (11) and (12).
Step 2: If $t < T$, go to step 1; otherwise, stop.
Output: G_t.

Algorithm 1: Network formulation of co-creation community.

$$P_{s_i \longrightarrow s_j} = \frac{1}{1 + e^{\left(\left(U_i - U_j\right)/\delta\right)}}, \tag{16}$$

where δ is the information noise in the decision-making process. $\delta \longrightarrow 0$ indicates that user i is completely rational. Node i is sure to imitate the strategy of j. When $\delta \longrightarrow \infty$, $P_{s_i \longrightarrow s_j}$ approximately equals 1/2. It indicates that user i is bounded rational. User i study the strategy of node j at random.

The strategy (e.g., active or passive participation) chosen by users and the related returns are the focus of this paper. Hence, denote $\overline{P}$ and $\overline{U}$ to, respectively, represent them.

$$\overline{P} = \lim_{t \longrightarrow +\infty} \sum_i \frac{p_i(t)}{n}, \tag{17}$$

$$\overline{U} = \lim_{t \longrightarrow +\infty} \sum_i \frac{U_i}{\overline{k}n}. \tag{18}$$

4. Computational Experiments of NEGM

In this section, NEGM of user participation behavior is analyzed by some computational experiments. Specially, internal motivations (e.g., self-development, reputation, and rewards) are discussed by return-cost analysis presented in Section 4.1. External motivations (e.g., network size, information noise, and network structure) are discussed in Section 4.2.

Motivated by [18], the computational system is divided into four modules:

(a) Develop the network, and define the game parameters and model parameters.

(b) Graphically present the network evolution, and show the change of nodes and the created relationships.

(c) Display the results, and show the degree of a network.

(d) Draw the evolutionary strategies ($\overline{P}$) and the related returns ($\overline{U}$).

Computational experiments are implemented by Matlab R2018b. The Algorithm 2 is given next.

4.1. Computational Results Discussed from Internal Motivations. The impact of internal motivations on user participation can be understood as the influence of game parameter (e.g., returns of R_1 and R_2, costs of C_1 and C_2,

Input: (1) *Network setting:* the graph G_t generated by Algorithm 1; (2) *the iteration number*: $t = T$; (3) *Evolutionary game model*: returns of R_1 and R_2, costs of C_1 and C_2, reward W, the probability p and $1 - p$ for choosing active behavior or passive behavior; (3) *Network properties*: the degree of each node k_i.
Initially: For $t = 0$, $M * m_0$ number of nodes within the inner group are connected as a complete graph. M number of nodes which are, respectively, selected from different groups are connected as a complete graph. Next, denote $p_i(0) = 0.5$ as the initial proportion that users choose active strategy.
For $1 \le t \le N$, do
Step 1: Calculate the expected returns of each node U_i based on equations (13) and (14).
Step 2: For each node i, select a neighbor j at random. Update $p_i(t+1)$ of node i based on equation (15).
Step 3: Calculate $\overline{P}$ and $\overline{U}$ according to equations (17) and (18).
If $t < T$, go to step 1; otherwise, stop.
Output: $\overline{P}$ and $\overline{U}$.

Algorithm 2: Simulation of NEGM.

reward W) on evolutionary strategies ($\overline{P}$) and the average returns ($\overline{U}$) in NEGM. Therefore, to provide a better comparison among different situations, parameters regarding external motivations are set as the same value in this section. Specially, network size is set as 2000, which indicates that there are 2000 users in this community. User interactive network has scale-free feature and some small groups. To reduce the influence of network structure on decision of user participation behavior, number of small groups and size of each small group are, respectively, denoted as 4 groups and 500 users. Given that two-player games occur many times (e.g., when users take part in communities and scan posts published in community, they need to make decisions on whether to active or passive participate), the iteration time of evolutionary game (t) is set as 500 times. To reduce the influence of information noise, δ is denoted as 100. Next, different situations regarding the influence of internal motivations on their decisions are detailed .

Situation 1. When $R_1 > C_1$, $R_2 > C_2$, $R_1 > R_2$, $C_1 > C_2$, and $R_1 - C_1 > R_2 - C_2$, both users choosing active strategy and users choosing passive strategy can obtain reasonable returns from their actions. Specially, returns obtained by active users are bigger than returns gained by passive users. This is the best situation of "more work for more pay." To describe this situation, parameters are set as $R_1 = 20$, $R_2 = 15$, $C_1 = 8$, and $C_2 = 6$. For observing the influence of rewards (W) on their decisions, five cases are, respectively, set as 0, 3, 6, 9, and 12. Figure 3 depicts the evolutionary strategies path ($\overline{P}$) and average returns ($\overline{U}$). Hierarchical axes, respectively, represent $\overline{P}$ and $\overline{U}$ in Figures 3(a) and 3(b). Specially, $\overline{P} = 1$ indicates that active participation is chosen by all users. $\overline{P} = 0$ indicates that passive participation is chosen by all users. $\overline{P} = (0, 1)$ indicates that the mixed strategies are chosen by users. Horizontal axis represents the iteration number of evolutionary game (t).

From Figure 3(a), users' strategies evolve as active participation with a probability of 0.56 in a short time ($t = 32$), although there is no reward paid for active users. When community managers reward active users, the probability for choosing active strategies continuously increases. Specially, $w = 12$ plays a significant role in promoting active participation (e.g., $\overline{P}$ reaches to 0.76). From Figure 3(b), average returns earned by users are located at interval [8.3, 12.9], which indicates that all users can obtain satisfied returns in this situation.

Situation 2. When $R_1 > C_1, R_2 > C_2, R_1 > R_2, C_1 > C_2$, and $R_1 - C_1 < R_2 - C_2$, a bad situation of "less work for more pay" occurs, which indicates that passive users obtain excessive returns from volunteer behavior of active users. To present this situation, parameters are set as $R_1 = 16$, $R_2 = 12$, $C_1 = 13$, $C_2 = 6$, $W = [0, 3, 6, 9, 12]$. Figure 4 shows the results of $\overline{P}$ and $\overline{U}$.

From Figure 4(a), populations prefer to choose passive strategy ($1 - \overline{P} \approx 0.53$) without consideration of rewards, which results from the unreasonable return-cost allocation. From Figure 4(b), average returns continuously increase as the growth of rewards. However, the probability for choosing active participation does not increase all the time, although the increasing caused by the growth of rewards is obvious (e.g., compared with situation of $W = 0$, other four situations considering rewards promote strategy selection of active participation). For example, $W = 3$ promotes active participation, while $W = 9$ inhibits active participation. This result differs from result discovered in TEGM, which indicates that rewards do not always promote active participation. According to [38], this phenomenon can be explained by the special network structure (e.g., scale-free feature). When a strategy is popular in a group, this strategy will spread in these populations. In other words, evolutionary strategies depend on the initial strategy which is spread in populations.

Situation 3. When $R_1 > C_1$ and $R_2 < C_2$, three possible cases including $\{R_1 > R_2, C_1 > C_2\}$, $\{R_1 > R_2, C_1 < C_2\}$, and $\{R_1 < R_2, C_1 < C_2\}$ bring opportunities for the establishment of best situation regarding $R_1 - C_1 > R_2 - C_2$. To simulate these three situations and compare with them, parameter setting is denoted in Table 5.

Figure 5 concludes three cases mentioned in situation 3.

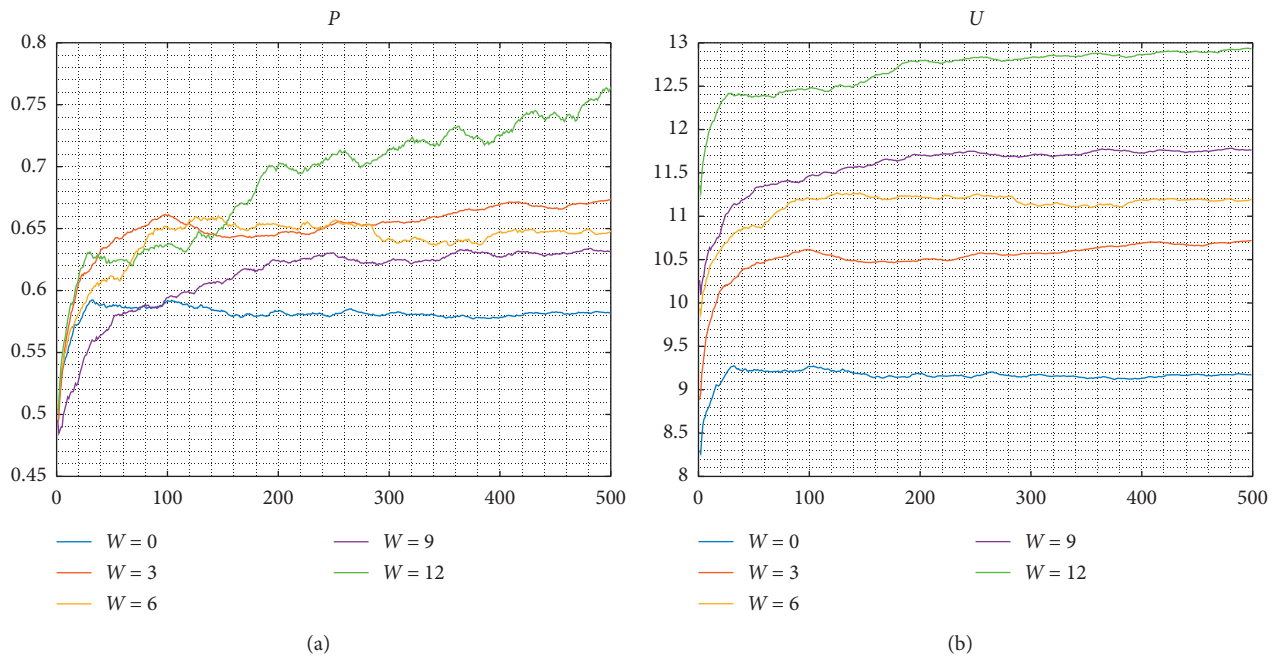


Figure 3: The simulated results of situation 1. (a) $\overline{P}$. (b) $\overline{U}$.

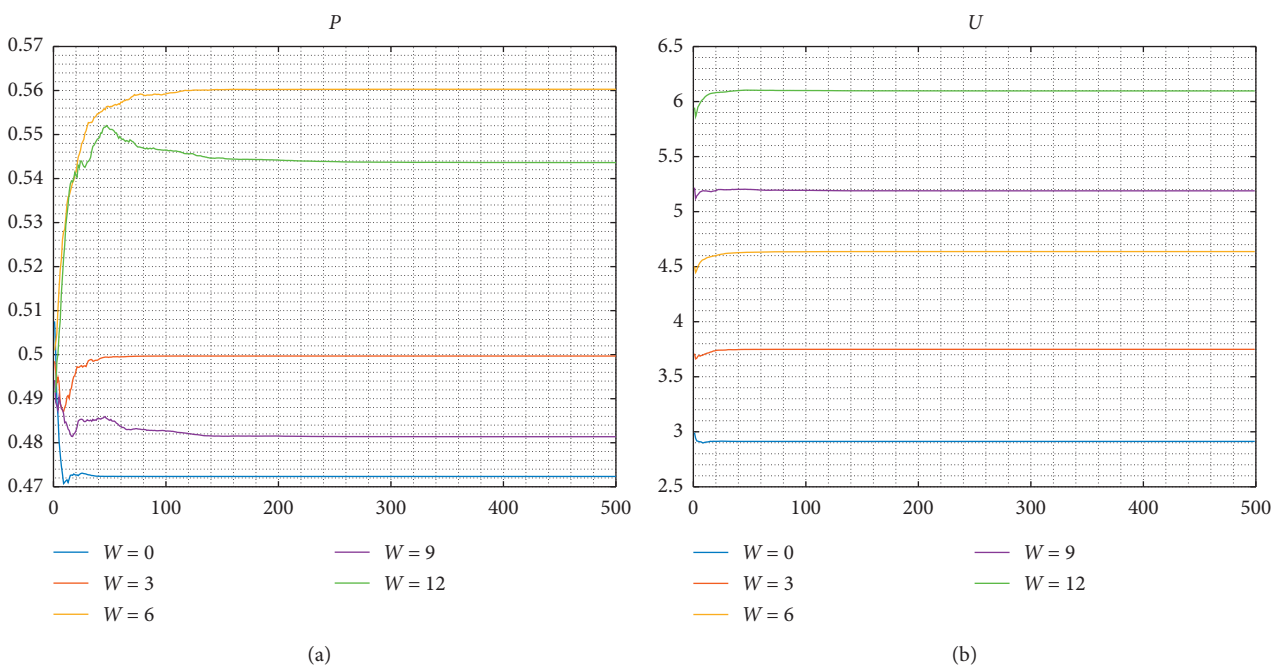


Figure 4: The simulated results of situation 2. (a) $\overline{P}$. (b) $\overline{U}$.

Table 5: Parameter setting of situation 3.

	R_1	R_2	C_1	C_2	W
Case 1: $R_1 > R_2$, $C_1 > C_2$	18	6	10	8	[0, 3, 6, 9, 12]
Case 2: $R_1 > R_2$, $C_1 < C_2$	22	10	10	12	[0, 3, 6, 9, 12]
Case 3: $R_1 < R_2$, $C_1 < C_2$	16	18	13	20	[0, 3, 6, 9, 12]

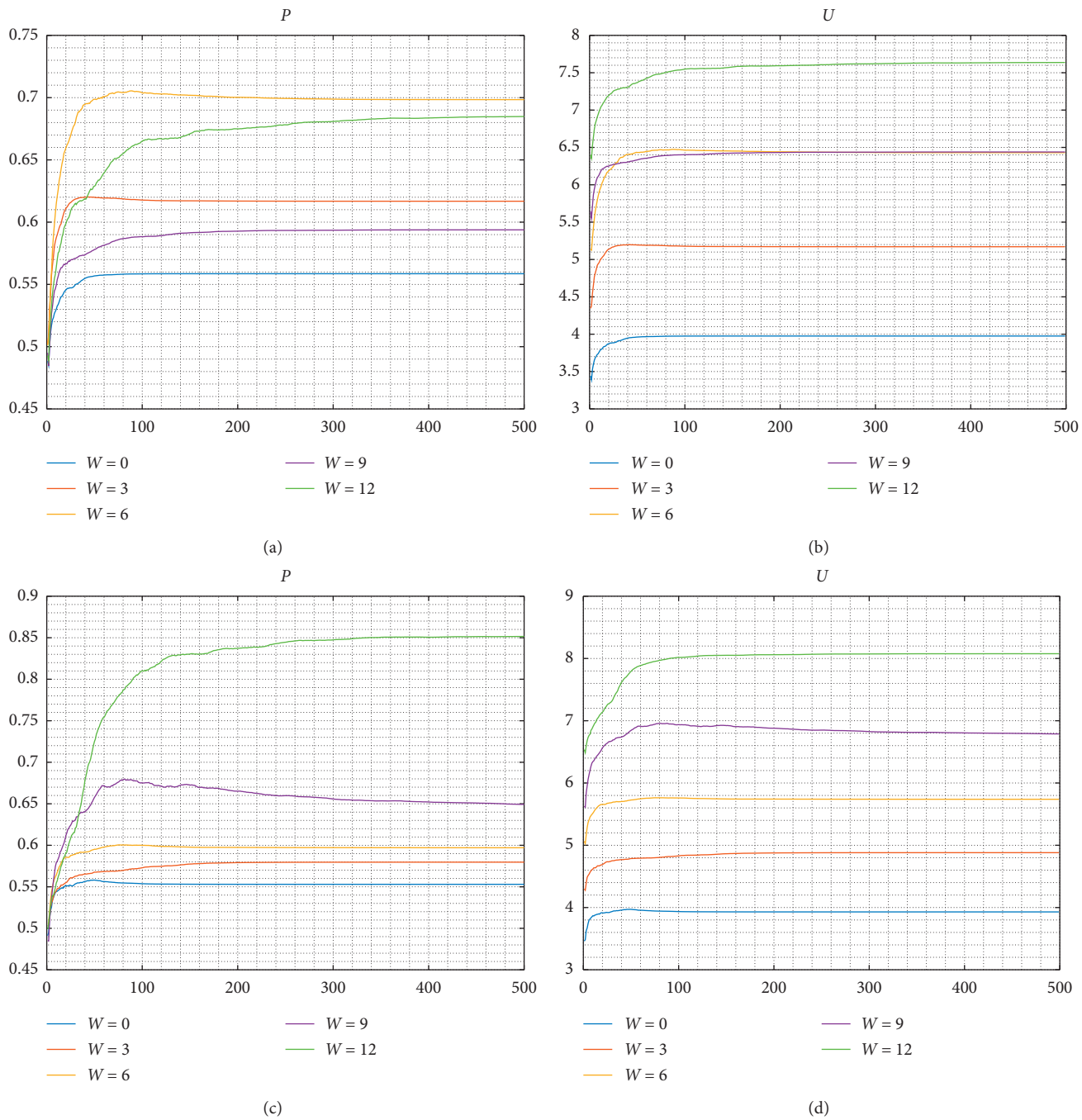


Figure 5: Continued.

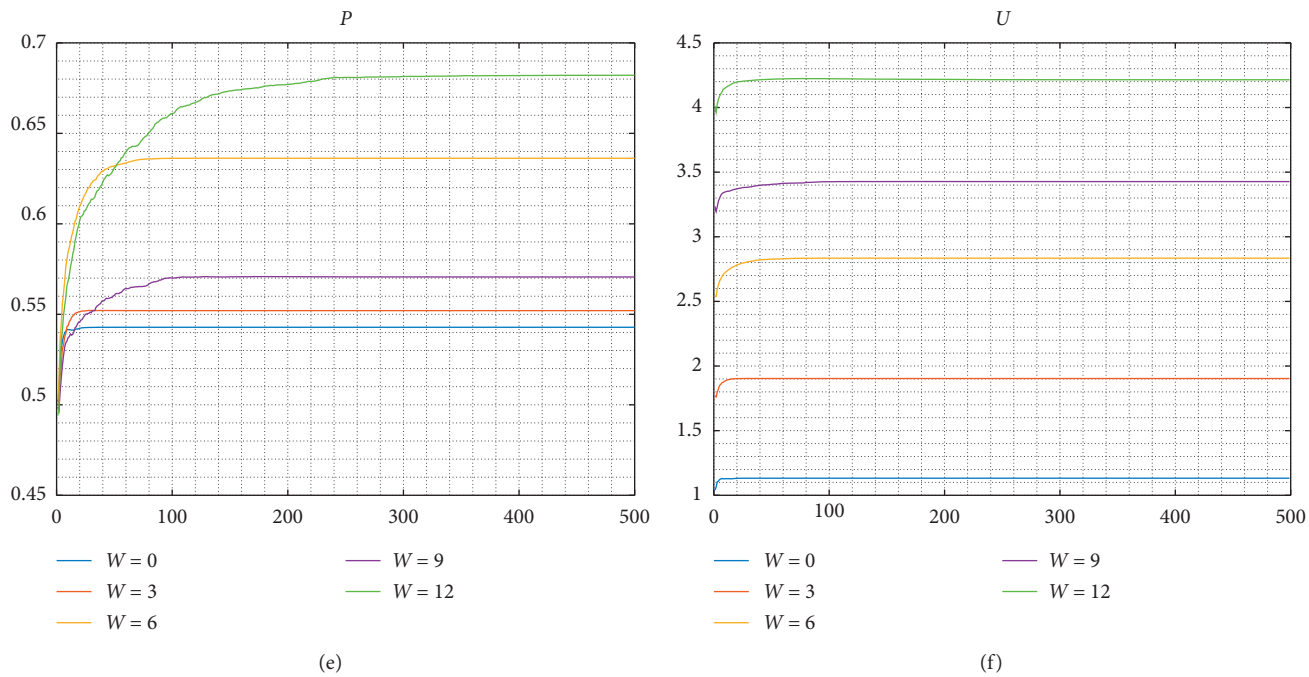


Figure 5: The simulated results of situation 3. (a) $\overline{P}$(Case 1). (b) $\overline{U}$(Case 1). (c) $\overline{P}$(Case 2). (d) $\overline{U}$(Case 2). (e) $\overline{P}$(Case 3). (f) $\overline{U}$(Case 3).

Table 6: Parameter setting of situation 4.

	R_1	R_2	C_1	C_2	W
Case 1: $R_1 > R_2$, $C_1 > C_2$	18	16	22	14	[0, 3, 6, 9, 12]
Case 2: $R_1 < R_2$, $C_1 > C_2$	16	20	18	16	[0, 3, 6, 9, 12]
Case 3: $R_1 < R_2$, $C_1 < C_2$	6	18	8	10	[0, 3, 6, 9, 12]

From Figure 5, some observations are concluded:

(a) Due to $R_1 - C_1 > R_2 - C_2$, users prefer to choose active participation ($\overline{P} > 0.5$) without considering rewards.

(b) Rewards sometimes promote active participation (e.g., red line in Figure 5) but sometimes inhibits active participation (purple line in Figures 5(a) and 5(b)).

(c) Case 3 usually occurs in the maturation of community's operation. Rules (e.g., how many rewards should be paid for active users, how to encrypt the valuable knowledges which are contributed by active users, and how much money should be paid by passive users for acknowledging valuable knowledges) are well designed and implemented for a long time. These rules lead to expensive costs for users who passively participate in communities. And the rules also avoid the occurrence of knowledge spillover. In this context, to acknowledge valuable knowledges from community, passive users have no choice but to contribute their knowledges, experiences, and skills for exchange.

(d) From Figures 5(b), 5(d), and 5(f), average returns are located at interval [1, 8]. The value is much smaller than the value in the same situation of "more work for more pay" (situation 1). Situation 1 describes that both active users and passive users can obtain high returns from communities. This situation usually occurs in the communities which include lots of valuable knowledges and many active users. For example, valuable knowledges attract users to participate in communities, and they obtain returns from learning knowledges. Besides, many active users indicate that users obtain returns from social communications via interaction with other persons. In situation 3, the limited valuable knowledges created by active users will be locked by managers. Active users can easily obtain these knowledges. However, passive users need to pay some money for acknowledging knowledges. This situation results in the outcome that few returns can be obtained by passive users. Therefore, although situations 1 and 3 present the same situation of "more work for more pay", the influential path on the decision of user participation is different.

Situation 4. When $R_1 < C_1$ and $R_2 > C_2$, three possible cases including $\{R_1 > R_2, C_1 > C_2\}$, $\{R_1 < R_2, C_1 > C_2\}$, and $\{R_1 < R_2, C_1 < C_2\}$ lead to the worse situation of $R_1 - C_1 < R_2 - C_2$. To simulate these three situations and compare with them, parameter setting is denoted in Table 6.

Figure 6 concludes three cases mentioned in situation 4.

From Figure 6, some observations are concluded:

(a) Due to $R_1 - C_1 < R_2 - C_2$, users prefer to choose passive strategy $(1 - \overline{P} > 0.5)$ without consideration of rewards.
(b) Situation 4 relates to the worse situation of "less work for more pay", which indicates that returns earned by active users are smaller than returns obtained by passive users. This situation inhibits active participation and thus leads to non-sustainability of communities. From Figures 6(a), 6(c), and 6(e), rewards play a slight role in promoting active participation. Specially, the probability for choosing active strategy is bigger than 0.5 only when $W = 12$.
(c) From Figures 6(b), 6(d), and 6(f), average returns are located at interval [−1.4, 3.9]. This value is much smaller than the value generated by other situations. Particularly, Figure 6(b) shows that average returns are smaller than zero, which indicates that lots of users (especially active users) cannot obtain reasonable returns from their volunteer behavior (e.g., experience sharing, knowledge sharing and helping others). This situation often occurs in the early stage of community's operation or the recession of community's operation.

On the one hand, in the early stage of community's operation, only a few active users contribute their knowledges and experiences in the co-creation community. Other users act as lurkers to scan the knowledges and experiences. In this situation, community managers need to maintenance the current active users via rewards (e.g., financial benefits and non-economic rewards). Meanwhile, rewards should also be paid for passive users who attempt to choose active strategy.

On the other hand, in the recession of community's operation, managers should decide whether to set rewards for users. From Figures 6(a), 6(c), and 6(e), only when $W = 12$, populations will choose active strategy with a probability of 0.5. It indicates that lots of costs (e.g., money) should be paid to maintenance community's operation. In this situation, managers should evaluate the value of community for company, before they decide whether to pay more money for community's operation.

Situation 5. When $R_1 < C_1$, $R_2 < C_2$, $R_1 < R_2$, and $C_1 < C_2$, two possible cases regarding $R_1 - C_1 < R_2 - C_2$ and $R_1 - C_1 > R_2 - C_2$ occur in user participation system. To simulate these two situations and compare with them, parameter setting is presented in Table 7.

Figure 7 concludes two cases mentioned in situation 5.

