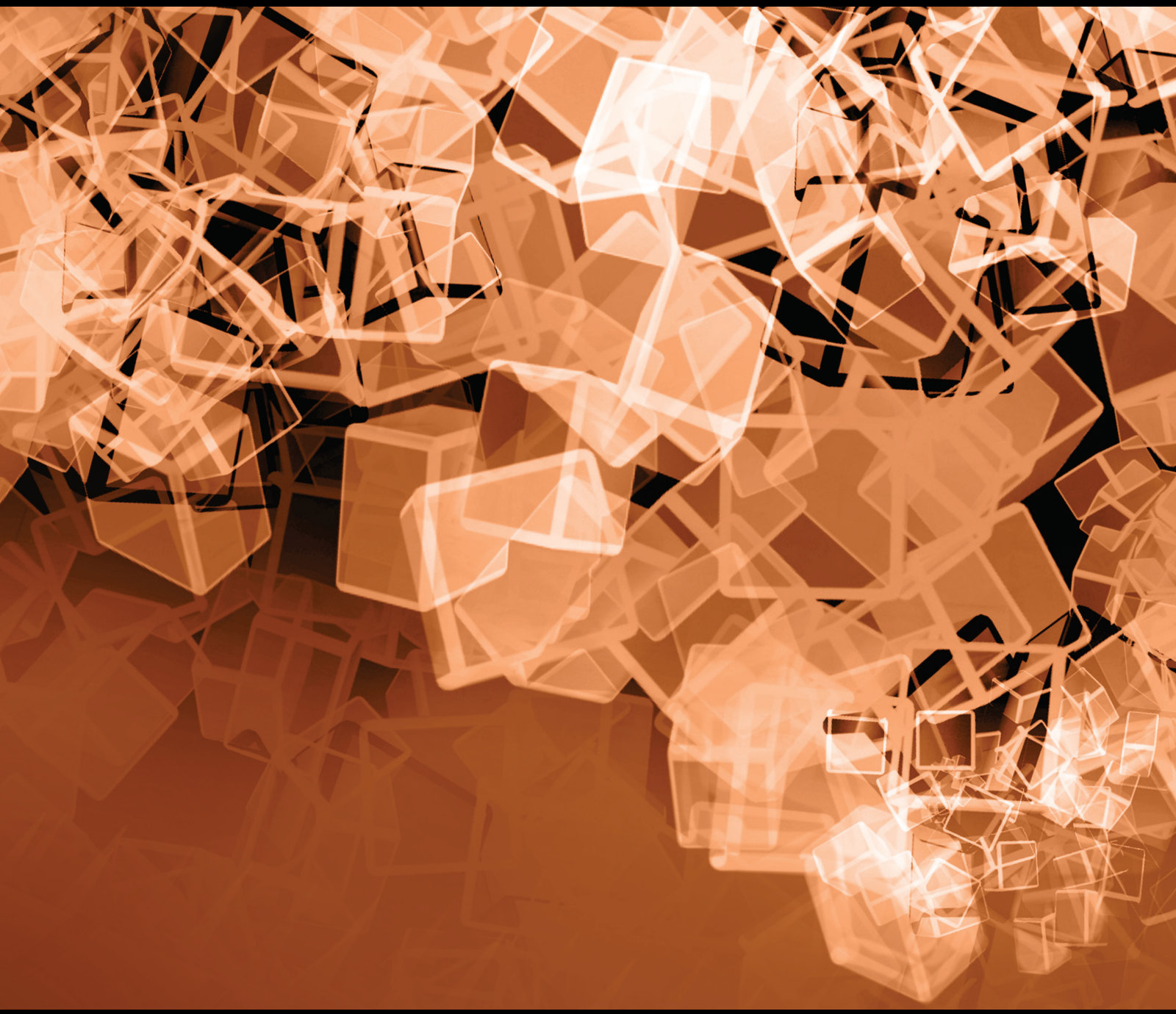
Discrete Dynamics in Nature and Society

Fintech and Financial Risk Analysis in the Era of Big Data 2021

Lead Guest Editor: Dehua Shen
Guest Editors: Xiao Li, Xiong Xiong, and Silvano Cincotti



Fintech and Financial Risk Analysis in the Era of Big Data 2021

Discrete Dynamics in Nature and Society

Fintech and Financial Risk Analysis in the Era of Big Data 2021

Lead Guest Editor: Dehua Shen
Guest Editors: Xiao Li, Xiong Xiong, and Silvano Cincotti

Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in “Discrete Dynamics in Nature and Society.” 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

Paolo Renna, Italy

Associate Editors

Cengiz Çinar, Turkey
Seenith Sivasundaram, USA
J. R. Torregrosa, Spain
Guang Zhang, China
Lu Zhen, China

Academic Editors

Douglas R. Anderson, USA
Viktor Avrutin, Germany
Stefan Balint, Romania
Kamel Barkaoui, France
Abdellatif Ben Makhlouf, Saudi Arabia
Gabriele Bonanno, Italy
Florentino Borondo, Spain
Jose Luis Calvo-Rolle, Spain
Pasquale Candito, Italy
Giulio E. Cantarella, Italy
Giancarlo Consolo, Italy
Anibal Coronel, Chile
Binxiang Dai, China
Luisa Di Paola, Italy
Xiaohua Ding, China
Tien Van Do, Hungary
Hassan A. El-Morshedy, Egypt
Elmetwally Elabbasy, Egypt
Marek Galewski, Poland
Bapan Ghosh, India
Caristi Giuseppe, Italy
Gisèle R Goldstein, USA
Vladimir Gontar, Israel
Pilar R. Gordoa, Spain
Luca Guerrini, Italy
Chengming Huang, China
Giuseppe Izzo, Italy
Sarangapani Jagannathan, USA
Ya Jia, China
Emilio Jiménez Macías, Spain
Polinpapiliñho F. Katina, USA
Eric R. Kaufmann, USA
Mehmet emir Koksal, Turkey
Junqing Li, China
Li Li, China
Wei Li, China
Ricardo López-Ruiz, Spain
Rodica Luca, Romania
Palanivel M, India
A. E. Matouk, Saudi Arabia
Rigoberto Medina, Chile
Vicenç Méndez, Spain
Dorota Mozyrska, Poland
Jesus Manuel Munoz-Pacheco, Mexico
Yukihiko Nakata, Japan
Luca Pancioni, Italy
Ewa Pawluszewicz, Poland
Alfred Peris, Spain
Adrian Petrusel, Romania
Andrew Pickering, Spain
Tiago Pinto, Spain
Chuanxi Qian, USA
Youssef N. Raffoul, USA
Maria Alessandra Ragusa, Italy
Aura Reggiani, Italy
Marko Robnik, Slovenia
Priyan S, Uzbekistan
Mouquan SHEN, China
Aceng Sambas, Indonesia
Christos J. Schinas, Greece
Mijanur Rahaman Seikh, India
Tapan Senapati, China
Kamal Shah, Saudi Arabia
Leonid Shaikhet, Israel
Piergiulio Tempesta, Spain
Fabio Tramontana, Italy
Cruz Vargas-De-León, Mexico
Francisco R. Villatoro, Spain
Junwei Wang, China
Kang-Jia Wang, China
Rui Wang, China
Xiaoquan Wang, China
Chun Wei, China
Bo Yang, USA
Zaoli Yang, China
Chunrui Zhang, China
Ying Zhang, USA
Zhengqiu Zhang, China
Yong Zhou, China
Zuonong Zhu, China
Mingcheng Zuo, China

Contents

Quantifying Cross-Correlations between Economic Policy Uncertainty and Bitcoin Market: Evidence from Multifractal Analysis
Junjun Ma, Tingting Wang, and Ruwei Zhao
Research Article (9 pages), Article ID 1072836, Volume 2022 (2022)

An Evolution Model of Fintech Supports Technology-Based SMEs Innovation and Growth
Lei Zhuang, Han Zhou, and Tingqiang Chen
Research Article (16 pages), Article ID 6867404, Volume 2022 (2022)

Liquidity, Credit Risk, and Their Interaction on the Spreads in China's Corporate Bond Market
Zijian Wu, Baochen Yang, and Yunpeng Su
Research Article (19 pages), Article ID 2996704, Volume 2022 (2022)

Promoting Variable Effect Consistency in Mixture Cure Model for Credit Scoring
Chenlu Zheng, Jianping Zhu, Xinyan Fan, Song Chen, and Zhiyuan Zhang
Research Article (13 pages), Article ID 3112987, Volume 2022 (2022)

On the Safe-Haven Ability of Bitcoin, Gold, and Commodities for International Stock Markets: Evidence from Spillover Index Analysis
Qian Wang, Yu Wei, Yao Wang, and Yuntong Liu
Research Article (16 pages), Article ID 9520486, Volume 2022 (2022)

Research on a Dynamic Evaluation of Financial Risks of Intellectual Property Pledge Financing of High-Tech Enterprises
Xiangyu Ge, Ying Jin, Qing Li, Xiaofang Yao, and Shican Liu
Research Article (11 pages), Article ID 2048811, Volume 2021 (2021)

Taming the Factor Zoo: New Evidence from China
Dongxu Chen, Xieyang Shen, and Tao Liu
Research Article (16 pages), Article ID 9592906, Volume 2021 (2021)

The Cross-Correlations between Foreign Flows in Chinese A-Share Markets and Uncertainties in Home Markets
Tao Bing, Fei Hu, and Hongkun Ma
Research Article (18 pages), Article ID 6541902, Volume 2021 (2021)

Socially Responsible Investment Portfolio Construction with a Double-Screening Mechanism considering Machine Learning Prediction
Jun Zhang and Xuedong Chen
Research Article (15 pages), Article ID 7390887, Volume 2021 (2021)

Incentive and Coordination in the Two-Sided Market: Evidence from the P2P Lending Market
Yingxiu Zhao and Baojuan Shi
Research Article (15 pages), Article ID 9240374, Volume 2021 (2021)

Dependence and Risk Spillover among Hedging Assets: Evidence from Bitcoin, Gold, and USD
Jiang Yu, Yue Shang, and Xiafei Li
Research Article (20 pages), Article ID 2010705, Volume 2021 (2021)

The Impacts of the Infectious Disease Epidemic on the Permanent Volatility of Precious Metal and Crude Oil Futures Markets: A Long-Term Perspective
Yue Shang, Xiaodan Chen, Yifeng Zhang, and Yu Wei
Research Article (12 pages), Article ID 2800671, Volume 2021 (2021)

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 1072836, 9 pages
https://doi.org/10.1155/2022/1072836

Research Article

Quantifying Cross-Correlations between Economic Policy Uncertainty and Bitcoin Market: Evidence from Multifractal Analysis

Junjun Ma,[1] **Tingting Wang**,[2] **and Ruwei Zhao**[3]

[1]*School of Economics and Management, Beijing University of Technology, Beijing 100124, China*
[2]*Tianjin Bohai Sea Monitoring and Management Center, Tianjin 300456, China*
[3]*School of Business, Jiangnan University, Wuxi, Jiangsu 214122, China*

Correspondence should be addressed to Tingting Wang; tjghwtt@163.com

Received 21 January 2022; Accepted 13 May 2022; Published 6 June 2022

Academic Editor: Dehua Shen

Copyright © 2022 Junjun Ma 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.

We investigate the dynamic correlation between the Bitcoin price (BTC) and the U.S. economic policy uncertainty index (USEPU) from the perspective of multifractality. Utilizing the multifractal detrended cross-correlation analysis (MF-DCCA), we confirm a long-range cross-correlation between BTC and USEPU. Moreover, the empirical results of MF-DCCA show that the power-law properties and multifractal characteristics between BTC and USEPU are significant. We further examine the long-range dependency of cross-correlation between BTC and USEPU series via the Hurst exponent test and confirm the durable cross-correlation. Finally, we introduce another multifractal indicator and examine the extent of multifractality among time series. The empirical results indicate that the BTC series, USEPU series, and the cross-correlation of BTC-USEPU present apparent multifractality, where BTC shows the strongest degree of multifractality.

1. Introduction

In the context of the subprime mortgage crisis and the public's growing concerns over government policy changes, Bitcoin was introduced by Nakamoto [1]. Bitcoin is an adequately decentralized cryptocurrency depending on sophisticated protocols. Since then, the cryptocurrency market represented by Bitcoin has developed rapidly and aroused the widespread interest of regulators, scholars, and investors. A series of studies have conducted in-depth research on price formation [2–4], market efficiency [5–7], stylized facts [8, 9], and price dynamics of Bitcoin [10–14]. Additionally, because of Bitcoin's independence over authority and significant fluctuations in its prices, the debates over whether it is a safe-haven asset and its relationship with government policy changes have caused controversy [14–19]. Especially during the trade war and COVID-19, there has been a sudden increase in economic policy uncertainty (EPU).

Recent literature examines whether Bitcoin can hedge from uncertainty based on the wavelet-based quantile-on-quantile regressions, ordinary least-squares regressions, and the Bayesian Graphical Structural VAR model [20, 21]. They argue that Bitcoin has safe haven properties and can serve as a suitable hedge against different kinds of uncertainty. Wang et al. [22] also confirm this argument and find that EPU (this paper takes three indexes as proxies for EPU, including VIX, equity market uncertainty, and US EPU) exhibits generally weak risk spillover effects on Bitcoin. In contrast, Conlon and McGee [23] document that the downside risk of a portfolio containing Bitcoin increased significantly relative to investing in the S&P 500 alone during the Covid-19 pandemic. These studies usually focus on the linear properties between variables or simple nonlinear relationships. However, it has been widely accepted that the dynamic relationships between asset prices are inherently nonlinear and even multifractal [24–32]. Furthermore, various nonlinear cross-correlations between different time series have

been confirmed, including EPU and stock market volatility [33], online searches and Bitcoin [34, 35], different capital markets [36–41], and global stock markets [24, 26]. With this view, we attempt to study the dynamic correlation between Bitcoin price (BTC) and EPU from the perspective of multifractality.

Since the financial crisis, the impact of EPU on the economy and finance has also become a hot topic for scholars [42–46]. Based on the frequency of newspaper coverage, Baker et al. [47] proposed a novel EPU index and found that the U.S. EPU index (USEPU) is strongly linked to macro and micro variables such as stock market volatility and employment. At present, North America and Western Europe are the regions with the highest recognition of Bitcoin. The daily EPU data of the United States are publicly available. Therefore, we connect USEPU (the uncertainty of the U.S. economy and policy) with BTC to explore the joint dynamics between them. Some recent literature studies the impact of EPU on the Bitcoin market from the perspectives of risk spillovers, volatility effects, and dynamic correlations [48–51]. On the one hand, the empirical results of the impact of EPU on the Bitcoin market are relatively mixed. On the other hand, Wang et al. [50] demonstrated that there is a dynamic correlation between EPU and bit using the DCC-GARCH model. Inspired by these works of literature, we attempt to use the MF-DCCA method (traditional econometric methods mainly study the correlation between time series from a linear perspective; however, the characteristics of multifractality are common in socioeconomic systems. Also, it is not reliable and comprehensive to analyze financial markets with traditional linear analysis methods. The long memory of financial time series makes it unable to meet the application conditions of the traditional correlation coefficient method. On the contrary, by calculating the multifractal spectrum of the time series, the purpose of understanding the internal complex structure of the fractal can be achieved. Therefore, this paper adopts the widely recognized MF-DCCA method to study multifractal features) to examine the dynamic cross-correlations between them. As we know, the advantage of the MF-DCCCA method is to measure the fractal characteristics of time series under different scales on the basis of eliminating local trends and further explore the cross-correlation and linear complex features between different time series [52–54].

Hence, we mainly adopt multifractal cross-correlation analysis (MF-DCCA) to investigate the multifractal characteristics between the USEPU series and the BTC series. The empirical evidence shows that there exists a long-range cross-correlation between the USEPU series and the BTC series. Moreover, using a log-log plot, we also find that no matter how the scaling order changes, the cross-correlations within the series demonstrate power-law character with the growing value of the fluctuation function. Furthermore, we calculate the Hurst exponent to test whether the cross-correlation between the BTC and USEPU series is enduring and find that it is persistent. Additionally, the results show that none of the Hurst exponent lines is constant, providing the preliminary evidence for multifractality. To confirm the multifractality between BTC and USEPU, we calculate another proxy for multifractality and perform the spectrum singularity check. The empirical results indicate that neither the Rényi exponent nor the singularity spectral curve exhibits a typical linear shape, which confirms the existence of multifractality between BTC and USEPU. Finally, we calculate the Hölder exponent difference and find that the BTC series shows the strongest multifractality character with the largest value of the Hölder exponent difference.

Our main contributions to the growing literature are twofold: first, different from the existing literature on the impact of EPU on BTC [48–51], our study is the first to study the dynamic relationship between BTC and EPU from the perspective of multifractality. Based on the MF-DCCA, we provide strong evidence for power-law cross-correlation and apparent multifractal features between USEPU and BTC. Our findings complement the existing studies on the impact of uncertainty on the Bitcoin market and provide a comprehensive understanding of their joint dynamics. Second, our study enriches the literature on multifractality analysis. The previous studies suggest that nonlinear correlations between EPU and traditional assets (stock, oil, and gold) are persistent [55, 56]. Unlike traditional assets, Bitcoin is a decentralized currency that is not dominated by any authority. However, contrary to intuitive expectations, EPU also exhibits robust multifractal characteristics with BTC. Moreover, BTC shows a continuous upward trend, especially after 2020, with a sharp rise and violent fluctuations. Our findings inform the dynamics, long-range autocorrelation, and multifractal characteristics of BTC. We expect that our study could shed light on the risk management of Bitcoin and help investors have a more comprehensive understanding of the innovative financial asset.

The remainder of this study is organized as follows. Section 2 introduces the data and discusses summary statistics of the BTC and USEPU series. In Section 3, we describe our research methodology, including cross-correlation analysis and MF-DCCA. Section 4 presents the detailed results and discusses the main empirical findings. Finally, we conclude in Section 5.

2. Data

Our sample is mainly composed of two-time series: the U.S. economic policy uncertainty index (USEPU) and Bitcoin price (BTC). Our selection of USEPU follows Baker et al. [47]. The index currently covers 27 major economies globally and has been adopted by many well-known institutions and academic research. As stated by the website, USEPU quantifies the coverage of various newspapers collected from Access World New's NewsBank source according to three sets of terms: (a) the terms related to "economic" or "economy" (e.g., "monetary policy"), (b) the terms related to "uncertain" or "uncertainty" (e.g., "COVID-19"), and (c) the terms related to "government actions" (e.g., "deficit"). We obtain data on the daily USEPU from the public website https://policyuncertainty.com, while the daily BTC is crawled from the website https://finance.yahoo.com,

which provides Bitcoin news, transaction data, and professional message. Since we study the relationship between U.S. economic policy uncertainty and Bitcoin price, the U.S. dollar is chosen as the exchange currency for Bitcoin (Bitcoin price data are usually the latest transaction price of Bitcoin against the US dollar). In this study, we conduct all of our analyses from September 2014 to August 2021.

Table 1 presents descriptive statistics of our data sample, which consists of 2,527 observations. Both USEPU and BTC series show right-skew characters, exhibiting greater averages of the USEPU and BTC (124.90 and 8551.19) compared with their medians (91.44 and 5067.11, respectively), and implying distributions within the series is right-skewed. The skewness (2.63 and 2.48) also shows more extreme values at the right end of the series. For the kurtosis, it is accessible that the values of USEPU and BTC are well over 3 (11.52 and 8.81). In other words, the data distributions are steeper than the normal distribution. Then, we conduct Jarque-Bera (JB) tests to test whether the series of USEPU and BTC demonstrates normality properties. The results reject the null hypothesis that the time series obeys a normal distribution since their JB statistics are much greater than 0 (10,540.42 and 6,148.18). The statistical results of the BTC and USEPU series provide strong evidence that neither series was normally distributed during our sample period.

Figures 1 and 2 document the overall fluctuations of BTC and USEPU, respectively. As a representative of virtual currency, BTC has shown a general upward trend, especially after 2017. As shown in Figure 1, from January 2020 to July 2021, BTC fluctuates significantly, which may be affected by government regulation of digital currencies, transaction demand, investor sentiment, etc. For USEPU in Figure 2, the series exhibits relatively stable fluctuations on average from 2014 to 2019. However, in early 2020, USEPU rises rapidly and then gradually normalizes. The sudden outbreak of COVID-19 and the US election may be the drivers of the violent fluctuations. From the evolution patterns of the BTC and USEPU series, we find that both series present fat-tailed characteristics, especially for the USEPU series, which is consistent with the results in descriptive statistics.

Table 1: Descriptive statistics.

Variable	USEPU	BTC
Mean	124.90	8,551.19
Minimum	3.32	178.10
Median	91.44	5,067.11
Maximum	861.10	63,503.46
Std. Dev	107.19	12,625.76
Skewness	2.62	2.48
Kurtosis	11.52	8.81
Jarque-Bera	10,540.42	6,148.18
Probability	≤0.01	≤0.01
N	2527	2527

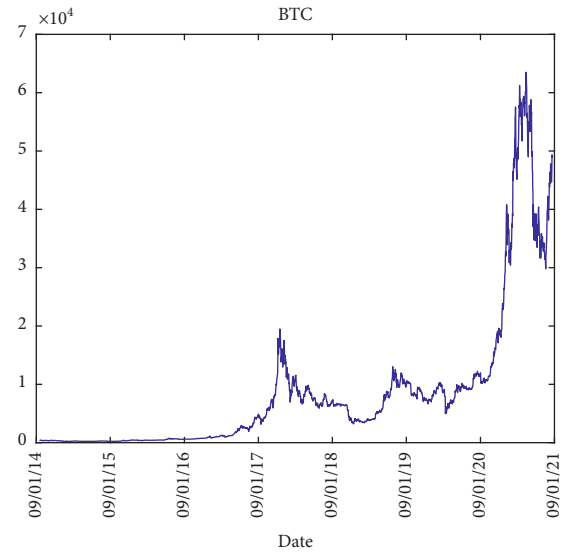


Figure 1: Evolution pattern of Bitcoin price (BTC).

3. Methodology

To examine whether there is multifractality between BTC and USEPU, we follow Zhou [52] and conduct the prevailing MF-DCCA (multifractal cross-correlation analysis), which combines the advantages of DFA and DCCA. First, we conduct the DCCA proposed by Podobnik and Stanley [57] to examine the long-range characteristics of cross-correlations between BTC and USEPU series. Then, we apply the MF-DCCA to quantitatively analyze the power-law cross-correlation and multifractal features within BTC and USEPU series.

3.1. Cross-Correlation Analysis. Bitcoin is often seen as a safe haven due to its decentralized nature, and at the same time, Bitcoin has been an attractive risk asset in recent years. Hence, is there a linkage between Bitcoin's price and economic policy uncertainty? To acquire a universal understanding of the cross-correlation properties, we initially introduce the cross-correlation statistic $Q_{cc}(m)$. Consider two equal-length series $\{x_k\}$ and $\{y_k\}$, where $k = 1, 2, \ldots, N$. Specifically, $Q_{cc}(m)$ is defined as

$$Q_{cc}(m) = N^2 \sum_{i=1}^{m} \frac{C_i^2}{N - i}, \tag{1}$$

Then, C_i represents the cross-correlation function and can be constructed through the following equation:

$$C_i = \frac{\sum_{k=i+1}^{N} x_k y_{k-i}}{\sqrt{\sum_{k=1}^{n} x_k^2 \sum_{k=1}^{n} y_k^2}}, \tag{2}$$

where m represents the degree of freedom; N is the length of the time series $\{x_k\}$ and $\{y_k\}$. If the value of $Q_{cc}(m)$ exactly matches the critical value of $\chi^2(m)$, a statistical cross-correlation cannot be found. Otherwise, we can infer that there is a statistically significant correlation.

3.2. MF-DCCA. Next, we apply the MF-DCCA to quantitatively analyze the power-law cross-correlation and multifractal features within BTC and USEPU series. Wang et al.

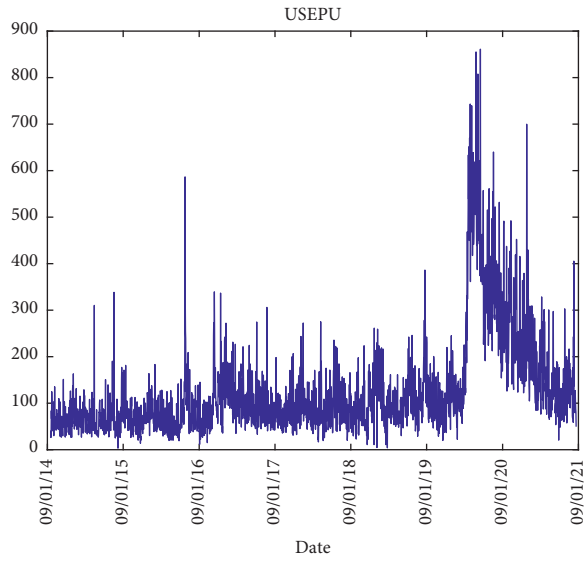


Figure 2: Evolution pattern of US economic policy uncertainty index (USEPU).

[41] established the widespread MF-DCCA method to investigate the multifractal power-law cross-correlation features between two nonstationary sequences with autocorrelation. Since the MF-DCCA method has been widely recognized, we only briefly introduce the steps as follows.

First, to remove the drift feature of each time series, we transform authentic series, $\{x_k\}$ and $\{y_k\}$, with a deduction of the constant shift ($\overline{x}$ and $\overline{y}$), respectively. The full calculation operation has been demonstrated in the following equations:

$$X(i) = \sum_{k=1}^{i} (x_k - \overline{x}), \overline{x} = \frac{1}{N} \sum_{k=1}^{N} x(k),$$
$$Y(i) = \sum_{k=1}^{i} (y_k - \overline{y}), \overline{y} = \frac{1}{N} \sum_{k=1}^{N} y(k). \tag{3}$$

Second, we break the detrended time series $X(i)$ and $Y(i)$ into N_s nonoverlapping equal parts with s observations to acquire a micro understanding of every single series. Taking N as the whole number of sample observations, the formation process of N_s is as follows:

$$Ns = \text{int}\left(\frac{N}{s}\right), \tag{4}$$

int(·) speaks on behalf of the integer function (because the time series length N may not always be an integer multiple of scale s, a relatively short portion will be discarded at the end of each profile. Hence, to include the remaining small segments, we do the same for the reverse order of the time series).

Third, with split N_s parts, we examine the local trend of every single part, ν. Specifically, the detrended covariance of every single part is marked as $F^2(s, \nu)$ and can be obtained from the following equations.