From Figure 7, some observations are concluded:

(a) Due to the unreasonable returns of both active and passive users (e.g., $R_1 < C_1$, $R_2 < C_2$), users prefer to choose passive strategy $(\overline{P} < 0.5)$ without considering rewards (no matter $R_1 - C_1 < R_2 - C_2$ or $R_1 - C_1 > R_2 - C_2$).
(b) From Figure 7, the probability for choosing active strategy (from 0.483 to 0.597) and average returns (from −2.4 to 3.8) continuously increase as the growth of rewards. This situation particularly occurs in the early stage of community's operation. As a new co-creation community opens, since there are no existing resources (e.g., knowledges and experiences) for users to exchange, both active and passive users cannot obtain enough returns. Thus, managers may set some rewards to attract more persons. For example, Pollen Club rewards their electronic equipment (e.g., phones and watch) for active users. Similar actions are also implemented by Xiaomi. These actions bring opportunities for creating a thriving community in the early stage of community's operation.

4.2. Computational Results Discussed from External Motivations. As Section 1 presents, the decision of user participation is also influenced by external motivations (e.g., number of users and group preference). These motivations can be treated as characteristics of user interactive network (e.g., network size and network structure). This section is to test the influence of external motivations on user participation behavior via analysis of user interactive network in NEGM. Specially, next sections, respectively, examine the effect of information noise in decision process (δ), network size (N), number of small groups (M), and size of small group (Group_Size). Further, four types of network (e.g., regular network, small-world network, scale-free network, and network with small groups) are employed to test the influence of network structure on the decision of user participation.

4.2.1. The Effect of Information Noise in Decision Process (δ). Certain external events may fascinate the community and trigger off unexpected bursts of participation [20]. Here, information noise (δ) is set to test the effect of unexpected bursts on user participation. As Section 4.1 analyses, two main cases including $R_1 - C_1 < R_2 - C_2$ and $R_1 - C_1 > R_2 - C_2$ exist in the user participation system. Thus, parameters of evolutionary game model are set based on these two cases. Besides, to exclude the effect of user interactive network, the related parameters are denoted as the same value. Bases on the characteristic of information noise mentioned in Section 3.2.3, the value of information noise increases 10 times at each change. Table 8 shows parameter setting.

Figure 8 concludes two cases mentioned in Table 8.

From Figure 8, some observations are concluded:

(a) Information noise has a great influence on the decision of user participation. To further verify the

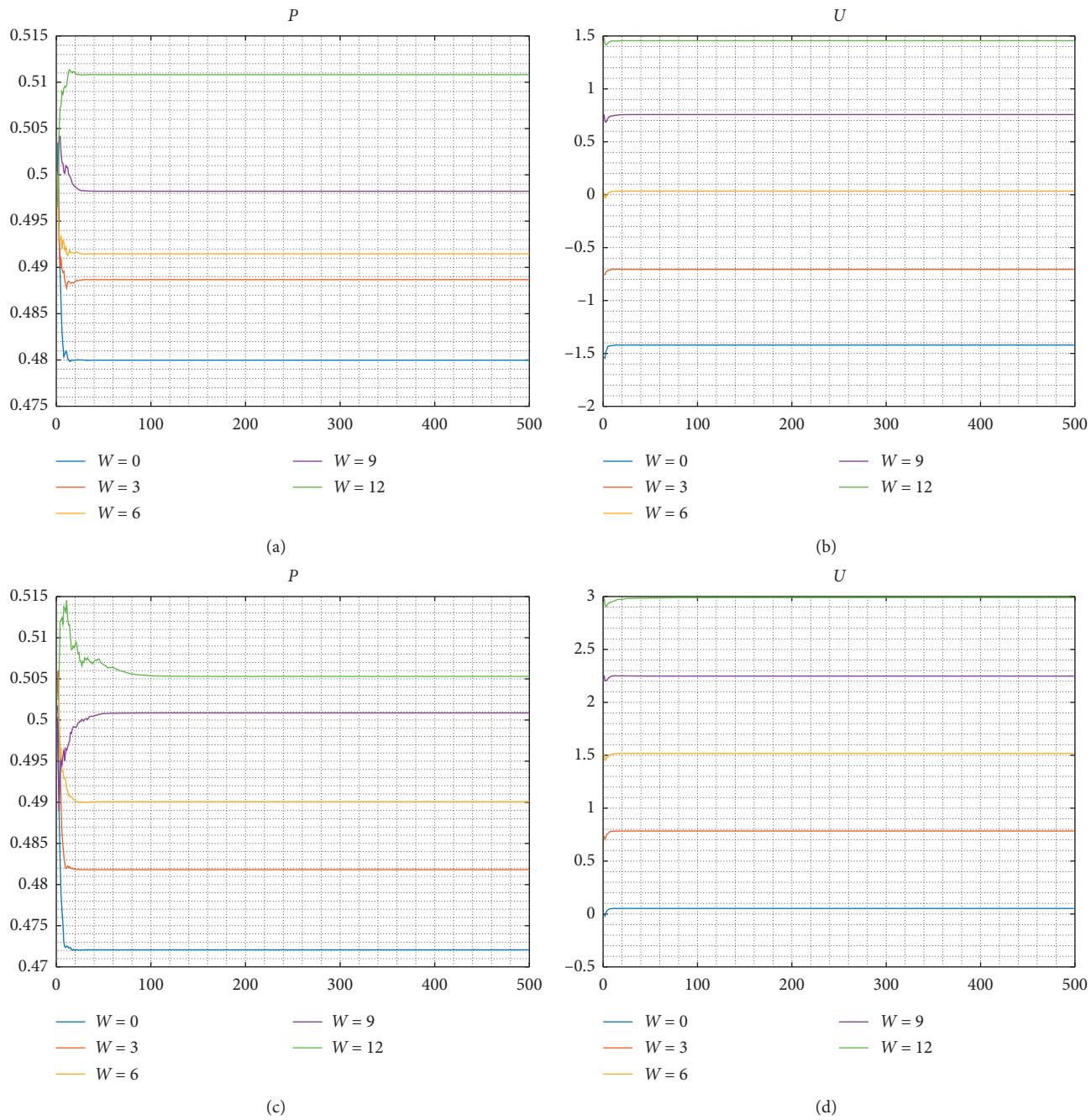


Figure 6: Continued.

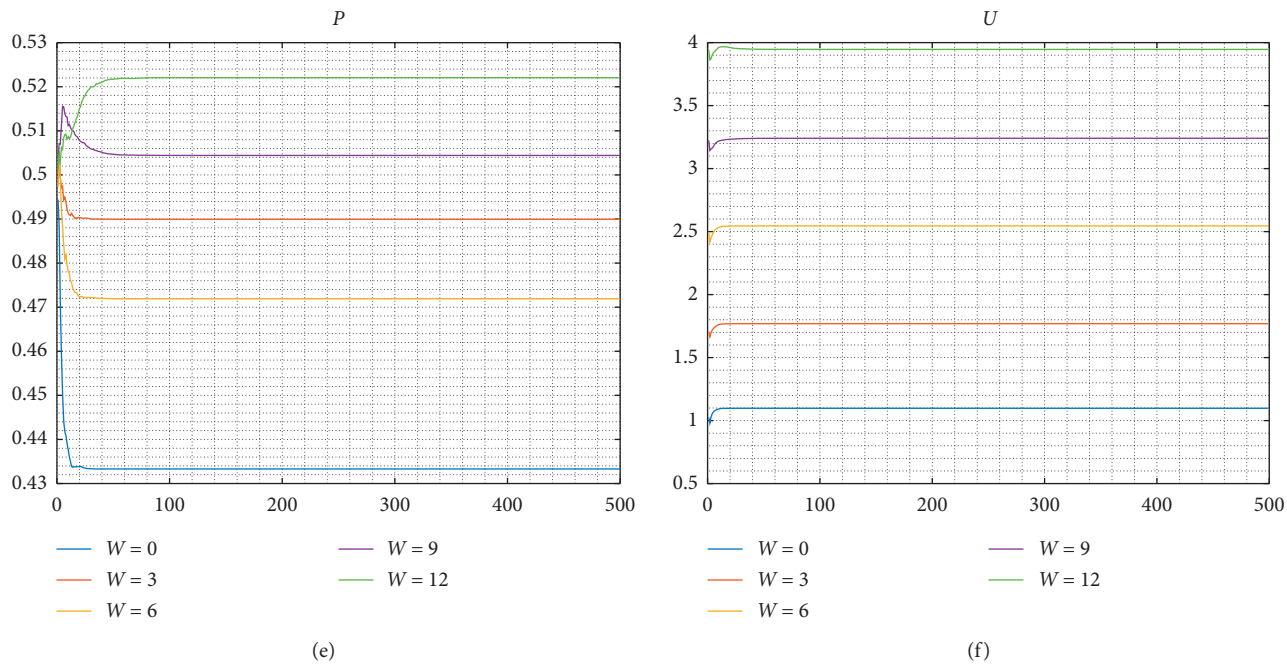


Figure 6: The simulated results of situation 4. (a) $\overline{P}$(Case 1). (b) $\overline{U}$(Case 1). (c) $\overline{P}$(Case 2). (d) $\overline{U}$(Case 2). (e) $\overline{P}$(Case 3). (f) $\overline{U}$(Case 3).

Table 7: Parameter setting of situation 5.

	R_1	R_2	C_1	C_2	W
Case 1: $R_1 - C_1 < R_2 - C_2$	12	16	16	18	[0, 5, 10, 15, 20]
Case 2: $R_1 - C_1 > R_2 - C_2$	12	16	14	20	[0, 5, 10, 15, 20]

significant influence, ANOVA analysis is employed to test the differences in means and variances at a 95% confidence level. When $p-\text{value} < 0.05$, it indicates that there is a critical difference between the test populations. All values (both Case 1 and 2) show significant differences.

(b) From Figure 8(a), when $\delta = 0.1$, it indicates that users are completely rational, and they are sure to imitate their neighbors' strategies. In this case, users who choose active strategy obtain good returns from community. Thus, users will imitate their strategy and choose active participation with a probability of 0.9. However, strategy selection of active participation is inhibited by the growth of information noise. For example, when $\delta = 100$ or $\delta = 1000$, users show the bounded rationality, and they imitate their neighbors' strategies based on their character and mood. Thus, the probabilities for choosing active strategy, respectively, decline as 0.63 and 0.49.

(c) Compared with results discovered in Figure 8(a), the completely reverse result that information noise promotes active participation is found. By reading the study of [39], the same result was also found in their study regarding human cooperation in networks. They explained that the proper information noise was helpful for cooperation promotion, especially in the context with many defectors.

4.2.2. The Effect of Network Size (N). According to empirical study created by [13], the number of users positively affects the intention of active participation in online communities. In this section, the number of users is treated as the network size to test the influence of number of users on the decision of user participation. Four types of network size (e.g., $N = [400, 800, 1200, 1600]$) are considered here. Besides, to reduce the effect of information noise, δ is set as 0.1. Other parameters are set as the same values which are detailed in Table 8. The related results are presented in Table 9.

From Table 9, *pp*-value of Case 1Case 1 (≤0.001) and Case 2 (≤0.001) indicates that there is a significant difference among different network sizes. In other words, network size has a critical role on the decision of user participation. This result also verifies the hypothesis portrayed by [13].

4.2.3. The Effect of M and Group_Size. As [35] argued, all online communities (including co-creation community) had their own special structures. For example, number of small groups is different. Pollen Club is divided as seven groups including experience sharing, suggestion for product improvement, feedback on product using, official activities,

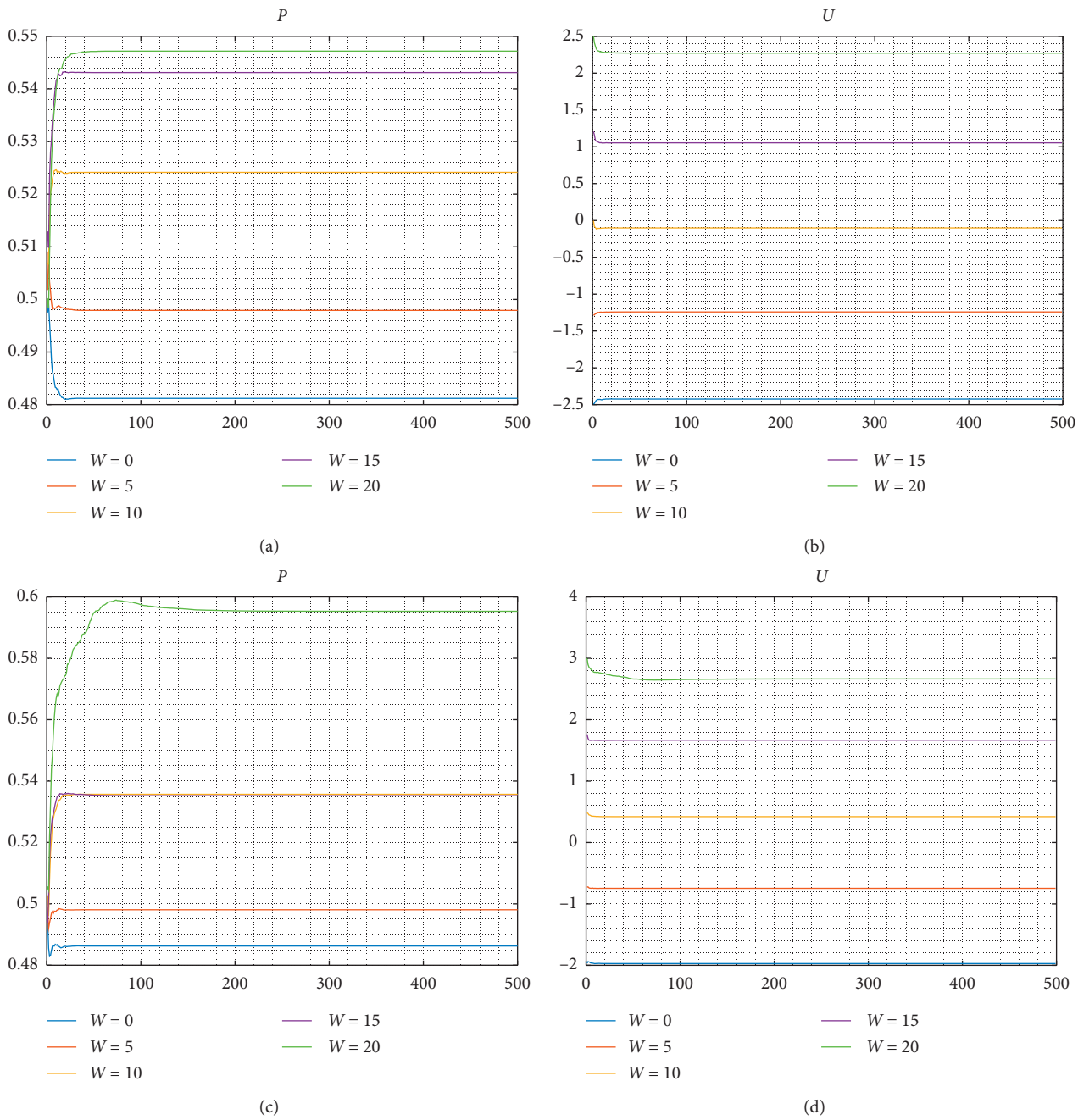


Figure 7: The simulated results of situation 5. (a) $\overline{P}$(Case 1). (b) $\overline{U}$(Case 1). (c) $\overline{P}$(Case 2). (d) $\overline{U}$(Case 2).

product evaluation, resources for better using, and others. Xiaomi Club only has 5 groups (e.g., the best posts, photo sharing, proposal for activities, suggestion for product improvement, and asking for help) in their community. Thus, it is necessary to test the influence of special structures on user participation. According to user interactive network proposed in this paper, number of small groups and size of small groups are considered.

Refer to size of small groups; two types of networks (e.g., uniform network and non-uniform network) should be introduced here. Main differences between them are presented in Figure 9. Specially, Figure 9(a) is a uniform network which 10 nodes, respectively, exist in each small group. Figure 9(b) is a non-uniform network which different number of nodes (e.g., 5, 10, and 15), respectively, exist in each small group. Motivated by the result observed in Section 4.2.2, this section assumes that size of small groups also has a great impact on decision-making.

To test the influence of M and Group_Size on the decision of user participation, three types of computational experiments are designed. Specially, experiment 1 and experiment 2 aim to analyze the influence of M on user participation from perspectives of uniform and nonuniform networks. Experiment 3 aims to make a comparison between

Table 8: Parameter setting considering the change of δ

	R_1	R_2	C_1	C_2	W	N	T	δ
Case 1: $R_1 - C_1 > R_2 - C_2$	18	6	10	8	0	2000	500	[0.1, 1, 10, 100, 1000]
Case 2: $R_1 - C_1 < R_2 - C_2$	18	16	22	14	0	2000	500	[0.1, 1, 10, 100, 1000]

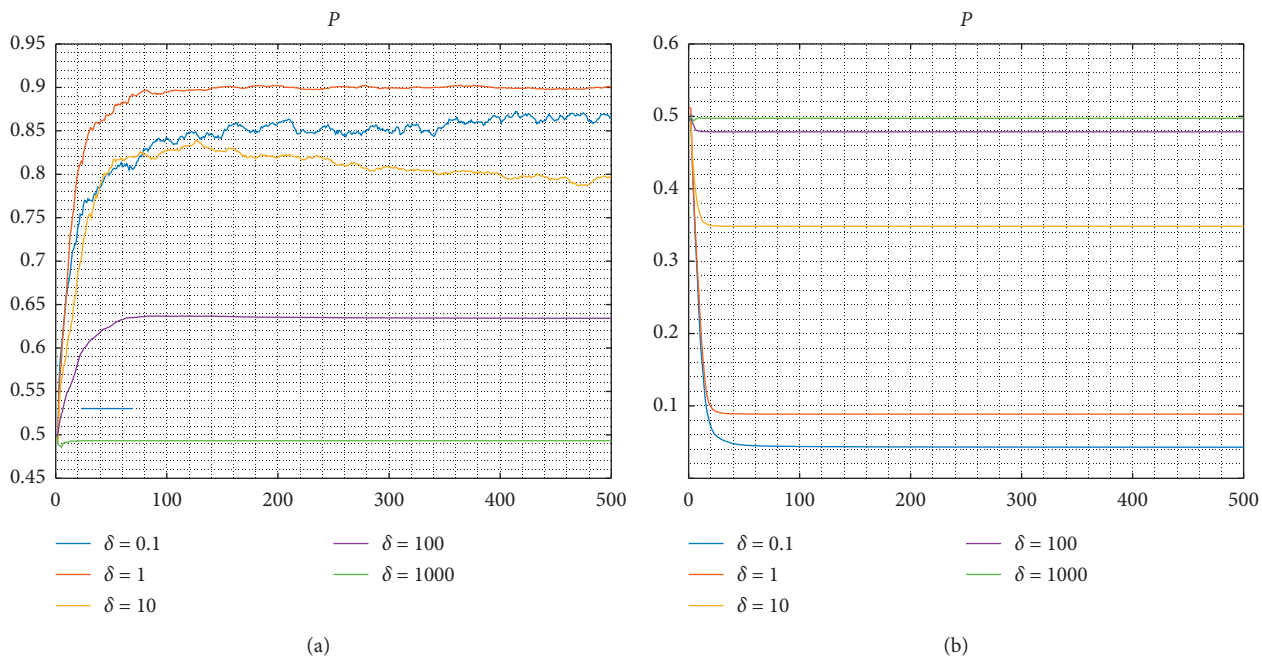


Figure 8: The simulated result considering the change of information noise. (a) $\overline{P}$(Case 1). (b) $\overline{P}$(Case 2).

Table 9: Summary of simulated result considering the change of network size.

	Network size	Convergence result of $\overline{P}$	p-value of ANOVA	Homogeneity tests	p-value regarding nonparametric tests
Case 1	$N = 400$	0.849	≤0.001	≤0.001	**≤0.001**
	$N = 800$	0.848			
	$N = 1200$	0.870			
	$N = 1600$	0.901			
Case 2	$N = 400$	0.046	≤0.001	0.945	—
	$N = 800$	0.041			
	$N = 1200$	0.027			
	$N = 1600$	0.044			

* Case 1 and Case 2, respectively, indicates that $R_1 - C_1 > R_2 - C_2$ and $R_1 - C_1 < R_2 - C_2$, which is detailed in Table 8.

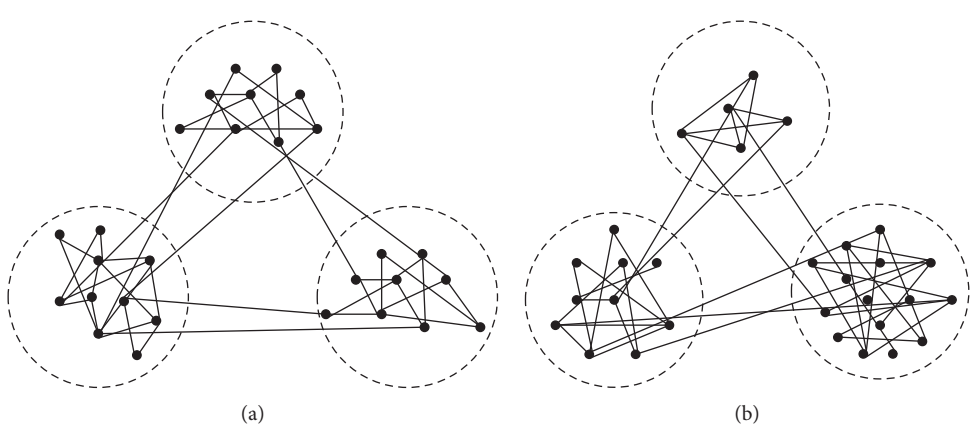


Figure 9: Two types of networks regarding uniform network and nonuniform network, (a) uniform network, (b) nonuniform network.

uniform and nonuniform networks. To exclude the impact of other parameters, all experiments of other parameters are denoted as $R_1 = 18$, $R_2 = 6$, $C_1 = 10$, $C_2 = 8$, $W = 0$, $\delta = 0.1$, $N = 2000$ and $t = 500$. Table 10 shows parameter setting of M and Group_Size. Specially, when $M = 3$, Group_Size of uniform network is set as 666, 666, and 668. Besides, Group_Size of non-uniform network is denoted as 50, 50, and 1900.

Figure 10 concludes all computational results mentioned in three experiments.

From Figure 10, some observations are concluded:

(a) Figures 10(a)–10(d) shows that significant different exists in the nonuniform network but not in the uniform network. The same result is also found by ANOVA analysis of their values. This phenomenon can be explained by the non-uniform distribution of users. When the biggest small group obtains good returns via choosing active strategy, the successful strategy will spread in populations. Otherwise, the defection strategy (e.g., passive strategy) is popular. In other words, group selection depends on the successful strategy adopted by the biggest small group.

(b) From Figures 10(e) and 10(f), uniform network shows a better performance on promotion of active participation. The same result also can be discovered from comparison between Figures 10(a) and 10(c). This result can be illustrated by the study of [40]. They argued that it caused by the detailed evaluation function of the players' success and the different evolutionary rules.

4.2.4. The Effect of Different Network Structures. Although empirical study discussed in Section 3.2.1 presents that user interactive network of co-creation community has scale-free features and the occurrence of small groups, other studies on user interactive network of online communities have identified that other types of networks (e.g., regular network, small-world network, and purely scale-free network) can also be employed to describe user interaction in online communities. Thus, this section aims to analyze the effect of different network structures on the decision of user participation.

As Section 4.2.1 presents, two types of cases regarding $R_1 - C_1 < R_2 - C_2$ and $R_1 - C_1 > R_2 - C_2$ exist in the user participation system. Thus, parameters of evolutionary game model are set as the same value detailed in Table 8. Besides, to exclude the effect of network size and information noise, the fixed value of $N = 2000$ and $\delta = 0.1$ is denoted here. The variable of this section is network structure which is generated by different network generation models [41]. Figure 11 shows the tendency of evolutionary strategies and average returns.

From Figure 11, some observations are concluded:

(a) Compared with regular network and small-world network, purely scale-free network and network with small groups inhibit active participation in Figure 11(a). Similar result was also found in the study of human cooperation [38]. Wu et al. [38] explained that long-tail degree distribution was the main reason. This result indicates that uniform distribution (e.g., regular network and small-world network) is more suitable for promoting active participation, which enhances the result found in Section 4.2.3. From Figure 11(b), users in network with small groups obtain the smallest average returns ($\overline{U} < 1$) from communities. However, the probability for choosing active strategy ($\overline{P} = 0.78$) is not obviously smaller than the probability in other network structures.

(b) There are no significant differences among lines depicted in Figure 11(c). It indicates that network structure has a limited role under the circumstance of $R_1 - C_1 < R_2 - C_2$.

4.2.5. Discussions and Managerial Insights. Based on computational experiments, some discussions and managerial insights are summarized as follows:

(a) Section 4.1 details all possible situations in real practice. The decision of user participation is significantly influenced by return-cost analysis based on the evaluation of several internal motivations. The probability for choosing active strategy depends on return allocation for active and passive users. Managers should carefully focus on return allocation for active users and ensure that returns obtained by active users are more substantial than those gained by passive users. In other words, the rule of "more pay for more work" is necessary in the user participation system. Similar result was also suggested by [17].

Experimental results indicate that rewards effectively promote active participation in many situations. However, the probability for choosing active participation does not continuously increase with the growth of rewards, which results from the special structure of interactive network (e.g., scale-free feature and the occurrence of small groups). For example, users are generally divided into several parts based on different small groups. The decisions of them are mainly affected by the strategy which is widely adopted by their own small groups. If a passive strategy is initially adopted by populations, it is hard to change their initial strategy via rewards.