For $\nu = 1, 2, 3, \ldots, N_s$, that is,

$$F^2(s, \nu) = \frac{1}{s} \sum_{i=1}^{s} \left[X_{(\nu-1)s+i} - \widetilde{X}_{(\nu-1)s+i}\right] \cdot \left[Y_{(\nu-1)s+i} - \widetilde{Y}_{(\nu-1)s+i}\right]. \tag{5}$$

For $\nu = N_s + 1, N_s + 2, N_s + 3, \ldots, 2N_s$, we have

$$F^2(s, \nu) = \frac{1}{s} \sum_{i=1}^{s} \left[X_{N-(\nu-N_s)s+i} - \widetilde{X}_{N-(\nu-N_s)s+i}\right] \cdot \left[Y_{N-(\nu-N_s)s+i} - \widetilde{Y}_{N-(\nu-N_s)s+i}\right], \tag{6}$$

where $\widetilde{X}$ and $\widetilde{Y}$ is the nth order polynomial fitness check of part ν.

Fourth, we average the local covariances of all detrended segments to get the qth order fluctuation function, $F_q(s)$. Thus, the fluctuation function can be constructed by the following equations:

$$F_q(s) = \left\{\frac{1}{2N_s} \sum_{\upsilon=1}^{2N_s} \left[F^2(s, \nu)\right]^{q/2}\right\}^{1/q}, \tag{7}$$

when $q = 0$ and

$$F_0(s) = \exp\left\{\frac{1}{4N_s} \sum_{\upsilon=1}^{2N_s} \ln\left[F^2(s, \nu)\right]\right\}. \tag{8}$$

It is evident that the movement of $F_q(s)$ is affected by the selection of q and s. To acquire a full perception of the multifractality within series, we carry out calculation processes from steps 2 to 4 on different scale s and then continue to the fifth and ending procedure.

Fifth, with the multiselection of scale s, the scaling character of the fluctuation function $F_q(s)$ could be found through the gradient analysis for every single q. If a long-range multifractality feature can be discovered within the detrended series, the power-law relationship can be depicted by the following equation:

$$F_q(s) \sim s^{H_{xy}(q)}, \tag{9}$$

where $H_{xy}(q)$ is the generalized Hurst exponent, which is the slope of the log($Fq(s)$)~log(s) function graph (it is evident from equation (9) that the gradient of function varies with the scaling order q. Specifically, considering the case where q takes a value of 2, if the value of gradient $H_{xy}(2)$ is above 0.5, the cross-correlation between time series $X(i)$ and $Y(i)$ is persistent. If $H_{xy}(2)$ is lower than 0.5, the cross-correlation would be antipersistent).

To have a further study of multifractality within series, we begin to introduce another proxy, Rényi exponent. The proxy is designated as $\tau_{xy}(q)$ presented in the following equation:

$$\tau_{xy}(q) = qH_{xy}(q) - 1. \tag{10}$$

Finally, we execute singularity spectrum analysis as an extra investigation to improve multifractality reliability.

Through the Legendre transformation, the singularity spectrum, $f_{xy}(\alpha)$, could be acquired and presented by the subsequent formulas, based on the involvement of the Rényi ($\tau_{xy}(q)$) and Hölder (α_{xy}) exponents, respectively.

$$\begin{aligned} \alpha_{xy} &= \tau'_{xy}(q), \\ f_{xy}(\alpha) &= q\alpha_{xy} - \tau_{xy}(q), \end{aligned} \tag{11}$$

where α_{xy} is defined as the Hölder exponent, representing the strength of the singularity spectrum. Therefore, we utilize the strength difference or the spectrum width of the Hölder exponent $\Delta\alpha_{xy}$ ($\Delta\alpha_{xy} = \alpha_{xy,\max} - \alpha_{xy,\min}$), as the measurement for multifractality degree. It is easy to learn from the equation that the greater value of $\Delta\alpha_{xy}$ would represent higher levels of multifractality.

4. Empirical Results

4.1. Cross-Correlation Test. Following the steps described in the previous section, we first conduct the cross-correlation test introduced in Section 3.1 and obtain general knowledge of the relationship between BTC and USEPU series. The results are presented in Figure 3. Specifically, the significance level is 5%, and the degrees of freedom range from 1 to $N-1$. The green and red lines represent cross-correlation statistics $Q_{cc}(m)$ and the chi-square critical value $\chi^2(m)$. The pattern of Figure 3 implies that $Q_{cc}(m)$ of BTC and USEPU is always higher than $\chi^2(m)$ with a distinct gap. Thus, the pattern of Figure 3 preliminarily demonstrates a long-range cross-correlation between BTC series and USEPU series.

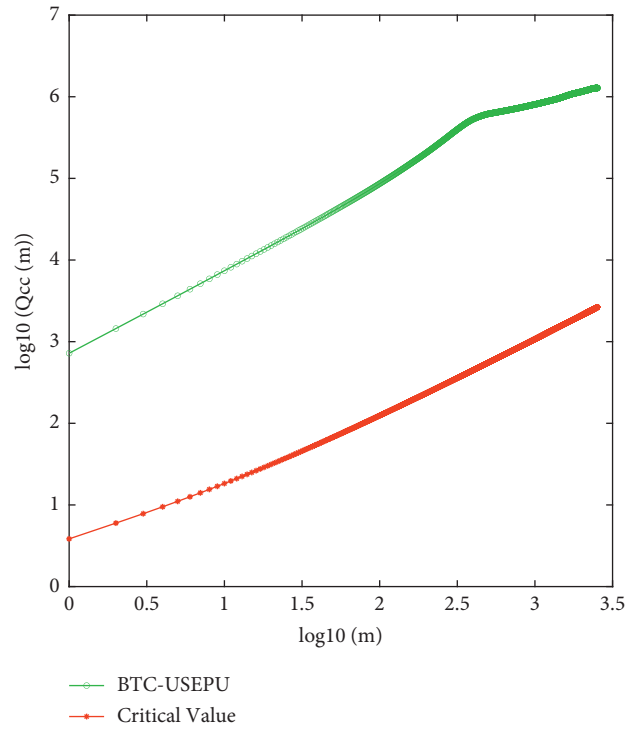


Figure 3: Log–log plot of $Q_{cc}(m)$ for BTC and USEPU.

4.2. Multifractal Detrended Cross-Correlation Analysis. So far, our study has proven the long-range cross-correlation between BTC and USEPU through the cross-correlation test. Next, we utilize the MF-DCCA [52] to perform a comprehensive quantitative examination of the nonlinear characteristics. Following Zhang et al. [34], we assign values to q at intervals of 1, with a minimum value of −10 and a maximum of 10. According to the statement in Section 3.2, we can infer that if $q<0$, the time series exhibits weak volatility. In contrast, strong fluctuations are highlighted. The evolution trend of $F_q(s)$ for BTC and USEPU is plotted in Figure 4. The results show that no matter how q changes, all lines show a similar evolution pattern, moving upward as time length s gradually increases, implying a long-range power-law cross-correlation between BTC and USEPU.

Furthermore, to examine whether the cross-relationship between the BTC and USEPU series has persistent and multifractal characteristics, we further compute the Hurst exponent $H_{xy}(q)$. As shown in Figure 5, the Hurst exponents of the BTC-USEPU, BTC, and USEPU are significantly greater than 0.5, which is the critical value. In other words, the correlation within the BTC series, the correlation within USEPU series, and the cross-correlation of BTC-USEPU are evidently persistent. From the trend of the curve, we can learn that Hurst exponents for the BTC-USEPU and the correlation within BTC decrease with the increase of q. For USEPU, the overall trend shows an S-shape curve that gradually increases as q varies from −10 to −2, peaks at $q=6$, and then decreases. We can infer that the cross-correlations among the three pairs above are strictly multifractal. Specifically, when q is less than 0, the Hurst exponent of the correlation within BTC is greater than BTC-USEPU and the correlation within USEPU. That is, correlation within BTC exhibits the highest degree of durability. When $q>0$, the blue line (the correlation within USEPU) climbs abruptly, exceeding the Hurst index of BTC-USEPU. In general, the results of Figure 5 demonstrate the persistence of the cross-correlation between BTC and USEPU and provide preliminary evidence for multifractality.

To further demonstrate the existence of multifractality, we then investigate the pattern of Rényi exponents, including the cross-correlation of BTC-USEPU (red line), the correlation within the BTC series (green line), and the correlation within USEPU series (blue line). The results are presented in Figure 6. It is obvious that all three curves are distinct from typical linear shapes, confirming that BTC-USEPU cross-correlation conforms to multifractal characteristics.

Finally, we measure the degree of multifractality of the cross-correlation between BTC and USEPU using the singularity spectrum test, and the results are documented in Figure 7. The patterns indicate that all singularity spectrums are distinct from linear shapes, which is also consistent with the characteristics of multifractality. Furthermore, as shown in Table 2, we calculate the gap between minimum and maximum Hölder exponents ($\Delta\alpha$) to represent the degree of multifractality. The delta values of BTC-USEPU, USEPU, and BTC are 0.417, 0.603, and 1.008, respectively. That is, BTC shows the strongest multifractality.

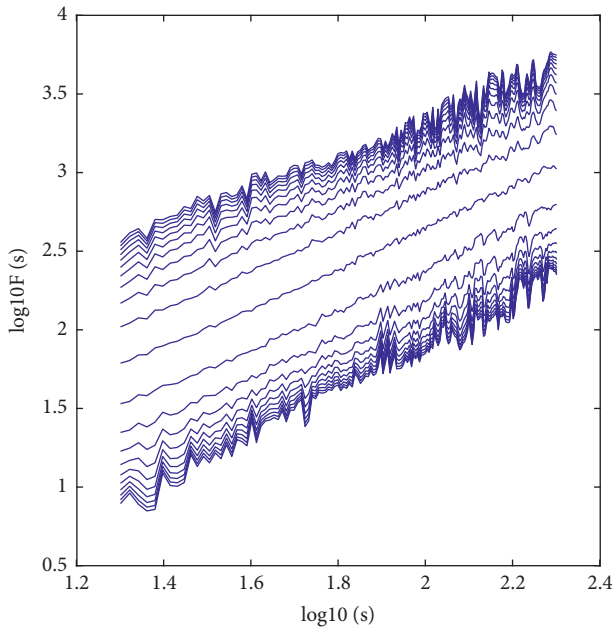


Figure 4: Log–log plot of $F_q(s)$ versus s for BTC and USEPU.

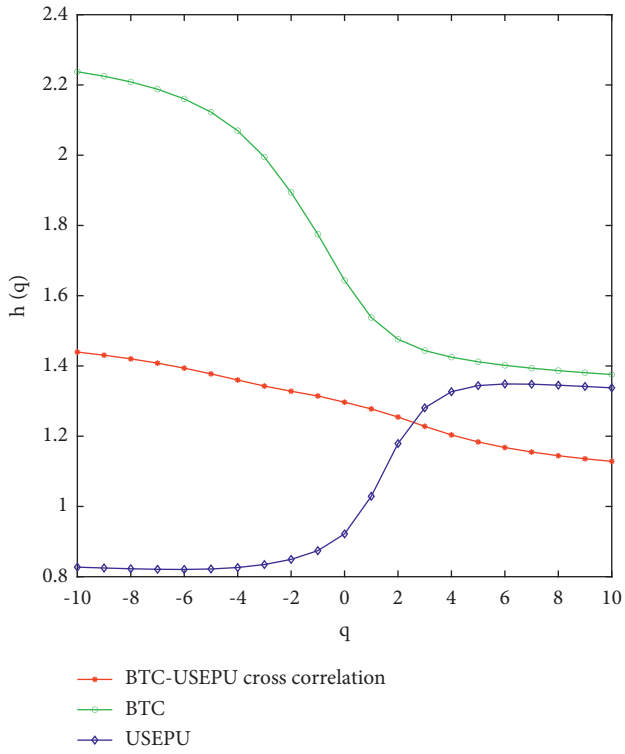


Figure 5: Generalized Hurst exponent of $h(q)$ versus q for BTC and USEPU cross-correlation, the correlation within BTC, and the correlation within USEPU.

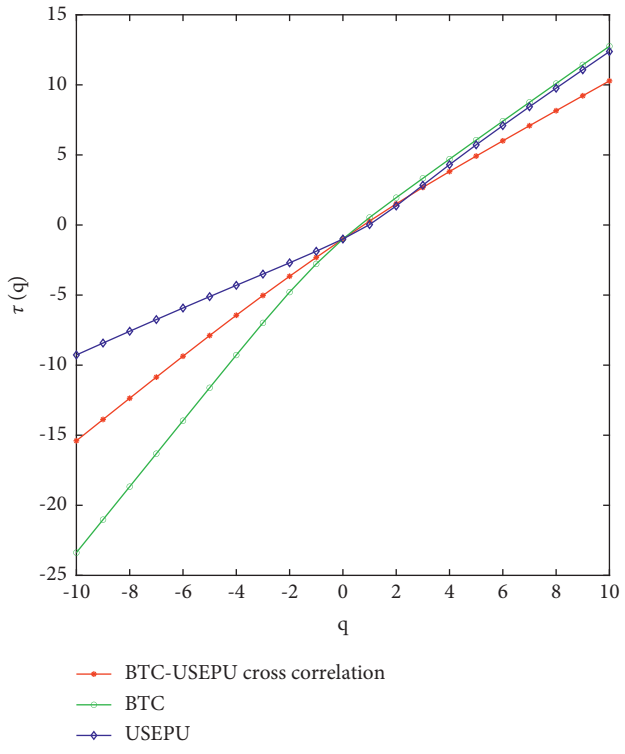


Figure 6: Rényi exponent of $\tau(q)$ versus q for BTC and USEPU cross-correlation, the correlation within BTC, and the correlation within USEPU.

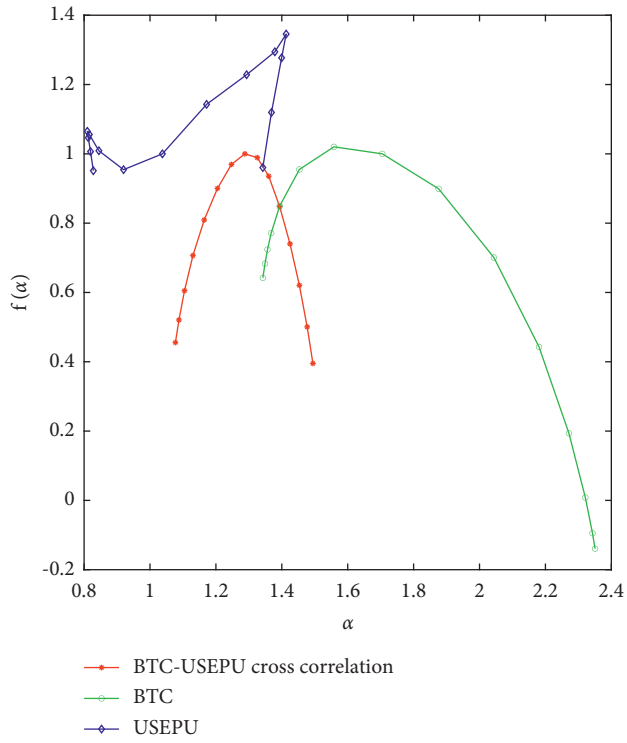


Figure 7: Spectrum singularity of f(α) versus α for BTC and USEPU cross-correlation, the correlation of BTC, and the correlation of USEPU.

Table 2: Max, min, and delta values of the Hölder exponent.

	α_{max}	α_{min}	$\Delta\alpha = \alpha min_{max}$
BTC	2.351	1.343	1.008
USEPU	1.413	0.810	0.603
BTC-USEPU cross-correlation	1.494	1.077	0.417

5. Conclusions

Our study examines the dynamic correlations between BTC and USEPU from September 2014 to August 2021. We mainly employ MF-DCCA to obtain a comprehensive insight into the long-range cross-correlation properties. First, we utilize cross-correlation analysis to preliminarily prove the long-range cross-correlation between BTC and USEPU. To further examine the nonlinear cross-correlation properties, we calculate the qth order fluctuation functions with an interval of one. All the curves demonstrate a similar evolution pattern, moving upward while time length s gradually increased, indicating the power-law properties and multifractal characteristics between BTC and USEPU. Then, we calculate the Hurst exponents to test the persistence of the cross-correlation, and the results support the existence of persistence. Moreover, we calculated another indicator of multifractality and performed a spectral singularity check to revalidate the multifractality between BTC and USEPU. Due to the unique design of Bitcoin's decentralization, the existing research on whether Bitcoin can be used as a safe-haven asset is still under discussion. Our findings inform the dynamics, long-range autocorrelation, and multifractal characteristics of BTC. Also, we confirm the dynamic cross-correlation between BTC and USEPU. Based on the multifractal analysis of BTC and USEPU, our findings can provide valuable insights into the understanding of Bitcoin price dynamics and potential suggestions to market participants on the risk management of Bitcoin.

Data Availability

We affirm that the Bitcoin price and economic policy uncertainty data used to support the findings of this study are available at https://finance.yahoo.com and https://policyuncertainty.com/ which is widely subscribed by the financial researchers, respectively.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are very grateful for comments and suggestions from the editor and reviewer. The authors acknowledge the financial support sponsored by National Natural Science Foundation of China (72101009 and 71901107) and Beijing Postdoctoral Research Activity Funding Project (Q6011011202101).

References

[1] S. Nakamoto and A. Bitcoin, "A peer-to-peer electronic cash system, bitcoin," 2008, https://bitcoin.org/bitcoin.pdf.
[2] P. Ciaian, M. Rajcaniova, and D. A. Kancs, "The economics of BitCoin price formation," *Applied Economics*, vol. 48, no. 19, pp. 1799–1815, 2016.
[3] X. Li and C. A. Wang, "The technology and economic determinants of cryptocurrency exchange rates: the case of Bitcoin," *Decision Support Systems*, vol. 95, pp. 49–60, 2017.
[4] A. Brauneis and R. Mestel, "Price discovery of cryptocurrencies: bitcoin and beyond," *Economics Letters*, vol. 165, pp. 58–61, 2018.
[5] A. Urquhart, "The inefficiency of Bitcoin," *Economics Letters*, vol. 148, pp. 80–82, 2016.
[6] S. Nadarajah and J. Chu, "On the inefficiency of Bitcoin," *Economics Letters*, vol. 150, pp. 6–9, 2017.
[7] A. K. Tiwari, R. K. Jana, D. Das, and D. Roubaud, "Informational efficiency of Bitcoin-An extension," *Economics Letters*, vol. 163, pp. 106–109, 2018.
[8] A. F. Bariviera, M. J. Basgall, W. Hasperué, and M. Naiouf, "Some stylized facts of the Bitcoin market," *Physica A: Statistical Mechanics and Its Applications*, vol. 484, pp. 82–90, 2017.
[9] W. Zhang, P. Wang, X. Li, and D. Shen, "Some stylized facts of the cryptocurrency market," *Applied Economics*, vol. 50, no. 55, pp. 5950–5965, 2018.
[10] B. M. Blau, "Price dynamics and speculative trading in bitcoin," *Research in International Business and Finance*, vol. 41, pp. 493–499, 2017.
[11] W. Zhang, P. Wang, X. Li, and D. Shen, "Multifractal detrended cross-correlation analysis of the return-volume relationship of bitcoin market," *Complexity*, vol. 2018, Article ID 8691420, 2018.

[12] A. Cretarola and G. Figà-Talamanca, "Detecting bubbles in Bitcoin price dynamics via market exuberance," *Annals of Operations Research*, vol. 299, no. 1, pp. 459–479, 2021.
[13] B. Jia, J. W. Goodell, and D. Shen, "Momentum or reversal: which is the appropriate third factor for cryptocurrencies?" *Finance Research Letters*, vol. 45, Article ID 102139, 2022.
[14] D. Shen, A. Urquhart, and P. Wang, "Bitcoin intraday time series momentum," *Financial Review*, vol. 57, pp. 319–344, 2021.
[15] S. Wu, M. Tong, Z. Yang, and A. Derbali, "Does gold or Bitcoin hedge economic policy uncertainty?" *Finance Research Letters*, vol. 31, pp. 171–178, 2019.
[16] M. Fama, A. Fumagalli, and S. Lucarelli, "Cryptocurrencies, monetary policy, and new forms of monetary sovereignty," *International Journal of Political Economy*, vol. 48, no. 2, pp. 174–194, 2019.
[17] T. V. H. Nguyen, B. T. Nguyen, K. S. Nguyen, and H. Pham, "Asymmetric monetary policy effects on cryptocurrency markets," *Research in International Business and Finance*, vol. 48, pp. 335–339, 2019.
[18] S. Corbet, C. Larkin, B. Lucey, A. Meegan, and L. Yarovaya, "Cryptocurrency reaction to fomc announcements: evidence of heterogeneity based on blockchain stack position," *Journal of Financial Stability*, vol. 46, Article ID 100706, 2020.
[19] C. Chen, L. Liu, and N. Zhao, "Fear sentiment, uncertainty, and bitcoin price dynamics: the case of COVID-19," *Emerging Markets Finance and Trade*, vol. 56, no. 10, pp. 2298–2309, 2020.
[20] E. Bouri, P. Molnár, G. Azzi, D. Roubaud, and L. I. Hagfors, "On the hedge and safe haven properties of Bitcoin: is it really more than a diversifier?" *Finance Research Letters*, vol. 20, pp. 192–198, 2017.
[21] E. Demir, G. Gozgor, C. K. M. Lau, and S. A. Vigne, "Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation," *Finance Research Letters*, vol. 26, pp. 145–149, 2018.
[22] G. J. Wang, C. Xie, D. Wen, and L. Zhao, "When Bitcoin meets economic policy uncertainty (EPU): measuring risk spillover effect from EPU to Bitcoin," *Finance Research Letters*, vol. 31, 2019.
[23] T. Conlon and R. McGee, "Safe haven or risky hazard? Bitcoin during the COVID-19 bear market," *Finance Research Letters*, vol. 35, Article ID 101607, 2020.
[24] F. Ma, Y. Wei, and D. Huang, "Multifractal detrended cross-correlation analysis between the Chinese stock market and surrounding stock markets," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 7, pp. 1659–1670, 2013.
[25] X. Zhuang, Y. Wei, and B. Zhang, "Multifractal detrended cross-correlation analysis of carbon and crude oil markets," *Physica A: Statistical Mechanics and Its Applications*, vol. 399, pp. 113–125, 2014.
[26] M. Dai, J. Hou, J. Gao, W. Su, L. Xi, and D. Ye, "Mixed multifractal analysis of China and US stock index series," *Chaos, Solitons & Fractals*, vol. 87, pp. 268–275, 2016.
[27] W. Zhang, Y. Li, Z. Zhang, and D. Shen, "The dynamic cross-correlations between foreign news, local news and stock returns," *Physica A: Statistical Mechanics and Its Applications*, vol. 509, pp. 861–872, 2018.
[28] X. Xiong, K. Xu, and D. Shen, "Dynamic cross-correlations between investors' attention and CSI300 index futures," *Fluctuation and Noise Letters*, vol. 18, no. 04, Article ID 1950022, 2019.
[29] S. Zou and T. Zhang, "Multifractal detrended cross-correlation analysis of the relation between price and volume in European carbon futures markets," *Physica A: Statistical Mechanics and Its Applications*, vol. 537, Article ID 122310, 2020.
[30] R. Zhao, "Economic policy uncertainty and local carbon emission trading: a multifractal analysis from US and guangdong," *Complexity*, vol. 2021, Article ID 8091394, 2021.
[31] X. Ge and A. Lin, "Multiscale multifractal detrended partial cross-correlation analysis of Chinese and American stock markets," *Chaos, Solitons & Fractals*, vol. 145, Article ID 110731, 2021.
[32] Y. Li, J. Goodell, and D. Shen, "Does happiness forecast implied volatility? Evidence from nonparametric wave-based Granger causality testing," *The Quarterly Review of Economics and Finance*, vol. 81, pp. 113–122, 2021.
[33] Z. Liu, Y. Ye, F. Ma, and J. Liu, "Can economic policy uncertainty help to forecast the volatility: a multifractal perspective," *Physica A: Statistical Mechanics and its Applications*, vol. 482, pp. 181–188, 2017.
[34] Z. Zhang, Y. Zhang, D. Shen, and W. Zhang, "The cross-correlations between online sentiment proxies: evidence from Google Trends and Twitter," *Physica A: Statistical Mechanics and Its Applications*, vol. 508, pp. 67–75, 2018.
[35] Y. Li, J. W. Goodell, and D. Shen, "Comparing search-engine and social-media attentions in finance research: evidence from cryptocurrencies," *International Review of Economics & Finance*, vol. 75, pp. 723–746, 2021.
[36] G. Cao, L. Xu, and J. Cao, "Multifractal detrended cross-correlations between the Chinese exchange market and stock market," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 20, pp. 4855–4866, 2012.
[37] H.-Y. Wang and T.-T. Wang, "Multifractal analysis of the Chinese stock, bond and fund markets," *Physica A: Statistical Mechanics and Its Applications*, vol. 512, pp. 280–292, 2018.
[38] R. Zhao and Y. Cui, "Dynamic cross-correlations analysis on economic policy uncertainty and US dollar exchange rate: AMF-DCCA perspective," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 6668912, 2021.
[39] G. Gajardo, W. D. Kristjanpoller, and M. Minutolo, "Does Bitcoin exhibit the same asymmetric multifractal cross-correlations with crude oil, gold and DJIA as the Euro, Great British Pound and Yen?" *Chaos, Solitons & Fractals*, vol. 109, pp. 195–205, 2018.
[40] M. A. Naeem, S. Farid, R. Ferrer, and S. J. H. Shahzad, "Comparative efficiency of green and conventional bonds pre- and during COVID-19: an asymmetric multifractal detrended fluctuation analysis," *Energy Policy*, vol. 153, Article ID 112285, 2021.
[41] G. Wang, Y. Tang, C. Xie, and S. Chen, "Is bitcoin a safe haven or a hedging asset? Evidence from China," *Journal of Management Science and Engineering*, vol. 4, no. 3, pp. 173–188, 2019.
[42] Ľ. Pástor and P. Veronesi, "Political uncertainty and risk premia," *Journal of Financial Economics*, vol. 110, no. 3, pp. 520–545, 2013.
[43] J. Brogaard and A. Detzel, "The asset-pricing implications of government economic policy uncertainty," *Management Science*, vol. 61, no. 1, pp. 3–18, 2015.
[44] W. Kang, K. Lee, and R. A. Ratti, "Economic policy uncertainty and firm-level investment," *Journal of Macroeconomics*, vol. 39, pp. 42–53, 2014.
[45] H. Gulen and M. Ion, "Policy uncertainty and corporate investment," *Review of Financial Studies*, vol. 29, no. 3, pp. 523–564, 2016.