Therefore, the influence of rewards on the promotion of active participation does not make sense on every occasion. Managers should design specific rewards based on different situations. For example, in the early stage of the community's operation, rewards should be appropriated among both active and passive users. By contrast, rewards are unnecessary in the recession of the community's operation, especially when the budget for the community's maintenance is not sufficient.

(b) Information noise has a great influence on user participation. Specially, when active participation is

Table 10: Parameter setting of M and Group_Size

	M	Group_Size
Experiment 1: uniform network	[3, 4, 5]	[666, 666, 668]; [500, 500, 500, 500]; [400, 400, 400, 400, 400].
Experiment 2: nonuniform network	[3, 4, 5]	[50, 50, 1900]; [50, 50, 50, 1850]; [50, 50, 50, 50, 1800].
Experiment 3: comparison between uniform network and nonuniform networks	4	[500, 500, 500, 500]; [50, 50, 50, 1850].

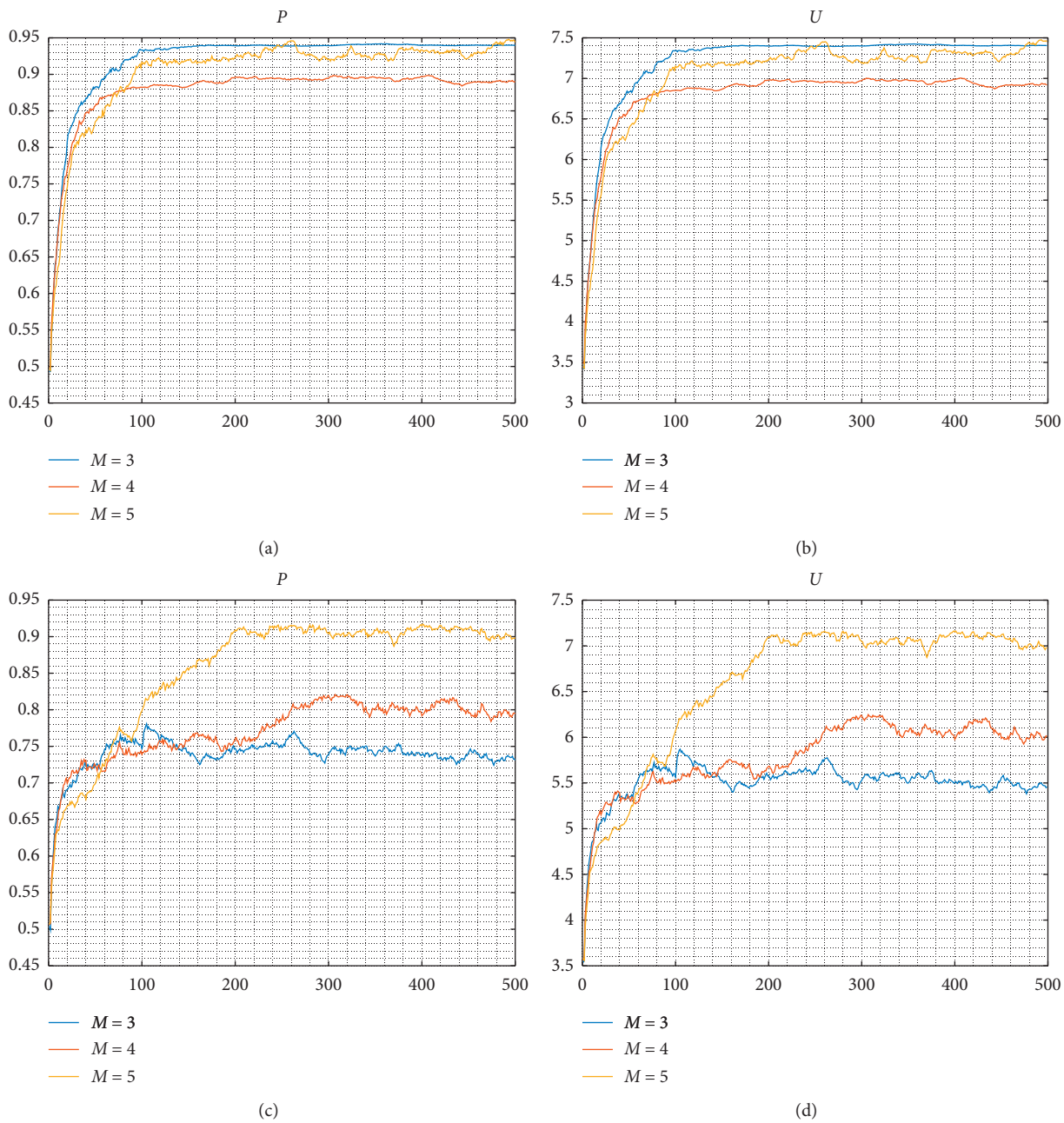


Figure 10: Continued.

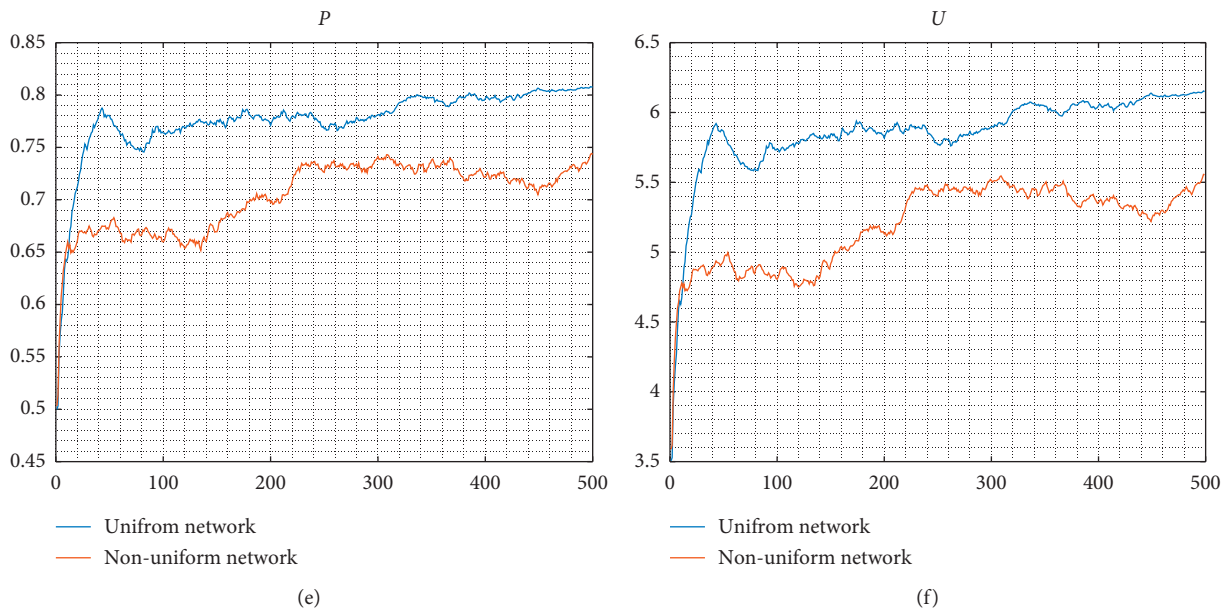


Figure 10: The simulated result considering the change of M and Group_Size. (a) $\overline{P}$(Experiment 1). (b) $\overline{U}$(Experiment 1). (c) $\overline{P}$(Experiment 2). (d) $\overline{U}$(Experiment 2). (e) $\overline{P}$(Experiment 3). (f) $\overline{U}$(Experiment 3).

the dominant strategy in populations, the falling tendency of active participation occurs in decision system with the growth of information noise. By contrast, when a passive strategy is adopted by many users, the increasing tendency of active participation occurs with the growth of information noise.

According to equation (16), information noise closely associates with the characteristic of users. The small value of information noise indicates that users are completely rational, while the big value of information noise indicates that users show the bounded rationality. When passive participation is a dominant strategy in the early stage of the community's operation, managers should make some attempts (e.g., employing some persons to share experience and knowledge in a community, and organizing some activities to attract more users) to promote active participation via enhancing information noise. Influenced by these information noises, users showing bounded relationality will make decisions based on their moods and characteristics. Thus, the dominant strategy of passive participation will be changed next time.

(c) Network size plays a critical role in promoting active participation. This result is also found in the empirical study of [13]. According to their investigation, lurkers' decisions are significantly influenced by network size. Lurkers prefer to share experience and knowledge in a thriving community. At first, they prefer to respond to posts rather than to publish posts. In this context, mangers can create some hot topics and recommend these topics for lurkers. This action will attract more lurkers to discuss their ideas in communities. Besides, it is also necessary to expand user base in a short time, which is also suggested in [5].

(d) Due to the non-uniform degree distribution of user interactive network, active participation is significantly promoted with the growing number of small groups. Besides, uniform network performs well on the promotion of active participation, compared with the result of nonuniform network. In a non-uniform network, the dominant strategy is easily controlled by small groups with many participants. The decisions of other users are significantly affected by their decisions. Thus, it is necessary to detect the number of users in each small group and maintain similar users in each group. If the user interactive network shows nonuniform characteristic, it is necessary to promote active participation via other actions (e.g., rewards setting, enhancing information noise, and expanding user base).

(e) Compared with the regular network and small-world network, purely scale-free network and network with small groups inhibit active participation when the dominant strategy is active participation. Almost 80 percent of online communicating networks have scale-free features, and these characteristics are not conductive to the promotion of active participation. Fortunately, the small-world feature (e.g., the clustering coefficient is high, and the average shortest path is short) is also identified in many real networks, which can change the current situation from passive participation to active participation. Online communications usually follow the principle of the "three degrees of influence rule." It indicates that two strangers can make the interaction via three

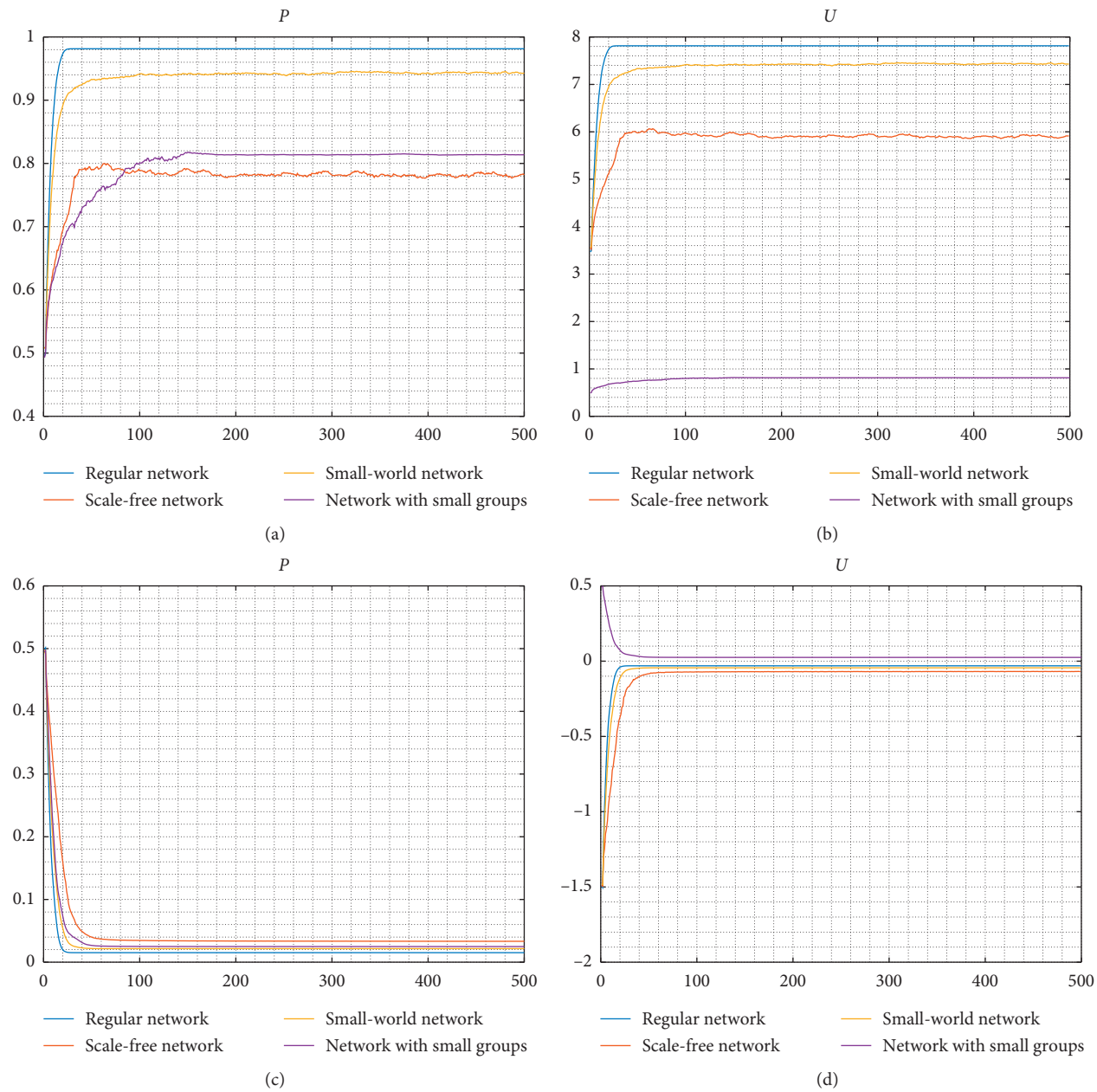


Figure 11: The simulated result considering the change of M and Group_Size. (a) $\overline{P}$(Case 1). (b) $\overline{U}$(Case 1). (c) $\overline{P}$(Case 2). (d) $\overline{U}$(Case 2).

relationships. Based on this finding, actions (e.g., publishing some hot-topic posts to attract more discussions, looking for the lead users to closely connect many persons, and organizing group-based competitions) should be taken to shorten the path among several users.

5. Conclusions

In this paper, the mechanism regarding how internal and external motivations influence the decision of users in participation in co-creation community is analyzed by using NEGM. User behavior (e.g., active participation and passive participation) and the evolutionary process of strategy selection are examined via computational experiments. Two main findings are concluded. (a) User participation behavior is significantly influenced by return-cost analysis based on the evaluation of internal motivations. Rewards sometimes enhance and sometimes inhibit user active participation. (b) User participation behavior is closely associated with network size, network structure, and information noise in the decision-making process.

This paper employs NEGM and computational experiments to investigate the evolution of user participation behavior in co-creation communities. It contributes to the recent literature in threefold manner. (a) Different from extensive studies focusing on identifying motivation of user participation, this paper concentrates on the motive of continuous user

participation behavior and the mechanism regarding whether users should actively participate or not, when to actively participate and how such a strategy can perform well. It fills up the research gap argued by [18]. (b) The influence of external motivations (e.g., network size, network structure, and information noise) on user participation behavior is analyzed by using NEGM, which transforms the research focus from individual-level to group-level effects [16]. (c) The computational experiments proposed in this paper are helpful for visualizing the evolution of users' strategies, grasping the dynamic evolution features in user participation games over time, and exploring the change of strategies in different situations.

This paper analyses the mechanism regarding how different motivations affect the decision of user participation. The practical recommendations for the sustainable operation of co-creation community are summarized. Firstly, for individual users, some proposed strategy should be taken, such as maximizing the returns of active participates, minimizing the returns of passive users via free-riding behavior, and motivating passive users to actively participate in order to maximize the group returns rather than eliminating them from communities. Secondly, if the strategies mentioned onwards are to be enticed to participates, external motivations, such as information noise, network size, and network structure, should be taken into consideration. Thirdly, rewards do not make sense all the time, which differs from recent studies on incentive design. Managers should implement the specific policy to promote active participation.

Limitations and future directions are concluded as follows. Firstly, although external motivations (e.g., network structure and information noise) are identified as critical factors influencing the decision of user participation, the question regarding how to manage user interactive network in a proper approach is not clear. The valuable direction for future research is suggested to explore the reconfiguration of user interactive network [42]. Secondly, all data used in computational experiments are hypothetical. Real data (particularly, how to abstract real data from practices is a difficult work) are recommended for the further research, although hypothetical data are also suitable for describing all possible situations in real practices. Thirdly, in this paper, all users are assumed as homogeneous stakeholders in a co-creation community. It is suggested that the characteristics of different users should be further distinguished in the future, so as to develop an asymmetrical evolutionary game model for describing their returns and costs in the user participation system, and thus making a comparison with the results observed in this paper. Fourthly, the improved NEGM should be developed to fit the real practice. For example, a memory of the actions and a genetic evolution (not just to imitate the strategy of the neighbor with a better reward) is recommended for future research The trade-off among different users can be optimized by the multi-objective programming model [43]..

Data Availability

The data used to support the finding of this paper are included within this paper.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant nos. 71672074, 72072072, and 71904021), Natural Science Foundation of Guangdong Province of China (No. 2019A1515010045), 2018 Guangzhou Leading Innovation Team Program (China) (Grant no. 201909010006) and Social Science Planning Project in Chongqing (Grant no. 2019QNGL27).

References

[1] H. J. Lee, D.-H. Lee, C. R. Taylor, and J.-H. Lee, "Do online brand communities help build and maintain relationships with consumers? A network theory approach," *Journal of Brand Management*, vol. 19, no. 3, pp. 213–227, 2011.

[2] Y. Yi and T. Gong, "Customer value co-creation behavior: scale development and validation," *Journal of Business Research*, vol. 66, no. 9, pp. 1279–1284, 2013.

[3] N. Ind, N. Coates, and K. Lerman, "The gift of co-creation: what motivates customers to participate," *Journal of Brand Management*, vol. 27, no. 2, pp. 181–194, 2020.

[4] S. Kamboj and Z. Rahman, "Understanding customer participation in online brand communities," *Qualitative Market Research: An International Journal*, vol. 20, no. 3, pp. 306–334, 2017.

[5] X. Lu, C. W. Phang, and J. Yu, "Encouraging participation in virtual communities through usability and sociability development," *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, vol. 42, no. 3, pp. 96–114, 2011.

[6] H.-T. Tsai, H.-C. Huang, and Y.-L. Chiu, "Brand community participation in Taiwan: examining the roles of individual-, group-, and relationship-level antecedents," *Journal of Business Research*, vol. 65, no. 5, pp. 676–684, 2012.

[7] J. Kang, L. Tang, and A. M. Fiore, "Enhancing consumer-brand relationships on restaurant Facebook fan pages: maximizing consumer benefits and increasing active participation," *International Journal of Hospitality Management*, vol. 36, pp. 145–155, 2014.

[8] B. Xu and D. Li, "An empirical study of the motivations for content contribution and community participation in Wikipedia," *Information & Management*, vol. 52, no. 3, pp. 275–286, 2015.

[9] J. Nielsen, "Participation inequality: encouraging more users to contribute," 2006, https://www.researchgate.net/publication/200772620_Participation_Inequality_Encouraging_More_Users_to_Contribute.

[10] T. Zhou, "Understanding users' participation in online health communities: a social capital perspective," *Information Development*, vol. 36, no. 3, pp. 403–413, 2020.

[11] D. Roberts, M. Hughes, and K. Kertbo, "Exploring consumers' motivations to engage in innovation through co-creation activities," *European Journal of Marketing*, vol. 48, no. 1-2, pp. 147–169, 2014.

[12] C. Chen, R. Du, J. Li, and W. Fan, "The impacts of knowledge sharing-based value co-creation on user continuance in online communities," *Information Discovery and Delivery*, vol. 45, no. 4, pp. 227–239, 2017.

[13] C. Fang and J. Zhang, "Users' continued participation behavior in social Q&A communities: a motivation perspective," *Computers in Human Behavior*, vol. 92, pp. 87–109, 2019.

[14] S. Malinen, "Understanding user participation in online communities: a systematic literature review of empirical studies," *Computers in Human Behavior*, vol. 46, pp. 228–238, 2015.

[15] K. Sigmund and M. A. Nowak, "Tides of tolerance," *Nature*, vol. 414, no. 6862, pp. 403–405, 2001.

[16] P. Shukla and J. Drennan, "Interactive effects of individual- and group-level variables on virtual purchase behavior in online communities," *Information & Management*, vol. 55, no. 5, pp. 598–607, 2018.

[17] A. Al-Dhanhani, R. Mizouni, H. Otrok, and A. Al-Rubaie, "A game theoretical model for collaborative groups in social applications," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5056–5065, 2014.

[18] G. Jiang, F. Ma, J. Shang, and P. Y. K. Chau, "Evolution of knowledge sharing behavior in social commerce: an agent-based computational approach," *Information Sciences*, vol. 278, pp. 250–266, 2014.

[19] K. Musial, M. Budka, and K. Juszczyszyn, "Creation and growth of online social network," *World Wide Web*, vol. 16, no. 4, pp. 421–447, 2013.

[20] J. Vassileva, "Motivating participation in social computing applications: a user modeling perspective," *User Modeling and User-Adapted Interaction*, vol. 22, no. 1-2, pp. 177–201, 2012.

[21] T. Kohler, J. Fueller, D. Stieger, and K. Matzler, "Avatar-based innovation: consequences of the virtual co-creation experience," *Computers in Human Behavior*, vol. 27, no. 1, pp. 160–168, 2011.

[22] H. Akman, C. Plewa, and J. Conduit, "Co-creating value in online innovation communities," *European Journal of Marketing*, vol. 53, no. 6, pp. 1205–1233, 2019.

[23] H.-H. M. Lee and W. van Dolen, "Creative participation: collective sentiment in online co-creation communities," *Information & Management*, vol. 52, no. 8, pp. 951–964, 2015.

[24] A. R. Lee and K. K. Kim, "Customer benefits and value co-creation activities in corporate social networking services," *Behaviour & Information Technology*, vol. 37, no. 7, pp. 675–692, 2018.

[25] M. Chepurna and J. R. Criado, "Identification of barriers to co-create on-line: the perspectives of customers and companies," *Journal of Research in Interactive Marketing*, vol. 12, no. 4, pp. 452–471, 2018.

[26] H. Shen, L. Wu, S. Yi, and L. Xue, "The effect of online interaction and trust on consumers' value co-creation behavior in the online travel community," *Journal of Travel & Tourism Marketing*, vol. 37, no. 4, pp. 418–428, 2020.

[27] T. Chen, J. Drennan, L. Andrews et al., "User experience sharing: understanding customer initiation of value co-creation in online communities," *European Journal of Marketing*, vol. 52, no. 5-6, pp. 1154–1184, 2018.

[28] C. Li, F. Zhang, C. Cao, Y. Liu, and T. Qu, "Organizational coordination in sustainable humanitarian supply chain: an evolutionary game approach," *Journal of Cleaner Production*, vol. 219, pp. 291–303, 2019.

[29] L. Xie and H. Han, "Capacity sharing and capacity investment of environment-friendly manufacturing: strategy selection and performance analysis," *International Journal of Environmental Research and Public Health*, vol. 17, no. 16, p. 5790, 2020.

[30] Z. H. Rong, X. R. Xu, and Z. X. Wu, "Experiment research on the evolution of cooperation and network game theory," *Scientia Sinica (Physica, Mechanica and Astronomica*, vol. 1, pp. 118–132, 2020, in Chinese.

[31] R. Esmaeilyfard, F. Hendessi, M. H. Manshaei et al., "A game-theoretic model for users' participation in ephemeral social vehicular networks," *International Journal of Communication Systems*, vol. 32, no. 12, Article ID e3998, 2019.

[32] L. Xu, C. Jiang, Y. Chen, Y. Ren, and K. J. R. Liu, "User participation in collaborative filtering-based recommendation systems: a game theoretic approach," *IEEE Transactions on Cybernetics*, vol. 49, no. 4, pp. 1339–1352, 2019.

[33] M. Liu, Y. Ma, Z. Liu, and X. You, "An IUR evolutionary game model on the patent cooperate of Shandong China," *Physica A: Statistical Mechanics and Its Applications*, vol. 475, no. 1, pp. 11–23, 2017.

[34] D. Friedman, "Evolutionary games in economics," *Econometrica*, vol. 59, no. 3, pp. 637–666, 1991.

[35] S. L. Johnson, S. Faraj, S. Faraj, and S. Kudaravalli, "Emergence of power laws in online communities: the role of social mechanisms and preferential attachment," *MIS Quarterly*, vol. 38, no. 3, pp. 795–808, 2014.

[36] C. Li and P. K. Maini, "An evolving network model with community structure," *Journal of Physics A: Mathematical and General*, vol. 38, no. 45, p. 9741, 2005.

[37] W. Liang, Y. Wang, Y. Shi, and G. Chen, "Co-occurrence network analysis of modern Chinese poems," *Physica A: Statistical Mechanics and Its Applications*, vol. 420, pp. 284–293, 2015.

[38] Z.-X. Wu, J.-Y. Guan, X.-J. Xu, and Y.-H. Wang, "Evolutionary prisoner's dilemma game on Barabási-Albert scale-free networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 379, no. 2, pp. 672–680, 2007.