[46] D. Zhang, L. Lei, Q. Ji, and A. M. Kutan, "Economic policy uncertainty in the US and China and their impact on the global markets," *Economic Modelling*, vol. 79, pp. 47–56, 2019.
[47] S. R. Baker, N. Bloom, and S. J. Davis, "Measuring economic policy uncertainty," *Quarterly Journal of Economics*, vol. 131, no. 4, pp. 1593–1636, 2016.
[48] P. Wang, W. Zhang, X. Li, and D. Shen, "Is cryptocurrency a hedge or a safe haven for international indices? A comprehensive and dynamic perspective," *Finance Research Letters*, vol. 31, pp. 1–18, 2019.
[49] L. Fang, E. Bouri, R. Gupta, and D. Roubaud, "Does global economic uncertainty matter for the volatility and hedging effectiveness of Bitcoin?" *International Review of Financial Analysis*, vol. 61, pp. 29–36, 2019.
[50] P. Wang, X. Li, D. Shen, and W. Zhang, "How does economic policy uncertainty affect the bitcoin market?" *Research in International Business and Finance*, vol. 53, Article ID 101234, 2020.
[51] T. L. D. Huynh, M. Wang, and V. X. Vo, "Economic policy uncertainty and the Bitcoin market: an investigation in the COVID-19 pandemic with transfer entropy," *Singapore Economic Review*, vol. 66, pp. 1–27, 2021.
[52] W. X. Zhou, "Multifractal detrended cross-correlation analysis for two nonstationary signals," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 77, no. 6, Article ID 066211, 2008.
[53] G. Cao, Y. Han, Q. Li, and W. Xu, "Asymmetric MF-DCCA method based on risk conduction and its application in the Chinese and foreign stock markets," *Physica A: Statistical Mechanics and its Applications*, vol. 468, pp. 119–130, 2017.
[54] S. Li, X. Lu, and X. Liu, "Dynamic relationship between Chinese RMB exchange rate index and market anxiety: a new perspective based on MF-DCCA," *Physica A: Statistical Mechanics and its Applications*, vol. 541, Article ID 123405, 2020.
[55] C. Z. Yao, C. Liu, and W. J. Ju, "Multifractal analysis of the WTI crude oil market, US stock market and EPU," *Physica A: Statistical Mechanics and Its Applications*, vol. 550, Article ID 124096, 2020.
[56] S. Papadamou, N. A. Kyriazis, and P. G. Tzeremes, "Nonlinear causal linkages of EPU and gold with major cryptocurrencies during bull and bear markets," *The North American Journal of Economics and Finance*, vol. 56, Article ID 101343, 2021.
[57] B. Podobnik and H. E. Stanley, "Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series," *Physical Review Letters*, vol. 100, no. 8, Article ID 084102, 2008.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 6867404, 16 pages
https://doi.org/10.1155/2022/6867404

Research Article

An Evolution Model of Fintech Supports Technology-Based SMEs Innovation and Growth

Lei Zhuang, Han Zhou, and Tingqiang Chen

School of Economics and Management, Nanjing Tech University, Nanjing 211816, China

Correspondence should be addressed to Tingqiang Chen; tingqiang88888888@163.com

Received 18 January 2022; Accepted 22 February 2022; Published 18 April 2022

Academic Editor: Dehua Shen

Copyright © 2022 Lei Zhuang 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 survival and growth of technology-based small and medium-sized enterprises are faced with a complex and changeable market environment. Introducing the role of fintech, the paper builds an evolutionary model for growth of technology-based SMEs and analyzes fintech incentive mechanisms to guide corporate growth strategies based on some cases by the simulation. The research results show that the evolutionary game strategy that SMEs are willing to adopt is {cooperation, cooperation}. And, under the incentive mechanism of fintech, the evolutionary game strategy adopted by SMEs and fintech institutions is {cooperation, cooperation, participation}. In the growth of fintech institutions serving technology-based SMEs, the selection strategy of SMEs is more sensitive to changes in fintech institutions' willingness to participate in cooperation, cooperation costs, and cooperation benefits. The selection strategy of fintech institutions is more affected by changes in cooperation costs and incentive funds during the process of participating in cooperation.

1. Introduction

Technology-based small and medium-sized enterprises (SMEs) play an important role in economic development such as technological innovation, social employment and international exchanges in the economic development [1]. High-tech small and medium-sized enterprises are economic entities that take scientific research personnel as the main body and use innovation and technical knowledge as means to engage in the research and development, production, sales of high-tech products, and to realize the industrialization of scientific and technological achievements [2]. It shows that the sharing rates of technology-based SMEs in China's patent inventions, enterprise technological innovations and new product development are about 65%, 75% and 80%. However, as the number of high-tech SMEs has increased year by year and the level of technology has improved, the homogeneity of products and services has become more common, which has brought huge growth and survival pressure to high-tech SMEs, and the survival rate of SMEs is more low, and the ability to survive and develop is low [3]. They will suffer from financing problems in the process of growing. In addition, with the spread of COVID-19 epidemic and the disruption of the global supply chain, financial institutions have difficulty in lending to technology-based SMEs in consideration of factors such as risks and costs, and are unwilling to provide more services [4]. Many technology-based SMEs have lost contact with the financial institution industry and are facing suspension and bankruptcy at any time. Therefore, it is of great significance for how to promote the growth of technology-based SMEs to develop sustainably in the age of fintech [5].

Technology-based small and medium-sized enterprises are in a disadvantaged position in the market. There are the following points from some literature and practices: first, business growth. It illustrates the vulnerability of SMEs due to unsustainable business growth. Technology-based SMEs are knowledge-intensive industries, which technical talents are the source of technological innovation, including entrepreneurs and employees [6]. Many technology-based SMEs lack a large number of high-quality entrepreneurs, who have the strategic thinking, management decision-making abilities and management knowledge such as corporate culture, finance systems, and team cultivation [7]. In addition, in the face of large technology-based companies such as Google, Ali, Tencent, and other technology giants,

technology-based SMEs are small in scale and weak in market competitiveness, and there is a possibility of being annexed [8]. Second, financing constraints. Funds are an indispensable important basis for SMEs in carrying out technological innovation activities. With the improvement of technological innovation level of SMEs, the demand of funding is also expanding [9]. Compared with large-scale enterprises, SMEs have lower possibility of lending by financial institutions because they are small in scale, and lack of tangible property for valuation [10]. On the other hand, due to the loan business of SMEs is mostly multifrequency and low-amount, financial institutions need to spend and bear higher business costs and risks [11]. Therefore, most financial institutions are unwilling to provide financial services for SMEs. Third, policies and regulations: large technology companies have absolute advantages in terms of technology, capital and other resources, production costs, market competition, etc. [12]. Some oligarchs try to abuse their dominant market positions to monopolize, causing many technology-based SMEs to be eliminated. Compared with other countries such as Japan and Germany, the antimonopoly law-related penalties and regulations have not been fully effective in China [13], and it is difficult to create a free and fair market environment to protect the healthy development of technology-based SMEs.

In recent years, attention has been paid to the growth and innovation of SMEs. Shortage of funds, insufficient technical talents, and imperfect management systems have become the obstacles to development of SMEs [14]. Considering loan costs and risks of financial institutions, commercial banks more likely to refuse to lend because of the low credit level of SMEs, which includes many Internal factors [15, 16]. Enterprise agglomeration and collaboration are important factors for enterprise growth. Interrelated companies conduct cooperative research and development to enhance their core competitiveness, improve their own innovation level, and form a new supply chain value network [17]. In the network, many companies have begun to spontaneously adopt R&D cooperation methods to carry out knowledge diffusion and technology sharing within the clusters to enhance their own internal innovation capabilities [18]. Cluster enterprises choose the corresponding technological innovation model after considering factors such as the size of the enterprise, financial strength, and technical level [19]. For example, Silicon Valley has formed high-tech industrial clusters with small business clusters and unique geographical advantages. It has nurtured technology giants such as Intel, Apple, and Google and has promoted the United States to become a technology powerhouse. Through the exchange and interaction of resources such as technology, knowledge, and information [2], SMEs can obtain more cooperation opportunities and synergistic benefits in the cluster [20], which also enhance their competitive advantages and further enhance their development [21]. Government policy support is an important external factor influencing the growth of enterprises [22], policy supply, implementation, and support intensity and policy changes affect the success or failure of the cultivation of high-tech SMEs [23].

Facing the digital economy era and the innovation of fintech, some scholars have shown that it can be effectively promote the development of high-tech SMEs by exploring new financing models such as Internet finance, investment and loan linkages, and innovative financing models of banks [24]. Some large commercial banks have already taken actions to provide online financial business models based on big data, blockchain and other technologies to provide more convenient online credit financial services for SMEs [25, 26]. From the perspective of fintech innovation, technological finance mainly refers to the integration of technological industry and financial industry, which finance serving technological enterprises [27]. It is a series of system and mechanism arrangements that realize the close integration of technology and finance [28]. The continuous promotion and integration of technology and finance can improve the regional technological and economic strength [29]. After the empirical test of some scholars, it is verified that the development of technology finance not only has a significant correlation with the regional technological development level, but also affects the development of regional technological innovation capabilities [30], which can help SMEs improve the economy, ease financing constraints, and promote the high-quality development of technology-based small and medium-sized enterprises [31]. It shows that the efficiency of local technology finance policies, such as financial subsidies, financial supervision, and technology credit, has an uneven distribution of impacts on the resource allocation efficiency of technology-based small and medium-sized enterprises, especially the resource allocation of enterprises in the start-up period and transformation period [32]. However, financial development has not provided strong support for high-tech SMEs, and the corresponding financial mechanism is imperfect. It is difficult to solve the funding problems needed for the growth and development of high-tech SMEs under financial innovation [33].

Most research results and practical theories for technology-based SMEs are based on their own growth and external financing channels, which attempt to help technology-based SMEs develop by improving financing channels, management systems and government policy [34]. This paper builds an evolutionary game model for innovation and growth among technology-based SMEs and introduces fintech institutions mechanism to explore the final strategy adopted by technology-based SMEs. The remaining part of this paper proceeds as follows. Section2 spontaneous model of technology-based SMEs innovation and growth. Section 3 evolutionary model of fintech boost technology-based SMEs growth. Section 4 case and simulation analysis of fintech supports technology-based SMEs growth. Section 5 summarizes and concludes the study.

2. Spontaneous Model of Technology-Based SMEs Innovation and Growth

Technology-based SMEs take knowledge and technological innovation as their core competitiveness, and provide the society with knowledge, technology, and high-tech products to obtain large amounts of funds to achieve sustainable

development of enterprises [4]. For technology-based SMEs, insufficient resources are restricting the innovation and growth of enterprises. Each technology-based enterprise has independent core technology and resources, and the entire process of knowledge innovation and technology productization requires a variety of technical knowledge, technological talents and innovative resources to succeed. Compared with large-scale enterprises, it is difficult to retain high-knowledge and innovative talents to serve technology-based SMEs. However, cooperative technological innovation has become a new development model in the high-tech field, bringing more resource sharing for SMEs, thereby enabling SMEs to better carry out innovative activities [35]. In the network of technology-based enterprises clusters, some companies with mutual individuality can share knowledge, relationship transfer, and mutual learning of management systems [36], reduce the friction and obstacles in the process of enterprise development, and promote the further development of each other. Based on the above content, the following hypotheses are proposed.

Hypothesis 1. There are technology-based SME A and B, both of which are subject to the game with bounded rationality.

Hypothesis 2. In the cluster network, regardless of the external environmental impact, both technology-based SME A and B choose to cooperate or not to cooperate. Cooperation strategy is information sharing, including management system, financing channels and technical knowledge etc. Noncooperation means the company has exclusive knowledge and technology, commodity production channels, financing methods and marketing methods, and is unwilling to cooperate with the other company, to find ways of enterprise development on their own.

Hypothesis 3. The initial income of enterprise A and B is A_1 and A_2 respectively. When enterprise A and B have the willingness to cooperate and reach an agreement to become a partner, the additional benefits for both are ΔV, the additional income of enterprise A is $\theta\Delta V$, the additional income of enterprise B is $(1-\theta)\Delta V$, Where $\theta(0<\theta<1)$ is the additional income distribution coefficient; When the cooperation breaks down, due to the needs of its own business development, one of the parties will not be able to immediately interrupt the connection between the two, and there will still be some business exchanges. Among them, the party who chooses to withdraw from the cooperation early will earn speculative gains. Enterprise A and B are respectively T_1 and T_2. When both parties have no willingness to cooperate, the net income of the two business activities is the initial income.

Hypothesis 4. Enterprise A and B need to pay a certain cooperation cost (technology sharing, human resource allocation) during the cooperation process. At this time, the total cooperation cost paid by enterprise A and B is c, the cost paid by enterprise A is tc, the cost paid by enterprise B is $(1-t)c$, $t(0<t<1)$ is the cost-sharing ratio coefficient. When the partnership breaks down, part of the business is still in contact, and the party who chooses to continue needs to pay additional loss costs, Enterprise A and B are respectively L_1 and L_2, and the exiting party needs to bear default costs W as a certain penalty.

The proportion of enterprise A choosing cooperation strategy is x $(0\leq x\leq 1)$, then the proportion of choosing noncooperative strategy is $(1\text{-}x)$; the proportion of enterprise B choosing cooperation strategy is y $(0\leq y\leq 1)$, then the proportion of choosing noncooperative strategy For $(1\text{-}y)$. The income matrix of the game between enterprises A and B is shown in Table 1 below:

At the initial stage, due to the influence of their internal environmental factors, the selected game strategy of technology-based SME A and B is not necessarily optimal. Instead, they constantly adjust their own strategies according to the process of growth and cultivation of the enterprise and the passage of time, which they can obtain strategies for both sides to stabilize. According to the above hypothesis and Table 1, calculate the income of the technology-based SME A to choose the information cooperation strategy U_A^1, the income of the noncooperation strategy U_A^2 and the average income $\overline{U}_A$.

Benefits of enterprise A's choice of cooperation strategy:

$$U_A^1 = y(A_1 + \theta\Delta V - tc) + (1-y)(A_1 + W - tc - L_1). \quad (1)$$

Benefits of noncooperative strategy of enterprise A:

$$U_A^2 = y(A_1 + T_1 - W) + (1-y)A_1. \quad (2)$$

Average income of enterprise A:

$$\overline{U}_A = xU_A^1 + (1-x)U_A^2. \quad (3)$$

From this, we can obtain the dynamic equation for replication of enterprise A:

$$\frac{dx}{dt} = x(1-x)[y(\theta\Delta V - T_1 + L_1) + W - tc - L_1]. \quad (4)$$

In the same way, the dynamic equation for replication of enterprise B can be obtained:

$$\frac{dy}{dt} = y(1-y)\{x[(1-\theta)\Delta V - T_2 + L_2] + W - (1-t)c - L_2\}. \quad (5)$$

According toequations (1) and (2), the two-dimensional power systems of technology-based SME A and B are:

$$\begin{cases} F_x = x(1-x)[y(\theta\Delta V - T_1 + L_1) + W - tc - L_1], \\ F_y = y(1-y)\{x[(1-\theta)\Delta V - T_2 + L_2] + W - (1-t)c - L_2\}. \end{cases} \quad (6)$$

The solution can get the local stable equilibrium points of the system, which are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(1-tc+L_2-W/(1-\theta)\Delta V - T_2 + L_2$, and $tc + L_1 - W/\theta\Delta V - T_1 + L_1)$.

According to Friedman's method of constructing the Jacobian matrix, the first-order partial derivatives of x and y of the copied dynamic equation are obtained, respectively. The Jacobian matrix is as follows:

Table 1: The income matrix of the game between technology-based SME A and B.

		Enterprise B	
		Cooperation	Noncooperation
Enterprise A	Cooperation	$A_1 + \theta\Delta V - tc, A_2 + (1-\theta)\Delta V - (1-t)c$	$A_1 + W - tc - L_1$, $A_2 + T_2 - W$
	Noncooperation	$A_1 + T_1 - W, A_2 + W - (1-t)c - L_2$	A_1 , A_2

$$J = \begin{bmatrix} (1-2x)\left[y\left(\theta\Delta V - T_1 + L_1\right) + W - tc - L_1\right], & x(1-x)\left(\theta\Delta V - T_1 + L_1\right), \\ y(1-y)\left[(1-\theta)\Delta V - T_2 + L_2\right], & (1-2y)\left\{x\left[(1-\theta)\Delta V - T_2 + L_2\right] + W - (1-t)c - L_2\right\} \end{bmatrix}. \quad (7)$$

And calculate the eigenvalues of the Jacobian matrix and the signs of the trace to judge the local stability in various situations, and obtain the evolutionary stability strategy [37], as shown in Table 2, the equilibrium points (0, 0) and (1, 1) is the local stable point, the saddle point is $\left(1 - tc + L_2 - W/(1-\theta)\Delta V - T_2 + L_2\right.$, $\left. tc + L_1 - W/\theta\Delta V - T_1 + L_1\right)$, The evolutionary stable strategy of the system is {cooperation, cooperation} or {no cooperation, no cooperation}.

It can be known that in the game process of technology-based SME A and B, the final strategy is affected by the saddle point. The straight line formed by the points (0,0), (1,0) and the saddle point determines the upper and lower squares. For part of the area, the more the coordinate value corresponding to the saddle point converges to (0, 0), the larger the area of the upper area, and the more likely it is for the SME A and B to choose a cooperation strategy. On the contrary, the SME A and B are more likely to choose a noncooperative strategy. In the game model of technology-based SME A and B, the position of the saddle point is affected by the income distribution coefficient, the cooperation cost sharing coefficient between the two parties, the cooperation income, the loss cost, the liquidated damages and the speculative income. Some scholars have shown that information technology sharing can promote collaborative innovation among technology-based SMEs. When the initial parameters reach a stable state, the cycle of collaborative innovation alliances among technology-based SMEs will become longer, which will boost the growth of technology-based SMEs [38].

3. Evolutionary Model of Fintech Boost Technology-Based SMEs Growth

Finance is an important force supporting the innovation and growth of technology-based SMEs. The development of technology-based SMEs not only requires internal technical support, but also requires a large amount of external financial support. In the process of research and development of new technology products for these enterprises, financing difficulties are the bottleneck restricting the development of enterprises. In the wave of fintech innovation, many financial institutions cooperate with technology-based SMEs, and use new technological tools to improve their business development and pay corresponding remunerations or give loan discounts. From the perspective of financing needs, financial institutions can solve the problem of insufficient funds for technology-based SMEs at different growth stages. Fintech institutions and technology-based enterprises can upgrade their business models through technological cooperation and fintech model in the digital age [39]. Based on the above content, this article adds the third-party game subject fintech institution C to the abovementioned technology-based SME A and B game model, and proposes the following hypotheses:

Hypothesis 5. Fintech institutions mainly meet the financing needs of technology-based SMEs in various stages of innovation and growth, and obtain technological tools for SMEs to improve their own business development. The collection of selection strategies of fintech institutions is {participate, not participate}. The three parties participating in the game are all bounded rationally, and the optimal strategy is selected through multiple games.

Hypothesis 6. When fintech institutions participate in the development of technology-based enterprises, the preferential policies provided by fintech institutions can reduce the total cost invested by enterprises A and B in the process of cooperation. The reduction is s, the interenterprise cost allocation coefficient remains unchanged. SME A and B can obtain cooperation benefits as R with the help of financial institutions, the distribution coefficient of cooperation income is $a\,(0 < a < 1)$, the income of enterprise A and B are aR, $(1-a)R$ respectively. Fintech institutions choose the strategy of "participating" in the development of technology-based enterprises, and their gains are represented by A_3.The fintech institution C chooses the strategy of "no participating," and the benefits it obtains are bA_3, $b(0 < b < 1)$ is the ratio of the income obtained by the fintech institution choosing the strategy of "no participating" to the income obtained by the fintech institution choosing the strategy of "participating."

Hypothesis 7. In participating in the collaborative innovation and growth of technology-based SMEs, fintech institutions must invest corresponding human and financial resources. They also need to supervise the business activities of technology-based SMEs. The total cost incurred is K_1. Fintech institutions give certain incentives to technology-based SMEs that actively participate in cooperation as K_2.

Table 2: Evolution and stability analysis of system equilibrium.

	det$J$	tr$J$	Stability
(0,0)	+	−	ESS
(0,1)	+	+	Unstable
(1,0)	+	+	Unstable
(1,1)	+	−	ESS
$(1-tc+L_2-W/(1-\theta)\Delta V-T_2+L_2\ ,\ \ tc+L_1-W/\theta\Delta V-T_1+L_1)$	−		Saddle point

The proportion of fintech institution C choosing to participate in the strategy is z $(0 \le z \le 1)$, then the proportion of choosing the noncooperative strategy is $(1-z)$. And according to the above 7 assumptions, the payment matrix of the game between fintech institutions and technology-based SMEs is shown in Tables 3 and 4.

According to Tables 3 and 4, the expected benefits of technology-based SME A when adopting the strategy of "cooperation" U_A and the expected benefits when adopting the strategy "noncooperation" U_{1-A} and average return $\overline{U}_A$, respectively:

$$\begin{aligned} U_A &= yz\left[A_1+\theta\Delta V-t(c-s)+K_2+aR\right]+(1-y)z\left[A_1-t(c-s)+K_2+aR+W-L_1\right] \\ &\quad + y(1-z)\left(A_1+\theta\Delta V-tc\right)+(1-y)(1-z)\left(A_1+W-tc-L_1\right), \\ U_{1-A} &= yz\left(A_1+T_1-W\right)+(1-y)zA_1+y(1-z)\left(A_1+T_1-W\right)+(1-y)(1-z)A_1, \\ \overline{U}_A &= xU_A+(1-x)U_{1-A}. \end{aligned} \tag{8}$$

The expected benefits and average benefits of technology-based SME B when adopting the strategy of "cooperation" U_B and the expected return U_{1-B} and average return $\overline{U}_B$ when adopting the strategy of "noncooperation" are as follows:

$$\begin{aligned} U_B &= xz\left[A_2+(1-\theta)\Delta V-(1-t)(c-s)+K_2+(1-a)R\right]+(1-x)z\left[A_2-(1-t)(c-s)+K_2+(1-a)R+W-L_2\right] \\ &\quad + x(1-z)\left[A_2+(1-\theta)\Delta V-(1-t)c\right]+(1-x)(1-z)\left[A_2+W-(1-t)c-L_2\right], \\ U_{1-B} &= xz\left(A_2+T_2-W\right)+(1-x)zA_2+x(1-z)\left(A_2+T_2-W\right)+(1-x)(1-z)A_2, \\ \overline{U}_B &= yU_B+(1-y)U_{1-B}. \end{aligned} \tag{9}$$

The expected return of fintech institution C U_C when adopting the strategy of "participation" and the expected return U_{1-C} and average return $\overline{U}_C$ when adopting the strategy of "nonparticipation" are as follows:

$$\begin{aligned} U_C &= xy\left(A_3-K_1-2K_2\right)+x(1-y)\left(A_3-K_1-K_2\right)+x(1-y)\left(A_3-K_1-K_2\right)+(1-x)(1-y)\left(A_3-K_1\right), \\ U_{1-C} &= xybA_3+x(1-y)bA_3+x(1-y)bA_3+(1-x)(1-y)bA_3, \\ \overline{U}_C &= zU_C+(1-z)U_{1-C}. \end{aligned} \tag{10}$$

It can be obtained from this that the dynamic equation for replication of technology-based SME A is as follows:

$$F_{(x)}=\frac{\mathrm{d}x}{\mathrm{d}t}=x(1-x)\left[z\left(aR+K_2+ts\right)y\left(\theta\Delta V-T_1+L_1\right)+W-tc-L_1\right]. \tag{11}$$

The dynamic equation for replication of technology-based SME B is as follows:

Table 3: Payment matrix of fintech institutions participating in the three-party game.

Fintech C		Enterprise B	
		Cooperation	Noncooperation
Enterprise A	Cooperation	$A_1 + \theta\Delta V - tc - s + K_2 + aR$, $A_2 + 1 - \theta\Delta V - (1-t)(c-s) + K_2 + (1-a)R$, $A_3 - K_1 - 2K_2$	$A_1 - tc - s - L_1 + K_2 + aR + W$, $A_2 + T_2 - W$, $A_3 - K_1 - K_2$
	Noncooperation	$A_1 + T_1 - W$, $A_2 - (1-t)(c-s) - L_2 + K_2 + (1-a)R + W$, $A_3 - K_1 - K_2$	A_1, A_2, $A_3 - K_1$

Table 4: Payment matrix of financial institutions nonparticipating in the three-party game.

Fintech C		Enterprise B	
		Cooperation	Noncooperation
Enterprise A	Cooperation	$A_1 + \theta\Delta V - tc$, $A_2 + 1 - \theta\Delta V - (1-t)c$, bA_3	$A_1 + W - tc - L_1$, $A_2 + T_2 - W$, bA_3
	Noncooperation	$A_1 + T_1 - W$, $A_2 + W - (1-t)c - L_2$, bA_3	A_1, A_2, bA_3

$$F_{(y)} = \frac{dy}{dt} = y(1-y)\{z[(1-a)R + K_2 + (1-t)s] + x[(1-\theta)\Delta V - T_2 + L_2] + W - (1-t)c - L_2\}. \quad (12)$$

The dynamic equation for replication of fintech institution C is as follows:

$$F_{(z)} = \frac{dz}{dt} = z(1-z)[(1-b)A_3 - K_1 - (x+y)K_2]. \quad (13)$$

Thus, the replication power system of technology-based SMEs A, B and fintech institution C is established as

$$\begin{cases} F_{(x)} = x(1-x)[z(aR + K_2 + ts) + y(\theta\Delta V - T_1 + L_1) + W - tc - L_1], \\ F_{(y)} = y(1-y)\{z[(1-a)R + K_2 + (1-t)s] + x[(1-\theta)\Delta V - T_2 + L_2] + W - (1-t)c - L_2\}, \\ F_{(z)} = z(1-z)[(1-b)A_3 - K_1 - (x+y)K_2]. \end{cases} \quad (14)$$

Proposition 1. *The equilibrium points of the dynamic evolutionary game of technology-based SME A, B and fintech institution C in the same equilibrium state are* $E_1 = (0,0,0)$, $E_2 = (0,0,0)$, $E_3 = (0,0,1)$, $E_4 = (0,1,0)$, $E_5 = (0,1,1)$, $E_6 = (1,0,0)$, $E_7 = (1,1,0)$, $E_8 = (1,1,1)$, *respectively.*

Proof 1. Separate orders $dx/dt = 0$, $dy/dt = 0$, $dz/dt = 0$ for the replication dynamic system of technology-based SME A, B and fintech institution C. It can be concluded that $E_1 = (0,0,0)$, $E_2 = (0,0,1)$, $E_3 = (0,1,0)$, $E_4 = (0,1,1)$, $E_5 = (1,0,0)$, $E_6 = (1,0,1)$, $E_7 = (1,1,0)$, $E_8 = (1,1,1)$ are the balance points of the system, and the proof is completed.