[39] G. Szabo, J. Vukov, and A. Szolnoki, "Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices," *Physical Review E*, vol. 72, Article ID 047107, 2005.

[40] D. Wu, H. Liu, Y. Bi, and H. Zhu, "Evolutionary game theoretic modeling and repetition of media distributed shared in P2P-based VANET," *International Journal of Distributed Sensor Networks*, vol. 10, no. 6, Article ID 718639, 2014.

[41] L. D. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, "Characterization of complex networks: a survey of measurements," *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.

[42] H.-J. Li, Z. Bu, Z. Wang, and J. Cao, "Dynamical clustering in electronic commerce systems via optimization and leadership expansion," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5327–5334, 2020.

[43] C. J. Cao, C. D. Li, Q. Yang, Y. Liu, and T. Qu, "A novel multi-objective programming model of relief distribution for sustainable disaster supply chain in large-scale natural disasters," *Journal of Cleaner Production*, vol. 174, pp. 1422–1435, 2018.

Hindawi
Complexity
Volume 2020, Article ID 6616988, 16 pages
https://doi.org/10.1155/2020/6616988

Review Article

Evaluation of the Urban Low-Carbon Sustainable Development Capability Based on the TOPSIS-BP Neural Network and Grey Relational Analysis

Wei Zhang,[1] Xinxin Zhang,[1] Fan Liu,[2] Yan Huang,[2] and Yuwei Xie[2]

[1]*School of Public Administration, Central China Normal University, Wuhan 430079, China*
[2]*School of Business Administration, Zhongnan University of Economics and Law, Wuhan 430073, China*

Correspondence should be addressed to Fan Liu; liufan_zuel@163.com

Received 1 November 2020; Revised 18 November 2020; Accepted 10 December 2020; Published 22 December 2020

Academic Editor: Baogui Xin

Copyright © 2020 Wei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of industrialization and urbanization, cities have become the main carriers of economic activities. However, the long-term development of cities has also caused damage to resources and the environment. Hence, objective and scientific evaluation of urban low-carbon sustainable development capacity is very important. An index system of urban low-carbon sustainable development capability is constructed in this paper, and a TOPSIS-BP neural network model is established to evaluate the low-carbon sustainable development capability of Beijing, Shanghai, Shenzhen, and Guangzhou in China. At the same time, the difference degree of low-carbon sustainable development level in these four cities is analyzed by standard deviation and coefficient of variation, and the influencing factors of urban low-carbon sustainable development ability are extracted by grey correlation analysis. The results show that (1) the capability of low-carbon sustainable development in four cities is rising and the difference of low-carbon sustainable development capability is decreasing; (2) the general view that the higher the general investment in low-carbon sustainable development, the higher the level of low-carbon sustainable development in cities has not been verified; (3) with the change of time series, the factors affecting the capability of low-carbon sustainable development in the same city are different and the influence of the same factor on the capability of low-carbon sustainable development in different cities is different.

1. Introduction

Green development emphasizes the coordination and mutual benefit of economic development, resource utilization, and ecological environment. Campbell [1] used a triangle model to study the priority of urban development planning. He believes that urban and rural development must take sustainable development as a vision and coordinate the conflicts between economic, environmental, and social interests. At present, green development has become an important trend in social and economic development. As the main carrier of economic activities, cities play an important role in the process of promoting economic development. According to the report of the National Bureau of Statistics, from 2010 to 2019, China's urbanization rate increased from 49.95% to 60.6%, an average annual increase of 1.065%. The rapid development of urbanization has brought a lot of employment opportunities and promoted the rapid development of real estate, infrastructure construction, and other related industries. This has greatly promoted the growth of GDP. According to a survey conducted by the National Bureau of Statistics, in 2019, China's top 100 cities have 13% of the country's land and 50% of the population, creating about 73% of GDP, accounting for about 62% of the country's commercial housing sales. However, the rapid development of cities has also brought about a large amount of energy consumption and energy and carbon dioxide emissions, and the contradiction between economic development and resources and the environment has become more and more intense [2]. Discharge of industrial waste and waste of water resources has brought severe challenges to the sustainable development of urban economic

environment. In 2019, the average PM2.5 concentration of 337 prefecture-level and above cities across the country was 36 micrograms/m^3, and cities with up to standard ambient air quality accounted for only 46.6% of all cities. Under the new economic normal, how to achieve energy conservation and emission reduction, enhance the sustainability of resource utilization, and reduce environmental pressure in urban development has become an important challenge for urban green development.

Therefore, there are more and more researches on urban green sustainable development. In terms of research content, part of the research focuses on the planning of urban green spaces. Van Herzele and Wiedemann [3] believed that urban green space has largely affected the quality of urban life. Therefore, they proposed a comprehensive index and used GIS technology to study the use of urban green space [4]; Horwood [5] uses the green infrastructure (GI) method to study the construction of urban green spaces and points out that the way of economic development affects the construction of urban green spaces. In addition, some studies mainly focus on the urban ecological environment [5–7]. Thorén [8] emphasized the importance of the environment and established an indicator system to evaluate the sustainability of the urban green structure. Brundtland et al. [9] believe that the sustainable development of the urban environment needs to meet the economic and social welfare conditions to improve the quality of the residents' living environment. Some scholars study the green development of cities from the aspects of technological innovation and technology [10, 11]. By designing an index system, evaluating the degree of green development of a city has also become an aspect of research [12]. In terms of research methods, Zhao et al. [13] is based on analytic hierarchy process (AHP) and uses data cluster analysis to study China's regional innovation and development capabilities; Li and Lin [14] used the DEA model to analyze the green growth rate of China's manufacturing industry; Duan et al. [15] based on the AHP-entropy method established an index system to evaluate the development level of Dalian's low-carbon economy. The main focus of the abovementioned literature is the sustainable development of the urban ecological environment. It pays less attention to the sustainable development of the urban economy, and the research method is relatively simple. In the case of more data, the amount of calculation is large and error-prone. However, in the new economic situation, the coordinated promotion of high-quality economic development and ecological environmental protection is the key to improving urban green development capabilities. In addition, the evaluation of urban green development capabilities also requires more complete evaluation methods to be able to more objectively evaluate urban green development capabilities and analyze the factors that affect the level of urban green development.

Beijing, Shanghai, Shenzhen, and Guangzhou are the four most economically powerful cities in mainland China. They are all national central cities and are also known as the "first-tier cities" of China's three major gateways. Among them, Beijing is the political center of China. As of the end of 2019, Beijing's urbanization rate reached 86.6%. "2016 China Urban Sustainability Report: Measuring Ecological Investment and Human Development" pointed out that Beijing's Human Development Index ranked second among Chinese cities. Shanghai is located in the Yangtze River Delta, in the east of China, at the mouth of the Yangtze River. In 2014, Shanghai's GDP ranked first among Chinese cities and second in Asia. In 2019, the "Top 100 Chinese Cities Green Competitiveness Ranking" was released, and Shanghai ranked 10th. Shenzhen is the window of China's reform and opening up. In the 2012 "World's Most Economically Competitive Cities" list by the Economist, Shenzhen ranked second; in 2019, it ranked third in the China City Creative Index. Guangzhou is an important central city and international trade center in China. Guangzhou has more than 8,700 national high-tech enterprises, ranking top three in the country in total. The four cities of Beijing, Shanghai, Shenzhen, and Guangzhou are relatively at the leading level in terms of comprehensive strength and competitiveness among the cities in mainland China, with a strong economic foundation and strong scientific research capabilities. Therefore, compared with other cities in China, the four cities of Beijing, Shanghai, Shenzhen, and Guangzhou have relatively high urban green development capabilities and levels. Assess the green development level of these four cities and evaluate the factors affecting the urban green development level. The analysis can serve as a model and reference for the green development of other cities in China.

This paper takes Beijing, Shanghai, Shenzhen, and Guangzhou as the research objects and designs an index system for evaluating the level of urban green development. Based on the entropy TOPSIS-BP neural network model, the green development levels of Beijing, Shanghai, Shenzhen, and Guangzhou are evaluated. At the same time, this paper uses the gray correlation analysis method to analyze the factors that affect the level of urban green development and provides references for the green development of other cities.

2. Literature Review

2.1. Research on the Content of Urban Green Development. The green development and sustainable development of a city are in the same line [16]; early research on urban green development focused on environmental sustainability [17]. Therefore, some scholars have done a lot of research on urban green space and ecological environment. For example, Haq [18] conducted empirical research on different cities and proposed to maintain urban green space to develop sustainable cities that improve the environment; Budruk et al. [19] investigates the use of green space in Indian cities and analyzes the importance of urban green space to sustainable urban green development; Wolch et al. [20] compares the development of green cities in the United States and China and supports sustainable urban green development strategies to protect the urban ecology; Dou et al. [21] emphasized the important role of urban ecological environment in urban green development and believed that implementing an ecological ecological city development strategy is the most effective way to achieve sustainable

urban development; Wu et al. [22] reviewed the research on urbanization and urban ecology in China and believed that the sustained and rapid development of Chinese cities requires the establishment of a scientific urban ecosystem; Niemelä et al. [23, 24] believe that the ecological service system is essential to urban life, and the green and sustainable development of cities requires reasonable planning of land resources, increasing vegetation coverage, and maintaining biodiversity.

With the development of the economy and society and the in-depth understanding of the concept of sustainable development, the research content of urban green development is no longer limited to the sustainable ecological environment and should be coordinated with environmental protection in the process of economic development [25]. Rees [26] believes that the study of urban economics has neglected the urban ecological environment and pointed out that it is necessary to formulate sustainable development strategies for the urban economy based on changes in the ecological environment; Campbell [1] believes that researchers should take sustainable urban development as the main goal. At the same time, it is combined with technology to solve the problem of incoordination between economic development and ecological environment; Dempsey et al. [27] believe that the green development of cities should not only consider the ecological environment but also the sustainable development of the urban economy; Rees et al. [26, 27] took British urban development as the research object, incorporated economic and social aspects into the content of sustainable urban development, and discussed the relationship between urban form and social sustainability. With the rapid development of urbanization and rapid changes in the ecological environment, it is increasingly important to explore the synergistic relationship between urban politics, economy, and ecological environment [28]. Urban green building development also plays an important role in urban green development [29]. The development of urban green space systems can effectively respond to the challenges of urbanization, such as protecting biodiversity and adapting to climate change [30, 31]. Therefore, Barles [32] reviews the process of urban development and believes that while the city consumes energy and resources in the development process; it must take into account the impact on the ecological environment. This is an unavoidable problem for the green development and sustainable development of cities.

The green development of the city mainly emphasizes the coordinated development of economy, ecology and society, and optimization of resource utilization. It can be seen from the abovementioned literature research that the research direction and focus of different scholars are different. These studies all involve the green development of cities, such as the planning and management of urban green space, the utilization of urban land resources, and the importance of the sustainable development of the urban ecological environment. However, the above literature mainly emphasizes the importance of urban sustainable development from a macroperspective, and there is little analysis of the factors that realize urban green development. Because urban development needs to take up a lot of natural resources and other resources, it is also necessary to achieve coordinated development with the economy, environment, and society and maximize the benefits of urban development. Therefore, there are still many problems to be solved on how to improve the city's green development ability and level.

2.2. Research on the Evaluation of Urban Green Development. With the rapid development of urbanization, there are more and more researches on urban green development. How to evaluate the ability and level of urban green development has also become the research focus of scholars. In terms of evaluation content, Chen and Wang [33] used panel data of 285 Chinese cities to conduct quantitative analysis to evaluate urban green spaces. This allows decision-makers to better weigh the relationship between economic development and natural facilities. Jin et al. [11] use the two influencing factors of macroeconomics and high-level innovation ability to evaluate the performance of urban green development. Li et al. [34] designed and developed 52 indicators based on Jining, Shandong, China. These indicators are economic growth and economic efficiency, ecological environment protection, and urban infrastructure construction. And, they developed a comprehensive index method of completely replacing polygons to evaluate the sustainable development ability of the city. Zhang et al. [10] used 103 cities in China as research objects, selected two indicators of knowledge innovation and product innovation, and used a spatial autoregressive model to verify the impact of innovation on the level of urban green development. Guan et al. [35] took Chongqing, China, as an example, combined system dynamics (SD) and geographic information system (GIS), and proposed a dynamic combination method of SD-GIS. The sustainable development level of Chongqing through modeling is evaluated. In terms of evaluation methods, Thinh et al. [36] created an ARC/INFO database of land use patterns. And through modeling, a cluster analysis of 116 cities in Germany was carried out to find ways to promote the sustainable development of cities. Thinh et al. [36, 37] used an improved entropy method combining experts and entropy weights to propose an indicator system for the evaluation of urban circular economy development. And put forward suggestions to promote the development of urban circular economy. Ding et al. [38] constructed an indicator system from three aspects: society, economy, and environment. The TOPSIS-entropy method is used to evaluate the sustainable development level of 287 prefecture-level and above cities in China, and the spatial distribution of the urban sustainable development level is analyzed. Meng et al. [39] used panel data of 31 provinces in China as the object to establish an indicator system and used the catastrophe progression method to measure the level of green economy development in 31 provinces in China. Lin and Ying [40] used the DEA model to evaluate the green development efficiency of urban agglomerations in the Middle Delta and the Yangtze River Delta and compared their green development efficiency. Zheng et al. [41] measured the green development

level of 78 cities in China from the perspective of urban household carbon emissions.

From the above evaluation research on urban green development, it can be found that most scholars use the method of establishing an index system to evaluate the level of urban green development. However, most of these indicators focus on urban green space planning and ecological environment, while there are few studies that take into account economic development, ecological environment, and social development. And, it is lack of attention to scientific research and innovation. In improving the ability of urban green development, the innovation of green technology plays an important role in sustainable economic growth and alleviating environmental pressure. In addition, most of the existing studies use AHP, entropy method, and TOPSIS to determine indicator weights and calculate the comprehensive level of urban green development. These research methods are subjective. Kahn [42] believes that selecting multiple indicators and adding weights through objective methods can solve subjective problems to a certain extent. This article summarizes the research of existing scholars and adds the indicator of green innovation technology. This article adopts the TOPSIS-BP neural network method to evaluate the green development review of the four cities of Beijing, Shanghai, Shenzhen, and Guangzhou, which is more objective. In addition, this paper also uses methods such as grey correlation analysis to extract the factors that affect the green development of cities, analyze which are the important factors, and provide certain references for the green development of other cities.

3. Research Design

3.1. Model Construction. The entropy TOPSIS method can eliminate subjective errors and relatively objectively measure the weight of each evaluation index and the level of green development of the city. The BP neural network can learn the standards. After training, it can stably simulate the evaluation of experts, determine the weight of each index, reduce the influence of subjective factors in different situations, and also get the green development level of each city. And, the BP neural network can verify the results obtained by the entropy TOPSIS method, so as to more truly reflect the level of urban green development.

However, the entropy TOPSIS method and BP neural network can measure the green development level of different cities. However, different cities have differences in resource endowments, resource allocation, location conditions, and external environment, which will cause differences in the level of green development and influencing factors between cities. Therefore, it is necessary to use the standard deviation and the coefficient of variation to quantitatively measure the degree of difference and use the grey correlation analysis to identify the factors that cause the difference.

3.1.1. Entropy Weight TOPSIS Method. The entropy method is an objective weight determination method that can be used to measure the weight of known data [43]. Rubinstein [44] proposes an entropy weight evaluation method is to determine the index weight, which is used to evaluate the water quality of the Three Gorges Reservoir. Zhao et al. [45] also use the entropy method to determine the index weight to assess environmental vulnerability; Sun et al. [46] use the entropy TOPSIS evaluation model to evaluate the impact of green technological innovation on the ecological economic efficiency of strategic emerging industries. In order to truly reflect the green development capabilities of China's first-tier cities, this article uses a relatively objective entropy method to determine the weight of each indicator. The specific steps are as follows.

Firstly, the raw data has to be standardized. The indicators that measure the green development capabilities of the first-tier cities in China have differences in dimensions and orientations. In order to avoid the influence between data, all indicators should be normalized according to the orientation of the data. Since different indicators have different impacts on urban green development, the indicators y_{ij} can be divided into positive indicators y_{ij}^{+} and negative indicators y_{ij}^{-}. The formulas are as follows:

Positive indicators:

$$y_{ij}^{+} = \frac{x_{ij} - \min\{x_{1j}, x_{2j}, \ldots, x_{nj}\}}{\max\{x_{1j}, x_{2j}, \ldots, x_{nj}\} - \min\{x_{1j}, x_{2j}, \ldots, x_{nj}\}}. \tag{1}$$

Negative indicators:

$$y_{ij}^{-} = \frac{\max\{x_{1j}, x_{2j}, \ldots, x_{nj}\} - x_{ij}}{\max\{x_{1j}, x_{2j}, \ldots, x_{nj}\} - \min\{x_{1j}, x_{2j}, \ldots, x_{nj}\}}, \tag{2}$$

x_{ij} is the jth index of i area, $\min(x_{ij})$ is the minimum value of the sample data, and $\max(x_{ij})$ is the maximum value of the sample data.

Secondly, the entropy H_j is calculated. The formula is

$$H_j = -k \sum f_{ij} \times \ln f_{ij}, \tag{3}$$

where $k = (\ln m)^{-1}$ and f_{ij} is the the normalized value of y_{ij}. Because the scope of y_{ij} is $[0, 1)$, we normalized y_{ij} by formula (4) to guarantee that $\ln f_{ij}$ is meaningful:

$$f_{ij} = \frac{1 + y_{ij}}{\sum_{i=1}^{m}\left(1 + y_{ij}\right)}. \tag{4}$$

Finally, calculate the weights:

$$W_j = \frac{1 - H_j}{\left(n - \sum_{j=1}^{m} H_j\right)}. \tag{5}$$

Then, it is combined with the weighted Topsis method to calculate the green development capacity of the first-tier

cities in mainland China from 2009 to 2018. Weighted Topsis is also known as the distance method of superior and inferior solution. Since it can make full use of the original data to reflect the superiority and inferiority of each evaluation plan, it is used to calculate the green development capabilities of Beijing, Shanghai, Shenzhen, and Guangzhou. Rashidi and Cullinane [47] applying the Topsis method to sustainable supplier evaluation, it is found that TOPSIS is superior to DEA in terms of computational complexity and sensitivity to changes in the number of suppliers. Dos Santos et al. [48] use TOPSIS to determine environmental standards for evaluating and selecting green suppliers in the furniture industry; In [49], combined with TOPSIS, the influence of multiattribute decision-making on the efficiency of the reverse logistics industry of scrapped automobiles is analyzed to improve the efficiency of resource utilization.

Specific steps are as follows:

Firstly, weigh the preprocessed data. Multiply the preprocessed data with the corresponding index weight to obtain a weighted normalized decision matrix:

$$A = a_{ij} = \left(Z_{ij} \times W_j\right) \tag{6}$$

Secondly, build positive ideal solution vectors separately a_j^+ and negative ideal solution vector a_j^-:

$$\begin{aligned} a_j^+ &= \max\left(a_{1j}, a_{2j}, a_{3j}, \ldots, a_{mj}\right), \\ a_j^- &= \min\left(a_{1j}, a_{2j}, a_{3j}, \ldots, a_{mj}\right). \end{aligned} \tag{7}$$

Thirdly, calculate the optimal distance and the worst distance. The Euclidean calculation formula is used to calculate the optimal distance from the evaluation value vector to the positive ideal solution and the worst distance to the negative ideal solution. The optimal and worst distance calculation formulas are, respectively,

$$\begin{aligned} d_j^+ &= \sqrt{\sum_{j=1}^{n}\left(a_{ij} - a_j^+\right)^2}, \\ d_j^- &= \sqrt{\sum_{j=1}^{n}\left(a_{ij} - a_j^-\right)^2}. \end{aligned} \tag{8}$$

Finally, calculate the closeness of each evaluation object. The closeness indicates how close each evaluation object is to the optimal distance. The larger the value, the closer the evaluation object is to the optimal level.

The calculation formula is

$$B_i = \frac{d_i^-}{d_i^- + d_i^+}, \quad 0 \le B_i \le 1. \tag{9}$$

3.1.2. Neural Network Model Construction. BP neural network is a multilayer feedforward neural network trained according to the error back propagation algorithm, which can make important contributions to the development of faster learning algorithms and research [50]. The neuron is the basic unit of the neural network. For the ith neuron, X_1, $X_2, \ldots, X_j$ are the inputs of the neuron, and the input is often the independent variable that has a key impact on the system model, and $W_1, W_2, \ldots, W_j$ are the connections the weight adjusts the proportion of each input. Structurally, the BP network has an input layer, a hidden layer, and an output layer; in essence, the BP algorithm uses the square of the network error as the objective function and uses the gradient descent method to calculate the minimum value of the objective function. Therefore, using BP neural network to predict and evaluate is more objective and scientific. Peng and Lai [51] use neural network model to evaluate the effectiveness of tourism e-commerce service innovation. Li and Wang [52] use the BP correction model to predict India's dependence on foreign oil. Zhang et al. [53] based on the BP neural network established an evaluation model to evaluate the smart growth plan of the city.

The calculation process of the BP neural network model is as follows:

(1) The model is build with full connected layers:

$$\alpha_j = \sum_{i=1}^{d} w_{ij} * x_i, \tag{10}$$

w_{ij} is the weight and x_i is the input of each layer.

(2) The activation function is

$$b_j = f\left(\alpha_j - \gamma_j\right). \tag{11}$$

(3) Calculate the error:

$$E_k = \frac{1}{2}\sum_{h=1}^{l}\left(y_h^k - y_h^k\right)^2. \tag{12}$$

(4) Update each weight with the chain rule:

$$w_{jh}' = w_{jh} - \eta * \frac{\partial E}{\partial w_{jh}}. \tag{13}$$

(5) Repeat execution to minimize the cost function value.

3.2. Analysis Method of Urban Green Development

3.2.1. Standard Deviation and Variation Coefficient. This paper used standard deviation and variation to analyze the difference characteristics of urban green development capabilities. The larger the standard deviation and the coefficient of variation, the greater the difference in green development capabilities between cities. Bobinaite et al. [54] combined the standard deviation and the coefficient of variation to analyze the price characteristics (volatility and

peak value) of the Lithuanian and Polish day-a-day power market in order to better understand the process of price formation in the Polish and Lithuanian power market, so as to provide a theoretical basis for the country to implement national energy policies and measures. The calculation formulas for standard deviation and coefficient of variation are

$$\delta^2 = \left[\frac{1}{n}\sum_{1}^{n}\left(X_{ij} - \frac{1}{n}\sum_{1}^{n} X_{ij}\right)\right]^{1/2}, \tag{14}$$

where δ^2 indicates the standard deviation of the city's green development capability during the study period, $(1/n)\sum_1^n X_{ij}$ is the average level of green development capacity of n cities in the ith year, and n is the number of cities, where X_{ij} is the green development capacity of the j city in the i year.

Record the mean value $M = (1/n)\sum_1^n X_{ij}$, which represents the average green development capacity of n cities in the i year. The coefficient of variation (C·V) is the ratio of the standard deviation to the mean. The larger the value, the larger the green development capacity gap between cities.

3.2.2. Grey Relational Analysis. The development of the system is often affected by many factors, and the influence of each factor on the system is different. Grey relational analysis has no requirements on the size and regularity of the sample and can judge whether the connection is close according to the similarity of the geometric shape of the sequence curve. The closer the curve is, the greater the correlation between the corresponding sequences, and vice versa. Related scholars have also applied the gray relational analysis method to practical research. The grey correlation analysis method is used to determine the factors that have a strong correlation with carbon emissions, which provides a theoretical basis for reducing carbon emissions [55]. Sun and Tang [56] taking 12 cities in Hunan Province as the empirical analysis object and using the gray correlation analysis method, it is found that the correlation between scientific and technical personnel and economic development is greater than the correlation between scientific and technological expenditure and economic development. Therefore, the grey correlation analysis method is used to extract the factors that affect the city's green development ability.

Specific steps are as follows:

(1) Determine the comparison sequence and the reference sequence. Take the index to measure the green development capacity of the first-tier cities in China as the comparison sequence $X_1(k)$, taking the green development capabilities of Beijing, Shanghai, Shenzhen, and Guangzhou from 2009 to 2018 as a reference sequence $X_0(k)$.

(2) Preprocess the comparison sequence and the reference sequence. The formula is

$$X_i'(k) = \frac{X_{ij}}{(1/n)\sum_i^n X_{ij}}, \qquad X_0'(k) = \frac{X_0(k)}{(1/n)\sum_i^n X_0(k)}. \tag{15}$$

Get a new comparison sequence $X_i'(k)$ with reference sequence $X_0'(k)$.

(3) Calculate the grey correlation coefficient. Find the absolute difference between the preprocessed comparison sequence and the reference sequence $\Delta_{0i}(k)$; among them, $\Delta_{0i}(k) = |X_i'(k) - X_0'(k)|$. And, get the maximum difference M and the minimum difference m. According to the formula,

$$\delta_{0i}(k) = \frac{m + \rho M}{\Delta_{0i}(k) + \rho m}. \tag{16}$$

The grey correlation coefficient is calculated, ρ is the resolution coefficient, and 0.5 is better.