According to Friedman's method of constructing the Jacobian matrix, the first-order partial derivatives of x, y, and z of the dynamic equation are respectively calculated to determine the evolution and stability strategy of the system. The corresponding Jacobian matrix is as follows:

$$J = \begin{bmatrix} a_1 & x(1-x)(L_1 + \theta\Delta V - T_1) & x(1-x)(R_1 + K_2 + ts) \\ y(1-y)[L_2 + (1-\theta)\Delta V - T_2] & a_2 & y(1-y)[K_2 + R_2 + (1-t)s] \\ -z(1-z)K_2 & -z(1-z)K_2 & a_3 \end{bmatrix}, \quad (15)$$

where $a_1 = (1-2x)[z(aR + K_2 + ts)y \quad (\theta\Delta V - T_1 + L_1) + W - tc - L_1]$,

$$a_2 = (1-2y)\{z[(1-a)R + K_2 + (1-t)s] + x[(1-\theta)\Delta V - T_2 + L_2] + W - (1-t)c - L_2\} \quad (16)$$
$$a_3 = (1-2z)[(1-b)A_3 - K_1 - (x+y)K_2].$$

Bring the above eight equilibrium points into the Jacobian matrix, and the eigenvalues obtained by the solution are shown in Table 5.

In order to facilitate the analysis of the signs of the eigenvalues corresponding to each equilibrium point, without loss of generality, assuming the participation of financial institutions, the net income of three partner SME A, B and fintech institution C is greater than the net income of the three parties' independent development. i.e. $W - tc - T_1 > 0$, $W - (1-t)c - T_2 > 0$, $(1-b)A_3 - K_1 - 2K_2 > 0$. Considering the changes in parameters such as costs, cooperation benefits, and incentives paid by various participating entities when fintech institutions services technology-based SMEs, this article puts forward the following three propositions for discussion. □

Proposition 2. *When $K_2 + aR + W - t(c-s) - L_1 < 0$ and $K_2 + (1-a)R + W - (1-t)(c-s) - L_2 < 0$, the sum of the rewards and cooperation benefits of the fintech institution that the technology-based SME A adopts a cooperation strategy is less than the cooperation cost paid by the technology-based SME A with the participation of fintech institution C and the additional payment paid by the counterparty's midway exit. The sum of the cost of loss, and the sum of rewards and cooperation gains from fintech institutions that technology-based SME B adopts a cooperation strategy is less than the cooperation cost paid by SME B with the participation of fintech institutions and the additional loss cost paid by the counterparty's midway exit sum. At this time, the equilibrium point of the system is $E_2(0,0,1)$ and $E_8(1,1,1)$, then the evolution strategy of the system is {no cooperation, no cooperation, participation} and {cooperation, cooperation, participation}.*

Proposition 3. *When $W - tc - L_1 > 0$ or $W - (1-t)c - L_2 > 0$, the scientific and technology-based SME A adopts a cooperative strategy to obtain fintech institution rewards greater than the sum of the cost paid by SME A when the fintech institution C does not participate and the additional loss cost paid by the counterparty's midway exit. Or the SME B adopts a cooperative strategy to obtain fintech institution rewards greater than the sum of the cost paid by the SME B when the fintech institution C does not participate and the additional loss cost paid by the counterparty's enterprise withdrawing midway. At this time, the system equilibrium point is $E_8(1,1,1)$, then the system evolution strategy is {cooperation, cooperation, participation}.*

Proposition 4. *When $K_2 + aR + W - t(c-s) - L_1 > 0$ and $W - tc - L_1 < 0$ or $K_2 + (1-a)R + W - (1-t)(c-s) - L_2 > 0$ and $W - (1-t)c - L_2 < 0$, the sum of the rewards and cooperation benefits of the fintech institution that the technology-based SME A adopts a cooperation strategy is greater than the cooperation cost paid by the technology-based SME A with the participation of fintech institutions and the counterparty's half-way exit payment. The sum of the additional loss cost of the technology-based SME A, and the rewards of the fintech institution that the technology-based SME A adopts a cooperative strategy is less than the sum of the cost paid by the SME A when the fintech institution C does not participate and the additional loss cost paid by the counterparty's midway exit. Or technology-based SME B adopts a cooperation strategy to obtain fintech institution rewards and the sum of cooperation benefits is greater than the sum of the cooperation cost paid by the SME B with the participation of fintech institutions and the additional loss cost paid by the other party's company withdrawing midway, and SME B adopts a cooperative strategy to obtain fintech institution rewards that are less than the sum of the cost paid SME B when the fintech institution does not participate and the additional loss cost paid by the counterparty's enterprise withdrawing midway. At this time, the system equilibrium point is $E_8(1,1,1)$, then the system evolution strategy is {cooperation, cooperation, participation}.*

Proof 2. According to the Lyapunov stability theory, the asymptotic stability at the equilibrium point is judged by analyzing the eigenvalues of the Jacobian matrix. When the eigenvalues are all nonpositive, the equilibrium point is the stable point of the evolutionary game. These judges the system equilibrium point of the above three propositions, as shown in Table 6, and the proof is complete. □

4. Case and Simulation Analysis of Fintech Supports Enterprises Growth

In recent years, many governments have paid more and more attention to the development of technology-based SMEs and adopted a series of policies and actions to support the development of SMEs. In response to the innovation and financing of technology-based SMEs, more cities centered on the themes of high-tech enterprises and fintech, and encouraged banks to increase loans to technology-based enterprises through diversified fintech policies, so as to smooth the capital chain and solve problems. It is difficult for enterprises to obtain financing, so ensure sufficient funds. This article selects cities for research to understand the specific situation of technology-based SMEs. Then, we conduct case scenario simulations based on research. There are technology-based SME A, B and financial institution C in the innovation network. In order to judge the evolution path of each participant more intuitively, numerical experiments are used to analyze the influence of the selection of A and B evolution strategies of fintech institutions and SMEs and the changes of parameters on the evolution path. According to the above assumptions, the parameters in the payout matrix need to satisfy $W - tc - T_1 > 0$, $W - (1-t)c - T_2 > 0$, $(1-b)A_3 - K_1 - 2K_2 > 0$. Suppose the initial value of each parameter in the subject of the three-party game: $\Delta V = 20$, $\theta = t = a = b = 0.5$, $A_1 = 25$, $A_2 = 30$, $A_3 = 60$, $W = 30$, $R = 16$, $s = 16$, $c = 50$, $T_1 = 5$, $T_2 = 10$, $L_1 = 40$, $L_2 = 45$, $K_1 = 10$, $K_2 = 8$. In addition, the initial willingness to participate in fintech institutions and SME A and B is $x = y = z = 0.5$.

Table 5: Eigenvalues of Jacobian matrix.

Equilibrium point	λ_1	λ_2	λ_3
$E_1(0,0,0)$	$W - tc - L_1$	$W - (1-t)c - L_2$	$(1-b)A_3 - K_1$
$E_2(0,0,1)$	$K_2 + aR + W - t(c-s) - L_1$	$K_2 + (1-a)R + W - (1-t)(c-s) - L_2$	$-[(1-b)A_3 - K_1]$
$E_3(0,1,0)$	$\theta\Delta V + W - T_1 - tc$	$-[W - (1-t)c - L_2]$	$(1-b)A_3 - K_1 - K_2$
$E_4(0,1,1)$	$K_2 + aR + \theta\Delta V + W - t(c-s) - T_1$	$-[K_2 + (1-a)R + W - (1-t)(c-s) - L_2]$	$-[(1-b)A_3 - K_1 - K_2]$
$E_5(1,0,0)$	$-(W - tc - L_1)$	$(1-\theta)\Delta V - T_2 + W - (1-t)c$	$(1-b)A_3 - K_1 - K_2$
$E_6(1,0,1)$	$-[K_2 + aR + W - t(c-s) - L_1]$	$K_2 + (1-a)R + W - (1-t)(c-s) + (1-\theta)\Delta V - T_2$	$-[(1-b)A_3 - K_1 - K_2]$
$E_7(1,1,0)$	$-[\theta\Delta V - T_1 + W - tc]$	$-[(1-\theta)\Delta V - T_2 + W - (1-t)c]$	$(1-b)A_3 - K_1 - 2K_2$
$E_8(1,1,1)$	$-[K_2 + aR + W] - t(c-s) + \theta\Delta V - T_1$	$-[K_2 + (1-a)R + W - (1-t)(c-s) + (1-\theta)\Delta V - T_2]$	$-[(1-b)A_3 - K_1 - 2K_2]$

Table 6: The local stability of the equilibrium point of the system (Proposition 2, 3, 4).

Equilibrium point	Proposition 2			
	λ_1	λ_2	λ_3	Stability
$E_1(0,0,0)$	—	—	+	Unstable
$E_2(0,0,1)$	—	—	—	ESS
$E_3(0,1,0)$	+	+	+	Saddle point
$E_4(0,1,1)$	+	+	—	Unstable
$E_5(1,0,0)$	+	+	+	Saddle point
$E_6(1,0,1)$	—	+	—	Unstable
$E_7(1,1,0)$	—	—	+	Unstable
$E_8(1,1,1)$	—	—	—	ESS
Equilibrium point	Proposition 3			
	λ_1	λ_2	λ_3	Stability
$E_1(0,0,0)$	+	+	+	Saddle point
$E_2(0,0,1)$	+	+	—	Unstable
$E_3(0,1,0)$	+	—	+	Unstable
$E_4(0,1,1)$	+	—	—	Unstable
$E_5(1,0,0)$	—	+	+	Unstable
$E_6(1,0,1)$	—	+	—	Unstable
$E_7(1,1,0)$	—	—	+	Unstable
$E_8(1,1,1)$	—	—	—	ESS
Equilibrium point	Proposition 4			
	λ_1	λ_2	λ_3	Stability
$E_1(0,0,0)$	—	—	+	Unstable
$E_2(0,0,1)$	+	+	—	Unstable
$E_3(0,1,0)$	+	+	+	Saddle point
$E_4(0,1,1)$	+	—	—	Unstable
$E_5(1,0,0)$	+	+	+	Saddle point
$E_6(1,0,1)$	—	+	—	Unstable
$E_7(1,1,0)$	—	—	+	Unstable
$E_8(1,1,1)$	—	—	—	ESS

According to the above numerical settings, the dynamic changes of fintech institution C and technology-based SME A and B's different initial willingness change strategy choices are simulated. On this basis, it further explores the influence of the reduction in cooperation cost of technology-based SMEs, the distribution coefficient of cooperation benefits, the cooperation cost of financial institutions and the changes in incentive parameters on the evolution path of technology-based SMEs when fintech institutions participate in cooperation.

4.1. Scenario Analysis of the Main Body's Initial Willingness. In the growth process of technology-based SMEs, the willingness of enterprises and fintech institutions to participate in cooperation affects the choice of their development strategies. Figure 1 is a simulation analysis of the impact of changes in the initial willingness of fintech institution C and SME A and B on the adoption of strategies under the condition that other initial parameters remain unchanged. Assuming that the initial willingness of the three parties in the game is the same, that is, as shown in Figure 1, the initial willingness thresholds of fintech institutions and SME A and B are between 0.4 and 0.5.When the initial willingness of the three parties to participate in cooperation is less than the critical value of 0.4, x and y converge to 0, z converges to 1, and the final equilibrium point tends to (0, 0, 1). At this time, well-developed SME B's willingness to participate is faster than enterprise A's convergence rate; when the initial willingness of the three parties to participate

in cooperation is greater than the critical value 0.5, x, y, z converge to 1, and the final equilibrium point tends to (1, 1, 1). At this time, the larger the parameter value of the initial willingness to participate in cooperation of SME A, the faster its convergence speed; when the initial willingness of the three parties to participate in cooperation is at a moderate level, the initial willingness of financial institutions to participate in cooperation slowly rises. The willingness of SME A to participate in cooperation is also rising, while the willingness of SME B to cooperate is slowly decreasing. With the increasing willingness of fintech institution C and SME A to participate in cooperation, the willingness of SME B to cooperate continues to rise, eventually all choose to participate. When the willingness of the three parties to participate in cooperation is high, the willingness of fintech institutions and technology-based SMEs to participate in cooperation will rise directly, and finally converge to the equilibrium point (1, 1, 1). The results show that as the initial willingness of the three parties to participate in cooperation continues to increase, x and y converge to 1 faster, and z converges to 1 slower. Eventually, the three parties will tend to participate in the cooperation. In the process, when financial institutions have a strong willingness to participate in cooperation, technology-based SMEs will eventually choose to cooperate and reach long-term cooperation with financial institutions, obtain financial support, and achieve their own stable development.

Whether technology-based SMEs cooperate with fintech institutions is affected by similar technology-based SMEs. Figure 2 is a simulation analysis of the impact of changes in the initial willingness of technology-based SME A on the strategies adopted by the three parties under the condition that other parameters remain unchanged. It can be seen from Figure 2 that the initial willingness to cooperate between technology-based SME B and fintech institution C is in a neutral state, and the initial willingness of technology-based SME A is between 0.3 and 0.4. When the initial willingness value of technology-based SME A is less than 0.3, x, y all converge to 0, z converges to 1, and the final equilibrium point tends to (0, 0, 1). At this time, the technology-based SME B with a higher initial willingness to participate is affected by the willingness of enterprise A to participate, and it converges that the speed is faster; when the initial willingness value of technology-based SME A is greater than 0.4, x, y, and z all converge to 1, and the final equilibrium point tends to (1, 1, 1). At this time, the enterprise with a higher initial willingness value A's convergence to 1 is faster than that of technology-based SME B. When technology-based SME A's initial willingness to cooperate is less than enterprise B, as enterprise A's initial willingness continues to decrease, enterprise B's initial willingness declines faster, and the willingness to cooperate with fintech institutions has risen slowly. When the initial value of willingness to cooperate with technology-based SME A is greater than that of enterprise B, the willingness to cooperate with enterprise A continues to increase, and the willingness to cooperate with enterprise B first slowly decreases and then continues to rise. The willingness to cooperate with financial institutions still rises slowly, and finally converges to Equilibrium point (1, 1, 1). The results show that the willingness of technology-based SMEs to participate in cooperation continues to rise, and the willingness to participate in the same type of enterprise B and financial institutions will gradually become stronger. This is due to the fintech institutions that provide certain financial incentives to participating technology-based SMEs and the influence of market behavior, whether small and medium-sized enterprises adopt cooperation strategies is greatly affected by the willingness of peer companies to participate. When one party has less willingness to participate in cooperation, they will eventually choose a noncooperative strategy.

Whether technology-based SMEs cooperate with fintech institutions is affected by the willingness of financial institutions to cooperate. Figure 3 is a simulation analysis of the three parties' willingness to participate in cooperation under the condition that other parameters remain unchanged. It can be seen from Figure 3 that both technology-based SMEs A and B are in a neutral state, and the willingness of financial institutions to participate in cooperation is between 0.3 and 0.4. When the willingness of financial institutions to cooperate is less than the critical value of 0.3, x and y both converge to 0, z Converges to 1, and the final equilibrium point tends to (0, 0, 1). At this time, the willingness of technology-based SMEs A and B to participate in cooperation is greatly reduced. When the willingness of fintech institutions to participate in cooperation is greater than the critical value of 0.4, x, y, and z are all converges to 1, and the final equilibrium point tends to (1, 1, 1). At this time, the willingness of technology-based SMEs A and B to participate in cooperation increases sharply with the increase of financial institutions' willingness to cooperate, and finally converges to 1. The willingness of financial institutions to participate in cooperation continues to increase. The willingness of small and medium-sized technological enterprises to participate in cooperation has decreased sharply and then increased sharply, and finally reached cooperation with financial institutions. The results show that when financial institutions' willingness to cooperate is low, technology-based SMEs remain on the sidelines, and their willingness to cooperate sharply declines over time; the greater the willingness of financial institutions to cooperate, the greater the willingness of technology-based SMEs to participate in cooperation will increase. This is also due to the fact that cooperation with low- and middle- and low-tech enterprises can bring itself the development of financial technology and additional cooperation benefits, and financial institutions will change from a lower willingness to participate in cooperation to a higher willingness to cooperate.

Figure 4 is a simulation analysis of the three parties' final cooperation strategy when the willingness to participate in cooperation changes at the same time when other parameters remain unchanged. It can be seen from Figure 4 that when the willingness of technology-based SME A and company B to participate in cooperation is very low, even if the willingness of financial institutions to participate in cooperation is high, both x and y converge to 0, and z converges to 1, neither technology-based SMEs A nor B will adopt a cooperation strategy. When the willingness of technology-based SME A and enterprise B to participate in

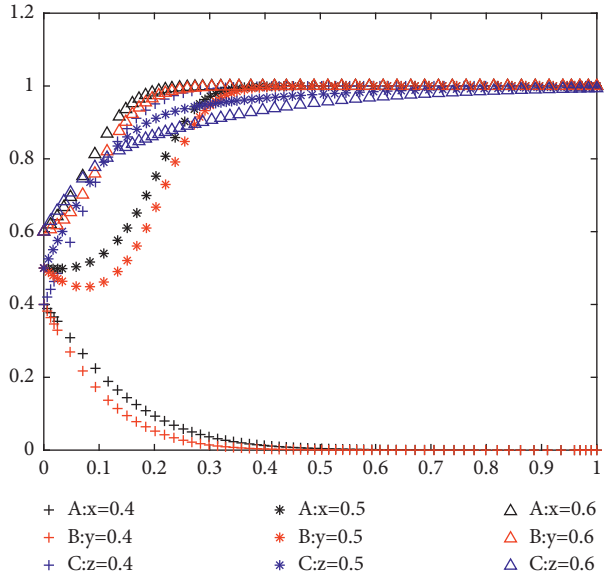


Figure 1: Evolutionary analysis of the simultaneous changes of the initial willingness to participate in cooperation (x), (y), and (z).

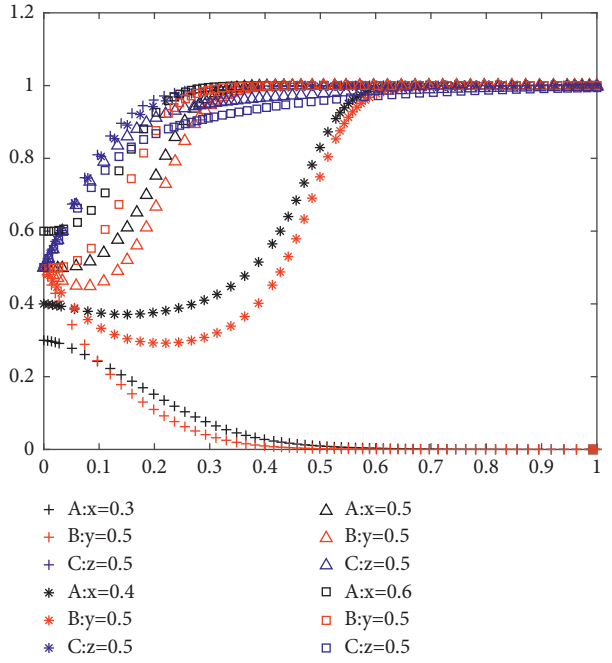


Figure 2: Evolutionary analysis of initial willingness to participate in cooperation (x) changes.

cooperation is high, even if the willingness of financial institutions to participate in cooperation is low, x and y will converge to 1, and z will converge to 1. That is, technology-based SMEs A and B are willing to cooperate with financial institution C, and the weaker technology-based SME A's willingness to participate in cooperation converges faster than company B.

4.2. Scenario Analysis of Changes in Enterprise Cooperation Costs. The changes in cost reduction brought about by the establishment of cooperation between fintech institutions and technology-based SMEs affect the strategies adopted by enterprises A and B in their growth and development. Figure 5 is a simulation analysis of the reduction in cooperation costs brought by fintech institutions to technology-based enterprises when other parameters remain unchanged. It can be seen from Figure 5 that the critical value of the cooperation cost reduction S is between 8 and 12. When the cooperation cost reduction is less than the critical value, x and y converge to 0, and the final equilibrium point converges to (0, 0, 1), With the continuous increase of S, the convergence speed of x and y slows down, and the reduction of technology-based small and medium-sized enterprises A

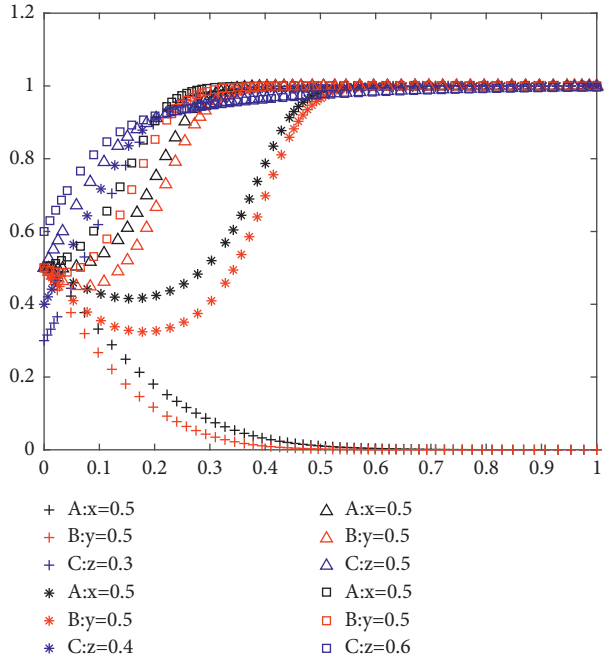


Figure 3: Evolutionary analysis of changes in initial willingness to participate in cooperation (z).

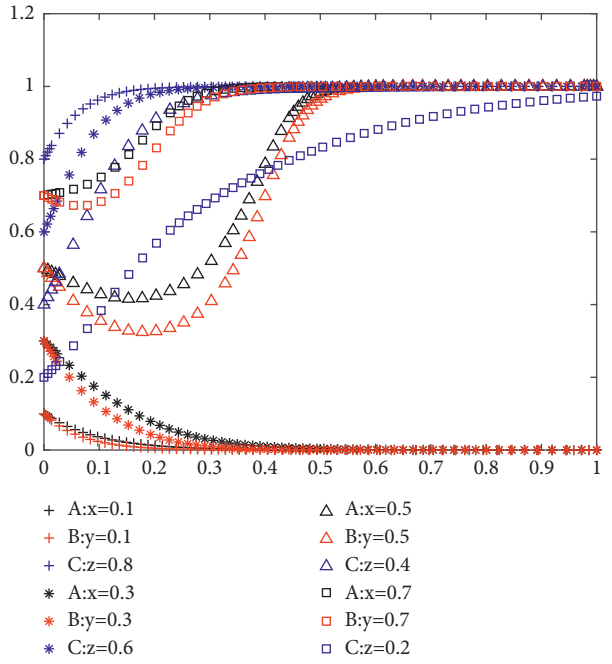


Figure 4: Evolutionary analysis of the initial willingness to participate in cooperation x and y simultaneously change.

is obvious. When the reduction in cooperation cost is greater than the critical value, x and y converge to 1, and the final equilibrium point converges to (1, 1, 1). As S continues to increase, the convergence speed of x and y accelerates. Similarly, the convergence speed of enterprise A is higher than that of enterprise B. The results show that when the reduction in cooperation costs is not enough to bring more benefits to the cooperation between technology-based SMEs, technology-based SMEs will not reach cooperation with fintech institutions. The cost reduction brought about by cooperation with fintech institutions affects whether technology-based SMEs adopt cooperation strategies, and it has a greater impact on the weaker technology-based SME A.

4.3. Scenario Analysis of Changes in Enterprise Revenue. The cooperation benefits brought by the collaborative innovation between fintech institutions and technology-based

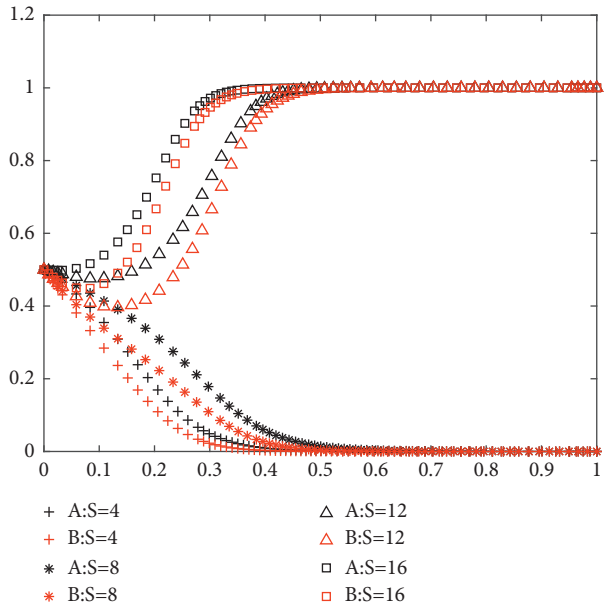


Figure 5: Evolutionary analysis of the change in the reduction of cooperation costs of SMEs with the participation of fintech institutions.

SMEs influence the strategies adopted by SMEs A and B in their growth and development. Figure 6 is a simulation analysis of the change in the amount of cooperation income R brought by financial institutions to scientific and technological enterprises when other parameters remain unchanged. It can be seen from Figure 6 that under the condition that other parameters remain unchanged, the critical value of the cooperation income R with fintech institutions is between 10 and 13. When the cooperation income is lower than the critical value, both x and y converge to 0, and the final equilibrium point Converges to (0, 0, 1), and R increases continuously, the speed at which x and y converge to 0 slows down. When the cooperative gain is higher than the critical value, x and y both converge to 1, and the final equilibrium point converges to (1, 1, 1), and with the continuous increase of R, the speed at which x and y converge to 1 accelerates, small and medium-sized enterprises are willing to adopt cooperative strategies. The results show that the higher the cooperation benefits brought by the collaborative innovation between fintech institutions and technology-based SMEs, the greater the willingness of technology-based SMEs to participate in cooperation. This is also due to the fact that cooperation with fintech institutions can bring externalities to technology-based SMEs. The financing channels reduce the pressure of shortage of funds, so as to better realize their own growth and development.