(4) Calculate the correlation coefficient of each indicator to the overall system $r(X_0, X_i)$, where $(X_0, X_i) = (\sum_{k=1}^n \delta_{0i}(k)/n)$. At this time, the correlation degree between each element in the indicator and the corresponding element in the parent sequence can be obtained. The higher the correlation degree, the stronger the influence on the green development ability.

3.3. Indicator Determination

3.3.1. Construction of Indicator System. The construction of the index system needs to be systematic, objective, and scientific. According to the concept and connotation of urban green development, this paper constructs an index system for comprehensive evaluation of urban green development. Singh et al. [57] compile the indicators related to sustainable development, from the three aspects of economy, environment, and society and put forward the use of total wastewater, solid waste, network communications, and other indicators to evaluate the city's sustainable development level. 24 indicators were determined to evaluate the level of sustainable development, including observation indicators for soil, water, air, and other factors [58]. Oh et al. [59] found that the sustainable development capacity of cities is related to human activities, but the carrying capacity of cities is limited. Therefore, maintaining sustainable urban development requires coordination of input and output. For example, while using land, water, and other resources for production and development, increase the investment in technology management to control the generation of waste water and waste gas and improve the city's carrying capacity. Feng et al. [60] established an indicator system based on resource consumption and economic and social development and evaluated the green development transformation of Chinese cities in two stages of input and output.

The research in Table 1 constructs a green development evaluation index system mainly from the perspective of input and output. The input indicators include capital, labor, and energy consumption, and the output indicators include GDP, CO_2 emissions, and SO_2 emissions. Therefore, referring to the above research results, this article mainly starts from the perspective of input and output to construct the evaluation index of green development of Chinese cities.

Table 1: Research on evaluation index system of green development ability.

Number	Author	Research area	Index
1	(Yuan et al.) [61]	Shandong peninsula City group, China	Resource input, nonresource input, ideal output and nonideal output
2	(Feng and Wang) [62]	China	Energy, labor, capital, GDP, and carbon dioxide emissions
3	(Qiu et al.) [63]	Xuzhou City group, China	Investment in fixed assets, number of employees in the whole society, total energy consumption, GDP, and industrial sulfur dioxide emissions
4	(Zhu et al.) [64]	31 provinces and cities in mainland China	Capital stock, labor, energy consumption, GDP, and carbon dioxide emissions
5	(Guo et al.) [65]	34 cities in northeast China	Construct evaluation indicators from the perspective of input-output, which mainly include indicators such as capital, labor, resource consumption, technological progress, economic development, industrial wastewater, industrial SO_2, and industrial dust
6	(Ma et al.) [66]	285 prefecture-level cities in China	From the perspective of input-output, the construction includes capital stock, number of employees, total water supply, annual electricity consumption, GDP, per capita disposable income, green coverage, public financial expenditure, industrial wastewater discharge, $PM_{2.5}$, SO_2, the evaluation index system including urban registered unemployment rate
7	(Feng et al.) [67]	165 countries worldwide	Energy consumption, labor, capital, GDP, SO_2 emissions, CO_2 emissions
8	(Pan et al.) [68]	China	Labor, energy consumption and capital stock, GDP, CO_2 emissions
9	(Chen et al.) [69]	China	Constructed green development evaluation indicators including labor, capital, energy consumption, expected output, and bad output.
10	(Shao et al.) [70]	Shanghai, China	Constructed input-output indicators including gross industrial output, capital, labor, energy consumption, carbon emissions, etc.

Among them, green input evaluation indicators include two aspects: technology input and resource utilization; green output includes two aspects: environmental impact and economic impact.

The green investment evaluation index system includes technology investment and resource utilization. The technical input includes R&D expenditure, fixed asset investment in the tertiary industry, number of research and development institutions, research and experimental development personnel, national science and technology awards, and amount of patent authorization. The green development of a city is related to the economic structure, and the increase in fixed investment in the tertiary industry can reflect the improvement of the economic development structure. At the same time, green and sustainable development is inseparable from technological innovation. R&D expenditures, the number of research and development institutions, research and test development personnel, national science and technology awards, and number of patent authorizations all reflect the city's scientific research and development in the development process. Support for technological innovation. Resource utilization includes total energy consumption, energy consumption per 10,000 yuan of GDP, comprehensive utilization of industrial solid waste, comprehensive utilization of industrial waste, and total water consumption. The total energy consumption, the energy consumption of 10,000 yuan of regional GDP, and the total water consumption can reflect the resource utilization and utilization efficiency of the city in development. The comprehensive utilization of industrial solid waste and the comprehensive utilization rate of industrial waste can reflect the city's ability in resource recycling.

The green output evaluation index system includes environmental and economic aspects. The city will produce economic benefits in the process of development, but it also affects the ecological environment. In terms of environment, the amount of industrial solid waste generated, the total amount of waste water discharge, and the amount of industrial waste gas emissions will damage the ecological environment, while increasing the amount of industrial solid waste disposal and forest coverage will have a positive impact on the ecological environment. The economic impact includes total GDP, technology market turnover, the added value of the tertiary industry, and the rate of decline in energy consumption for 10,000 yuan of regional GDP. The total GDP can reflect the level of economic development as a whole. The technological market turnover, the added value of the tertiary industry, and the rate of decline in energy consumption per 10,000 yuan of regional GDP can reflect the structure of economic growth and reflect the degree of urban green economy development. Table 2 summarizes the above evaluation indicators.

3.3.2. Data Sources. In order to ensure the authority and availability of the data, the data in this article come from "China City Statistical Yearbook (2010–2019)" and "China Regional Economic Statistics Yearbook (2010–2019)." The data for Beijing are also from the "Beijing Statistical Yearbook (2010–2019)," the data for Shanghai are from the "Shanghai Statistical Yearbook (2010–2019)," and the data for Shenzhen are from the "Shenzhen Statistical Yearbook (2010–2019)," Guangzhou City Data come from "Guangzhou Statistical Yearbook (2010–2019)." In addition, it also

Table 2: Index system.

System	Indicator	Variables
Green investment	Technology investment	R&D expenditure (100 million yuan)
		Fixed assets investment in tertiary industry (100 million yuan)
		Number of research and development institutions (a)
		Research and experimental development staff (person)
		National science and technology awards (items)
		Number of patents granted (items)
	Resource utilization	Total energy consumption (10,000 tons of standard coal)
		Energy consumption per 10,000 yuan of regional GDP (ton of standard coal)
		Comprehensive utilization of industrial solid waste (10,000 tons)
		Comprehensive utilization rate of industrial waste (%)
		Total water consumption (100 million cubic meters)
Green output	Environmental impact	Amount of industrial solid waste produced (10,000 tons)
		Industrial solid waste disposal (10,000 tons)
		Forest cover rate (%)
		Total wastewater discharge (10,000 tons)
		Industrial waste gas emissions (tons)
	Economic impact	GDP (100 million yuan)
		Technology market turnover (100 million yuan)
		Added value of tertiary industry (100 million yuan)
		Reduction rate of energy consumption per 10,000 yuan of GDP (%)

includes the statistical bulletins of economic and social development of various regions.

4. Empirical Research

4.1. Results of TOPSIS-BP Neural Network Model. According to the entropy TOPSIS method introduced above, this paper uses the panel data of Beijing, Shanghai, Shenzhen, and Guangzhou from 2009 to 2018 to calculate the green development level of Beijing, Shanghai, Shenzhen, and Guangzhou from 2009 to 2018. The calculation results are shown in Table 3.

After using the TOPSIS method to obtain the green development level of the four cities of Beijing, Shanghai, Shenzhen, and Guangzhou, this paper uses BP neural network to remodel and verify the results. Sun and Gao [71] used the Adaboost-BP neural network model when evaluating the potential energy-saving capabilities of China's power industry and believed that it was more accurate than conventional methods. Li et al. [72] used principal component analysis (PCA) and backpropagation (BP) neural network models to process factors and indicators to evaluate and analyze the development models of 35 smart cities in China. Therefore, using the TOPSIS-BP neural network model to evaluate and analyze the level of urban green development is more accurate and objective.

This paper has 40 samples of urban green development data, of which 36 samples are used as training sample data, and the other 4 sets of data are used as verification data. Firstly, use PyCharm as a coding tool and pytorch as a deep learning framework to build a BP neural network model. Perform data cleaning on the collected 40 data, delete the data with more null values, use the median method to insert some of the null values, and standardize the data to eliminate the differences between different characteristics. Second, build a model. This model uses 4 layers of neurons, of which the hidden layer is 2. The input neurons are 19, the number of neurons is 30 and 10, and the output neuron is 1, and the ReLU function is used as the activation function of the hidden layer. Increase the discreteness of the data, the learning rate is 0.0001, and the number of training samples is 1000.

Table 3: Test results based on the topsis model.

Years	Test results			
	Beijing	Shanghai	Guangzhou	Shenzhen
2009	0.192	0.398	0.319	0.307
2010	0.155	0.379	0.302	0.298
2011	0.193	0.335	0.218	0.379
2012	0.238	0.304	0.205	0.474
2013	0.697	0.332	0.228	0.384
2014	0.280	0.423	0.244	0.447
2015	0.293	0.441	0.262	0.439
2016	0.308	0.509	0.261	0.459
2017	0.313	0.514	0.605	0.544
2018	0.358	0.569	0.692	0.545

Accurate and objective training samples can ensure the accuracy of output results to a greater extent. In order to ensure an objective and accurate evaluation of the regional green development capabilities, select the first 36 sample data among the 40 data as the training set. The input vector is $X = [X\_1, X\_2, \ldots, X\_36]$. According to the calculation result of the TOPSIS method as the output result of the neural network, the target output vector is $Y = [Y\_1, Y\_2, \ldots, Y\_36]$. After outputting the results, use Python's Matplotlib to compare the expected results with the real results. Table 3 shows the comparison between the actual value and the predicted value of the TOPSIS-BP neural network model.

Based on the measurement results of the TOPSIS-BP neural network model, the degree of fit between the actual value and the predicted value can be observed more intuitively from Figure 1.

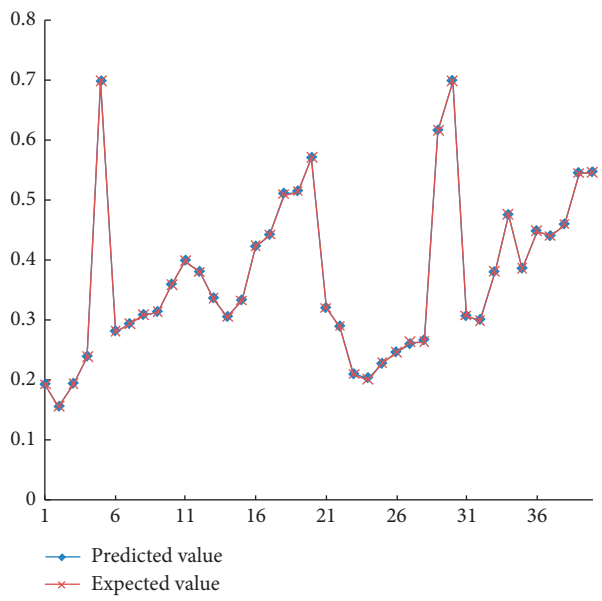


Figure 1: Fitting curve of the TOPSIS-BP neural network model.

Combining Table 4 and Figure 2, it can be seen that the actual value of the test sample fits the predicted value very well. The data calculated by TOPSIS and BP neural network models are consistent. Although there is a certain error, the error is small and the maximum error is 0.00423. In addition, the data of the verification data are the actual value (0.43888223, 0.45845163, 0.5440791, 0.5455292), the predicted value (0.438819, 0.458668, 0.543993, 0.545479), and the error is small. It can be seen that when the BP neural network is used for evaluation, the influence of subjective factors is reduced to a certain extent, and it can effectively evaluate and analyze complex problems. Moreover, the trained BP neural network model can be retained, as long as the indicator data is input, the green development level of other cities can be tested.

4.2. Difference in Green Development Capabilities. Based on the collected sample data, after standardizing the original data, combined with the weighted Topsis-BP neural network model, the green development capabilities of Beijing, Shanghai, Guangzhou, and Shenzhen in 2009–2018 can be calculated. At the same time, the standard deviation and coefficient of variation are combined to measure the degree of difference in green development capabilities between different cities.

On the whole, the green development capabilities of Beijing, Shanghai, Shenzhen, and Guangzhou are improving in fluctuations, but different cities show different characteristics. From the perspective of development level, compared with Beijing and Guangzhou, Shanghai, and Shenzhen have higher green development capabilities, with the highest value appearing in 2018, reaching 0.569 and 0.545, respectively. However, during the study period, Shanghai's green development ability experienced a process of decline first and then rise. The lowest value appeared in 2012, reaching 0.304. In fact, Shanghai is China's financial, trade, and technological innovation center, with high-tech industries dominated by integrated circuits, software, and biomedicine, and strong green development capabilities. Shenzhen is a national logistics hub and an international technological industry innovation center, and the total number of PCT international patent applications has ranked first in the country for 14 consecutive years, which shows that technological innovation has a significant effect on improving the city's green development capabilities.

The green development capacity of Guangzhou has gone through three obvious stages. Among them, from 2009 to 2012, the green development capacity declined year by year, but the rate of decline was relatively slow. From 2013 to 2016,

Table 4: Comparison of actual value and predicted value of the TOPSIS-BP neural network model.

Years	Expected value	Predicted value	Relative error (%)
Beijing			
2009	0.1918	0.1918	0.0074
2010	0.1548	0.1548	0.0258
2011	0.1931	0.1931	0.0199
2012	0.2381	0.2383	0.0498
2013	0.6968	0.6968	0.0003
2014	0.2803	0.2805	0.0739
2015	0.2927	0.2925	0.0480
2016	0.3077	0.3075	0.0416
2017	0.3127	0.3126	0.0427
2018	0.3584	0.3584	0.0009
Shanghai			
2009	0.3977	0.3982	0.1151
2010	0.3789	0.3794	0.1469
2011	0.3352	0.3354	0.0616
2012	0.3042	0.3043	0.0549
2013	0.3317	0.3320	0.1143
2014	0.4226	0.4216	0.2325
2015	0.4409	0.4410	0.0220
2016	0.5086	0.5097	0.2247
2017	0.5137	0.5137	0.0076
2018	0.5694	0.5695	0.0161
Guangzhou			
2009	0.3190	0.3191	0.0399
2010	0.2885	0.2885	0.0075
2011	0.2085	0.2087	0.0970
2012	0.2008	0.2025	0.8326
2013	0.2274	0.2263	0.4907
2014	0.2450	0.2453	0.1252
2015	0.2632	0.2590	1.6083
2016	0.2621	0.2655	1.3161
2017	0.6150	0.6151	0.0240
2018	0.6978	0.6977	0.0277
Shenzhen			
2009	0.3068	0.3057	0.3457
2010	0.2981	0.2994	0.4331
2011	0.3793	0.3790	0.0806
2012	0.4743	0.4743	0.0099
2013	0.3843	0.3849	0.1608
2014	0.4474	0.4473	0.0106
2015	0.4388	0.4389	0.0144
2016	0.4587	0.4585	0.0472
2017	0.5440	0.5441	0.0158
2018	0.5455	0.5455	0.0092

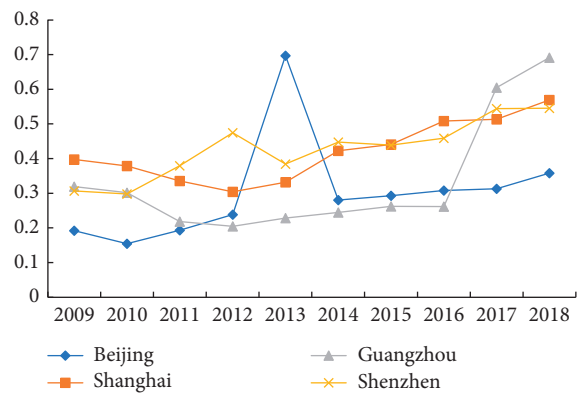


Figure 2: 2009–2018 Beijing, Shanghai, Shenzhen, and Guangzhou urban green development capabilities.

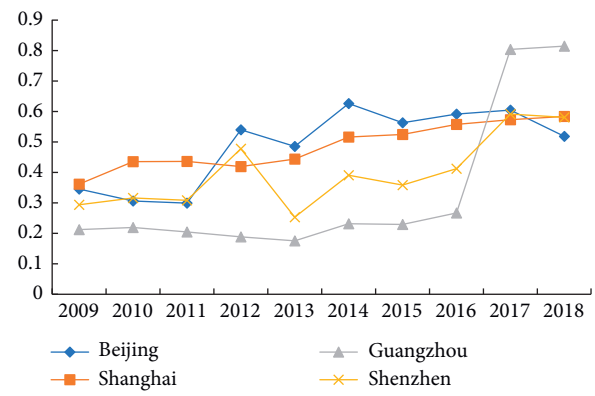


Figure 3: 2009–2018 Beijing, Shanghai, Shenzhen, and Guangzhou urban green investment.

the green development capacity of Guangzhou was slowly improved, only from 0.228 to 0.261. At this time, among the four cities, Guangzhou's green development capability is the last. After 2016, Guangzhou's green development capabilities have rapidly improved, with an average annual growth rate much higher than the other three cities, and in 2017-2018, the green development capabilities of the four cities ranked first. Beijing's green development capacity is slowly improving, reaching its peak in 2013, but its green development capacity is relatively low. In 2009–2011 and 2017-2018, the green development capacity was at the bottom of the first-tier cities.

It can be seen from Figure 3 that, on the whole, the urban green investment levels of Beijing, Shanghai, Shenzhen, and Guangzhou have risen in volatility, showing certain differences in different cities. During the study period, Beijing and Shanghai maintained a relatively high level of green investment, ranking first among the four cities in 2009–2011 and 2012–2016. Among these four cities, Guangzhou's green investment level was at the bottom between 2009 and 2016, and the improvement was extremely slow. There is a big gap with other cities. However, since 2016, Guangzhou's green investment level. It has been rapidly improved, and the level of green investment ranks first in the four cities. The level of green investment in Shenzhen has experienced a process of rising first, falling, and then slowly rising, showing a trend of volatility and rising overall. After 2017, the green investment level of the four cities has entered a flat phase.

In general, from 2009 to 2018, the green output levels of the four cities rose in fluctuations, and the green output levels of different cities have their own evolutionary trends in Figure 4. The level of green output in Beijing has always been lower than that of other cities, being the last of the four cities. But, in 2013, Beijing's green output level reached 0.751, ranking first among the four cities. The green output curve of Shanghai and Guangzhou showed a "U" shape and experienced a process of first decline and then slowly increase, and both reached the maximum in 2018. Shenzhen's green output level was higher than other cities from 2011 to 2016, ranking first among the four cities, but after 2016, it showed a certain degree of downward trend.

Both the standard deviation and the coefficient of variation are indicators to measure the difference in green

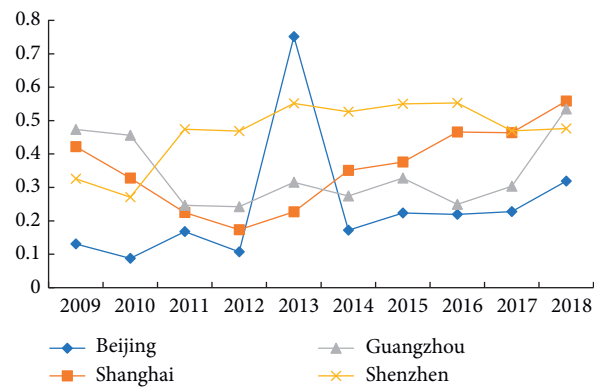


Figure 4: 2009–2018 Beijing, Shanghai, Shenzhen, and Guangzhou urban green output levels.

development capabilities between cities. On the whole, the standard deviation and the coefficient of variation showed a similar evolutionary trend, and both experienced a process of first becoming larger and then becoming smaller in Figure 5. From 2009 to 2013, the standard deviation and the coefficient of variation increased year by year, indicating that the difference in urban green development capabilities is gradually expanding. From 2014 to 2018, the standard deviation and coefficient of variation decreased year by year, indicating that the difference in green development capabilities between cities is gradually shrinking. In fact, since 2014, China's economy has gradually entered the "new normal," from factor-driven and investment-driven to innovation-driven, not only the role of emerging industries, service industries, and small and microenterprises, but also the manufacturing of steel, cement, electrolytic aluminum, etc. The industry has gradually realized "de-capacity," and the problems of resource waste and environmental pollution have been gradually solved, which has gradually reduced the green development gap between cities.

4.3. Influencing Factors of the Urban Green Development Level. Studying the influencing factors of the green development capacity of Beijing, Shanghai, Shenzhen, and Guangzhou is of great significance for how to improve the

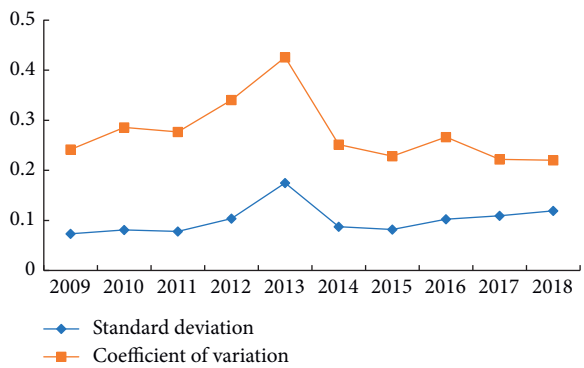


Figure 5: 2009–2018 Beijing, Shanghai, Shenzhen, and Guangzhou urban green development level difference index.

green development capacity and realize the sustainable development of the region. Therefore, using the grey correlation analysis method, according to the size of the gray correlation coefficient of each element, filter out the factors that affect the green development ability. It can be found that there are obvious temporal and spatial differences in the factors affecting the green development capacity in 2009–2018.

Under the same time series, the factors affecting the green development capabilities of Beijing, Shanghai, Guangzhou, and Shenzhen show certain differences. From Table 5, it can be found that, from 2009 to 2013, the factors affecting the green development capacity of the four major cities were quite different. Among them, the amount of patent authorization and industrial waste gas emissions are the primary factors affecting the improvement of the green development capabilities of Beijing and Shenzhen. The total energy consumption is the primary factor that affects the improvement of the green development capacity of Shanghai and Guangzhou, and the last influential factors have different degrees of influence on the green development capacity of different cities. Forest coverage, fixed asset investment in the tertiary industry, research and experimental development personnel, and number of research and development institutions are the last factors affecting the green development capabilities of Beijing, Shanghai, Guangzhou, and Shenzhen, respectively. At the same time, the same influencing factor has different degrees of importance to different cities. The number of research and development institutions has a second place in the influence of Guangzhou, but the influence on Shenzhen is in the 20th place. The impact of industrial waste gas emissions on Shenzhen ranks first, but the impact on Shanghai ranks 19th.

It can be seen from Table 6, compared with 2009–2013, the factors affecting the green development capabilities of the four major cities in 2014–2018 are still quite different. Among them, research and experimental development personnel are the primary factor affecting the improvement of Beijing and Shanghai's green development capabilities. The number of patent authorizations and the transaction volume of technology contracts are the factors affecting the improvement of the green development capabilities of Guangzhou and Shenzhen. The national science and technology awards and the comprehensive utilization of industrial solid waste are the last factors affecting the green development capabilities of Beijing and Shanghai, respectively. Industrial waste gas emissions are the last factors affecting the green development capabilities of Guangzhou and Shenzhen. At the same time, the same influencing factor has different degrees of importance to different cities. Research and experimental development personnel are the first factor that affects the green development capabilities of Beijing and Shanghai, but the impact on Guangzhou ranks 15th. The influence of technology contract turnover on Shenzhen ranks first, but the influence on Shanghai ranks 12th.

As the time series change, the factors that affect the green development capability of the same city are also changing. From 2009 to 2018, the primary factor affecting Beijing's green development capabilities changed from the number of patents granted to the number of research and experimental development personnel; the primary factor influencing Shanghai's green development capability has changed from total energy consumption to the number of research and experimental development personnel; the primary factor affecting Guangzhou's green development capabilities has changed from total energy consumption to the number of patents granted; the primary factor affecting Shenzhen's green development capabilities has changed from industrial waste gas emissions to technical contract turnover.

In the two time periods of 2009–2013 and 2014–2018, the number of patent grants, research and experimental development personnel, technology contract turnover, tertiary industry fixed asset investment, tertiary industry added value, and total energy consumption have appeared more than 3 times in the top 5 influencing factors in 4 cities. This shows that attaching importance to scientific and technological innovation, reducing unit energy consumption, and improving resource utilization are particularly important for improving green development capabilities.

5. Discussion

This article builds on the evaluation index system for the green development capability of four cities in Beijing, Shanghai, Shenzhen, and Guangzhou, based on the panel data published by each city in 2009–2018. The article combines entropy method, weighted topsis, and BP neural network to measure the green development capacity, green input level, and green output level of these four cities from 2009 to 2018. And, on this basis, the green development level of the four cities and the green development differences between the cities are analyzed, and then the grey correlation analysis method is used to identify the factors that affect the green development of the city.