4.4. Scenario Analysis of Cooperation Costs in Fintech Services. The cost of collaborative cooperation between fintech institutions and technology-based SMEs affects the strategies adopted by technology-based SMEs A and B and fintech institution C. Figure 7 is a simulation analysis of the strategies adopted by the three parties to the cost changes of fintech institutions participating in the cooperation of technology-based SMEs. It can be seen from Figure 7 that the critical value of the cost K_1 paid by financial institutions to participate in cooperation is between 10 and 22. When the cost of participating in cooperation K_1 of fintech institutions is less than the critical value, x, y, and z all converge to 1, and the final equilibrium point converges to (1, 1, 1), fintech institutions are willing to cooperate with technology-based small and medium-sized enterprises A and B. When the cost of fintech institutions' participation in cooperation K_1 is greater than the critical value, x, y, and z all converge to 0, and the final equilibrium point converges to (0, 0, 0), fintech institutions and SMEs A and B are unwilling to cooperate, and the willingness of fintech institutions to participate has converged to 0 faster than technology-based SMEs A and B. The results show that when the cost of participating in cooperation is high, the benefits of fintech institutions are far less than those of noncooperative strategies. Fintech institutions will not be willing to cooperate, and high-tech SMEs will not seek help from fintech institutions.

4.5. Scenario Analysis of Incentive Changes in Fintech Cooperation. The financial rewards given by fintech institutions for participating in cooperative technology-based SMEs affect the strategies adopted by technology-based SMEs A and B and fintech institution C. Figure 8 is a simulation analysis of the strategy adopted by the three parties for changes in the financial reward K_2 given by fintech institutions to participating in cooperative SMEs. It can be seen from Figure 8 that the critical value of the financial reward K_2 given by fintech institutions for participating in cooperative technological SMEs is between 8 and 14. When the financial reward given by fintech institutions to participating technological SMEs is less than the critical value, x, y Both converge to 0, z converges to 1, and the final equilibrium point converges to (0, 0, 1), technology-based SMEs are unwilling to adopt cooperation strategies. When

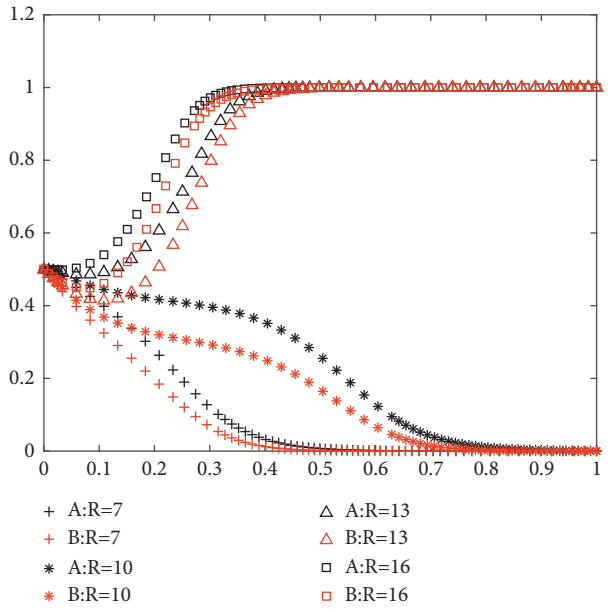


Figure 6: Evolutionary analysis of the changes in the cooperative income of SMEs with the participation of fintech institutions.

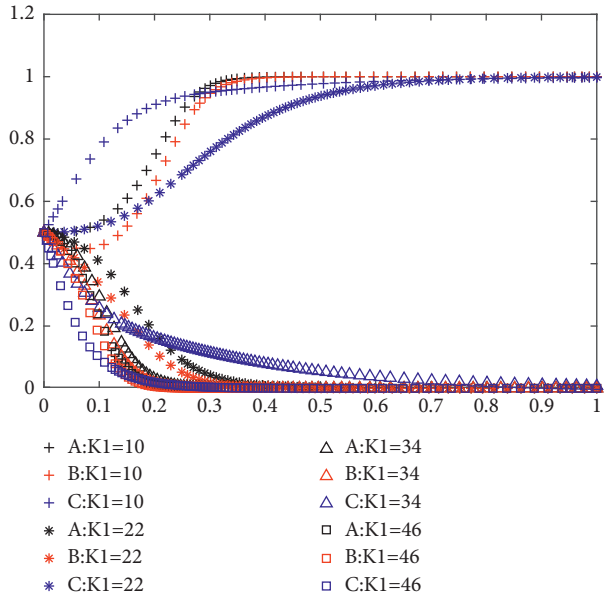


Figure 7: Evolutionary analysis of the cost changes of financial institutions participating in the cooperation of technology-based SMEs.

fintech institutions give financial rewards for participating in cooperative technology-based SMEs greater than the threshold value, x, y both converge to 1, z converges to 0, and the final equilibrium point converges to (1, 1, 0), financial institutions are unwilling to adopt cooperative strategies. The financial reward is between the critical values, x, y, z all converge to 1, and the final equilibrium point converges to (1, 1, 1), the three parties are willing to cooperate, and as the financial reward K_2 continues to increase, small and medium-sized technology-based Enterprises' willingness to participate in cooperation converges to 1 faster than fintech institutions. The results show that when financial institutions give relatively low financial rewards for participating in cooperative technology-based SMEs, due to the choice of peer companies and market behavior, the subjective willingness of technology-based SMEs to participate in cooperation is small, and when the financial reward reaches a certain standard, Technology-based SMEs will reach a

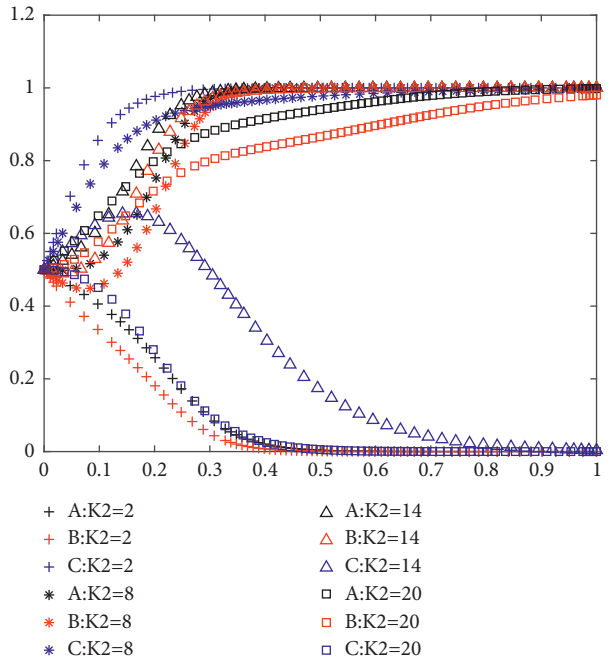


Figure 8: The evolution analysis of fintech institutions' funding support parameters for technology-based SMEs.

cooperation agreement with fintech institutions. Once fintech institutions give financial rewards to participating technology-based SMEs beyond the scope of fintech institutions' commitments, fintech institutions' willingness to participate in cooperation will show a slow increase and then a sharp decline. The final choice is noncooperative strategy.

5. Conclusions

Technology-based SMEs with better upstream development and strong strength provide technical knowledge, financing channels and corporate management methods to downstream companies of the same type with weaker downstream innovation and growth. Technology cooperation between technology-based SMEs cannot meet the funding gap needed in the later stages of enterprise growth, and technology-based SMEs have to seek support from other financing institutions. With the participation of fintech institutions in cooperation, technology-based SMEs can provide them with technical knowledge to obtain financial support, and financial institutions can also use the technology of the other company to complete their own business transformation. This paper establishes the evolutionary game model of technology-based SMEs A, B and fintech institution C, systematically analyzes the strategies adopted by the three-party game of technology-based SMEs A, B and fintech institution C, and conducts case simulation to analyze its influencing factors, and draws the following in conclusion:

(1) Technology-based SMEs A and B and financial institution C's willingness to participate in cooperation have different influences on each other's choice of strategies. The strategy adopted by technology-based SMEs is greatly affected by the willingness of peer companies to participate in cooperation. When companies with stronger development capabilities are less willing to participate, companies of the same type will not actively seek cooperation from fintech institutions. The influence of technology-based SMEs adopted strategies that are affected by fintech institutions' willingness to participate in cooperation is significant. When fintech institutions are more willing to participate in cooperation, SMEs will seek cooperation. Therefore, fintech institutions should increase their enthusiasm for establishing cooperation with SMEs, cooperate extensively with enterprises, provide corresponding financial support, and make full use of existing technical knowledge to realize digital business transformation.

(2) With the participation of fintech institutions, the reduction of cooperation costs and the changes in cooperation gains of technology-based SMEs have a significant impact on the strategies adopted by the tripartite entities. The reduction of the cooperation cost of fintech institutions is higher than the range that they can bear, and the willingness of fintech institutions to participate in cooperation is reduced; the profit value of SMEs after the participation of fintech institutions is lower than the expected range, and the establishment of SMEs and fintech institutions that the willingness to cooperate is reduced. Therefore, in order to promote cooperation between technology-based SMEs and fintech institutions, the distribution of cooperation costs and cooperation benefits must be fair and reasonable, so as to achieve mutual benefit and win-win results for multiple parties, and promote the sustainable growth of enterprises.

(3) Fintech institutions are more sensitive to the cost and financial support of participating in cooperation than high-tech SMEs. The cost and the value of financial support paid by fintech institutions in participating in the cooperation are within a certain range. Fintech institutions are willing to provide cooperation support to technology-based SMEs. Once they exceed expectations, fintech institutions will no longer establish cooperative relations. Therefore, when establishing cooperation between fintech institutions and high-tech SMEs, they should do what they can, formulate reasonable cooperation costs, appropriately increase support, guide enterprises to actively participate in the cooperation, and to a certain extent alleviate the financing pressure of high-tech SMEs.

In the growth process of technology-based SMEs, learning from the development experience and practice paths of stronger enterprises and reaching cooperative technological innovations can reduce the problems encountered in the growth of enterprises. Fintech institutions launch innovative technology credit products based on the development needs and characteristics of technology-based enterprises in the cluster network to improve the service strength of finance. In the fintech innovation, in addition to the impact of financial institutions on the development of technology-based SMEs, relevant government measures will also change the direction of development. In order to promote the development of fintech and promote the industrialization of scientific and technological achievements of SMEs, some local governments will issue relevant policies to help them establish a relationship, promote the cooperation between fintech institutions and SMEs, and realize "technology + policy + finance." The tripartite dialogue mechanism and docking platform of "enterprise creation" provides a full range of services for regional scientific and technological innovation.

Data Availability

The method in this article is computer mathematical simulation. Model parameter data are set based on enterprise interviews.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (no. 71871115), Provincial Social Science Foundation of Jiangsu (no. 21EYC013), Provincial Soft Science Program of Jiangsu (no. BR2021063), and Research Project of Philosophy and Society of Universities in Jiangsu Province (no. 2021SJZDA019).

References

[1] M. Bodlaj, S. Kadic-Maglajlic, and I. Vida, "Disentangling the impact of different innovation types, financial constraints and geographic diversification on SMEs' export growth," *Journal of Business Research*, vol. 108, no. 1, pp. 466–475, 2020.

[2] H. Karadag, "Financial management challenges in small and medium-sized enterprises: a strategic management approach," *Emerging Markets Journal*, vol. 5, no. 1, pp. 26–40, 2015.

[3] G. N. Chandler and S. H. Hanks, "Founder competence, the environment, and venture performance," *Entrepreneurship: Theory and Practice*, vol. 18, no. 3, pp. 77–89, 1994.

[4] S. B. Moore and S. L. Manring, "Strategy development in small and medium sized enterprises for sustainability and increased value creation," *Journal of Cleaner Production*, vol. 17, no. 2, pp. 276–282, 2009.

[5] J. Sun, "On the defects of administrative monopoly in China's "anti-monopoly law" and its improvement," *Canadian Social Science*, vol. 6, no. 2, pp. 1–19, 2010.

[6] M. E. Porter, *Competitive Advantage of Nations: Creating and Sustaining superior Performance*, Simon & Schuster, Manhattan, NY , USA, 2011.

[7] L. Owen, C. Goldwasser, and K. Choate, "Collaborative innovation throughout the extended enterprise," *Strategy & Leadership*, vol. 36, no. 1, pp. 39–45, 2008.

[8] B. B. Yu and L. Yu, "Study the innovation chain docking industry chain of industry cluster based on evolutionary game model," *Science Research Management*, vol. 36, no. 4, pp. 30–38, 2015.

[9] M. Tian, P. Deng, Y. Zhang, and M. P. Salmador, "How does culture influence innovation? A systematic literature review," *Management Decision*, vol. 56, no. 5, pp. 1088–1107, 2018.

[10] L. D. Propris, "Innovation and inter-firm Co-operation: the case of the west midlands," *Economics of Innovation and New Technology*, vol. 9, no. 5, pp. 421–446, 2000.

[11] G. Saint-Paul, "Technological choice, financial markets and economic development," *European Economic Review*, vol. 36, no. 4, pp. 763–781, 1992.

[12] L. Chen, "From fintech to finlife: the case of fintech development in China," *China Economic Journal*, vol. 9, no. 3, pp. 225–239, 2016.

[13] R. G. King and R. Levine, "Finance, entrepreneurship and growth," *Journal of Monetary Economics*, vol. 32, no. 3, pp. 513–542, 1993.

[14] C. Loderer, R. Stulz, and U. Waelchli, "Firm rigidities and the decline in growth opportunities," *Management Science*, vol. 63, no. 9, pp. 3000–3020, 2017.

[15] D. R. Gnyawali and B.-J. Park, "Co-opetition and technological innovation in small and medium-sized enterprises: a multilevel conceptual model," *Journal of Small Business Management*, vol. 47, no. 3, pp. 308–330, 2009.

[16] T. Khanna and Y. Yafeh, "Business groups in emerging markets: paragons or parasites?" *Journal of Economic Literature*, vol. 45, no. 2, pp. 331–372, 2007.

[17] J. Guan and R. C. M. Yam, "Effects of government financial incentives on firms' innovation performance in China: e," *Research Policy*, vol. 44, no. 1, pp. 273–282, 2015.
[18] V. Revest and A. Sapio, "Financing technology-based small firms in europe: what do we know?" *Small Business Economics*, vol. 39, no. 1, pp. 179–205, 2012.
[19] M. Halme and M. Korpela, "Responsible innovation toward sustainable development in small and medium-sized enterprises: a resource perspective," *Business Strategy and the Environment*, vol. 23, no. 8, pp. 547–566, 2014.
[20] E. B. Bayarçelik, F. Taşel, and S. Apak, "A research on determining innovation factors for SMEs," *Procedia-Social and Behavioral Sciences*, vol. 150, pp. 202–211, 2014.
[21] K. Li, J. Niskanen, M. Kolehmainen, and M. Niskanen, "Financial innovation: credit default hybrid model for SME lending," *Expert Systems with Applications*, vol. 61, no. 11, pp. 343–355, 2016.
[22] B. A. Minton, R. M. Stulz, and A. G. Taboada, "Are the largest banks valued more highly?" *Review of Financial Studies*, vol. 32, no. 12, pp. 4604–4652, 2019.
[23] D. Wójcik, "Financial geography II: the impacts of FinTech–Financial sector and centres, regulation and stability, inclusion and governance," *Progress in Human Geography*, vol. 45, no. 4, pp. 878–889, 2021.
[24] J. Wang, "From aperture satellite to "Internet finance": i," *Telecommunications Policy*, vol. 42, no. 7, pp. 566–574, 2018.
[25] C. Milligan, "": developments in the geography of mental illhealth - the influence of the informal sector," *Health & Place*, vol. 6, no. 3, pp. 189–200, 2000.
[26] K. P. Gallagher, R. Kamal, J. Jin, Y. Chen, and X. Ma, "Energizing development finance? The benefits and risks of China's development finance in the global energy sector," *Energy Policy*, vol. 122, no. 11, pp. 313–321, 2018.
[27] J. E Stiglitz, *Freefall: America, Free Markets, and the Sinking of the World Economy*, WW Norton & Company, NY, USA, 2010.
[28] C. Su, "A study of policies for improving the technological innovation capacity of small and medium-sized enterprises," *TEM Journal*, vol. 7, no. 2, pp. 268–280, 2018.
[29] M. Nasiri, J. Ukko, M. Saunila, T. Rantala, and H. Rantanen, "Digital-related capabilities and financial performance: the mediating effect of performance measurement systems," *Technology Analysis & Strategic Management*, vol. 32, no. 12, pp. 1393–1406, 2020.
[30] W. Younas and K. R. Kalimuthu, "Telecom microfinance banking versus commercial banking: a battle in the financial services sector," *Journal of Financial Services Marketing*, vol. 26, no. 2, pp. 67–80, 2021.
[31] J. Lin, "Incumbent firm invention in emerging fields: evidence from the semiconductor industry," *Strategic Management Journal*, vol. 32, no. 1, pp. 55–75, 2011.
[32] P. Gamage, "Bank finance for small and medium-sized enterprises in Sri Lanka: issues and policy r," *Studies in Business and Economics*, vol. 10, no. 2, pp. 32–43, 2015.
[33] J. Kang, "Mobile payment in Fintech environment: trends, security challenges, and services," *Human-centric Computing and Information sciences*, vol. 8, no. 1, pp. 1–16, 2018.
[34] Y. Wen, M. Zhao, L. Zheng, Y. Yang, and X. Yang, "Impacts of financial agglomeration on technological innovation: a spatial and nonlinear perspective," *Technology Analysis & Strategic Management*, vol. 33, no. 8, pp. 1–13, 2021.
[35] P. Gomber, R. J. Kauffman, C. Parker, and B. W. Weber, "On the fintech revolution: interpreting the forces of innovation, disruption, and transformation in financial services," *Journal of Management Information Systems*, vol. 35, no. 1, pp. 220–265, 2018.
[36] R. C. Basole and S. S. Patel, "Transformation through unbundling: visualizing the global fintech ecosystem," *Service Science*, vol. 10, no. 4, pp. 379–396, 2018.
[37] D. Friedman, "Evolutionary geconomics," *Econometrica*, vol. 59, no. 3, pp. 637–666, 1991.
[38] W. Currie, D. Gozman, and J. Seddon, "Dialectic tensions in the financial markets: a longitudinal study of pre- and post-crisis regulatory technology," *Journal of Information Technology*, vol. 33, no. 4, pp. 304–325, 2018.
[39] M. René, "Stulz. FinTech, BigTech, and the future of banks," *The Journal of Applied Corporate Finance*, vol. 31, no. 4, pp. 86–97, 2019.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 2996704, 19 pages
https://doi.org/10.1155/2022/2996704

Research Article

Liquidity, Credit Risk, and Their Interaction on the Spreads in China's Corporate Bond Market

Zijian Wu, Baochen Yang, and Yunpeng Su

College of Management and Economics, Tianjin University, Tianjin 300072, China

Correspondence should be addressed to Baochen Yang; bchyang@tju.edu.cn

Received 20 October 2021; Revised 25 February 2022; Accepted 3 March 2022; Published 12 April 2022

Academic Editor: Leonid Shaikhet

Copyright © 2022 Zijian Wu 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 investigates the impact of multidimension liquidity, credit risk, and the interaction between liquidity and credit risk on corporate bond spreads based on a large transaction data set from July, 2006 to June, 2016, including the monthly data of 3716 bonds in China. Our main findings reveal that liquidity premiums are the main parts of corporate bond spreads. The interaction between liquidity and credit risk plays a significant role in determining corporate bond spreads. In addition, the differences between the interbank market and the exchange market have a significant impact on corporate bond spreads in normal period, and the interaction between liquidity and credit risk has an enhanced impact on corporate bond spreads during financial crisis. We also find that the interaction between liquidity and credit risk will increase with the increase of liquidity risk and credit risk and it is a time-varying dynamic process.

1. Introduction

Several literature studies investigate the determinants of corporate yield spreads and link them to credit risk and liquidity [1–3]. Huang and Huang [4] use structural bond pricing models and find that credit risk accounts for only a small fraction of bond spreads. Collin-Dufresn et al. [5] document a large unexplained portion of bond spreads which are driven mainly by factors that are independent of credit risk. Consequently, many researchers focus on the potential for liquidity to explain a large portion of bond yield spreads [6–8]. Lin et al. [9] and Acharya et al. [10] find that liquidity risk is priced in corporate bonds returns. The relevant literature can also refer to Zou et al. [11]; Dick-Nielsen et al. [12]; and Helwege et al. [13]. In addition, by proposing a new illiquidity measure, Longstaff et al. [7] discover that credit risk is the main determinant of corporate yield spreads. Covitz and Downing [14] report similar findings with Longstaff et al. [7] through investigating very short-term commercial paper issued by nonfinancial U.S. corporations.

The impact of the two types of risk on the spreads has been discussed in several literature studies. Longstaff and Schwartz [15] find that the correlation between default risk and the interest rate has a significant impact on the properties of the credit spreads. Several literature studies have discussed the impact of liquidity and credit risk on spreads [16–20]. Ericsson and Renault [16] analyze the interaction between liquidity and credit risk in theory. Studies by Chen et al. [8]; Covitz and Downing [14]; Rossi [21]; and Kalimipalli and Nayak [22] provide evidence that liquidity impacts are mingled with credit risk impacts on bond yield spreads. Wang and Wu [23] find strong evidence of a significant interactive impact of liquidity and credit risk, magnifying during the financial crisis period. Sperna Weiland et al. [24] propose a novel way of modeling credit-liquidity interactions through mutually exciting processes. Chen et al. [25] develop a structural credit risk model to examine how the interactions of liquidity and default risk affect corporate bond pricing. Li et al. [26] propose a generalized bond pricing model, accounting for all the impacts of credit risk, liquidity risk, and their correlation, and they extend the traditional bond pricing model with only credit risk by incorporating liquidity risk into the framework in which the probabilities of the two risk events are estimated by a joint distribution.

In fact, previous literature regarding the above tends to focus on how liquidity or credit component contributed to the yield spreads independently in China's corporate bond market. Ai et al. [27] discuss the impact of the interaction term on corporate bond spreads by using a single dimensional liquidity measure. Different from previous literature, this paper investigates liquidity in different dimensions from the perspective of speculative trading and investment trading and introduces dynamic credit risk and static credit risk. And the paper also studies different types of the interaction between liquidity and credit risk on corporate bond spreads and explains their affecting mechanism and market transmission channels. For example, dynamic credit risk represents the credit risk level of a firm over time, and it changes in resonance with the dynamic change of liquidity risk. When liquidity affects asset prices, dynamic credit risk usually interacts with liquidity risk in a dynamic spiral, forming the resonance enhancement interaction between liquidity and credit risk to dynamically affect asset prices. This study aims to provide a deeper understanding of the impact of multidimension liquidity, credit risk, and the interaction between liquidity and credit risk on the spreads in China. This paper uses a large transaction data set from July 2006 to June 2016, including the monthly data of 3716 bonds in China's bond market.

Our research contributes to the literature in four ways. First, we select multidimension liquidity measures from the Chinese corporate bond market, which can measure liquidity more comprehensively. We select bonds in China's interbank market and exchange market and construct many liquidity measures from three dimensions of trading activity, price shock and bond life time. Many scholars construct the liquidity measure of corporate bonds from a single dimension or several dimensions, which only reflects a certain dimension or several dimensions of liquidity. Our method of multidimension liquidity measures is more accurate and reliable than those used in the existing literature.

Second, we introduce the interaction between liquidity and credit risk into the spread model. Our results demonstrate that the interaction between liquidity and credit risk plays a significant role in determining yield spreads in China's bond market, and such findings are robust to the alternative proxies for liquidity and credit risk with various model specifications. It shows that the impact of the interaction between liquidity and credit risk on bond spreads is significant in China's bond market. Therefore, we note that liquidity and credit risk alone does not explain the entire yield-spread behavior. This study provides evidence regarding the significant impact of the interaction between liquidity and credit risk on the spreads.

Third, we investigates liquidity premiums from multiple dimensions and find that liquidity premiums are the main parts of corporate bond spreads in China's corporate bond market. This finding is consistent with Collin-Dufresn et al. [5] and Huang and Huang [4].

Finally, dummy variables are introduced to eliminate the impact of the differences in the structure of the interbank market and the exchange market on the results, because the differences between the two markets are taken into account. The research shows that the results are consistent when the characteristics of the two markets are taken into account, and the interaction between liquidity and credit risk has a significant impact on corporate bond spreads, and the impact is greater during financial crisis.

Our main findings are as follows. First, liquidity premiums are the main parts of corporate bond spreads in China's corporate bond market. Second, the interaction between liquidity and credit risk on bond spreads plays a significant role in determining corporate bond spreads and liquidity and credit risk alone does not explain the entire yield-spread behavior. The interaction between liquidity and credit risk will increase with the increase of liquidity risk and credit risk, and it is a time-varying dynamic process. Related research can also refer to Sperna Weiland et al. [24] and Chen et al. [25]. Third, the interaction between liquidity and credit risk has an enhanced impact on corporate bond spreads during financial crisis than normal period. And the differences between the interbank market and the exchange market have a significant impact on corporate bond spreads in normal period than financial crisis. Finally, dynamic credit risk changes in resonance with the dynamic change of liquidity risk, while the static credit risk does not change with the dynamic change of liquidity risk.

The rest of this paper is organized as follows. Section 2 describes the selected variables and the hypothesis; Section 3 elaborates on the empirical results; Section 4 gives the results of robustness test; and Section 5 presents our conclusions.

2. Data and Sample

2.1. Data and Sample Description. We use a sample of corporate bonds in WIND database during the period July, 2006 to June, 2016. We collect data on bond liquidity and credit risk from WIND database, no transaction of bonds is removed. When bonds are less than a year's remaining trading time, we drop the corresponding observations from the sample as these bonds are rarely traded. Finally, we have 47436 observations, a total of 3716 bonds, including 2281 interbank bonds and 1435 exchange bonds. In addition, we collect data for issuer characteristics and the trading volume, trading days, the yield to maturity, the highest price, the lowest price, the average price, the remaining life, age, credit rating, coupon rate and historical price, and the proxies of liquidity and credit risk are calculated based on the above data. In many cases, bonds with time-to-maturity of less than 12 months are especially unlikely to trade, so we exclude such bonds with time-to-maturity of less than 12 months from the sample [28].

2.2. Selected Variables and Hypothesis

2.2.1. Spread. The Spread is the difference between the yield-to-maturity on the corporate bond and the corresponding Treasury rate. When calculating the spread sequence, we selected corporate bonds and treasury bonds with the same maturity yield period to calculate spread by subtraction. We primarily use existing treasury data, but where unavailable we use estimated data from WIND database. Our analysis

employs the spread of each bond as a dependent variable in the regression model.