The research results show that, first, the urban green development level of Beijing, Shanghai, Shenzhen, and Guangzhou has improved significantly, but there are obvious temporal and spatial differences. Among them, the green development capacity of Shanghai and Shenzhen is relatively higher than that of Beijing and Guangzhou. The green development capacity of Guangzhou has improved the

Table 5: 2009–2013 grey correlation coefficient of urban green development ability of Beijing, Shanghai, Shenzhen, and Guangzhou.

Variables	Beijing	Rank	Shanghai	Rank	Guangzhou	Rank	Shenzhen	Rank
R&D expenditure	0.780	6	0.633	17	0.501	19	0.675	9
Fixed assets investment in tertiary industry	0.801	3	0.617	20	0.557	11	0.699	7
Number of research and development institutions	0.747	12	0.686	11	0.830	2	0.579	20
Research and experimental development staff	0.769	9	0.689	10	0.454	20	0.715	5
National science and technology awards	0.674	17	0.901	2	0.518	18	0.629	15
Number of patents granted	0.806	1	0.751	6	0.530	14	0.715	4
Total energy consumption	0.726	15	0.903	1	0.911	1	0.736	3
Energy consumption per 10,000 yuan of regional GDP	0.701	16	0.684	12	0.544	12	0.659	12
Comprehensive utilization of industrial solid waste	0.769	10	0.871	3	0.785	3	0.615	18
Comprehensive utilization rate of industrial waste	0.777	8	0.837	4	0.685	5	0.663	11
Total water consumption	0.778	7	0.743	7	0.618	10	0.765	2
Amount of industrial solid waste produced	0.733	14	0.630	18	0.527	15	0.649	13
Industrial solid waste disposal	0.667	18	0.740	8	0.666	6	0.617	17
Forest cover rate	0.611	20	0.761	5	0.643	8	0.580	19
Total wastewater discharge	0.737	13	0.695	9	0.745	4	0.685	8
Industrial waste gas emissions	0.756	11	0.623	19	0.634	9	0.781	1
GDP	0.784	4	0.656	15	0.524	17	0.703	6
Technology market turnover	0.802	2	0.660	13	0.536	13	0.623	16
Added value of tertiary industry	0.782	5	0.655	16	0.525	16	0.671	10
Reduction rate of energy consumption per 10,000 yuan of GDP	0.658	19	0.657	14	0.645	7	0.649	14

Table 6: 2014–2018 China's first-tier cities' green development capacity grey correlation coefficient.

Variables	Beijing	Rank	Shanghai	Rank	Guangzhou	Rank	Shenzhen	Rank
R&D expenditure	0.899	7	0.841	10	0.851	4	0.693	5
Fixed assets investment in tertiary industry	0.913	4	0.851	9	0.784	9	0.698	3
Number of research and development institutions	0.870	11	0.922	3	0.818	8	0.684	9
Research and experimental development staff	0.929	1	0.982	1	0.655	15	0.680	13
National science and technology awards	0.598	20	0.609	16	0.784	10	0.655	15
Number of patents granted	0.919	3	0.776	11	0.874	1	0.719	2
Total energy consumption	0.724	13	0.494	19	0.578	19	0.621	19
Energy consumption per 10,000 yuan of regional GDP	0.921	2	0.967	2	0.781	11	0.681	12
Comprehensive utilization of industrial solid waste	0.622	18	0.462	20	0.745	12	0.623	17
Comprehensive utilization rate of industrial waste	0.676	17	0.577	17	0.655	16	0.622	18
Total water consumption	0.902	5	0.691	14	0.839	6	0.688	8
Amount of industrial solid waste produced	0.870	10	0.912	4	0.662	14	0.684	10
Industrial solid waste disposal	0.677	16	0.620	15	0.627	18	0.665	14
Forest cover rate	0.763	12	0.878	6	0.851	3	0.693	6
Total wastewater discharge	0.690	15	0.894	5	0.727	13	0.684	11
Industrial waste gas emissions	0.715	14	0.508	18	0.504	20	0.566	20
GDP	0.900	6	0.866	8	0.828	7	0.692	7
Technology market turnover	0.886	9	0.723	12	0.855	2	0.749	1
Added value of tertiary industry	0.895	8	0.867	7	0.847	5	0.694	4
Reduction rate of energy consumption per 10,000 yuan of GDP	0.620	19	0.697	13	0.644	17	0.643	16

fastest, from the last in 2012–2016 to the first after 2017. Second, the green input level and green output level are increasing in fluctuation, but there are obvious differences. Among them, the green input level of Beijing is basically at the forefront of the four cities, but the green output level is basically at the bottom. Third, the differences in green development among Beijing, Shanghai, Shenzhen, and Guangzhou have experienced a process of first expanding and then shrinking. Among them, the entry of China's economy into a new normal has played an important role in reducing the differences in green development between cities. Fourth, under the same time series, the factors affecting the green development capabilities of Beijing, Shanghai, Guangzhou, and Shenzhen show certain differences. In addition, as the time series changes, the factors that affect the green development capability of the same city are also changing. Among them, the primary factors affecting the green development capabilities of Beijing and Shanghai have changed from the number of patent authorizations and total energy consumption to the number of research and experimental development personnel. The primary factors affecting the green development capabilities of Guangzhou and Shenzhen have changed from total energy consumption and industrial waste gas emissions to patent authorization

and technology contract turnover. It can be seen from these changes that technological innovation and resource utilization efficiency are becoming more and more important to the level of urban green development and have become the main influencing factors.

6. Conclusion

The sustained and rapid development of the national economy has driven the rapid development of cities. However, in the process of urban development, problems such as resource waste and environmental pollution will inevitably arise. Urban green development is conducive to improving the competitiveness of cities and coordinating the relationship between economic development and environmental society. Therefore, urban green development will become a trend. This paper establishes an index system to evaluate the level of urban green development, uses the TOPSIS-BP neural network model to model the green development levels of Beijing, Shanghai, Shenzhen, and Guangzhou, and use grey correlation to analyze the factors that affect the level of urban green development. It is found that, in the process of urban green development, green technology innovation has played an important role.

There are several main contributions of this article, first of all the theoretical significance: firstly, the current research on urban green and sustainable development mainly focuses on the planning of urban land resources and the protection of the ecological environment from a macroperspective [73–76]. This article analyzes the specific factors that affect the level of urban green development and to some extent makes up for the lack of analysis of specific factors that affect the level of urban green development in existing studies. Secondly, the connotation of regional differences enriched and expanded. As each city has different resource endowments, resource allocation, location conditions, and external environment, this will inevitably cause differences between regions [77–79]. In fact, this article uses the gray relational analysis method to find that, under the same time series, the factors affecting the green development capabilities of Beijing, Shanghai, Guangzhou, and Shenzhen show certain differences. This shows that there are differences in green development capabilities among different regions and further proves that regional differences exist objectively and enriches the connotation of regional differences.

At the same time, this article also has certain practical significance. Firstly, this article uses the topsis-bp neural network model to evaluate and analyze the level of urban green development, which is more accurate and objective. BP neural network has good learning ability and can provide certain theoretical and practical reference for evaluating the green development level of other cities. In addition, the article constructs an evaluation index system for urban green development from the perspective of input and output. This index system takes into account the economic and environmental impacts of technological input and resource utilization, and can scientifically measure the level of green development. First-tier cities such as Beijing, Shanghai, Guangzhou, and Shenzhen can formulate relevant policies based on this indicator system to improve their green development capabilities. Finally, the research samples are Beijing, Shanghai, Shenzhen, and Guangzhou, four developed cities in China. These four cities are in a leading position in mainland China in terms of economic strength and scientific research strength. This is also of important reference value for the formulation of green development policies in other regions.

In response to the above research conclusions, in order to promote the improvement of China's urban green development capabilities, this article makes the following recommendations:

First, increase investment in science and technology and green technology innovation. From the data analysis results of the article, compared with energy consumption and "three wastes" emissions, green technology innovation has increasingly become an important factor in the process of urban green development. In addition, technological innovation can effectively improve production efficiency and reduce energy consumption, which is of great significance to the development of green cities. Cai and Shang [80] believe that advances in science and technology are conducive to enhancing the carrying capacity of urban ecosystems, thereby improving the level of green development of cities. Li and Luo [81] believe that science and technology have become an important factor affecting sustainable development. Fang et al. [82] found that sufficient technical support plays a key role in improving the sustainable development of cities, improving the ecological environment, and developing green cities.

Second, establish and improve laws and policies related to urban green development and establish a sense of green development. Production and living activities have an important influence on the development of the city. Strengthening the supervision of production activities can reduce to a certain extent the behaviors of enterprises and production units that undermine the green development of cities. The improvement of the public's awareness of green development can to a certain extent enhance residents' willingness to live a green lifestyle and can also increase the public's participation in government green management and strengthen the supervision of economic activities. Hawkins and Wang [83] found on the basis of a national survey of the United States that government initiatives and increased public willingness to participate are essential to the urban ecological environment and sustainable development.

Finally, improve resource utilization and develop a green economy. To realize the green development of the city, it is necessary to coordinate the development of the city with the resources and the environment. That is to say, how to scientifically use limited resources, improve the utilization rate of resources, and develop a green economy is of vital importance to the green development of the city. Wang et al. [84] studied the green development of 9 cities in the Pearl River Delta and proposed that urban green development should improve the energy utilization structure and build a low-carbon recycling industry system.

This paper mainly uses the TOPSIS-BP neural network model to evaluate the green development level of four cities

in Beijing, Shanghai, Shenzhen, and Guangzhou and analyzes the factors that affect the city's green development level through grey correlation analysis methods. This article has certain rationality and objectivity in the research content and research methods. But, this article still has some shortcomings: first of all, the development of some cities has regional characteristics, and the green development of cities is also unique. The sample areas selected in this article may bring greater reference significance to economically developed areas and lack a certain degree of universal adaptation. Secondly, due to the high authenticity required for data collection, the indicators selected in this article are all quantitative indicators to ensure the authenticity and accuracy of the data. However, adding some qualitative indicators will help measure the level of urban green development more comprehensively. Future research will also focus on the impact of qualitative indicators on the level of urban green development, expand the scope of research, and increase the universality of research.

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 71804056, 71672111, 71932004, and 71803197), Humanities and Social Sciences Research Project of the Ministry of Education of China (18YJC630250 and 18YJC630094), China Postdoctoral Science Foundation (2018M642033), the Fundamental Research Funds for the Central Universities (31511910801), and Soft Science Project of Technological Innovation in Hubei Province (2019ADC029).

References

[1] S. Campbell, "Green cities, growing cities, just cities?: urban planning and the contradictions of sustainable development," *Journal of the American Planning Association*, vol. 62, no. 3, pp. 296–312, 1996.

[2] M. Yokohari, K. Takeuchi, T. Watanabe, and S. Yokota, "Beyond greenbelts and zoning: a new planning concept for the environment of Asian mega-cities," in *Urban Ecology*, pp. 783–796, Springer, Berlin, Germany, 2008.

[3] A. Van Herzele and T. Wiedemann, "A monitoring tool for the provision of accessible and attractive urban green spaces," *Landscape and Urban Planning*, vol. 63, no. 2, pp. 109–126, 2003.

[4] V. D. B. M. Annerstedt, P. Mudu, V. Uscila et al., "Development of an urban green space indicator and the public health rationale," *Scandinavian Journal of Public Health*, vol. 44, pp. 159–167, 2016.

[5] K. Horwood, "Green infrastructure: reconciling urban green space and regional economic development: lessons learnt from experience in England's north-west region," *Local Environment*, vol. 16, no. 10, pp. 963–975, 2011.

[6] E. Andersson, S. Barthel, S. Borgström et al., "Reconnecting cities to the biosphere: stewardship of green infrastructure and urban ecosystem services," *Ambio*, vol. 43, no. 4, pp. 445–453, 2014.

[7] U. G. Sandström, P. Angelstam, and A. Khakee, "Urban comprehensive planning—identifying barriers for the maintenance of functional habitat networks," *Landscape and Urban Planning*, vol. 75, no. 1-2, pp. 43–57, 2006.

[8] K. H. Thorén, ""The green poster" a method to evaluate the sustainability of the urban green structure," *Environmental Impact Assessment Review*, vol. 20, no. 3, pp. 359–371, 2000.

[9] G. H. Brundtland, M. Khalid, S. Agnelli, S. Al-Athel, and B. Chidzero, *Our Common Future*, Vol. 8, Oxford University Press, Oxford, UK, 1987.

[10] J. Zhang, Y. Chang, L. Zhang, and D. Li, "Do technological innovations promote urban green development?—a spatial econometric analysis of 105 cities in China," *Journal of Cleaner Production*, vol. 182, pp. 395–403, 2018.

[11] P. Jin, C. Peng, and M. Song, "Macroeconomic uncertainty, high-level innovation, and urban green development performance in China," *China Economic Review*, vol. 55, pp. 1–18, 2019.

[12] M. Artmann, L. Inostroza, and P. Fan, *Urban Sprawl, Compact Urban Development and Green Cities. How Much Do We Know, How Much Do We Agree?*, Elsevier, Amsterdam, Netherlands, 2019.

[13] S. L. Zhao, W. Song, D. Y. Zhu, X. B. Peng, and W. Cai, "Evaluating China's regional collaboration innovation capability from the innovation actors perspective—an AHP and cluster analytical approach," *Technology in Society*, vol. 35, no. 3, pp. 182–190, 2013.

[14] K. Li and B. Lin, "Impact of energy conservation policies on the green productivity in China's manufacturing sector: evidence from a three-stage DEA model," *Applied Energy*, vol. 168, pp. 351–363, 2016.

[15] Y. Duan, H. Mu, N. Li, L. Li, and Z. Xue, "Research on comprehensive evaluation of low carbon economy development level based on AHP-entropy method: a case study of Dalian," *Energy Procedia*, vol. 104, pp. 468–474, 2016.

[16] H. F. Tong, Y. Yang, J. Y. Wang, and Y. Feng, "Modeling China's green economy 2050: scenario analysis based on the system dynamics model," *China Soft Science*, vol. 6, pp. 20–34, 2015.

[17] M. Jacobs, *The Green Economy: Environment, Sustainable Development and the Politics of the Future*, UBC Press, Vancouver, Canada, 1993.

[18] S. M. A. Haq, "Urban green spaces and an integrative approach to sustainable environment," *Journal of Environmental Protection*, vol. 2, no. 5, p. 601, 2011.

[19] M. Budruk, H. Thomas, and T. Tyrrell, "Urban green spaces: a study of place attachment and environmental attitudes in India," *Society & Natural Resources*, vol. 22, no. 9, pp. 824–839, 2009.

[20] J. R. Wolch, J. Byrne, and J. P. Newell, "Urban green space, public health, and environmental justice: the challenge of making cities "just green enough," *Landscape and Urban Planning*, vol. 125, pp. 234–244, 2014.

[21] X. Dou, S. Li, and J. Wang, "Ecological strategy of city sustainable development," *APCBEE Procedia*, vol. 5, pp. 429–434, 2013.

[22] J. Wu, W. Xiang, and J. Zhao, "Urban ecology in China: historical developments and future directions," *Landscape and Urban Planning*, vol. 125, pp. 222–233, 2014.

[23] J. Niemelä, S.-R. Saarela, T. Söderman et al., "Using the ecosystem services approach for better planning and conservation of urban green spaces: a Finland case study,"

Biodiversity and Conservation, vol. 19, no. 11, pp. 3225–3243, 2010.

[24] M. F. Aronson, C. A. Lepczyk, K. L. Evans et al., "Biodiversity in the city: key challenges for urban green space management," *Frontiers in Ecology and the Environment*, vol. 15, no. 4, pp. 189–196, 2017.

[25] P. Dicken, *Global Shift: Reshaping the Global Economic Map in the 21st Century*, Sage, Thousand Oaks, CA, USA, 2003.

[26] W. E. Rees, "Ecological footprints and appropriated carrying capacity: what urban economics leaves out," *Environment and Urbanization*, vol. 4, no. 2, pp. 121–130, 1992.

[27] N. Dempsey, G. Bramley, S. Power, and C. Brown, "The social dimension of sustainable development: defining urban social sustainability," *Sustainable Development*, vol. 19, no. 5, pp. 289–300, 2011.

[28] N. Heynen, "Urban political ecology," *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, pp. 1–9, John Wiley & Sons, Hoboken, NJ, USA, 2016.

[29] S. Yin and B. Li, "Academic research institutes-construction enterprises linkages for the development of urban green building: selecting management of green building technologies innovation partner," *Sustainable Cities and Society*, vol. 48, Article ID 101555, 2019.

[30] S. Pauleit, B. Ambrose-Oji, E. Andersson et al., "Advancing urban green infrastructure in Europe: outcomes and reflections from the GREEN SURGE project," *Urban Forestry & Urban Greening*, vol. 40, pp. 4–16, 2019.

[31] Z. S. Venter, C. M. Shackleton, F. Van Staden, O. Selomane, and V. A. Masterson, "Green Apartheid: urban green infrastructure remains unequally distributed across income and race geographies in South Africa," *Landscape and Urban Planning*, vol. 203, Article ID 103889, 2020.

[32] S. Barles, "Society, energy and materials: the contribution of urban metabolism studies to sustainable urban development issues," *Journal of Environmental Planning and Management*, vol. 53, no. 4, pp. 439–455, 2010.

[33] W. Y. Chen and D. T. Wang, "Economic development and natural amenity: an econometric analysis of urban green spaces in China," *Urban Forestry & Urban Greening*, vol. 12, no. 4, pp. 435–442, 2013.

[34] F. Li, X. Liu, D. Hu et al., "Measurement indicators and an evaluation approach for assessing urban sustainable development: a case study for China's Jining city," *Landscape and Urban Planning*, vol. 90, no. 3-4, pp. 134–142, 2009.

[35] D. Guan, W. Gao, W. Su, H. Li, and K. Hokao, "Modeling and dynamic assessment of urban economy-resource-environment system with a coupled system dynamics—geographic information system model," *Ecological Indicators*, vol. 11, no. 5, pp. 1333–1344, 2011.

[36] N. X. Thinh, G. Arlt, B. Heber, J. Hennersdorf, and I. Lehmann, "Evaluation of urban land-use structures with a view to sustainable development," *Environmental Impact Assessment Review*, vol. 22, no. 5, pp. 475–492, 2002.

[37] N. Wang, J. C. K. Lee, J. Zhang, H. Chen, and H. Li, "Evaluation of urban circular economy development: an empirical research of 40 cities in China," *Journal of Cleaner Production*, vol. 180, pp. 876–887, 2018.

[38] L. Ding, Z. Shao, H. Zhang, C. Xu, and D. Wu, "A comprehensive evaluation of urban sustainable development in China based on the TOPSIS-entropy method," *Sustainability*, vol. 8, no. 8, p. 746, 2016.

[39] B. Meng, L. Zhou, L. Qu, and M. Z. Abedin, "Measurement of urban green economy development-an empirical analysis from 31 provinces in China," *Ekoloji*, vol. 28, pp. 2069–2082, 2019.

[40] L. Lin and L. Ying, "A dynamic evaluation on and comparison between the green efficiency of the central delta urban agglomeration and the Yangtze river delta urban agglomeration," *Journal of Jiangxi University of Finance and Economics*, vol. 3, pp. 3–12, 2015.

[41] S. Zheng, R. Wang, E. L. Glaeser, and M. E. Kahn, "The greenness of China: household carbon dioxide emissions and urban development," *Journal of Economic Geography*, vol. 11, no. 5, pp. 761–792, 2011.

[42] M. E. Kahn, *Green Cities: Urban Growth and the Environment*, Brookings Institution Press, Washington, DC, USA, 2007.

[43] Y. X. He, Z. Jiao, and J. Yang, "Comprehensive evaluation of global clean energy development index based on the improved entropy method," *Ecological Indicators*, vol. 88, pp. 305–321, 2018.

[44] R. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization," *Methodology and Computing in Applied Probability*, vol. 1, no. 2, pp. 127–190, 1999.

[45] J. Zhao, G. Ji, Y. Tian, Y. Chen, and Z. Wang, "Environmental vulnerability assessment for mainland China based on entropy method," *Ecological Indicators*, vol. 91, pp. 410–422, 2018.

[46] L.-y. Sun, C.-l. Miao, and L. Yang, "Ecological-economic efficiency evaluation of green technology innovation in strategic emerging industries based on entropy weighted TOPSIS method," *Ecological Indicators*, vol. 73, pp. 554–558, 2017.

[47] K. Rashidi and K. Cullinane, "A comparison of fuzzy DEA and fuzzy TOPSIS in sustainable supplier selection: implications for sourcing strategy," *Expert Systems with Applications*, vol. 121, pp. 266–281, 2019.

[48] B. M. Dos Santos, L. P. Godoy, and L. M. S. Campos, "Performance evaluation of green suppliers using entropy-TOPSIS-F," *Journal of Cleaner Production*, vol. 207, pp. 498–509, 2019.

[49] Z. Wang, H. Hao, F. Gao, Q. Zhang, J. Zhang, and Y. Zhou, "Multi-attribute decision making on reverse logistics based on DEA-TOPSIS: a study of the Shanghai end-of-life vehicles industry," *Journal of Cleaner Production*, vol. 214, pp. 730–737, 2019.

[50] S. E. Fahlman, *An Empirical Study of Learning Speed in Back-Propagation Networks*, Carnegie Mellon University, Computer Science Department, Pittsburgh, PA, USA, 1988.

[51] L. Peng and L. Lai, "A service innovation evaluation framework for tourism e-commerce in China based on BP neural network," *Electronic Markets*, vol. 24, no. 1, pp. 37–46, 2014.

[52] S. Li and Q. Wang, "India's dependence on foreign oil will exceed 90% around 2025—the forecasting results based on two hybridized NMGM-ARIMA and NMGM-BP models," *Journal of Cleaner Production*, vol. 232, pp. 137–153, 2019.

[53] R. Zhang, Y. Wang, K. Wang et al., "An evaluating model for smart growth plan based on BP neural network and set pair analysis," *Journal of Cleaner Production*, vol. 226, pp. 928–939, 2019.

[54] V. Bobinaite, A. Juozapaviciene, M. Staniewski, and P. Szczepankowski, "Comparative analysis of features of Polish and Lithuanian day-ahead electricity market prices," *Energy Policy*, vol. 63, pp. 181–196, 2013.

[55] Y. Huang, L. Shen, and H. Liu, "Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China," *Journal of Cleaner Production*, vol. 209, pp. 415–423, 2019.

[56] Q. Sun and Y. Tang, "The grey relational degree measurement of city's S&T input and sustainable economic development based on the data from Hunan province," *Procedia Engineering*, vol. 21, pp. 457–463, 2011.

[57] R. K. Singh, H. R. Murty, S. K. Gupta, and A. K. Dikshit, "An overview of sustainability assessment methodologies," *Ecological Indicators*, vol. 9, no. 2, pp. 189–212, 2009.

[58] Y. Xing, R. M. W. Horner, M. A. El-Haram, and J. Bebbington, "A framework model for assessing sustainability impacts of urban development," in *Accounting Forum*, pp. 209–224, Elsevier, Amsterdam, Netherlands, 2009.

[59] K. Oh, Y. Jeong, D. Lee, W. Lee, and J. Choi, "Determining development density using the urban carrying capacity assessment system," *Landscape and Urban Planning*, vol. 73, no. 1, pp. 1–15, 2005.

[60] Y. Feng, X. Dong, X. Zhao, and A. Zhu, "Evaluation of urban green development transformation process for Chinese cities during 2005–2016," *Journal of Cleaner Production*, vol. 266, Article ID 121707, 2020.

[61] W. Yuan, J. Li, L. Meng, X. Qin, and X. Qi, "Measuring the area green efficiency and the influencing factors in urban agglomeration," *Journal of Cleaner Production*, vol. 241, Article ID 118092, 2019.

[62] C. Feng and M. Wang, "The heterogeneity of China's pathways to economic growth, energy conservation and climate mitigation," *Journal of Cleaner Production*, vol. 228, pp. 594–605, 2019.

[63] F. Qiu, Y. Chen, J. Tan, J. Liu, Z. Zheng, and X. Zhang, "Spatial-temporal heterogeneity of green development efficiency and its influencing factors in growing metropolitan area: a case study for the Xuzhou metropolitan area," *Chinese Geographical Science*, vol. 30, pp. 352–365, 2020.

[64] B. Zhu, M. Zhang, Y. Zhou et al., "Exploring the effect of industrial structure adjustment on interprovincial green development efficiency in China: a novel integrated approach," *Energy Policy*, vol. 134, Article ID 110946, 2019.

[65] Y. Guo, L. Tong, and L. Mei, "The effect of industrial agglomeration on green development efficiency in northeast China since the revitalization," *Journal of Cleaner Production*, vol. 258, Article ID 120584, 2020.

[66] L. Ma, H. Long, K. Chen, S. Tu, Y. Zhang, and L. Liao, "Green growth efficiency of Chinese cities and its spatio-temporal pattern," *Resources, Conservation and Recycling*, vol. 146, pp. 441–451, 2019.

[67] C. Feng, M. Wang, G.-C. Liu, and J.-B. Huang, "Green development performance and its influencing factors: a global perspective," *Journal of Cleaner Production*, vol. 144, pp. 323–333, 2017.

[68] W. Pan, W. Pan, C. Hu et al., "Assessing the green economy in China: an improved framework," *Journal of Cleaner Production*, vol. 209, pp. 680–691, 2019.

[69] L. Chen, X. Zhang, F. He, and R. Yuan, "Regional green development level and its spatial relationship under the constraints of haze in China," *Journal of Cleaner Production*, vol. 210, pp. 376–387, 2019.

[70] S. Shao, R. Luan, Z. Yang, and C. Li, "Does directed technological change get greener: empirical evidence from Shanghai's industrial green development transformation," *Ecological Indicators*, vol. 69, pp. 758–770, 2016.

[71] W. Sun and Q. Gao, "Exploration of energy saving potential in China power industry based on Adaboost back propagation neural network," *Journal of Cleaner Production*, vol. 217, pp. 257–266, 2019.

[72] X. Li, P. S. W. Fong, S. Dai, and Y. Li, "Towards sustainable smart cities: an empirical comparative assessment and development pattern optimization in China," *Journal of Cleaner Production*, vol. 215, pp. 730–743, 2019.

[73] Z. Wu, R. Chen, M. E. Meadows, D. Sengupta, and D. Xu, "Changing urban green spaces in Shanghai: trends, drivers and policy implications," *Land Use Policy*, vol. 87, Article ID 104080, 2019.

[74] Q. Ji, C. Li, and P. Jones, "New green theories of urban development in China," *Sustainable Cities and Society*, vol. 30, pp. 248–253, 2017.

[75] M. L. Derkzen, A. J. van Teeffelen, H. Nagendra, and P. H. Verburg, "Shifting roles of urban green space in the context of urban development and global change," *Current Opinion in Environmental Sustainability*, vol. 29, pp. 32–39, 2017.

[76] M. Contesse, B. J. M. van Vliet, and J. Lenhart, "Is urban agriculture urban green space? A comparison of policy arrangements for urban green space and urban agriculture in Santiago de Chile," *Land Use Policy*, vol. 71, pp. 566–577, 2018.

[77] K. Lv, X. Feng, S. Kelly, L. Zhu, and M. Deng, "A study on embodied carbon transfer at the provincial level of China from a social network perspective," *Journal of Cleaner Production*, vol. 225, pp. 1089–1104, 2019.

[78] K. Feng, K. Hubacek, L. Sun, and Z. Liu, "Consumption-based CO_2 accounting of China's megacities: the case of Beijing, Tianjin, Shanghai and Chongqing," *Ecological Indicators*, vol. 47, pp. 26–31, 2014.

[79] W. Pan, W. Pan, Y. Shi et al., "China's inter-regional carbon emissions: an input-output analysis under considering national economic strategy," *Journal of Cleaner Production*, vol. 197, pp. 794–803, 2018.

[80] C. Cai and J. Shang, "Comprehensive evaluation on urban sustainable development of Harbin city in northeast China," *Chinese Geographical Science*, vol. 19, no. 2, pp. 144–150, 2009.

[81] S. Li and X. Luo, "Emergy assessment and sustainability of ecological-economic system using gis in China," *Acta Ecologica Sinica*, vol. 35, no. 5, pp. 160–167, 2015.

[82] C. Fang, X. Cui, G. Li et al., "Modeling regional sustainable development scenarios using the urbanization and eco-environment Coupler: case study of Beijing-Tianjin-Hebei urban agglomeration, China," *Science of The Total Environment*, vol. 689, pp. 820–830, 2019.

[83] C. V. Hawkins and X. Wang, "Sustainable development governance: citizen participation and support networks in local sustainability initiatives," *Public Works Management & Policy*, vol. 17, no. 1, pp. 7–29, 2012.

[84] M.-X. Wang, H.-H. Zhao, J.-X. Cui et al., "Evaluating green development level of nine cities within the Pearl river delta, China," *Journal of Cleaner Production*, vol. 174, pp. 315–323, 2018.

Hindawi
Complexity
Volume 2020, Article ID 6640000, 10 pages
https://doi.org/10.1155/2020/6640000

Research Article

Parameter Optimization on the Three-Parameter Whitenization Grey Model and Its Application in Simulation and Prediction of Gross Enrollment Rate of Higher Education in China

Jihong Sun,[1] **Hui Li**,[2] **Bo Zeng**,[2] **Xiaoyun Zhao**,[3] **and Chuanhui Wang**[3]

[1]*Center of High Education, National Institution of Education Science, Beijing 100088, China*
[2]*College of Mathematics and Statistics, Chongqing Technology and Bussiness University, Chongqing 400067, China*
[3]*College of Economics, Qufu Normal University, Rizhao 276800, China*

Correspondence should be addressed to Xiaoyun Zhao; niethy@163.com

Received 5 November 2020; Revised 20 November 2020; Accepted 2 December 2020; Published 19 December 2020

Academic Editor: Baogui Xin

Copyright © 2020 Jihong Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The gray prediction model, based on the GM(1,1) method, is an important branch of gray theory with the most active research and the most fruitful results, and it is the most widely used because of its small sample size, simple modeling process, and easy to use. Such advantages have been successfully applied in many fields such as transportation, agriculture, energy, medicine, and environment and have been gradually developed into a mainstream predictive modeling method. This study combines the Three-parameter Whitenization Grey Model (TWGM(1,1)), which fits the inhomogeneous exponential law sequence, and the Particle Swarm Algorithm (PSA) to optimize the order and background value coefficients under the condition of the minimum sum of squares of simulation errors, and hence, to solve the problem that the cumulative order is fixed to "1" and the background value coefficient value is fixed to "0.5." As a result, a parameter-optimized gray system model with flexibility, adaptability, and dynamic adjustment is designed to simulate and predict China's higher education gross enrollment rate. The application shows that the model has better overall simulation and prediction performance than others. On the one hand, the parametric optimization model significantly improves its own performance, and on the other hand, its intelligent and adjustable adaptivity improves the accuracy and further extends its application.

1. Introduction

Since Professor Ju-Long [1] established the gray system theory in 1982, the gray prediction model based on the GM(1,1) method is the most active, fruitful, and widely used branch of gray theory. With feature of the small sample size, simple modeling process, and low cost of learning, it has been successfully applied to many fields, such as transportation, agriculture, energy, medicine, and environment [2–6]. The model gradually developed into a mainstream predictive modeling method. The enrollment rate of higher education is influenced by many factors such as social, economic, cultural, and geographical factors, which are characterized by partly known and partly unknown information; thus, the GM(1,1) model, as a model to study the uncertainty system problem of "partly known and partly unknown," has been used to predict the enrollment rate of higher education. For example, Liu et al. [7] used the GM(1, 1) prediction model to make a mathematical analysis of the gross enrollment rate of higher education. It scientifically predicted and eased the employment situation of college students. It provided reference for promoting social stability and economic development. Dong [8] used the GM(1, 1) model to predict the fluctuation of China's medium- and long-term gross enrollment ratio in higher education, which provides a reference for government assessment and citizens' choice of higher education.

The GM(1, 1), as the first gray prediction model proposed to be applicable to univariate modeling, has shown favorable simulation and prediction performance for the

sequence of unity exponential growth characteristics. However, according to the Statistical Yearbook, the gross enrollment rate of higher education in China was 51.6% in 2019, which exceeded the 50% mark for the first time, indicating that China's higher education has moved from the mass stage to the generalized stage. The shift in development scale has undoubtedly increased the complexity and uncertainty of the higher education GER system, which makes the system behavior series more similar to nonuniform exponential growth characteristics, and the original system behavior series with similar unity exponential growth characteristics can no longer satisfy the existing system. How to develop a high-performance gray prediction model that is more flexible, adaptable, and suitable for modeling and predicting the gross enrollment ratio in higher education dynamically is the focus of this research.

In recent years, researchers have carried out a lot of research on the modeling ability of approximate nonhomogeneous exponential sequences. These studies can be divided into three categories according to different modeling ideas. The first is to optimize the structure of the GM(1,1) model to ensure that the final reduction formula of the model presents an approximate homogeneous exponential form. The NHGM(1, 1, k) [9] model extends the basic form $x^{(0)}(k) + az^{(1)}(k) = b$ of the GM(1,1) model to $x^{(0)}(k) + az^{(1)}(k) = 0.5(2k-1)b + c$. The SIGM model [10] expands the basic form $x^{(0)}(k) + az^{(1)}(k) = b$ of the GM(1,1) model to $x^{(0)}(k) + az^{(1)}(k) = kb + c$. The KRNGM model [11] introduces the nonlinear function $f(t)$ into the whitening differential equation of the GM(1,1) model. The second is to modify the adaptability of modeling objects. The conversion from a nonhomogeneous exponential sequence to a homogeneous exponential sequence can be realized by the difference between adjacent elements in the original sequence. The third is the direct modeling method. Wang [12] proposed a direct modeling method of the GM(1,1) model generated from the original data without accumulation, and it had been gradually optimized [13, 14]. Ye and Li [15] established a whitening weight gray prediction model to measure the influence of the probability of interval gray number in the prediction results. Zeng and Liu [16] established a direct modeling method of DGM(1, 1) based on the original sequence by omitting the accumulation and subtraction processes. In addition, the models mentioned above have been applied in variate fields. Xiao et al. [17–20] studied parameter optimization of Grey Riccati model and its application in the prediction on the energy consumption and carbon. A new structure of the Gray Verhulst model is proposed by Xiao et al. [21–24], which improves the ability of the gray model to model saturated S-sequences, and it is applied in China's tight gas production forecasting.

The above optimization models around approximate nonaligned exponential sequences have better properties, modeling capabilities, and a wider range of applications than the GM(1,1) model. The parameters (such as order, background value coefficients, and initial values.) of the gray prediction model are crucial to the performance of the model. However, these performance parameters are often simplified to a specific value in the above optimization model. Therefore, optimization of performance parameters is a key means to improve the stability, applicability, and flexibility of the model. A large number of studies have been launched for the optimization of performance parameters from many directions, such as initial value [25–27], order [28, 29], and background value [30, 31]. The optimal value of each performance parameter should meet the condition of the minimum sum of squared simulation errors of the model, and the optimization process requires a lot of calculations. The Particle Swarm Algorithm (PSA) provides an optional solution for the optimization of those parameters, which find the global optimum by following the currently searched optimum. A distinctive feature compared to other modern optimization methods is that the PSA requires few parameters to be adjusted, is simple and easy to implement, and converges quickly, which has become a research hotspot in the field of modern optimization methods [32–35]. Zeng and Liu [36] proposed the SAIGM_FO model and used the PSA to optimize the order of the model. Wang and Li [37] used the PSA to optimize the structural parameters (a, b) of the model. It has been proved that the PSA in the gray prediction model improves the modeling accuracy and model flexibility.

In this study, it is firstly proposed that the PSA is applied in the Three-parameter Whitenization Grey Model (TWGM(1,1)) [38] to optimize the order and background value coefficients under the condition of minimum simulation error squared. The solution is not only capable to solve the problem that the cumulative order is fixed to "1" and the background value coefficient value is fixed to "0.5" but also displays a feature of flexible, adaptable, and dynamically adjustable in the application in modeling China's higher education gross enrollment ratio (GER). The empirical analysis shows that the model has better overall simulation and prediction performance than others. On the one hand, the parametric optimization model significantly improves its own performance, and on the other hand, its intelligent and adjustable adaptivity improves the accuracy and further extends its application. The rest of the study is organized as below. Section 2 describes the original Three-parameter Whitenization Grey Model. Then, the parameter optimization process of the Three-parameter Whitenization Grey Model is followed in Section 3, and Section 4 gives an application of the optimization model on China's higher education gross enrollment rate. The conclusion of the study is revealed in the last section.

2. Original TWGM(1,1) Model

According to the derivation process of the time-response function of the classic GM(1,1) model, the whitenization equation of the Three-parameter Whitenization Grey Model is established, and then the time-response function is derived by solving the differential equation. This model is called the Three-parameter Whitenization Grey Model, or TWGM (1,1) model in short.

Definition 1. Suppose sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$ is the original data sequence, where $x^{(0)}(k) \geq 0$,

$k = 1, 2, \ldots n$, and $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n))$ is the 1-AGO sequence of $X^{(0)}$, where

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1, 2, \ldots n. \tag{1}$$

$Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n))$ is the immediate mean-generating sequence of $X^{(1)}$, where

$$z^{(1)}(k) = 0.5 \times \left[x^{(1)}(k) + x^{(1)}(k-1)\right], \quad k = 2, 3 \ldots, n. \tag{2}$$

Definition 2. Assuming that the sequences $X^{(0)}$ and $X^{(1)}$ are as shown in Definition 1, then

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t} + ax^{(1)} = bt + c. \tag{3}$$

Equation (3) is the whitening differential equation of the TWGM(1,1) model. The equation

$$x^{(0)}(k) + az^{(1)}(k) = 0.5(2k-1)b + c, \tag{4}$$

is the basic form of the TWGM(1,1) model.

According to the modeling ideas of the GM(1,1), the model parameters a, b, and c are estimated through the basic form of the TWGM(1,1) model. The time-response function of the TWGM(1,1) model is obtained by solving the differential equation.

Theorem 1. *Suppose the sequence $X^{(0)}$, $X^{(1)}$, and $Z^{(1)}$ are as shown in Definition 1, $\widehat{p} = (a, b, c)^T$ is the parameter list, and*

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix},$$

$$B = \begin{bmatrix} -z^{(1)}(2) & \frac{3}{2} & 1 \\ -z^{(1)}(3) & \frac{5}{2} & 1 \\ \vdots & \vdots & \vdots \\ -z^{(1)}(n) & \frac{(2n-1)}{2} & 1 \end{bmatrix}. \tag{5}$$

Then, the least-square estimation parameter list of the TWGM(1,1) model is satisfied.

Theorem 2. *The time-response function of the TWGM(1,1) model is*

$$x^{(1)}(t) = \left(x^{(1)}(1) - \frac{b}{a} + \frac{b}{a^2} - \frac{c}{a}\right)e^{-a(t-1)} + \frac{b}{a}t - \frac{b}{a^2} + \frac{c}{a}. \tag{6}$$

The inverse-accumulating reduction formula of the TWGM(1,1) model is

$$\widehat{x}^{(0)}(t) = \left(1 - e^{a}\right)\left(x^{(0)}(1) - \frac{b}{a} + \frac{b}{a^2} - \frac{c}{a}\right)e^{-a(t-1)} + \frac{b}{a}. \tag{7}$$

3. Parameter Optimization of the Three-Parameter Whitenization Gray Model

3.1. Background Value Optimization. In the TWGM(1,1) model modeling process, the sequence generated next to the mean is used as the background value, which is a common smoothing method to weaken the influence of extreme values or outliers in the 1-AGO sequence on the magnitude of the gray effect. Among them, the background value coefficient is the weight of adjacent elements in the process of constructing the adjacent mean sequence. The difference of its value will affect the calculation result of the adjacent mean value series and then affect the simulation and prediction effects of the model. In the above modeling process, the background value coefficient is set to 0.5. This processing method lacks flexibility and cannot guarantee that the model achieves the best simulation effect. So, it is necessary to optimize the background value of the original model. The optimization of the background value is mainly based on the optimization of the background value coefficient. Optimize the definition of the sequence immediately adjacent to the mean.

Definition 3. Suppose that $X^{(1)}$ is as shown in Definition 1, $Z_w^{(1)} = (z_w^{(1)}(2), z_w^{(1)}(3), \ldots, z_w^{(1)}(n))$ is the optimized sequence generated by the immediate mean of $X^{(1)}$, and w is the background value coefficient,$w \in (0, 1)$, where

$$z_w^{(1)}(k) = wx^{(1)}(k) + (1-w)x^{(1)}(k-1), \quad k = 2, 3, \ldots, n. \tag{8}$$

Then,

$$x^{(0)}(k) + a\left(wx^{(1)}(k) + (1-w)x^{(1)}(k-1)\right) = kb + c, \tag{9}$$

is the TWGM(1,1) model with the background value coefficient $w\,(0 < w < 1)$.

Among them, the optimal value of the background value coefficient w should satisfy the minimum sum of squared simulation errors of the model, namely,

$$\min f(w) = \frac{1}{n-1}\sum_{k=2}^{n}\left[x^{(0)}(k) - \widehat{x}^{(0)}(k)\right]^2, \quad 0 < w < 1. \tag{10}$$

In equation (10), $x^{(0)}(k)$ is the original modeling data and $\widehat{x}^{(0)}(k)$ is the simulated data of the model.

3.2. Order Optimization. In the original modeling process, the cumulative order is fixed at "1," which leads to poor flexibility and adaptability of the model. Therefore, Wu et al. [39, 40] introduced fractional order into gray modeling based on the "in between" idea. This realizes the expansion of the cumulative order of the gray prediction model from integer to fraction. Meng et al. [38, 41] used the Gamma function to give the functional expression of the fractional accumulation operator, which provided the basis for constructing the fractional gray prediction model. This section optimizes the order of the TWGM(1,1) model in order to improve model performance and model adaptability.

Definition 4. Suppose $X^{(0)}$ is as shown in Definition 1 and $X^{(r)}(r \in R^{+})$ is a new sequence; then, sequence $X^{(r)} = (x^{(r)}(1), x^{(r)}(2), \ldots, x^{(r)}(n))$ is called the r-order cumulative generating sequence of $X^{(0)}$, where

$$x^{(r)}(k) = \sum_{i=1}^{k} \frac{\Gamma(r+k-i)}{\Gamma(k-i+1)\Gamma(r)} x^{(0)}(i), \quad k = 1, 2, \ldots n. \tag{11}$$

Then, the sequence $Z^{(r)} = (z^{(r)}(2), z^{(r)}(3), \ldots, z^{(r)}(n))$ is called the adjacent mean-value generating sequence of $X^{(r)}$, where

$$z^{(r)}(k) = \frac{x^{(r)}(k) + x^{(r)}(k-1)}{2}, \quad k = 2, 3, \ldots, n. \tag{12}$$

Then,

$$x^{(r-1)}(k) + az^{(r)}(k) = 0.5(2k-1)b + c, \tag{13}$$

is the TWGM(1,1) model of r order.

Among them, the optimal value of the r order should satisfy the minimum sum of squared simulation errors of the model, namely,

$$\min f(r) = \frac{1}{n-1} \sum_{k=2}^{n} \left[x^{(0)}(k) - \widehat{x}^{(0)}(k) \right]^2, \quad r \in R^{+}. \tag{14}$$

3.3. Simultaneous Optimization of Background Value and Order. Based on the above optimization analysis, this section derives and constructs a new TWGM(1,1) model under the condition of simultaneous optimization of background value and order. The new model is called the TWGM$(1, 1|w, r)$ model.

3.3.1. Definition

Definition 5. Suppose the sequences $X^{(0)}$ and $X^{(r)}$ are as shown in Definitions 1 and 4; then,

$$\frac{\mathrm{d}x^{(r)}}{\mathrm{d}t} + ax^{(r)} = bt + c, \tag{15}$$

is called the whitening differential equation of the TWGM $(1, 1|w, r)$ model.

The sequence $Z_w^{(r)} = (z_w^{(r)}(2), z_w^{(r)}(3), \ldots, z_w^{(r)}(n))$ is called the optimized sequence generated by the adjacent mean value of $X^{(r)}$ under the condition of the background value coefficient w, where

$$z_w^{(r)}(k) = wx^{(r)}(k) + (1-w)x^{(r)}(k-1), \quad k = 2, 3, \ldots, n. \tag{16}$$

Integrating both ends of formula (15) in interval $[k-1, k]$, we can obtain

$$\int_{k-1}^{k} dx^{(r)}(t) + a\int_{k-1}^{k} x^{(r)}(t)\mathrm{d}t = \int_{k-1}^{k} bt\mathrm{d}t + \int_{k-1}^{k} c\mathrm{d}t. \tag{17}$$

Because

$$\begin{gathered} \int_{k-1}^{k} dx^{(r)}(t) = x^{(r)}(t) - x^{(r)}(t-1) = x^{(r-1)}(t), \\ \int_{k-1}^{k} bt\mathrm{d}t = 0.5(2k-1)b, \int_{k-1}^{k} c\mathrm{d}t = c, \end{gathered} \tag{18}$$

among them, the size of $\int_{k-1}^{k} x^{(r)}(t)\mathrm{d}t$ can be approximately replaced by the area represented by $z_w^{(r)}(t)$, namely,

$$\int_{k-1}^{k} x^{(r)}(t)\mathrm{d}t \approx wx^{(r)}(t) + (1-w)x^{(r)}(t-1) = z_w^{(r)}(t). \tag{19}$$

So, equation (15) can be converted to

$$x^{(r-1)}(k) + z_w^{(r)}(k) = 0.5(2k-1)b + c. \tag{20}$$

Equation (20) is the basic form of the TWGM$(1, 1|w, r)$ model.

3.3.2. Parameter Estimation

Theorem 3. *Suppose the sequences $X^{(0)}$, $X^{(r)}$, and $Z_w^{(r)}$ are as shown in Definitions 1, 4, and 5, $\widehat{p} = (a, b, c)^T$ is the parameter list, and*

$$Y = \begin{bmatrix} x^{(r-1)}(2) \\ x^{(r-1)}(3) \\ \vdots \\ x^{(r-1)}(n) \end{bmatrix} = \begin{bmatrix} x^{(r)}(2) - x^{(r)}(1) \\ x^{(r)}(3) - x^{(r)}(2) \\ \vdots \\ x^{(r)}(n) - x^{(r)}(n-1) \end{bmatrix},$$

$$B = \begin{bmatrix} -z_w^{(r)}(2) & \frac{3}{2} & 1 \\ -z_w^{(r)}(3) & \frac{5}{2} & 1 \\ \vdots & \vdots & \vdots \\ -z_w^{(r)}(n) & \frac{(2n-1)}{2} & 1 \end{bmatrix} = \begin{bmatrix} -wx^{(r)}(2) - (1-w)x^{(r)}(1) & \frac{3}{2} & 1 \\ -wx^{(r)}(3) - (1-w)x^{(r)}(2) & \frac{5}{2} & 1 \\ \vdots & \vdots & \vdots \\ -wx^{(r)}(n) - (1-w)x^{(r)}(n-1) & \frac{(2n-1)}{2} & 1 \end{bmatrix}. \tag{21}$$

Then, the least-square estimation parameter list of the TWGM(1, 1|w, r) model satisfies $\widehat{p} = (a, b, c)^T = (B^T B)^{-1} B^T Y$.

3.3.3. Time-Response Derivation. According to the whitening differential equation of the TWGM(1, 1|w, r) model to derive its time-response function and according to formula (15), the corresponding homogeneous equation is

$$\frac{\mathrm{d}x^{(r)}(t)}{\mathrm{d}t} + ax^{(r)}(t) = 0 \Rightarrow \frac{\mathrm{d}x^{(r)}(t)}{\mathrm{d}t} = -ax^{(r)}(t). \tag{22}$$

Then,

$$\ln\left|x^{(r)}\right| = -\mathrm{at} + \ln\left|C_1\right|. \tag{23}$$

The general solution of homogeneous equation (22) is

$$x^{(r)}(t) = C_1 e^{-\mathrm{at}}. \tag{24}$$

Using the constant variation method, replace formula (24) C_1 with $u(t)$, and let

$$x^{(r)}(t) = u(t)e^{-\mathrm{at}}. \tag{25}$$

Derivation from both ends of equation (25) with respect to t:

$$\frac{\mathrm{d}x^{(r)}(t)}{\mathrm{d}t} = u'(t)e^{-at} - au(t)e^{-\mathrm{at}}. \tag{26}$$

Substituting equation (26) into formula (15), we obtain

$$u\prime(t)e^{-\mathrm{at}} - au(t)e^{-at} = bt + c - ax^{(r)}. \tag{27}$$

According to equation (27), there is $u\prime(t) = (bt + c)e^{\mathrm{at}}$, so

$$u(t) = \int (bt + c)e^{\mathrm{at}}\mathrm{d}t = b\int te^{\mathrm{at}}\mathrm{d}t + c\int e^{\mathrm{at}}\mathrm{d}t = \frac{b}{a}te^{\mathrm{at}} - \frac{b}{a^2}e^{\mathrm{at}} + \frac{c}{a}e^{\mathrm{at}} + C. \tag{28}$$

Substituting equation (28) into (25), we obtain

$$x^{(r)}(t) = \left(\frac{b}{a}te^{\mathrm{at}} - \frac{b}{a^2}e^{\mathrm{at}} + \frac{c}{a}e^{\mathrm{at}} + C\right)e^{-\mathrm{at}}. \tag{29}$$

Calculating the above equation,

$$x^{(r)}(t) = \frac{b}{a}t - \frac{b}{a^2} + \frac{c}{a} + Ce^{-\mathrm{at}}. \tag{30}$$

Among them, $x^{(r)}(1) = x^{(0)}(1)$ is a known item. At that time $t = 1$, equation (30) can be obtained:

$$x^{(r)}(1) = Ce^{-a} + \frac{b}{a} - \frac{b}{a^2} + \frac{c}{a}. \tag{31}$$

We obtain

$$C = \frac{x^{(r)}(1) - (b/a) + \left(b/a^2\right) - (c/a)}{e^{-a}}. \tag{32}$$

Substituting equation (32) into (30), we obtain

$$x^{(r)}(t) = \left(x^{(r)}(1) - \frac{b}{a} + \frac{b}{a^2} - \frac{c}{a}\right)e^{-a(t-1)} + \frac{b}{a}t - \frac{b}{a^2} + \frac{c}{a}. \tag{33}$$

Among them, $t = 2, 3, \ldots, n$ and equation (33) is the time-response formula. At this point, the time-response derivation is over.