2.2.2. Liquidity Measures. There is no consensus on how to measure the liquidity of an asset so we examine a number of liquidity-related measures for corporate bonds. The following describes the measures and their implementation in more detail. In this study, we select multidimension liquidity measures, which can measure liquidity more comprehensively. The liquidity proxies used in this analysis include the following seven liquidity measures: turnover, trade vol, day, amihud, range, time-to-maturity and age.

Turnover is the frequency of market assets trading hands in a certain period. In general, therefore, this proxy is expected to be negatively related to yield spreads. The second liquidity measure is Vol, it is obtained by dividing the total trading volume with the number of months during which the issue is traded. The common expectation is that this variable is expected to have an inverse relation with the yield spreads. Day is defined as the number of trading days in the corresponding month. This variable is expected to be negatively related to yield spreads. Following the amihud measure [29], we construct a monthly liquidity measure proxy to describe the liquidity. We construct a monthly liquidity measure proxy of Range based on Han and Zhou [30]. The disadvantage of this measure is that it should be sensitive to outliers. Time-to-Maturity is the time left until maturity from the issuance date of a security and is a liquidity measure proxy [7, 14, 31]. Age is defined as the age of bond transactions since the issue in the corresponding month. Researchers frequently proxy for the liquidity of a corporate bond with its age [32–35]. A number of problems with studying liquidity in corporate bonds arise from the fact that corporate bonds trade only infrequently [33, 36]. Therefore, the liquidity is closely related to age.

2.2.3. Credit Risk Measures. This study considers three credit risk variables as explanatory variables. They are rating, coupon and volatility.

Rating is the credit rating assigned to each security that a firm issues [14, 37]. Our study adopts the coding method of Covitz and Downing [14] and Shin and Kim [37] for credit ratings: AAA = 1, AA+ = 2, . . ., and $C = 14$. We assign the lower credit rating to 15. Obviously, a positive relationship is expected between yield spreads and credit ratings. Coupon refers to the coupon rate of bonds [7, 38]. Longstaff et al. [7] and Bharath and Shumway [38] show that coupon rates have significantly positive correlation with yield spreads. See also Chen et al. [39] or Lin et al. [9]. As expected, the coupon of bond issued by firm with higher default risk is significantly high, other things being constant. Thus, we expect the coefficient of the coupon rate to be positive. Hull [40] uses the historical price volatility of corporate bonds as a measure of the bond's credit. Volatility is the historical price volatility of corporate bonds, and it is estimated from the past historical prices of 125 days. It is well known that as a firm approaches default, the risk associated with its debt also increases, and such risk is correlated to the bond volatility. Accordingly, we expect the Volatility to have a positive correlation with bond spreads.

2.2.4. Hypothesis. Turnover, Vol, Day are the proxies of trading activity [3, 12, 37]. The greater the values is, more active the bond transaction is, which shows that the bond liquidity risk is smaller, and the spreads are smaller. Therefore, Turnover, Vol, Day have negative relation with spreads in the corporate bond market. Time-to-Maturity represents the remaining life of the bond. These bonds close to maturity are rarely traded, and the bond holders often choose to hold them until maturity, and they will not be traded, so the coefficient of Time-to-Maturity is negative [14, 31]. New bonds are not more liquid than older ones. Because of information asymmetry, when new bonds are issued, investors will not buy and trade. Over time, for more than two years, investors will choose to buy the bonds. Then the bond transactions become active, and the spreads begin to become small. Therefore, Age and spreads are negative correlation [33, 34]. Based on the above analysis, we propose hypothesis H1.

H1: turnover, vol, day, age and time-to-maturity have negative relation with spreads in the corporate bond market.

Range and Amihud are the liquidity proxies of price impact [29, 30, 37]. When trading a certain amount of bonds, the greater the bond price changes caused by the transaction, which indicates that the liquidity is poorer, so the bond spreads are also larger. Therefore, Amihud and Range have positive correlation with the bond spreads. Rating, Coupon, and Volatility are the proxies of credit risk [14, 38–40]. The greater the value is, the worse credit is, and the greater the spreads are. So Rating, Coupon, and Volatility have positive relation with spreads in the corporate bond market. Based on the above analysis, we propose hypothesis H2.

H2: range, amihud, rating, coupon, and volatility have positive relation with spreads in the corporate bond market.

Ericsson and Renault [16] prove theoretically that the liquidity and credit risk are correlated, and the interaction of liquidity and credit risk is present. When liquidity is low, default risk of bonds tends to be high, suggesting the possibility of an interactive impact of liquidity and default risk. It directly draws on the implications that the interaction of liquidity and credit risk is present. See also Chen et al. [25] or Li et al. [26]. We test this hypothesis using data of bonds with different liquidity and credit characteristics such as Age, Vol, Day, Amihud and Coupon. We now turn to empirical tests. Based on the above analysis, we propose hypothesis H3.

H3: the interaction between liquidity and credit risk has a significant impact on corporate bond spreads.

3. Empirical Analysis

3.1. Summary Statistics

3.1.1. Description. Table 1 provides summary statistics for the variables used in empirical analysis. It summarizes trading data and bond characteristics of the sample used in our study. Mean, median, maximum, minimum, standard

Table 1: Summary statistics for the whole sample. The table shows the summary statistics for variables used in the regression analysis for the whole sample.

Variable	Mean	Median	Maximum	Minimum	Std. dev.	Jarque-Bera	Probability	Observations
Spread	2.5463	2.4940	35.5890	0.0082	1.1275	4558494.0000	0.0001	47436.0000
Turnover	0.0828	0.0362	2.8476	0.0001	0.1332	2378552.0000	0.0001	47436.0000
Vol	87.5034	42.3949	3072.2300	0.0102	133.5900	3004256.0000	0.0001	47436.0000
Day	4.6538	2.0000	23.0000	1.0000	5.3667	38023.8300	0.0001	47436.0000
Amihud	0.0122	0.0003	3.5210	0.0001	0.0798	667000000.0000	0.0001	47436.0000
Range	0.0238	0.0005	8.4091	0.0001	0.1410	685000000.0000	0.0001	47436.0000
Time-to-maturity	5.4085	5.2932	27.0027	1.0027	2.2759	334283.6000	0.0001	47436.0000
Age	2.2654	1.9123	13.9781	0.0247	1.7405	105440.7000	0.0001	47436.0000
Volatility	0.5004	0.4143	14.4579	0.0002	0.3617	6590932.0000	0.0001	47436.0000
Coupon	6.7952	6.9000	9.3000	3.1700	1.0530	1311.1280	0.0001	47436.0000
Rating	3.3323	4.0000	15.0000	1.0000	1.5065	734923.6000	0.0001	47436.0000

We report mean, median, maximum, minimum, standard deviation, Jarque-Bera, probability, and observations. Spread is the difference between the yield-to-maturity on the corporate bond and the corresponding treasury rate. Turnover is the frequency of market assets trading hands in a certain period of time. Vol is the average of the total trading volume in the corresponding month. Day is defined as the number of trading days in the corresponding month. Amihud and range are illiquidity measures and describe the liquidity measure of the bond price impact dimension. Time-to-maturity is the remaining years of the bond. Age denotes years since issuance. Volatility is a standard deviation of historical bond price. Coupon is the annual coupon interest. Rating is the credit rating assigned to each security, coded so that AAA = 1. The sample period is July 2006 to June 2016. The whole sample includes 3716 bond issues and 47436 observations.

deviation, Jarque-Bera, probability, and observations for each variable are reported for the whole sample in Table 1.

From the whole sample, as seen from Table 1, the observation is 47436. The average of the spread is 2.5463%, with a standard deviation of 1.1275%. As seen from the median, the distribution of the spread has a long right tail, indicating that some of the very wide yield spreads skew the distribution to the right. Turnover registers a value of 0.0828 every month per issue, reflecting some large values in the right tail of the distribution, while vol averages 87.5034, again with a long right tail. Day averages 4.6538, with a standard deviation of 5.3667. Amihud registers a value of 0.0122 every month, reflecting some large values in the right tail of the distribution, while range averages 0.0238, again with a long right tail. The average of time-to-maturity is 5.4085 years, and the median is 5.2932 years, while age averages 2.2654 years, and the median is 1.9123 years. The mean value of coupon is 6.7952, with a standard deviation 1.0530, while volatility averages 0.5004, with a standard deviation of 0.3617, where the lowest value is 0.0002 and the highest is 14.4579. Finally, The average credit rating score is 3.3323 (between AA− and AA), with a standard deviation of 1.5065.

3.1.2. Pairwise Correlation Test. Table 2 shows the pairwise correlation coefficients among the independent variables respectively in the whole sample. As confirmed from Table 2, turnover and vol are positively high correlated with each other, with a correlation coefficient of 0.8683, while turnover and day seem to be quite weakly correlated and vol and day are also like this; amihud is also positively correlated with range, with a correlation coefficient of 0.7673. Coupon, volatility, and rating are positively correlated with one another; the coefficient value between coupon and rating is relatively high, with a correlation coefficient of 0.3228, while volatility and coupon exhibit weaker correlation, with a correlation coefficient of 0.0234. On the other hand, liquidity and credit risk proxies also have correlation. Age is negatively correlated with Coupon, with a correlation coefficient of −0.2927.

3.1.3. Unit Root Test. In order to avoid spurious regressions, we first make unit root tests for the sample data. From the results of ADF test of Table 3, all variables are significant at the level of 1%, and the original hypothesis of unit root is rejected. That is to say, all the variables are stationary sequence, and the sample data can be analyzed by regression analysis. Then we will study the impact factors of the spreads, especially the interaction between liquidity and credit risk. The stable variable sequence can ensure the validity of the regression results.

3.2. The Impact of the Interaction between Liquidity and Credit Risk on Spreads. Some scholars study liquidity and explain why it affects asset prices [41–44]. Bagehot [43] proposes that due to the impact of information asymmetry, liquidity refers to the impact of a transaction on the market price. [44] believe that illiquidity is the risk that assets cannot be sold at a reasonable price within a given time. The above literature shows that liquidity is a potential cost of asset transaction, and investors with illiquid assets need liquidity risk compensation, namely liquidity risk premiums. Information asymmetry is an important factor affecting asset liquidity. The reason why liquidity affects asset price is that investors' information asymmetry leads to speculative transaction, investment transaction and investors' compensation for different liquidity risk.

Acharya and Pedersen [45] explain the affecting mechanism of liquidity on asset prices. They establish a theoretical model of Liquidity Capital Asset Pricing Model and identify two channels of risk transmission. The first channel is direct compensation for the transaction costs of assets. This is a premium for liquidity level. The second channel is the compensation of asset returns for systemic

Table 2: Pairwise correlations between independent variables for the whole sample.

	Turnover	Vol	Day	Amihud	Range	Time-to-maturity	Age	Volatility	Coupon	Rating
Turnover	1									
Vol	0.8683***	1								
Day	−0.0293***	−0.0268***	1							
Amihud	−0.0926***	−0.0976***	−0.0201***	1						
Range	−0.1030***	−0.1085***	0.0064***	0.7673***	1					
Time-to-maturity	0.0113***	0.0437***	0.1396***	0.0303	0.0710	1				
Age	−0.1541***	−0.1232***	0.0741***	0.0274	0.0423	−0.2945***	1			
Volatility	0.0490***	0.0798***	0.0900***	0.0640	0.1036***	0.1158***	0.0583***	1		
Coupon	0.0985	0.0357***	0.0921***	−0.0026	−0.0315	−0.2345***	−0.2927***	0.0234***	1	
Rating	0.1040***	0.0198***	0.0655***	0.0063	0.0090	−0.1691***	−0.1786***	0.0404***	0.3228***	1

The table shows the pairwise correlation coefficients for the variables used in the regression analysis for the whole sample. The meanings of the variables are given in Table 1. The sample period is July 2006 to June 2016. The whole sample includes 3716 bond issues and 47436 observations. "***" indicates significance at the 1% level.

liquidity. This is a premium for liquidity risk. Bongaerts et al. [46] use an asset pricing approach to compare the impacts of liquidity level and liquidity risk on expected U.S. corporate bond returns and analyze which of these two channels is more important. Liu [47] and Bervas [48] find that a liquid market presents many different dimensions. Díaz and Escribano [49] classify and describe the variety of the existing liquidity measures with the different characteristics and dimensions of liquidity. Previous literature suggests that liquidity has many dimensions. We divide liquidity into three dimensions based on the economic significance of liquidity, including trading activity, price shock, and bond survival time. Liquidity of different dimension has a different impact on corporate bond spreads. Liquidity of transaction activity dimension is the common result of speculative trading and investment trading, while liquidity of price shock dimension is mostly due to investors' speculative trading. In addition, liquidity of bond survival time dimension mainly represents investors' investment transaction. Different dimension liquidity premiums have different economic significance. Therefore, liquidity should be measured across the different dimensions in order to have a global view of liquidity. We study the mechanism and channels of liquidity's impact on asset price from the perspective of multidimension liquidity.

Bond credit risk is the risk of loss to investors due to bond default. Investors need some compensation for risk when holding defaultable assets, which is why credit risk affects asset prices. Referring to the existing literature, we select three credit risk indicators, including price volatility, coupon rate, and credit rating. We divide credit risk into two dimensions according to the economic significance, including dynamic credit risk and static credit risk. Relevant literature can also be referred to Li and Song [50]. Different dimension credit risk have different impacts on corporate bond spreads. Dynamic credit risk is mostly caused by investors' speculative trading of liquidity risk, so it often interacts with liquidity, while static credit risk is mainly determined by the default probability of credit bonds themselves, and it is often not linked with liquidity risk. Therefore, credit risk premium of different dimension has a different economic significance. Credit rating and coupon rate are static credit risk measures, while price volatility is a dynamic credit risk measure. The static credit risk represents the credit risk level of an firm in this period of time, and it does not change with the dynamic change of liquidity risk. When liquidity affects asset prices, static credit risk generally does not interact with liquidity risk in a dynamic spiral. Dynamic credit risk represents the credit risk level of an firm over time, and it changes in resonance with the dynamic change of liquidity risk. When liquidity affects asset prices, dynamic credit risk usually interacts with liquidity risk in a dynamic spiral, forming the resonance enhancement interaction between liquidity and credit risk to dynamically affect asset prices. This may be affecting mechanism of the interaction between liquidity and credit risk for asset prices.

From the perspective of the market transmission mechanism of risk, if investors hold assets with high credit risk, some informed traders will take the lead in selling assets when adverse information comes out. Massive asset sales can cause asset prices to fall rapidly and make some investors panic. In the case of information asymmetry, some investors will follow to sell assets at lower prices and suffer liquidity losses. Further decline of asset prices will increase liquidity risk. Information outflow and information asymmetry become the spiral transmission path of the interaction between liquidity and credit risk on asset prices.

From an economic point of view, investors holding illiquid assets need liquidity risk compensation, similarly, investors holding defaultable assets need credit risk compensation. Default risk and liquidity are linked through investors' transaction based on information asymmetry, which will inevitably lead to a series of dynamic economic transmission impacts of liquidity and credit risk, and then affect asset prices spirally. Therefore, we introduce interaction between liquidity and credit risk into corporate bond spread model to study their affecting mechanism on asset prices and risk transmission channels.

Based on the above analysis, the impact of the interaction between liquidity and credit risk on corporate bond spreads

Table 3: ADF unit root test.

	Spread	Turnover	Vol	Day	Amihud	Range	Time-to-maturity	Age	Volatility	Coupon	Rating
Panel A. The interbank market											
t-statistic	−9.9746***	−31.5022***	−27.1805***	−22.0632***	−157.0229***	−156.2083***	−9.7004***	−14.8666***	−13.1248***	−18.5285***	−33.6873***
1% level	−3.4304	−3.4304	−3.4304	−3.4304	−3.4304	−3.4304	−3.4304	−3.4304	−3.4304	−3.4304	−3.4304
Prob.*	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Panel B. The exchange market											
t-statistic	−9.9574***	−31.6225***	−33.1601***	−10.9301***	−53.5030***	−53.5321***	−6.8669***	−22.4293***	−8.6634***	−22.5326***	−75.8180***
1% level	−3.4305	−3.4305	−3.4305	−3.4305	−3.4305	−3.4305	−3.4305	−3.4305	−3.4305	−3.4305	−3.4305
Prob.*	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Panel C. The whole sample											
t-statistic	−11.5249***	−18.8832***	−20.0654***	−15.9010***	−43.0203***	−41.7255***	−10.0992***	−22.1840***	−15.6788***	−21.7887***	−58.7438***
1% level	−3.4303	−3.4303	−3.4303	−3.4303	−3.4303	−3.4303	−3.4303	−3.4303	−3.4303	−3.4303	−3.4303
Prob.*	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

The table shows the ADF unit root test results for the variables used in the regression analysis. The test results include the test of the *t* statistic, the 1% significance level, and the probability of receiving the original hypothesis. The meanings of the variables are given in Table 1. The sample period is July 2006 to June 2016. The subsample (panel A) includes 2281 bond issues and 25648 observations. The subsample (panel B) includes 1435 bond issues and 21788 observations. The whole sample (panel C) includes 3716 bond issues and 47436 observations. "*," "**," and "***' indicate significance at the 10%, 5%, and 1% levels, respectively.

is discussed in this part. The premium regression equation that contains the interaction is as follows

$$\begin{aligned}\text{Spread}_{it} &= \alpha + \beta_1 \times (\text{liquidity risk factors})_{it} + \beta_2 \times (\text{credit risk factors})_{it} \\ &+\beta_3 \times \text{interaction function}\,(\text{liquidity risk factors, credit risk factors})_{it} + \varepsilon_t.\end{aligned} \tag{1}$$

To investigate whether the impact of interaction on the spreads is present, the following assumption is given

$$\begin{aligned}\text{Spread}_{it} &= \alpha + \beta_1 \times (\text{liquidity risk factors})_{it} + \beta_2 \times (\text{credit risk factors})_{it} \\ &+\beta_3 \times (\text{liquidity risk factors})_{it} \times (\text{credit risk factors})_{it} + \varepsilon_t.\end{aligned} \tag{2}$$

If β_3 is significantly different from zero, the impact of the interaction on the spreads is present. For estimating the impact of the interaction, we run a pooled regression of bond spreads in the interbank market, exchange market and the whole sample. The term "Liquidity risk factors" represents a set of possible proxies for liquidity risk from three dimensions, including transaction activity (turnover, vol, and day), price impact (amihud and range), and bond survival time (time-to-maturity, age). Similarly, the term "credit risk factors" represents a set of possible proxies for the default risk, including coupon and volatility. The cross terms of the proxies of liquidity and credit risk represent their interaction. If the regression coefficient of the cross term is significantly different from zero, the interaction has a significant impact on the bond spreads. In order to guarantee the reliability of the research results, we have made empirical tests from many perspectives.

3.2.1. The Empirical Results Based on Liquidity Proxies of Transaction Activity. China's bond market is the second largest in the world after the America. According to the People's Bank of China, by the end of December 2021, the stock of bonds in China's bond market is about $21 trillion, while that in the US is about $50 trillion. In 2021, China's bond issuance reach about $9.6 trillion, accounting for about 54% of China's GDP ($17.7 trillion) and 11% of global GDP ($90 trillion). It shows that China's bond market has become an important factor affecting China's economy and global economy. According to the People's Bank of China, the stock of corporate debenture bonds is about $5 trillion by the end of December 2021. Corporate Bonds are important parts of China's bond market. Therefore, it is of great significance to study the impact factors of Chinese corporate bond prices.

China's corporate bond market consists of interbank market and exchange market. In the interbank market, commercial banks, insurance companies, securities firms and other financial institutions buy and sell bonds, and participants ask selected counterparties to complete the transaction. The exchange bond market is dominated by nonbank financial institutions and individuals. China has two types of corporate bonds. One is issued by central government departments or state-owned enterprises, and the other is issued by listed companies. The former can be traded on the exchange market and the interbank market, while the latter can only be traded on the exchange market. This shows that both of markets are important trading places for corporate bonds in China, but the types of corporate bonds traded in the two markets are different, and the structures of investors are different. Therefore, it is necessary to study the impact factors of corporate bond spreads in the two markets. In order to explore the impact of the two markets on the test results, we, respectively, analyze the sample of whole market and subsamples of the interbank market and exchange market. In the whole sample, the dummy variable is introduced to test the significance of the impact of the interbank market and the exchange market on the test results. Where, the dummy variable of the interbank market is set to 1, and the dummy variable of the exchange market is set to 0. The basic proxies of credit risk and liquidity are given, including Vol, Day, Range, Age, Coupon and Volatility. The empirical results are shown in Tables 4 and 5.

In Table 4, Model 1 gives the regression results that include liquidity, and Model 2 gives the regression results including credit risk and liquidity. Compared with Model 2, the results of Model 1 show that the liquidity premiums are the main components of corporate bond spreads. Corporate bonds are rarely traded in China's bond market, which results in poor liquidity. In Model 1 and Model 2, the regression coefficients of Vol are −0.0020 and −0.0018, respectively, and they are significant at the level of 1%, which is consistent with the hypothesis. The greater the values is, more active the bond transaction is. It shows that the bond liquidity risk is smaller, and the spreads are smaller [3, 12, 37]. The regression coefficients of Turnover and Day are positive, and they are inconsistent with the hypothesis. China's corporate bond market trades infrequently, and the increase of turnover rates and trading days may be due to speculative trading. In Model 1 and Model 2, the regression coefficients of Amihud are 0.3513 and 0.2517, respectively, and they are significant at the level of 1%, which is consistent with the hypothesis. The results of model 1 show that for

Table 4: The impact of interaction between liquidity and credit risk on spreads based on transaction activity in the whole sample.

	1	2	3	4	5	6
C	2.5912***	0.7928***	0.3340***	0.2715***	0.3530***	0.3298***
	(111.7038)	(11.0667)	(6.96)	(5.78)	(8.66)	(6.87)
Dummy	0.4812***	0.4367***	0.2687***	0.2065***	−0.1371***	0.2494***
	(20.8164)	(19.3486)	(22.38)	(17.42)	(−13.11)	(20.73)
Turnover	1.9158***	1.5967***		−0.0725	2.2141***	1.7381***
	(12.7908)	(10.8800)		(−0.30)	(5.24)	(4.28)
Vol	−0.0020***	−0.0018***	−0.0013***		−0.0018***	−0.0020***
	(−13.6127)	(−12.4066)	(−5.38)		(−4.49)	(−5.01)
Day	0.0737***	0.0650***	−0.0108**	−0.0122**		−0.0132**
	(37.6845)	(33.6513)	(−1.96)	(−2.22)		(−2.42)
Amihud	0.3513***	0.2517***	0.2506***	0.2480***	−0.1368**	0.2487***
	(4.0216)	(2.9561)	(4.41)	(4.36)	(−2.34)	(4.39)
Age	−0.2030***	−0.1739***	−0.1349***	−0.1314***	−0.1254***	−0.1313***
	(−32.1614)	(−27.7859)	(−49.35)	(−47.86)	(−44.52)	(−47.99)
Volatility		0.2141***	−0.2482***	−0.2266***	0.1438***	−0.2437***
		(8.8111)	(−13.70)	(−12.54)	(9.78)	(−13.42)
Coupon		0.2422***	0.3315***	0.3391***	0.3606***	0.3310***
		(25.0131)	(49.42)	(51.69)	(66.47)	(49.26)
Turnover × volatility				0.2046**	−0.3269**	−0.2896*
				(2.28)	(−1.97)	(−1.81)
Turnover × coupon				0.0081	−0.1077*	−0.0624
				(0.24)	(−1.84)	(−1.11)
Vol × volatility			0.0005***		0.0005***	0.0008***
			(6.33)		(3.69)	(6.05)
Vol × coupon			0.0001***		0.0001*	0.0000
			(2.93)		(1.72)	(0.57)
Day × volatility			0.0614***	0.0615***		0.0611***
			(26.65)	(26.65)		(26.58)
Day × coupon			0.0052***	0.0050***		0.0055***
			(7.18)	(6.94)		(7.60)
aj−R2	0.1752	0.2184	0.2682	0.2669	0.2117	0.2723

To investigate whether the impact of interaction on the spreads is present, the following regression equations are given. $\text{Spread}_{it} = \alpha + \beta_1 \times (\text{liquidity risk factors})it + \beta_2 \times (\text{credit risk factors})it + \beta_3 \times (\text{transaction activity liquidity risk factors})it \times (\text{credit risk factors})it + \epsilon t$. The meanings of the variables are given in Table 1. The sample period is July 2006 to June 2016. The whole sample includes 3716 bond issues and 47436 observations. The *t*-statistics are shown in parentheses below each coefficient estimate, and "*," "**," and "***" indicate significance at the 10%, 5%, and 1% levels, respectively.

every 1 unit increase in liquidity risk of price shock dimension, corporate bond spreads increase by 0.3513 unit. See also Shin and Kim [37] and Han and Zhou [30]. In Model 1 and Model 2, the regression coefficients of Age are −0.2030 and −0.1739, respectively, and they are significant at the level of 1%, which is consistent with the hypothesis. Because of information asymmetry, investors will not buy and trade new bonds. Over time, investors will choose to buy the bonds. Then the transactions become active, and the spreads begin to become small. Goldstein et al. [33] and Hotchkiss and Jostova [34] find similar results. In Model 2, the regression coefficients of Volatility and Coupon are 0.2141 and 0.2422, respectively, and they are significant at the level of 1%, which is consistent with the hypothesis. Volatility and Coupon are the proxies of credit risk. The greater the value is, the worse credit is, and the greater the spreads are. So Volatility and Coupon have positive relation with spreads [37, 39, 40].

The regression results containing interaction terms are given in Model 3. Model 4 and Model 5 gives the robustness analysis results. In addition, the regression analysis results of the all transaction activity dimension are given in Model 6. Cross term represents the corresponding interaction proxy. If the coefficient of cross term is significantly different from zero, the impact of the interaction on the spreads is significant. According to the test results of Table 4, the coefficient of dummy variable is significant at the level of 1%, which indicates that the differences between the interbank market and the exchange market cannot be ignored.

In Table 4, regression results of Model 3 show that the coefficient of vol is significant at the level of 1%, and the sign is in line with the hypothesis, while the regression coefficient of coupon is very significant. Meanwhile, the coefficient of the cross term of vol and coupon is very significant. The coefficient of the cross term of vol and coupon is significantly different from zero, which indicates that the interaction has impacts on the spreads. In Model 3, the regression coefficient of volatility is significant, but the sign is not in line with the hypothesis. It shows that when the trading volume of high-credit risk assets increases, the panic selling of some investors due to asymmetric information leads to increase asset price volatility. The increase of the volatility leads to the increase of speculative liquidity risk. And the spiral increase of speculative liquidity risk and static credit risk leads to the impact of the interaction between liquidity and credit risk on

asset prices. The coefficients of day, coupon and day × coupon have similar characteristics.