3.3.4. Derivation of Accumulative Reduction

Definition 6. Suppose sequence $X^{(0)}$ is as shown in Definition 1 and if $r \in \mathbf{R}^+$, then sequence $X^{(-r)} = (x^{(-r)}(1), x^{(-r)}(2), \ldots, x^{(-r)}(n))$ is called the r-order accumulative generating sequence of $X^{(0)}$, where

$$x^{(-r)}(k) = \sum_{i=0}^{k-1} (-1)^i \frac{\Gamma(r+1)}{\Gamma(i+1)\Gamma(r-i+1)} x^{(0)}(k-i), \quad k = 1, 2, \ldots, n. \tag{34}$$

Theorem 4. *Suppose sequence $X^{(0)}$ is as defined in Definition 1, $p \in \mathbf{R}^+$, $q \in \mathbf{R}^+$, $X^{(p)}$ is the qth order cumulative generation sequence of $X^{(0)}$, $X^{(q)}$ is the pth order cumulative generation sequence of $X^{(0)}$, $X^{(p+q)}$ is the $p+q$-order cumulative generation sequence of $X^{(0)}$, $(X^{(p)})^{(q)}$ is the q-order cumulative generation sequence of $X^{(p)}$, and $(X^{(q)})^{(p)}$ is the p-order cumulative generating sequence of $X^{(q)}$; then, the multiple cumulative generating operator satisfies the commutative law and exponential rate, namely,*

$$\left(X^{(p)}\right)^{(q)} = \left(X^{(q)}\right)^{(p)} = X^{(p+q)}. \tag{35}$$

Corollary 1. *According to Theorem 1, it can be derived as*

$$X^{(0)} = \left(X^{(r)}\right)^{(-r)} = \left(X^{(-r)}\right)^{(r)}. \tag{36}$$

According to Corollary 1, which is equation (36), the final reduction formula of the TWGM$(1, 1|w, r)$ model can be derived as

$$\hat{x}^{(0)}(k) = \left(\hat{x}^{(r)}(k)\right)^{(-r)}(k) = \sum_{i=0}^{k-1} (-1)^i \frac{\Gamma(r+1)}{\Gamma(i+1)\Gamma(r-i+1)} \hat{x}^{(r)}(k-i). \tag{37}$$

In equation (37), when $k = 2, 3, \ldots, n$, $\hat{x}^{(0)}(k)$ is called the simulated value and when $k = n+1, n+2, \ldots$, $\hat{x}^{(0)}(k)$ is called the predicted value.

3.4. TWGM(1,1) Model Parameter Optimization Based on PSO Algorithm. In the TWGM$(1, 1|w, r)$ model, there are two undetermined parameters (the background value coefficient and order). In order to obtain the best performance of the TWGM$(1, 1|w, r)$ model, the optimal values of these two undetermined parameters are required. The combined solution of the two parameters is easy to cause errors due to mutual influence and mutual interference of the parameters, and at the same time, it weakens the optimization effect of the independent parameter on the model. Therefore, in the process of solving the optimal value of the parameter, a step-by-step method is adopted. Specifically, first, in $0 < w < 1$, assuming that the order does not change, that is, $r = 1$, the optimal background value coefficient is solved. Secondly, on the basis of $w = w*$, in $r \in R^+$, the optimal order is solved. According to Definitions 3 and 4, the optimal background value coefficient w and the optimal order r are found within their respective value ranges and should be obtained under the condition that the sum of squared simulation errors of the model is minimized, that is, to satisfy equations (10) and (14).

Obviously, the optimization process of each parameter to be determined takes a lot of time and takes up limited computer resources. Various group optimization algorithms (such as Particle Swarm Optimization and ant colony algorithm.) provide good solutions to complex distributed optimization problems. The Particle Swarm Optimization (PSO) is a swarm-intelligent global optimization algorithm that simulates the predation behavior of birds. Its basic concept comes from the study of the foraging behavior of birds. The algorithm has the advantages of simple structure, few parameters, and easy programming. At the same time, the Particle Swarm Optimization algorithm based on adaptive mutation of the population fitness variance effectively solves the phenomenon of premature convergence and can significantly improve the global convergence performance. It has been used in function optimization and neural network. It is widely used in training, engineering, and other fields [28–31]. Therefore, this paper uses PSO to optimize the background value coefficient and order step by step. The PSO algorithm solving steps are as follows:

Step 1. Initialize randomly the position and velocity of the particles in the particle swarm.

Step 2. Set the particle in the current position to the position of the best particle in the initial population.

Step 3. Calculate the average relative simulation error of the TWGM$(1,1|w,r)$ model when $w = p\text{Best}$(or $r = p\text{Best}$):

Step 3.1. Calculate the cumulative sequence $X^{(r)}$ of order r.

Step 3.2. Calculate the sequence $Z_w^{(r)}$ immediately adjacent to the background value coefficient w.

Step 3.3. Construct matrix B and Y, and solve model parameter $\widehat{p} = (a,b,c)^T$.

Step 3.4. Calculate the simulation value $\widehat{X}^{(0)}$.

Step 3.5. Calculate the average relative simulation error $f(p\text{Best})$ of $\widehat{X}^{(0)}$.

Step 3.6. Judge whether $|f(p\text{Best}) - f(g\text{Best})|$ is less than the given convergence value δ; if it is satisfied, go to *Step 9*. Otherwise, execute *Step 4*.

Step 4. Perform the following operations for all particles in the particle swarm:

Step 4.1. Update particle position and velocity:

$$\begin{aligned} V &= \omega \times V + c_1 \times \text{rand} \times (p\text{Best} - \text{Present}) \\ &\quad + c_2 \times \text{rand} \times (g\text{Best} - \text{Present}), \\ \text{Present} &= \text{Present} + V, \\ \omega &= \omega_{\max} - \text{run} \times \frac{(\omega_{\max} - \omega_{\min})}{\text{runMax}}. \end{aligned} \tag{38}$$

Step 4.2. If the particle fitness is better than the fitness, set it to the new position.

Step 4.3. If the particle fitness is better than the fitness, set it to the new position.

Step 5. Calculate the variance of the group fitness σ^2, and calculate $f(p\text{Best})$:

$$\begin{aligned} \sigma^2 &= \sum_{i=1}^{n} \left(\frac{f_i - f_{\text{avg}}}{f} \right)^2, \\ f &= \begin{cases} \max\left\{ \left| f_i - f_{\text{avg}} \right| \right\}, & \max\left\{ \left| f_i - f_{\text{avg}} \right| \right\} > 1, \\ 1, & \text{others}. \end{cases} \end{aligned} \tag{39}$$

Step 6. Calculate the probability of mutation p_m.

Step 7. Generate random number $\varepsilon \in [0,1]$; if $\varepsilon < p_m$, press to execute mutation operation. Otherwise, go to *Step 8*.

Step 8. Determine whether the algorithm convergence criterion is satisfied; if it is satisfied, execute Step 9. Otherwise, turn to *Step 3*.

Step 9. Output the optimal value gBest of the background value coefficient w(or order r) and the simulation and prediction data of the TWGM$(1,1|w,r)$ model at this time, and the algorithm operation ends.

4. Model Application: Forecast of China's Higher Education Gross Enrollment Rate

To identify the optimized performance on each parameter, the variables (the background value coefficient w and order r) are simulated with four different groups. The original TWGM(1,1) model means that the model is not optimized for order and background value coefficients at all. The first group is the model with, $w = 0.5$ and $r = 1$, that is, the TWGM(1,1|0.5,1)model. The second one is the model with $w = 0.5$ and $r = r*$, that is, the TWGM(1,1|0.5,$r*$) model, which indicates that the model only optimizes the order. The third one is the model with, $w = w^*$ and $r = 1$, that is, the TWGM$(1,1|w^*,1)$ model, which means that the model only optimizes the background value coefficient. The last group is $w = w^*$ and $r = r*$, that is, the TWGM$(1,1|w^*,r^*)$ model, which suggests that the model optimizes the order and background value coefficients at the same time. The simulation results with Matlab for four parameter groups are shown in Table 1, where $x^{(0)}(k)$ is China's high education enrollment data from 1991 to 2019.

In order to intuitively reflect the simulation effects of different models, a comparison chart of the simulation curves of different models is drawn, as shown in Figure 1, where models with different parameter values show different simulation effects. The curve of the TWGM(1,1| 0.5, 1) and TWGM$(1,1|w^*,1)$ models overlap and show a same trend. The TWGM$(1,1|w^*,r^*)$ model initially displays better simulating and predicting effects for the gross enrollment ratio in higher education in China since it produces the most close result with raw data.

In order to measure the optimization performance of different parameters, the average relative simulation percentage error $\overline{\Delta}_S$, average relative prediction percentage error $\overline{\Delta}_F$, and comprehensive percentage error $\overline{\Delta}$ of the gray model of each model are calculated according to the results in Table 1. The results are shown in Table 2. The relevant calculation equation is as follows:

$$\begin{aligned} \overline{\Delta}_S &= \frac{1}{t-1} \sum_{k=2}^{t} \Delta k, \\ \overline{\Delta}_F &= \frac{1}{n-t} \sum_{k=t+1}^{n} \Delta k, \\ \overline{\Delta} &= \frac{1}{n-1} \sum_{k=2}^{n} \Delta k. \end{aligned} \tag{40}$$

Among them, Δk represents the simulation/prediction error of $\widehat{x}^{(0)}(k)$, n represents the number of experimental data, and t represents the number of experimental data used for simulation.

Based on the results of the calculations in Table 2, the relative percentage error comparison figure with different parameter values is plotted, as shown in Figure 2 in

Table 1: Simulation and prediction results of different models on China's higher education gross enrollment rate.

Serial number	$x^{(0)}(k)$	TWGM(1,1|0.5,1)		TWGM(1,1|0.5,$r*$)		TWGM$(1,1\|w^*,1)$		TWGM$(1,1\|w^*,r^*)$	
		$\hat{x}^{(0)}(k)$	$\Delta(k)$	$\hat{x}^{(0)}(k)$	$\Delta(k)$	$\hat{x}^{(0)}(k)$	$\Delta(k)$	$\hat{x}^{(0)}(k)$	$\Delta(k)$
Simulation data									
$k=2$	3.5	3.14	10.2188	3.13	10.6725	3.15	10.1019	3.10	11.5502
$k=3$	3.9	4.04	3.5078	4.06	4.1620	4.04	3.6783	4.05	3.7288
$k=4$	5	4.97	0.6278	5.00	0.0311	4.98	0.4414	4.98	0.4818
$k=5$	6	5.94	1.0113	5.96	0.6118	5.95	0.8097	5.93	1.1729
$k=6$	7.2	6.95	3.4645	6.96	3.2806	6.97	3.2561	6.92	3.9439
$k=7$	8.3	8.00	3.5661	8.00	3.5641	8.02	3.3488	7.94	4.3312
$k=8$	9.1	9.10	0.0159	9.09	0.1324	9.12	0.2488	9.01	1.0183
$k=9$	9.8	10.24	4.5378	10.22	4.2692	10.27	4.7877	10.12	3.2693
$k=10$	10.5	11.44	8.9113	11.40	8.5588	11.46	9.1774	11.28	7.4608
$k=11$	12.5	12.68	1.4111	12.63	1.0562	12.71	1.6634	12.50	0.0000
$k=12$	13.3	13.97	5.0293	13.92	4.6743	14.00	5.2945	13.77	3.5632
$k=13$	15	15.32	2.1024	15.27	1.8063	15.35	2.3637	15.11	0.7258
$k=14$	17	16.72	1.6586	16.68	1.8634	16.76	1.4040	16.51	2.8898
$k=15$	19	18.18	4.3196	18.16	4.4102	18.23	4.0694	17.98	5.3818
$k=16$	21	19.70	6.1832	19.71	6.1376	19.75	5.9357	19.52	7.0515
$k=17$	22	21.29	3.2394	21.33	3.0264	21.34	2.9821	21.14	3.9173
$k=18$	23	22.94	0.2639	23.04	0.1533	23.00	0.0032	22.84	0.7002
$k=19$	23.3	24.66	5.8381	24.82	6.5170	24.73	6.1234	24.63	5.6917
Forecast data									
$k=20$	24.2	26.45	9.3103	26.69	10.2816	26.52	9.6068	26.50	9.5236
$k=21$	26.5	28.32	6.8708	28.65	8.1088	28.40	7.1623	28.48	7.4706
$k=22$	26.9	30.27	12.5146	30.71	14.1460	30.35	12.8231	30.56	13.5932
$k=23$	30	32.29	7.6443	32.86	9.5415	32.38	7.9409	32.74	9.1365
$k=24$	34.5	34.40	0.2759	35.13	1.8131	34.50	0.0000	35.04	1.5616
$k=25$	37.5	36.60	2.3881	37.50	0.0000	36.71	2.1169	37.46	0.1167
$k=26$	40	38.90	2.7601	39.99	0.0210	39.00	2.4889	40.00	0.0000
$k=27$	42.7	41.28	3.3183	42.61	0.2195	41.40	3.0476	42.68	0.0541
$k=28$	45.7	43.77	4.2234	45.35	0.7650	43.89	3.9543	45.49	0.4502
$k=29$	48.1	46.36	3.6163	48.23	0.2714	46.49	3.3446	48.46	0.7477
$k=30$	51.6	49.06	4.9237	51.25	0.6712	49.20	4.6548	51.58	0.0365

Table 2: Error comparison of different models.

Error type	TWGM(1,1|0.5,1) (%)	TWGM(1,1|0.5,$r*$) (%)	TWGM$(1,1\|w^*,1)$ (%)	TWGM$(1,1\|w^*,r^*)$ (%)
Average relative simulation percentage error $\overline{\Delta}_s$	3.6615	3.6071	3.6494	3.7155
Average relative prediction percentage error $\overline{\Delta}_F$	5.2587	4.1672	5.1946	3.8810
Grey model comprehensive percentage error $\overline{\Delta}$	4.2673	3.8195	4.2355	3.7782

supplementary materials. From Table 2 and Figure 2 in supplementary materials, the following conclusions are drawn:

(1) The background value optimization effect: comparing the TWGM(1,1|0.5,1) model and TWGM$(1,1|w^*,1)$ model with a $w^* = 0.4538$, the graph shows that the integrated percent error of the gray model after the background value optimization is slightly lower than the original model, but the difference is not remarkable, which suggests that the background value optimization has an optimization effect on the comprehensive performance of China's higher education gross enrollment rate, but the effect is not significant.

(2) The order optimization effect: compare the TWGM(1,1|0.5,1) model with the TWGM(1,1|0.5,$r*$) model with a $r* = 1.1127$ in Figure 2 in supplementary materials, the average relative percentage prediction error and the comprehensive percentage error after the order optimization are significantly lower than the original model, which shows that the order optimization has a remarkable improvement on the model performance in simulating China's higher education gross enrollment rate.

(3) The TWGM$(1,1|w^*,r^*)$ model possesses the lowest error of 3.7782%, which indicates that the optimization of both the background value and the order is of the best option for improvement on the model performance. Thus, the TWGM$(1,1|w^*,r^*)$ model provides the best fitting effect and is preferred for simulation and prediction.

5. Conclusions

The study proposed a new optimization model by combining the Three-parameter Whitenization Grey Model with the PSA. To demonstrate the performance of this new model, an application analysis is carried out in the predication of China's high education enrollment ratio with Matlab simulation. The findings suggest that the new optimization model proposed in this study improves the performance of the original gray model remarkably. Specifically, compared with the original gray model, the average relative percentage prediction error and comprehensive percentage error after the optimization are significantly reduced, which indicates that the simultaneous optimization of the background value and order has a notable improvement on the overall performance in simulating and predicting China's higher education gross enrollment ratio. Besides, it is indicated that the intelligent and adjustable adaptivity improves the accuracy of optimization and, furthermore, extends its application in the empirical study in the future.

Data Availability

The data (High Education Gross Enrollment Ratio of China) can be found in the supplementary information files or in "The Education Statistic Yearbook of China," Chapter 1, Page 21: Gross Enrollment Ratio of Education by Level (https://data.cnki.net/trade/Yearbook/Single/N2020070382?z=Z017).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study is supported by Fundamental Research Funds for the Central Scientific Research Institutes of NIES, "Research on the evaluation of first-class undergraduate education," No. GYB2019002.

Supplementary Materials

Figure 1: Comparison of simulation curves of different models. Figure 2: Comparison of relative percentage errors of different models. (*Supplementary Materials*)

References

[1] D. Ju-Long, "Control problems of grey systems," *Systems & Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[2] N. Xu, S. Ding, Y. Gong, and J. Bai, "Forecasting Chinese greenhouse gas emissions from energy consumption using a novel grey rolling model," *Energy*, vol. 175, pp. 218–227, 2019.

[3] N.-T. Nguyen and T.-T. Tran, "Optimizing mathematical parameters of Grey system theory: an empirical forecasting case of Vietnamese tourism," *Neural Computing and Applications*, vol. 31, no. 2, pp. 1075–1089, 2019.

[4] B. Zeng, S. Liu, B. Yun, and M. Zhou, "Research on early prediction and warning of human diseases based on grey system modeling technology," *Chinese Management Science*, vol. 28, no. 1, pp. 144–152, 2020.

[5] Z. X. Wang, Z. W. Wang, and Q. Li, "Forecasting the industrial solar energy consumption using a novel seasonal GM(1,1) model with dynamic seasonal adjustment factors," *Energy*, vol. 200, 2020.

[6] H. M. Duan, X. P. Xiao, J. Long et al., "Tensor alternating least squares grey model and its application to short-term traffic flows," *Applied Soft Computing Journal*, vol. 89, 2020.

[7] L. Liu, Y. Feng, and C. Yang, "My country's higher education gross enrollment rate GM(1,1) prediction model and its application research," *Journal of Yunnan Normal University (Philosophy and Social Sciences Edition)*, vol. 41, no. 3, pp. 123–127, 2009.

[8] X. Dong, "Research on the grey prediction model of national medium and long-term higher education gross enrollment rate," *Journal of National Academy of Education Administration*, vol. 5, pp. 43–47, 2010.

[9] L. Zhan and H. Shi, "Grey modeling method and model for approximate inhomogeneous index data," *System Engineering Theory and Practice*, vol. 33, no. 3, pp. 689–694, 2013.

[10] B. Zeng and C. Li, "Forecasting the natural gas demand in China using a self-adapting intelligent grey model," *Energy*, vol. 112, pp. 810–825, 2016.

[11] X. Ma, Y.-S. Hu, and Z.-B. Liu, "A novel kernel regularized nonhomogeneous grey model and its applications," *Communications in Nonlinear Science and Numerical Simulation*, vol. 48, pp. 51–62, 2017.

[12] Y. Wang, "GM(1,1) Gradually optimizing the promotion of direct modeling method," *System Engineering Theory and Practice*, vol. 02, pp. 120–124, 2003.

[13] Y. Wang, G. Liu, and K. Liu, "A gradual optimization and direct modeling method of GM(1,1)," *Systems Engineering Theory and Practice*, vol. 09, pp. 99–104+144, 2000.

[14] Y. Wang, "The direct modeling method and properties of GM(1,1)," *System Engineering Theory and Practice*, vol. 1, pp. 27–31, 1988.

[15] L. Ye and J. Li, "Research on interval grey number prediction model considering whitening weight function," *Fuzzy Systems and Mathematiics*, vol. 33, no. 4, pp. 165–174, 2020.

[16] B. Zeng and S. Liu, "DGM(1,1) direct modeling method of approximate non-homogeneous exponential sequence," *System Engineering Theory and Practice*, vol. 31, no. 2, pp. 297–301, 2011.

[17] Q. Xiao, M. Gao, X. Xiao et al., "A novel grey Riccati-Bernoulli model and its application for the clean energy consumption prediction," *Engineering Applications of Artificial Intelligence*, vol. 95, Article ID 103863, 2020.

[18] Q. Xiao, M. Shan, M. Gao, X. Xiao, and M. Goh, "Parameter optimization for nonlinear grey Bernoulli model on biomass energy consumption prediction," *Applied Soft Computing*, vol. 95, Article ID 106538, 2020.

[19] M. Gao, H. Yang, Q. Xiao et al., "A novel fractional grey Riccati model for carbon emission prediction," *Journal of Cleaner Production*, vol. 10, Article ID 124471, 2020.

[20] Q. Xiao, M. Shan, M. Gao et al., "Grey information coverage interaction relational decision making and its application," *Journal of Systems Engineering and Electronics*, vol. 31, no. 2, pp. 359–369, 2020.

[21] B. Xiao, H. Li, and X. Ma, "A novel multi-variable grey forecasting model and its application in forecasting the grain production in China," *Computers & Industrial Engineering*, vol. 150, Article ID 106915, 2020.

[22] B. Zeng, X. Ma, and M. Zhou, "A new-structure grey Verhulst model for China's tight gas production forecasting," *Applied Soft Computing*, vol. 96, Article ID 106600, 2020.

[23] B. Zeng, M. Zhou, X. Liu et al., "Application of a new grey prediction model and grey average weakening buffer operator to forecast China's shale gas output," *Energy Reports*, vol. 6, pp. 1608–1618, 2020.

[24] B. Zhang, M. Tong, and X. Ma, "A new-structure grey Verhulst model: development and performance comparison," *Applied Mathematical Modelling*, vol. 81, pp. 522–537, 2020.

[25] J. J. Wang, Y. G. Dang, J. Ye et al., "An improved grey prediction model based on matrix representations of the optimized initial value," *Journal of Grey System*, vol. 30, no. 3, pp. 143–156, 2018.

[26] S. Ding, K. W. Hipel, and Y.-G. Dang, "Forecasting China's electricity consumption using a new grey prediction model," *Energy*, vol. 149, pp. 314–328, 2018.

[27] S. Ding, R. Li, and Y. Dang, "GM(1,1) power model based on optimization of initial conditions and its application," *China Management Science*, vol. 28, no. 1, pp. 153–161, 2020.

[28] S. H. Mao, M. Y. Gao, X. P. Xiao et al., "A novel fractional grey system model and its application," *Applied Mathematical Modelling*, vol. 40, no. 7-8, pp. 5063–5076, 2016.

[29] X. Ma, W. Wu, B. Zeng, Y. Wang, and X. Wu, "The conformable fractional grey system model," *ISA Transactions*, vol. 96, pp. 255–271, 2020.

[30] X. Zeng, S. Yan, F. He, and Y. Shi, "Multi-variable grey model based on dynamic background algorithm for forecasting the interval sequence," *Applied Mathematical Modelling*, vol. 80, pp. 99–114, 2020.

[31] W. Zhang, R. Xiao, B. Shi et al., "Forecasting slope deformation field using correlated grey model updated with time correction factor and background value optimization," *Engineering Geology*, vol. 260, 2019.

[32] K. Li, L. Liu, J. Zhai, T. M. Khoshgoftaar, and T. Li, "The improved grey model based on particle swarm optimization algorithm for time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 55, pp. 285–291, 2016.

[33] N. Xu, Y. Dang, and Y. Gong, "Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China," *Energy*, vol. 118, pp. 473–480, 2017.

[34] M. Nouiri, A. Bekrar, A. Jemai, S. Niar, and A. C. Ammari, "An effective and distributed particle swarm optimization algorithm for flexible job-shop scheduling problem," *Journal Of Intelligent Manufacturing*, vol. 29, no. 3, pp. 603–615, 2018.

[35] M. Navabi, A. Davoodi, and M. Reyhanoglu, "Optimum fuzzy sliding mode control of fuel sloshing in a spacecraft using PSO algorithm," *Acta Astronautica*, vol. 167, pp. 331–342, 2019.

[36] B. Zeng and S. Liu, "A self-adaptive intelligence gray prediction model with the optimal fractional order accumulating operator and its application," *Mathematical Methods in the Applied Sciences*, vol. 40, no. 18, pp. 7843–7857, 2017.

[37] Z.-X. Wang and Q. Li, "Modelling the nonlinear relationship between CO2 emissions and economic growth using a PSO algorithm-based grey Verhulst model," *Journal of Cleaner Production*, vol. 207, pp. 214–224, 2019.

[38] W. Meng, B. Zeng, and S. Li, *Grey Forecasting Theory and its Application*, Science Press, Beijing, China, 2020.

[39] L. F. Wu, S. F. Liu, and L. G. Yao, "Discrete gray model based on fractional order accumulation," *Systems Engineering Theory and Practice*, vol. 34, no. 7, pp. 1822–1827, 2014.

[40] L. Wu, S. Liu, L. Yao, S. Yan, and D. Liu, "Grey system model with the fractional order accumulation," *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 7, pp. 1775–1785, 2013.

[41] W. Meng and B. Zeng, "Discrete gray model and order optimization based on reciprocal fractional order operator," *Control and Decision Making*, vol. 31, no. 10, pp. 1903–1907, 2016.