In Model 4, the coefficients of day, coupon and day × coupon are significant, which shows that the interaction between day and coupon has a significant impact on corporate bond spreads. The reason is that static credit risk does not weaken when investment liquidity increases, and the interaction between liquidity and credit risk affects asset prices in the same direction as static credit risk.

In Model 5, the coefficient of the interaction between vol and volatility is greater than one of the interaction between vol and coupon. And they have enhanced impacts on bond spreads. It shows that dynamic credit risk changes in resonance with the dynamic change of liquidity risk, while the static credit risk does not change with the dynamic change of liquidity risk. When liquidity affects asset prices, dynamic credit risk interacts with liquidity risk in a dynamic spiral, forming a bigger resonance enhancement interaction between liquidity and credit risk to dynamically affect asset prices than static credit risk. Li and Song [50] study the pricing of China's convertible bonds and find that the dynamic credit risk is important in convertible bond pricing.

Regression analysis results are given in Model 6, which contains all the liquidity risk proxies of the transaction activity. The coefficients of vol × volatility, day × volatility and day × coupon are significant, which shows that the interaction between liquidity and credit risk has a significant impact on spreads. The coefficients of turnover × coupon and vol × coupon are not significant, and the reason is that when liquidity affects asset prices, static credit risk generally does not interact with liquidity risk in a dynamic spiral. From Model 1 to Model 6, the coefficients of age are all significant at the level of 1%, and the signs are negative. Chinese investors tend to be cautious about newly listed bonds due to information asymmetry, so they prefer to buy corporate bonds that trade for many years, which causes bond liquidity to increase over time. Therefore, the high age bonds have lower bond spreads in China's corporate bond market. Relevant literature can also be referred to Zhu [51].

In Table 5, the empirical results of the interbank market and the exchange market are given. Regression results of Model 1 show that the regression coefficient of vol is significant at the level of 1% and the sign is in line with the hypothesis on the exchange market, while the regression coefficient of day is significant at the level of 1% and the sign is in line with the hypothesis in the interbank market. The reason is that exchange market is dominated by individual investors with many trading days but few trading volumes, so trading volume is a good proxy for liquidity in the exchange market. On the other hand, interbank market is dominated by institutional investors with few trading days and high trading volumes, so trading day is a better proxy for liquidity in the interbank market. In the interbank market, the assumption that the coefficient of day × coupon is equal to zero is rejected, which indicates that the interaction between liquidity and credit risk has a significant impact on the spreads in Model 1. When investors hold high credit risk bond, with the increase of trading days, some investors panic sell-off due to asymmetric information, that is to say, they are worried about potential default risk to sudden frequent trading bonds, which leads to spiral interaction between liquidity and credit risk, resulting in corporate bond spreads widen. The regression coefficient of vol × coupon has similar conclusion on the exchange market.

Robustness analysis are done in Model 2 and Model 3. The empirical results show that the regression coefficients of the interaction between liquidity and credit risk are all significant in the exchange market, while some of these coefficients are not significant in the interbank market. The reason is that the interbank market is mainly composed of institutional investors with infrequent transactions, leading to the discontinuous linkage between liquidity and credit risk and the failure to form continuous transmission and feedback of information. This shows that the establishment of a continuous market trading mechanism, which makes information flow and reduces information asymmetry, is conducive to price discovery.

Regression analysis results are given in Model 4, which contains all the liquidity risk proxies of the transaction activity. The regression coefficients of interaction between liquidity and credit risk are all significant in the exchange market, while the regression coefficients of Turnover × Coupon and Day × Volatility are not significant in the interbank market. The reason is that investor structure and transaction characteristic are different in the interbank market and the exchange market.

Through the above analysis, interaction between liquidity and credit risk cannot be ignored and liquidity premiums are the main components of corporate bond spreads in China's bond market. Dynamic credit risk changes in resonance with the dynamic change of liquidity risk, while the static credit risk does not change with the dynamic change of liquidity risk. When liquidity affects asset prices, dynamic credit risk interacts with liquidity risk in a dynamic spiral, forming a bigger resonance enhancement interaction between liquidity and credit risk to dynamically affect asset prices than static credit risk.

3.2.2. Subsample Results. In this part, we give sub samples regression analysis, and the liquidity and credit risk proxies are divided into high, medium and low in order to explore the impact of the interaction between liquidity and credit risk on the spreads.

In Table 6, there are nine subsamples. The regression results of the three subsamples are given according to the trading volume in the panel A. In the high-vol subsample, the regression coefficients of amihud × volatility, amihud × coupon, age × volatility and age × coupon are significant, while the regression coefficients of vol × volatility and vol × coupon are not significant. It suggests that the interactions between liquidity and credit risk in high-vol bonds are mainly the interaction between speculative liquidity of price shock dimension and credit risk and the interaction between investment liquidity and credit risk. In the low-vol subsample, the regression coefficient of amihud × volatility is not significant, suggesting that the spiral interaction between speculative liquidity of price shock dimension and dynamic

Table 5: The impact of interaction between liquidity and credit risk on spreads based on transaction activity in interbank market and exchange market.

	The inter bank market				The exchange market			
	1	2	3	4	1	2	3	4
C	0.3490***	0.2341***	0.2655***	0.3312***	0.8091***	0.6666***	0.4069***	0.7630***
	(5.63)	(3.83)	(4.80)	(5.32)	(10.59	(8.83)	(6.91)	(10.16)
Turnover		0.8623***	1.7225***	1.5566***		−1.8768	18.8014***	24.2606***
		(3.14)	(4.34)	(3.93)		(−1.39)	(6.97)	(9.39)
Vol	0.0003		−0.0012***	−0.0004	−0.0114***		−0.0246***	−0.0252***
	(1.15)		(−2.93)	(−0.99)	(−9.98)		(−10.86)	(−11.62)
Day	−0.0676***	−0.0324		−0.0766***	−0.0022	−0.0070		−0.0018
	(−3.08)	(−1.54)		(−3.47)	(−0.31)	(−1.00)		(−0.26)
Amihud	12.8027***	13.4242***	13.4778***	12.9596***	0.2895***	0.3265***	0.0028	0.2956***
	(8.80)	(9.22)	(9.36)	(8.93)	(4.96)	(5.65)	(0.05)	(5.16)
Age	−0.1474***	−0.1443***	−0.1417***	−0.1420***	−0.1309***	−0.1248***	−0.1087***	−0.1252***
	(−37.34)	(−36.25)	(−35.89)	(−35.79)	(−34.87)	(−33.47)	(−28.39)	(−33.83)
Volatility	−0.1971***	−0.1899***	−0.1939***	−0.1967***	−0.1253***	−0.0750**	0.1765***	−0.0762**
	(−8.06)	(−7.76)	(−9.88)	(−8.06)	(−3.83)	(−2.31)	(8.00)	(−2.37)
Coupon	0.3678***	0.3822***	0.3882***	0.3672***	0.2522***	0.2659***	0.3301***	0.2543***
	(41.14)	(43.26)	(49.53)	(40.86)	(23.62)	(25.18)	(43.42)	(24.23)
Turnover × volatility		0.2945**	−0.2244	−0.3161**		16.3455***	13.1480***	12.1640***
		(2.50)	(−1.43)	(−1.97)		(34.04)	(12.99)	(12.49)
Turnover × coupon		−0.1400***	−0.1075*	−0.0821		−0.4313**	−2.2377***	−3.2118***
		(−3.69)	(−1.96)	(−1.50)		(−2.38)	(−6.08)	(−9.09)
Vol × volatility	0.0007***		0.0009***	0.0010***	0.0138***		0.0042***	0.0046***
	(6.86)		(6.69)	(7.19)	(31.82)		(4.56)	(5.25)
Vol × coupon	−0.0002***		0.0001	−0.0002***	0.0008***		0.0026***	0.0026***
	(−4.15)		(−0.13)	(−3.10)	(5.40)		(8.55)	(8.73)
Day × volatility	−0.0124	0.0110		−0.0096	0.0215***	0.0181***		0.0144***
	(−1.09)	(0.99)		(−0.83)	(6.62)	(5.59)		(4.48)
Day × coupon	0.0207***	0.0094***		0.0215***	0.0065***	0.0068***		0.0065***
	(6.27)	(3.03)		(6.50)	(6.70)	(7.13)		(6.84)
aj−R2	0.2276	0.2247	0.2263	0.2318	0.3426	0.3546	0.3045	0.3649

To investigate whether the impact of interaction on the spreads is present, the following regression equations are given. $\text{Spread}_{it} = \alpha + \beta_1 \times$ (liquidity risk factors)it $+ \beta_2 \times$ (credit risk factors)it $+ \beta_3 \times$ (transaction activity liquidity risk factors)it $\times$ (credit risk factors)it $+ \epsilon$t. The meanings of the variables are given in Table 1. The sample period is July 2006 to June 2016. The subsample of the interbank market includes 2281 bond issues and 25648 observations. The subsample of the exchange market includes 1435 bond issues and 21788 observations. The t-statistics are shown in parentheses below each coefficient estimate, and "*," "**," and "***" indicate significance at the 10%, 5%, and 1% levels, respectively.

credit risk requires high trading volume as the basic condition.

In the high-amihud subsample of panel B, the regression coefficients of vol × volatility, age × volatility and age × coupon are significant, while the regression coefficients of vol × coupon, amihud × volatility, and amihud × coupon are not significant. It shows that high price shock is usually caused by a huge amount of investors' sudden trading. Generally, prices will recover calm immediately after a huge impact, which makes it difficult to form risk transmission and interaction between liquidity and credit risk in a very short period of time. In the subsamples of panel B, the regression coefficients of vol × coupon are all not significant. Coupon is static credit risk measure. The static credit risk represents the credit risk level of an firm in this period of time, and it does not change with the dynamic change of liquidity risk. When liquidity affects asset prices, static credit risk generally does not interact with liquidity risk in a dynamic spiral.

In the high-age subsample of panel C, the regression coefficients of vol × volatility, amihud × coupon, age × volatility and age × coupon are significant, while the regression coefficients of vol × coupon and amihud × volatility are not significant. It shows that high age bonds are actively traded, and sufficient trading volume forms the spiral interaction between speculative liquidity and dynamic credit risk and the interaction between liquidity of price shock dimension and static credit risk, thus affecting asset prices. In the low-age subsample of panel C, the regression coefficients of vol × coupon, age × volatility and age × coupon are significant, while the regression coefficients of vol × volatility, amihud × volatility, and amihud × coupon are not significant. The reason is that low age bonds are not actively traded, and the lack of trading volume only forms the interaction between investment liquidity and static credit risk, which indicates that the interaction between liquidity of price shock dimension and credit risk requires sufficient trading volume as the basic condition.

To sum up, the interaction between liquidity and credit risk will increase with the increase of liquidity risk and credit risk and it is a time-varying dynamic process. When market liquidity risk increases, some investors will take the lead in selling corporate bonds with high default risk, which causes the price of corporate bonds with high credit risk to fluctuate greatly. These behaviors of investors further cause more investors to sell these corporate bonds with high credit risk,

Table 6: The impact of interaction between liquidity and credit risk on spreads in subsample results.

	Panel A: vol			Panel B: amihud			Panel C: age		
	Hight	Middle	Low	Hight	Middle	Low	Hight	Middle	Low
c	−1.0268***	−0.4000***	−0.6494***	−0.6142***	−0.4396*	−0.6475***	0.2471	0.0498	−2.0406***
	(−5.84)	(−3.56)	(−6.10)	(−3.92)	(−1.84)	(−8.50)	(1.16)	(0.33)	(−19.95)
Dummy	0.3600***	0.3715***	0.3015***	0.4950***	0.3671***	0.3201***	0.1926***	0.2973***	0.2909***
	(5.80)	(21.04)	(6.43)	(4.70)	(8.28)	(20.60)	(5.90)	(19.01)	(17.81)
Vol	0.0003	−0.0101***	0.0110	−0.0652	−0.0091	−0.0010***	−0.0010	−0.0008**	0.0001
	(0.76)	(−6.49)	(0.52)	(−1.54)	(−1.24)	(−3.60)	(−1.46)	(−2.47)	(0.49)
Day	0.0947***	0.0616***	0.0341***	0.0259***	0.0466***	0.0621***	0.0583***	0.0608***	0.0373***
	(23.33)	(42.81)	(20.30)	(7.42)	(19.39)	(50.00)	(22.49)	(46.14)	(26.67)
Amihud	−1940.6200***	−83.3530**	0.8493**	0.6793	−12.5975	−280.7775***	1.7829**	−0.0439	−0.9569
	(−5.15)	(−2.24)	(2.27)	(1.58)	(−0.37)	(−3.86)	(2.03)	(−0.09)	(−1.16)
Age	0.2520***	0.2224***	0.2046***	0.2015***	0.1688***	0.1779***	−0.0557	−0.0841	1.0736***
	(6.72)	(10.34)	(9.73)	(5.75)	(5.18)	(10.33)	(−1.30)	(−1.18)	(6.90)
Volatility	0.0909	0.0558	0.1072**	0.1180*	0.1472	−0.0227	−0.4399***	0.8224***	−0.2797***
	(1.32)	(1.16)	(2.38)	(1.85)	(1.52)	(−0.62)	(−4.82)	(12.72)	(−6.19)
Coupon	0.4754***	0.4229***	0.4915***	0.4886***	0.4350***	0.4604***	0.3421***	0.3014***	0.7048***
	(20.58)	(25.74)	(32.98)	(22.75)	(12.88)	(41.73)	(9.34)	(14.22)	(48.18)
Vol × volatility	−0.0001	0.0039***	0.0248***	0.0188**	0.0003	0.0006***	0.0007***	0.0006***	0.0000
	(−1.06)	(5.90)	(2.95)	(2.34)	(0.17)	(7.15)	(3.51)	(6.76)	(−0.02)
Vol × coupon	−0.0001	0.0011***	−0.0054*	0.0015	0.0005	0.0000	0.0002	0.0000	−0.0001**
	−1.44	(5.02)	(−1.82)	(0.26)	(0.54)	(1.02)	(1.51	(−0.26)	(−2.47)
Amihud × volatility	400.6426***	15.0529*	−0.0999	−0.1028	19.2996	317.0865***	−0.3220	−0.3831**	0.0986
	(4.79)	(1.94	(−0.82)	(−0.74)	(1.30)	(10.76)	(−1.04)	(−2.21)	(0.63)
Amihud × coupon	238.7952***	9.6909*	−0.0996**	−0.0854	1.5526	23.8625**	−0.2105*	0.0782	0.1160
	(4.60)	(1.89)	(−1.96)	(−1.47)	(0.32)	(2.28)	(−1.83)	(1.13)	(1.10)
Age × volatility	0.0498**	−0.0736***	−0.0321***	−0.0301*	−0.0702***	−0.0186**	0.0840***	−0.2944***	0.7791***
	(2.50)	(−6.77)	(−3.11)	(−1.85)	(−4.14)	(−2.02)	(4.37)	(−10.13)	(10.95)
Age × coupon	−0.0648***	−0.0552***	−0.0563***	−0.0522***	−0.0478***	−0.0511***	−0.0175**	0.0096	−0.2350***
	(−11.01)	(−16.27)	(−16.63)	(−9.31)	(−9.00)	(−18.76)	(−2.19)	(0.95)	(−10.66)
aj−R2	0.2512	0.2789	0.2984	0.2518	0.3123	0.2679	0.1349	0.2178	0.4787

To investigate whether the impact of interaction on the spreads is present, the following regression equations are given. $\text{Spread}_{it} = \alpha + \beta_1 \times (\text{liquidity risk factors})it + \beta_2 \times (\text{credit risk factors})it + \beta_3 \times (\text{different dimension liquidity risk factors})it \times (\text{credit risk factors})it + \epsilon t$. The meanings of the variables are given in Table 1. The sample period is July 2006 to June 2016. The whole sample includes 3716 bond issues and 47436 observations. The t-statistics are shown in parentheses below each coefficient estimate, and "*," "**," and "***" indicate significance at the 10%, 5%, and 1% levels.

so the investors who hold these high credit risk bonds have to bear the gradually increasing liquidity risk. With the increase of market liquidity risk, firms with high default risk are affected due to the depletion of operating cash flow, and the probability of default will also increase. Because of information asymmetry, investors sell their bonds more quickly, creating a spiral interaction between liquidity and credit risk that affects asset prices. If market liquidity risk continues to increase, corporate bonds with low credit risk will gradually be affected by the interaction between liquidity and credit risk, and eventually more corporate bond prices will be affected by risk contagion.

4. Robustness Test

4.1. Robustness Test Based on Liquidity. Liquidity has many dimensions, and price impact is one of the most important dimensions of liquidity. Famous measures are amihud and range [29, 30, 37]. Controlling other liquidity and credit risk measure proxies, we use the measures of the liquidity of the price impact dimension to discuss the robustness of impact of the interaction between liquidity and credit risk on the spreads. The basic proxies of liquidity and credit risk are given, including vol, day, amihud, age, coupon and volatility. Amihud measure reflects the impact of a unit of trading volume on the price, while Range measure is the impact of a unit of trading volume on the highest and lowest price, and both of which represent the liquidity measure of the price shock dimension [20, 30]. Therefore, robustness analysis is done by range to replace amihud, and coupon and volatility are proxies of credit risk. The cross terms are the proxies of the interaction between liquidity and credit risk.

To investigate the robustness of impact of interaction between liquidity and credit risk on the spreads, the following regression equation is given

$$\begin{aligned}\text{Spread}_{it} &= \alpha + \beta_1 \times (\text{liquidity risk factors})_{it} + \beta_2 \times (\text{credit risk factors})_{it} \\ &+\beta_3 \times (\text{price impact liquidity risk factors})_{it} \times (\text{credit risk factors})_{it} + \varepsilon_{it}.\end{aligned} \tag{3}$$

If β_3 is significantly different from zero, the impact of the interaction on the spreads is present, and the result is robust. The term "liquidity risk factors" represents a set of basic proxies for liquidity from three dimensions, including transaction activity (vol, day), price impact (range) and bond survival time (age). The term "price impact Liquidity risk factors" represents price impact, including amihud or range or both. The term "credit risk factors" represents the proxies for the default risk, including coupon and volatility. In Table 7, the regression results of the interaction between liquidity and credit risk are given in the whole sample and subsamples.

In the whole sample, the dummy variable is introduced to test the significance of the impact of the interbank market and the exchange market on the test results, where the dummy variable of the interbank market is set to 1, and the one of the exchange market is set to 0. As shown in the Model 1 to Model 3, the regression coefficients of the dummy variables are all significant. In order to ensure the robustness of the test results, we do regression analysis of the interbank market and the exchange market independently in Table 7.

In the whole sample, the regression results of Model 1 show that the regression coefficients of amihud, volatility and coupon are significant at the level of 1%, and the coefficients of amihud × volatility and amihud × coupon are significant at the level of 5%. This shows that the liquidity of price shock dimension interacts with both dynamic credit risk and static credit risk, thus affecting asset prices. In Model 2, amihud is replaced by range to test the robustness of the results. The regression coefficients of range, volatility, coupon and range × volatility are significant at the level of 1%, while the regression coefficient of range × coupon is not significant. The reason is that the static credit risk represents the credit risk level of an firm in this period of time, and it does not easily interact with liquidity risk. In Model 3, amihud and range are introduced into the regression equation. The regression coefficients of volatility and coupon are very significant, while the coefficients of amihud and amihud × coupon are significant at the level of 10% and the coefficient of range is not significant. The reason may be the high correlation between amihud and range. In spite of this, the results still prove that the impact of the interaction between liquidity and credit risk on the spreads is not negligible, and the results are robust in the Chinese corporate bond market.

In the interbank market, the regression results of Model 1 show that the regression coefficients of amihud, coupon and amihud × coupon are significant at the level of 1%, and the signs are in line with expectation. It shows that the impact of the interaction between liquidity and credit risk on the spreads is present. In Model 1, the regression coefficient of amihud × volatility is greater than one of amihud × coupon. It shows that the liquidity of price shock dimension has a greater interaction with dynamic credit risk than static credit risk. Dynamic credit risk represents the credit risk level of an firm over time, and it changes in resonance with the dynamic change of liquidity risk. When liquidity affects asset prices, dynamic credit risk interacts with liquidity risk in a dynamic spiral, forming the resonance enhancement interaction between liquidity and credit risk to dynamically affect asset prices. In Model 2, amihud is replaced by range to test the robustness of the results. The regression coefficient of range × volatility is also greater than one of Range × Coupon, and this result is consistent with Model 1. In Model 3, amihud and range are introduced into the regression equation. The regression coefficients of amihud × volatility and range × volatility are greater than ones of amihud × coupon and amihud × coupon, respectively. The

Table 7: The impact of interaction between liquidity and credit risk on spreads in the dimension of price impact.

	The whole sample			The interbank market			The exchange market		
	1	2	3	1	2	3	1	2	3
C	0.0555	0.0571*	0.0533	0.1860***	0.1894***	0.1768***	0.0273	0.0251	0.0246
	(1.61)	(1.65)	(1.54)	(4.40)	(4.49)	(4.19)	(0.50)	(0.46)	(0.45)
Dummy	0.2474***	0.2502***	0.2505***						
	(20.64)	(20.83)	(20.84)						
Vol	−0.0003***	−0.0003***	−0.0003***	−0.0003***	−0.0003***	−0.0003***	0.0013***	0.0013***	0.0013***
	(−6.79)	(−6.78)	(−6.77)	(−5.89)	(−5.97)	(−5.65)	(7.84)	(8.01)	(8.01)
Day	0.0569***	0.0570***	0.0570***	0.0479***	0.0478***	0.0468***	0.0523***	0.0524***	0.0524***
	(56.54)	(56.70)	(56.63)	(12.34)	(12.29)	(12.02)	(45.24)	(45.39)	(45.29)
Amihud	1.1900***		0.9831*	51.1388***		41.3188***	0.9672**		0.6071
	(3.14)		(1.70)	(6.45)		(3.33)	(2.41)		(1.00)
Range		0.5741***	0.2260		34.0069***	19.6406**		0.7470***	0.4286
		(2.81)	(0.71)		(7.06)	(2.50)		(3.44)	(1.28)
Age	−0.1399***	−0.1399***	−0.1399***	−0.1500***	−0.1494***	−0.1495***	−0.1293***	−0.1297***	−0.1295***
	(−51.33)	(−51.33)	(−51.32)	(−38.31)	(−38.18)	(−38.23)	(−33.62)	(−33.73)	(−33.68)
Volatility	0.0891***	0.0918***	0.0903***	−0.1565***	−0.1581***	−0.1550***	0.4001***	0.4092***	0.4119***
	(7.00)	(7.19)	(7.05)	(−9.78)	(−9.92)	(−9.71)	(19.07)	(19.34)	(19.36)
Coupon	0.3508***	0.3500***	0.3506***	0.3938***	0.3936***	0.3949***	0.3286***	0.3278***	0.3277***
	(77.25)	(76.96)	(76.92)	(67.41)	(67.62)	(67.66)	(45.77)	(45.50)	(45.32)
Amihud × volatility	−0.2631**		0.0396	38.8328***		−23.1397***	−0.8894***		−0.3006
	(−2.17)		(0.22)	(17.27)		(−3.87)	(−6.90)		(−1.60)
Amihud × coupon	−0.1123**		−0.1464*	−14.1722***		−2.2058	−0.0304		−0.0677
	(−2.20)		(−1.90)	(−10.47)		(−0.98)	(−0.56)		(−0.84)
Range × volatility		−0.2660***	−0.2518**		36.2733***	52.1587***		−0.5962***	−0.4727***
		(−3.86)	(−2.40)		(19.31)	(11.11)		(−8.12)	(−4.29)
Range × coupon		−0.0311	0.0197		−11.6695***	−11.8540***		−0.0264	0.0111
		(−1.20)	(0.49)		(−10.64)	(−6.35)		(−0.95)	(0.26)
aj−R2	0.2564	0.2565	0.2566	0.2334	0.2364	0.2371	0.3004	0.3013	0.3013

To investigate whether the impact of interaction on the spreads is present, the following regression equations are given. $\text{Spread}_{it} = \alpha + \beta_1 \times (\text{liquidity risk factors})_{it} + \beta_2 \times (\text{credit risk factors})_{it} + \beta_3 \times (\text{price impact liquidity risk factors})_{it} \times (\text{credit risk factors})_{it} + \epsilon t$. The meanings of the variables are given in Table 1. The sample period is July 2006 to June 2016. The subsample (the interbank bond market) includes 2281 bond issues and 25648 observations. The subsample (the exchange market) includes 1435 bond issues and 21788 observations. The whole sample includes 3716 bond issues and 47436 observations. The *t*-statistics are shown in parentheses below each coefficient estimate, and "*," "**," and "***" indicate significance at the 10%, 5%, and 1% levels, respectively.

results prove that the liquidity of price shock dimension has a greater interaction with dynamic credit risk than static credit risk, and the impact of the interaction between liquidity and credit risk on the spreads is present in the interbank bond market.

Empirical studies show that similar results are found in the exchange market. In the exchange market, the regression coefficients of coupon, and volatility are significant at the level of 1%, and the signs are in line with the hypothesis. In Model 1, the regression coefficient of amihud × volatility is significant at the level of 1%, while the regression coefficient of amihud × coupon is not significant. It shows that the liquidity of price shock dimension has a greater interaction with dynamic credit risk than static credit risk. In Model 2, amihud is replaced by range to test the robustness of the results, and the similar results are found. This shows that the results are robust. In Model 3, the coefficients of amihud and range are not significant. Because the correlation between amihud and range is high.

The above analysis shows that the liquidity of price shock dimension has a greater interaction with dynamic credit risk than static credit risk, and the impact of the interaction between liquidity and credit risk on the spreads is not negligible in the interbank market and the exchange market. And the results are robust. This suggests that similar results are found in the two markets. When investors encounter liquidity risk, they panic to sell assets with high credit risks due to information asymmetry, and the behaviors of investors are consistent in the two markets. Because the behaviors of investors are determined by human egoism.

4.2. Robustness Test Based on Credit Risk and Macrofactors. In order to investigate whether the results are stable, we perform robustness analysis based on credit risk and macrofactors. We have carried out a lot of robustness tests of liquidity risk, and the results are stable. Rating is the credit rating assigned to each security that a firm issues [14, 37]. The higher the credit rating, the lower the credit risk. Coupon refers to the coupon rate of bonds [20, 38]. The lower the coupon, the less credit risk. It shows that rating and coupon are proxy variables of credit risk. So we use rating to replace coupon as the proxy of credit risk, in order to explore whether the interaction between liquidity and credit risk is present on the spreads. We choose the basic liquidity and credit risk proxies, and use their cross terms to represent the interaction. The empirical tests of the whole period, the financial crisis and the normal period are done in Panel A, Panel B and Panel C of Table 8. And the empirical tests of macro factors are done in Panel D of Table 8.

In Panel A, Panel B and Panel C of Table 8, the dummy variable is introduced to test the significance of the impact of the interbank market and the exchange market on the test results. Where, the dummy variable of the interbank market

is set to 1, and the one of the exchange market is set to 0. As shown in Table 8, the regression coefficients of the dummy variables are significant in normal period, which indicates that the impact of the differences between the two markets on the spreads is significant in normal period. However, the coefficient of dummy variable is not significant during the financial crisis. The reason may be that liquidity and credit risk have major impact on the spreads during the financial crisis than the differences between the two markets. In the whole period, the regression coefficients of vol, amihud, volatility, rating, vol × volatility, vol × rating, amihud × volatility and amihud × rating are significant at the level of 1%, and the signs are in line with expectation, which shows that the interaction between liquidity and credit risk is present on the spreads and the results are robust in the Chinese corporate bond market.

As shown in Table 8, the coefficient of Vol is increased from −0.0009 to −0.0010 from the normal period to the financial crisis, while the coefficient of age is dropped from −0.1362 to −0.1141 from the normal period to the financial crisis. This shows that compared with normal period, speculative liquidity risk premiums are increased and investment liquidity risk premiums are reduced during the financial crisis. During the financial crisis, many investors have no confidence in the future and become less willing to hold assets for the long term. Some investors begin to implement the short-term investment strategy of quick entry and quick exit, which leads to the increase of the impact of speculative liquidity and the decrease of the impact of investment liquidity on asset prices. The coefficient of Volatility is increased from 0.3131 to −0.4597 from the normal period to the financial crisis, while the coefficient of Rating is dropped from −0.2076 to −0.0459 from the normal period to the financial crisis. This shows that compared with normal period, dynamic credit risk premiums are increased and static credit risk premiums are reduced during the financial crisis. During the financial crisis, some investors sell assets in a panic. In the case of insufficient liquidity, the price fluctuation of assets with high default risk will increase, and the increase of price fluctuation will lead more investors to sell the assets with high default risk. Therefore, during the financial crisis, the impact of dynamic credit risk on asset price will gradually increase compared with static credit risk.

During the financial crisis, the regression coefficients of vol, rating and vol × rating are significant at the level of 1%, and the signs are in line with expectation, while the regression coefficients of vol, amihud, volatility, rating, vol × volatility, and amihud × rating are significant at the level of 1%, and the signs are in line with expectation in normal period. This shows that the impact of the interaction between liquidity and credit risk cannot be ignored. In addition, the coefficients of age × volatility and age × rating are increased from −0.0473 and −0.0059 to 0.1181 and −0.0351 from the normal period to the financial crisis, respectively. It shows that the interaction between liquidity and credit risk has greater impact on the spreads during the financial crisis. And the results are robust. During financial crises, some investors prefer to sell assets with high default risk, which leads to large price swings. Due to information asymmetry, this will cause more investors to continue to sell assets with high default risk, further forming a bigger liquidity crisis. Finally, it leads to the incrementally increasing spiral interaction between liquidity and credit risk during financial crises.

In Panel D of Table 8, we perform robustness analysis based on macro factors. Some scholars find that macro factors affect corporate bond spreads. Han and Zhou [30] focus on the linkage between the corporate bond spreads and the macroeconomic conditions. See also Yang et al. [20]. In order to study macro factors, we construct matched pairs of bonds by stripping out credit risk. Relevant studies refer to Dick-Nielsen et al. [12] and Helwege et al. [13]. We study the impact of the interaction between liquidity and credit risk on corporate bond spreads under macro factors. BML_Vol is the monthly trading volume of the corporate bond market. BML_Amihud is illiquidity measures of the corporate bond market based on Amihud. Year102 is the difference between the 10-year treasury rate and the 2-year Treasury rate. S10 year is the spread between the bond market index yield and the 10-year Treasury rate. CPI is the consumer price index, and GDP is gross domestic product. M0 is an increase in cash in circulation. In Panel D of Table 8, vol, turnover and amihud are liquidity measures of matched pairs of bonds. The cross-term is the interaction between liquidity and credit risk. The regression coefficients of Vol and Turnover are significant at the level of 1%, and the signs are in line with expectation, while the regression coefficients of vol × coupon and turnover × rating are significant at the level of 1%. This shows that under the condition of controlling macrofactors, the interaction between liquidity and credit risk has a significant impact on corporate bond spreads. In Panel D of Table 8, the regression coefficients of BML_Vol, Year102, S10 year, CPI, GDP, and M0 are significant at the level of 1%. Macrofactors affect corporate bond spreads by common liquidity factors in China's corporate bond market. During the financial crisis, the impact of the interaction between liquidity and credit risk on corporate bond spreads is closely related to macro factors.

4.3. Robustness Test Based on Credit and Liquidity Levels. In Table 9, we divide the sample into six subsamples to investigate the robustness based on credit and liquidity levels. The sample is divided into three subsamples according to the three levels of coupon, and then each subsample is divided into two subsamples according to the two levels of Vol. The cross terms of vol, amihud, age and coupon are used as measures of the interaction between liquidity and credit risk, and the regression results are shown in Table 9. These cross terms are vol × coupon, amihud × coupon and age × coupon.

In the high-coupon subsample, the regression coefficient of vol × coupon in the low-vol sample is significant at the level of 1%, while the regression coefficient of vol × coupon in the high-vol sample is not significant. In middle-coupon and low-coupon sub sample, the regression coefficient of vol × coupon also have similar results. When trading volume

Table 8: Robustness test of impact of interaction between liquidity and credit risk on spreads based on credit risk and macrofactors.

C	*Dummy*	*Vol*	*Day*	*Amihud*	*Age*	*Volatility*	*Rating*	*Vol × volatility*	*Vol × rating*	*Amihud × volatility*	*Amihud × rating*	*Age × volatility*	*Age × rating*	*aj-R2*
Panel A: the whole period														
1.8930***	0.2326***	−0.0010***	0.0596***	1.8066***	−0.1658***	0.0823***	0.1884***	0.0006***	0.0001***	−0.5708***	−0.3607***	−0.0145**	−0.0002	0.2262
(76.10)	(18.85)	(−9.22)	(57.82)	(10.05)	(−28.17)	(3.20)	(32.61)	(7.37)	(4.59)	(−4.64)	(−8.63)	(−2.03)	(−0.15)	
Panel B: during the financial crisis														
1.8560***	0.0122	−0.0010***	0.0042	−0.3237	−0.1141***	−0.4597***	0.0459***	0.0004***	0.0008***	0.0218	0.0925	0.1181***	−0.0351***	0.2031
(22.04)	(0.20)	(−3.47)	(1.02)	(−0.59)	(−5.09)	(−7.99)	(3.07)	(2.86)	(8.12)	(0.08)	(1.29)	(5.44)	(−3.79)	
Panel C: normal period														
1.7215***	0.2609***	−0.0009***	0.0652***	1.3577***	−0.1362***	0.3131***	0.2076***	0.0006***	0.0001	−0.3170**	−0.2607***	−0.0473***	−0.0059***	0.2419
(66.63)	(21.14)	(−7.83)	(62.26)	(7.14)	(−22.73)	(11.54)	(33.83)	(7.18)	(1.16)	(−2.38)	(−5.78)	(−6.45)	(−4.06)	
Panel D: macrofactors														
C	*Vol*	*Turnover*	*Amihud*	*BML_Vol*	*BML_Amihud*	*Year102*	*S10 year*	*CPI*	*GDP*	*M0*	*Vol × Coupon*	*Turnover × Coupon*	*Day × Coupon*	*Adj−R2(%)*
1.2203 ***	−0.0015***		0.1846***	−0.0078***	1.0654*	0.1760***	0.1487***	0.1081***	−0.0969***	−0.0205***	0.0002***		0.0004**	2.9413
(7.51)	(−3.78)		(2.74)	(−6.23)	(1.90)	(4.91)	(6.37)	(6.23)	(−3.54)	(−3.90)	(3.23)		(2.09)	
1.2328***		−1.8679***	0.1659**	−0.0079***	1.0650*	0.1738***	0.1495***	0.1079***	−0.0978***	−0.0208***		0.2129***	0.0005**	3.0494
(7.59)		(−4.35)	(2.45)	(−6.31)	(1.90)	(4.85)	(6.41)	(6.22)	(−3.58)	(−3.95)		(3.76)	(2.17)	

To investigate the robustness of impact of interaction on the spreads, the following regression equation is given. $\text{Spread}_{it} = \alpha + \beta 1 \times (\text{liquidity risk factors})_{it} + \beta 2 \times (\text{credit risk factors})_{it} + \beta 3 \times (\text{different dimension liquidity risk factors})_{it} \times (\text{credit risk factors})_{it} + \epsilon t$. In panel A, panel B and panel C, the meanings of the variables are given in Table 1. In panel D, BML_Vol is the monthly trading volume of the corporate bond market. BML_Amihud is illiquidity measures of the corporate bond market based on amihud. Year102 is the difference between the 10-year treasury rate and the 2-year treasury rate. S10 year is the spread between the bond market index yield and the 10-year treasury rate. CPI is the consumer price index, and GDP is gross domestic product. M0 is an increase in cash in circulation. In panel D, vol, turnover and amihud are liquidity measures of matched pairs of bonds. The cross-term is the interaction between liquidity and credit risk. The whole period is July 2006 to June 2016. The sample period of the financial crisis is May 2007 to December 2009. The sample period of normal period is from July 2006 to April 2007 and from January 2010 to June 2016. The t-statistics are shown in parentheses below each coefficient estimate, and "*," "**," and "***" indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 9: The impact of interaction between liquidity and credit risk on spreads in subsample results.

Coupon	Vol	C	Dummy	Vol	Day	Amihud	Age	Volatility	Coupon	Vol × coupon	Amihud × coupon	Age × coupon	aj-R2
High	High	−2.0665***	0.4177***	−0.0012	0.0793***	−832.9094	0.9520***	0.7036***	0.5790***	0.0001	109.0945	−0.1525***	0.1360
		(−3.58)	(5.94)	(−0.77)	(17.58)	(−1.34)	(4.32)	(15.89)	(7.95)	(0.36)	(1.40)	(−5.45)	
	Low	−0.1327	0.4796***	−0.0576***	0.0532***	−0.5420	1.3198***	0.1143***	0.4020***	0.0071***	0.0987	−0.2061***	0.2191
		(−0.33)	(13.71)	(−4.10)	(25.81)	(−0.37)	(8.64)	(3.58)	(7.84)	(3.99)	(0.52)	(−10.63)	
Middle	High	−2.2082***	0.3998***	−0.0017	0.0794***	−561.0267	0.9547***	0.7029***	0.5978***	0.0001	75.9640	−0.1528***	0.1386
		(−3.80)	(5.66)	(−1.11)	(17.52)	(−0.90)	(4.33)	(15.85)	(8.15)	(0.70)	(0.97)	(−5.46)	
	Low	−0.1795	0.4800***	−0.0585***	0.0538***	−0.5347	1.3125***	0.1103***	0.4073***	0.0072***	0.0984	−0.2052***	0.2211
		(−0.44)	(13.69)	(−4.15)	(26.14)	(−0.36)	(8.63)	(3.46)	(7.92)	(4.03)	(0.52)	(−10.63)	
Low	High	−1.9248***	0.3986***	0.0006	0.0728***	184.3172	0.2139***	−0.0110	0.6444***	−0.0001	−41.4980	−0.0524***	0.2195
		(−10.75)	(11.30)	(0.96)	(21.44)	(0.76)	(5.65)	(−0.45)	(21.10)	(−1.37)	(−0.99)	(−7.53)	
	Low	−2.6984***	0.2102***	0.0353***	0.0292***	−0.4641	0.2063***	0.1488***	0.8341***	−0.0068***	0.1374	−0.0509***	0.2321
		(−15.46)	(7.75)	(4.92)	(13.58)	(−0.63)	(6.71)	(5.81)	(27.52)	(−5.36)	(1.05)	(−8.53)	

To investigate whether the impact of interaction on the spreads is present, the following regression equations are given. $\text{Spread}_{it} = \alpha + \beta_1 \times (\text{liquidity risk factors})_{it} + \beta_2 \times (\text{credit risk factors})_{it} + \beta_3 \times (\text{different dimension liquidity risk factors})_{it} \times (\text{representative credit risk factor})_{it} + \epsilon t$. The meanings of the variables are given in Table 2. The sample period is July 2006 to June 2016. The whole sample includes 3716 bond issues and 47436 observations. The t-statistics are shown in parentheses below each coefficient estimate, and "*," "**," and "***" indicate significance at the 10%, 5%, and 1% levels, respectively.

is relatively high, liquidity risk is small, and small liquidity risk does not interact with credit risk. Liquidity risks are rising as trading volumes fall. When liquidity risk increases to a certain extent, it begins to interact spirally with credit risk, affecting asset prices. This shows that the interaction between liquidity and credit risk will start from scratch with the increase of liquidity risk and it is a time-varying dynamic process.

In high-coupon, middle-coupon and low-coupon subsamples, the regression coefficient of age × coupon in the low-vol sample is greater than one of age × coupon in the high-vol sample. For example, in the high-coupon sub sample, the regression coefficient of age × coupon in the low-vol sample is −0.2061, while the regression coefficient of age × coupon in the high-vol sample is −0.1525. It shows that when credit risk is controlled, the interaction between liquidity and credit risk will increase with the increase of liquidity risk. In low-vol sample, the regression coefficients of age × coupon in the low-coupon sample, the middle-coupon sample and the high-coupon sample are −0.0509, −0.2052, and −0.2061, respectively. It suggests that when liquidity risk is controlled, the interaction between liquidity and credit risk will increase with the increase of credit risk.

The above analysis shows that the interaction between liquidity and credit risk will increase with the increase of liquidity risk and credit risk and it is a time-varying dynamic process. The impact of the interaction between liquidity and credit risk on the spreads is not negligible in China's corporate bond market. And, the results are robust.

5. Conclusion

This paper investigates the impact of multidimension liquidity, credit risk, and the interaction between liquidity and credit risk on corporate bond spreads based on a large transaction data set from July 2006 to June 2016, including the monthly data of 3716 bonds in China. Our main conclusions are as follows.

First, liquidity premiums are the main parts of corporate bond spreads in China's corporate bond market. We select multidimension liquidity measures from the Chinese corporate bond market, and construct many liquidity measures from three dimensions of trading activity, price shock and bond life time. This study shows that the results are robust. Related research can also refer to Longstaff et al. [7]. Our method of multidimension liquidity measures is more accurate and reliable than those used in the existing literature.

Second, the interaction between liquidity and credit risk on bond spreads plays a significant role in determining corporate bond spreads and liquidity and credit risk alone does not explain the entire yield-spread behavior. Related research can also refer to Sperna Weiland et al. [24] and Chen et al. [25]. The interaction between liquidity and credit risk will increase with the increase of liquidity risk and credit risk and it is a time-varying dynamic process. When credit risk is controlled, the interaction between liquidity and credit risk will increase with the increase of liquidity risk. When liquidity risk is controlled, the interaction between liquidity and credit risk will increase with the increase of credit risk.

The liquidity of price shock dimension has a greater interaction with dynamic credit risk than static credit risk. And, these results are robust.

Third, we find that the interaction between liquidity and credit risk has an enhanced impact on corporate bond spreads during financial crisis than normal period. The reason is that liquidity and credit risk interact during financial crisis. We also find that the differences between the interbank market and the exchange market have impacts on corporate bond spreads in normal period than financial crisis. The reason may be that liquidity and credit risk have major impact on the spreads during the financial crisis than the differences between the two markets. In addition, compared with normal period, speculative liquidity risk premiums are increased and investment liquidity risk premiums are reduced, while dynamic credit risk premiums are increased and static credit risk premiums are reduced during the financial crisis.

Finally, dynamic credit risk changes in resonance with the dynamic change of liquidity risk, while the static credit risk does not change with the dynamics change of liquidity risk. When liquidity affects asset prices, dynamic credit risk interacts with liquidity risk in a dynamic spiral, forming a bigger resonance enhancement interaction between liquidity and credit risk to dynamically affect asset prices than static credit risk.

This study provides new interpretation channels for corporate bond pricing. These findings are robust. Investor structure differences between the interbank market and the exchange market may be a direction for further research.

Data Availability

This study uses the data from the WIND database.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 71471129 and 71501140) and Tianjin Philosophy and Social Science Planning Project (Grant nos. TJGL19-018).

References

[1] E. J. Elton, M. J. Gruber, and D. Agrawal, "Explaining the rate spread on corporate bonds," *The Journal of Finance*, vol. 56, no. 1, pp. 247–277, 2001.

[2] N. Friewald, R. Jankowitsch, and M. G. Subrahmanyam, "Illiquidity or credit deterioration: a study of liquidity in the US corporate bond market during financial crises," *Journal of Financial Economics*, vol. 105, no. 1, pp. 18–36, 2012.

[3] X. Luo and J. Zhang, "Pricing Chinese convertible bonds with default intensity by Monte Carlo method," *Discrete Dynamics in Nature and Society*, vol. 2019, pp. 1–8, 2019.

[4] J. Z. Huang and M. Huang, "How much of the corporate-treasury yield spread is due to credit risk?" *Review of Asset Pricing Studies*, vol. 2, no. 2, pp. 153–202, 2012.
[5] P. Collin-Dufresn, R. S. Goldstein, and J. S. Martin, "The determinants of credit spread changes," *The Journal of Finance*, vol. 56, no. 6, pp. 2177–2207, 2001.
[6] J. Driessen, "Is default event risk priced in corporate bonds?" *Review of Financial Studies*, vol. 18, no. 1, pp. 165–195, 2005.
[7] F. A. Longstaff, S. Mithal, and E. Neis, "Corporate yield spreads: default risk or liquidity? New evidence from the credit default swap market," *The Journal of Finance*, vol. 60, no. 5, pp. 2213–2253, 2005.
[8] L. Chen, D. A. Lesmond, and J. Wei, "Corporate yield spreads and bond liquidity," *The Journal of Finance*, vol. 62, no. 1, pp. 119–149, 2007.
[9] H. Lin, J. Wang, and C. Wu, "Liquidity risk and expected corporate bond returns," *Journal of Financial Economics*, vol. 99, no. 3, pp. 628–650, 2011.
[10] V. V. Acharya, Y. Amihud, and S. T. Bharath, "Liquidity risk of corporate bond returns: conditional approach," *Journal of Financial Economics*, vol. 110, no. 2, pp. 358–386, 2013.
[11] X. P. Zou, F. Huang, and C. J. Yang, "Liquidity risk, Liquidity demand of investors and asset pricing," *Journal of Management Science in China*, vol. 12, no. 6, pp. 139–149, 2009.
[12] J. Dick-Nielsen, P. Feldhütter, and D. Lando, "Corporate bond liquidity before and after the onset of the subprime crisis," *Journal of Financial Economics*, vol. 103, no. 3, pp. 471–492, 2012.
[13] J. Helwege, J. Z. Huang, and Y. Wang, "Liquidity effects in corporate bond spreads," *Journal of Banking & Finance*, vol. 45, pp. 105–116, 2014.
[14] D. Covitz and C. Downing, "Liquidity or credit risk? The determinants of very short-term corporate yield spreads," *The Journal of Finance*, vol. 62, no. 5, pp. 2303–2328, 2007.
[15] F. Longstaff and E. Schwartz, "A simple approach to valuing risky fixed and floating rate debt," *The Journal of Finance*, vol. 50, no. 3, pp. 789–819, 1995.
[16] J. Ericsson and O. Renault, "Liquidity and credit risk," *The Journal of Finance*, vol. 61, no. 5, pp. 2219–2250, 2006.
[17] D. Duffie and K. J. Singleton, "Modeling term structures of defaultable bonds," *Review of Financial Studies*, vol. 12, no. 4, pp. 687–720, 1999.
[18] A. X. Wang, W. Z. Xie, and W. L. Yu, "Empirical research on China's corporate bond yield spread," *Journal of Management Sciences in China*, vol. 15, no. 5, pp. 32–41, 2012.
[19] Q. Gao and H. F. Zou, "A comparative study on the secondary market pricing of China' enterprise bonds and corporate bonds," *Journal of Financial Research*, vol. 1, pp. 84–100, 2015.
[20] B. Yang, Z. Wu, and Y. Su, "The determinants of the non-defaultable spreads of corporate bonds: evidence from China," *Discrete Dynamics in Nature and Society*, vol. 2021, pp. 1–21, Article ID 5595099, 2021.
[21] M. Rossi, "Realized volatility, liquidity, and corporate yield spreads," *Quarterly Journal of Forestry*, vol. 4, no. 1, Article ID 1450004, 2014.
[22] M. Kalimipalli and S. Nayak, "Idiosyncratic volatility vs. liquidity? Evidence from the US corporate bond market," *Journal of Financial Intermediation*, vol. 21, no. 2, pp. 217–242, 2012.
[23] J. Wang and W. C. Wu, "Liquidity, credit quality, and the relation between volatility and trading activity: evidence from the corporate bond market," *Journal of Banking & Finance*, vol. 50, pp. 183–203, 2015.
[24] R. C. Sperna Weiland, R. J. Laeven, and F. De Jong, "Feedback between credit and liquidity risk in the US corporate bond market," in *Proceedings of the 30th Australasian Finance and Banking Conference*, Sydney NSW 2052 Australia, June 2017.
[25] H. Chen, R. Cui, Z. He, and K. Milbradt, "Quantifying liquidity and default risks of corporate bonds over the business cycle," *Review of Financial Studies*, vol. 31, no. 3, pp. 852–897, 2018.
[26] X. Li, B. Yang, Y. Su, and Y. An, "Pricing corporate bonds with credit risk, liquidity risk, and their correlation," *Discrete Dynamics in Nature and Society*, vol. 2021, pp. 1–14, Article ID 6681035, 2021.
[27] C. R. Ai, Y. Zhang, and C. F. Cui, "Bond's liquidity and default risk correlation premium and empirical test," *Journal of Management Sciences in China*, vol. 05, pp. 87–94, 2015.
[28] O. Sarig and A. Warga, "Bond price data and bond market liquidity," *Journal of Financial and Quantitative Analysis*, vol. 24, no. 3, pp. 367–378, 1989.
[29] Y. Amihud, "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, vol. 5, no. 1, pp. 31–56, 2002.
[30] S. Han and H. Zhou, "Effects of liquidity on the non-default component of corporate yield spreads: evidence from intraday transactions data," *Quarterly Journal of Forestry*, vol. 6, no. 3, Article ID 1650012, 2016.
[31] Z. He and K. Milbradt, "Endogenous liquidity and defaultable bonds," *Econometrica*, vol. 82, no. 4, pp. 1443–1508, 2014.
[32] G. J. Alexander, A. K. Edwards, and M. G. Ferri, "The determinants of trading volume of high-yield corporate bonds," *Journal of Financial Markets*, vol. 3, no. 2, pp. 177–204, 2000.
[33] M. A. Goldstein, E. S. Hotchkiss, and E. R. Sirri, "Transparency and liquidity: a controlled experiment on corporate bonds," *Review of Financial Studies*, vol. 20, no. 2, pp. 235–273, 2007.
[34] E. Hotchkiss and G. Jostova, "Determinants of corporate bond trading: a comprehensive analysis," *Quarterly Journal of Forestry*, vol. 7, no. 2, Article ID 1750003, 2017.
[35] S. Mahanti, A. Nashikkar, M. Subrahmanyam, G. Chacko, and G. Malik, "Latent liquidity: a new measure of liquidity, with an application to corporate bonds," *Journal of Financial Economics*, vol. 88, no. 2, pp. 272–298, 2008.
[36] A. K. Edwards, L. E. Harris, and M. S. Piwowar, "Corporate bond market transaction costs and transparency," *The Journal of Finance*, vol. 62, no. 3, pp. 1421–1451, 2007.
[37] D. Shin and B. Kim, "Liquidity and credit risk before and after the global financial crisis: evidence from the Korean corporate bond market," *Pacific-Basin Finance Journal*, vol. 33, pp. 38–61, 2015.
[38] S. T. Bharath and T. Shumway, "Forecasting default with the merton distance to default model," *Review of Financial Studies*, vol. 21, no. 3, pp. 1339–1369, 2008.
[39] T.-K. Chen, H.-H. Liao, and P.-L. Tsai, "Internal liquidity risk in corporate bond yield spreads," *Journal of Banking & Finance*, vol. 35, no. 4, pp. 978–987, 2011.
[40] J. C. Hull, *Options, Futures, and Other Derivatives*, Pearson Education India, Chennai, India, 2006.
[41] L. Fisher, "Determinants of risk premiums on corporate bonds," *Journal of Political Economy*, vol. 67, no. 3, pp. 217–237, 1959.
[42] J. R. Hicks, "Liquidity," *Economic Journal*, vol. 72, no. 288, pp. 787–802, 1962.
[43] W. Bagehot, "The only game in town," *Financial Analysts Journal*, vol. 27, no. 2, pp. 12–14, 1971.

[44] Y. Amihud and H. Mendelson, "Asset pricing and the bid-ask spread," *Journal of Financial Economics*, vol. 17, no. 2, pp. 223–249, 1986.
[45] V. Acharya and L. Pedersen, "Asset pricing with liquidity risk," *Journal of Financial Economics*, vol. 77, no. 2, pp. 375–410, 2005.
[46] D. Bongaerts, F. De Jong, and J. Driessen, "An asset pricing approach to liquidity effects in corporate bond markets," *Review of Financial Studies*, vol. 30, no. 4, pp. 1229–1269, 2017.
[47] W. Liu, "A liquidity-augmented capital asset pricing model," *Journal of Financial Economics*, vol. 82, no. 3, pp. 631–671, 2006.
[48] A. Bervas, "Market liquidity and its incorporation into risk management," *Financial Stability Review*, vol. 8, no. 5, pp. 63–79, 2006.
[49] A. Díaz and A. Escribano, "Measuring the multi-faceted dimension of liquidity in financial markets: a literature review," *Research in International Business and Finance*, vol. 51, Article ID 101079, 2020.
[50] P. Li and J. Song, "Pricing Chinese convertible bonds with dynamic credit risk," *Discrete Dynamics in Nature and Society*, vol. 2014, pp. 1–5, Article ID 492134, 2014.
[51] R. F. Zhu, "Illiquidity and risk premium in corporate bonds: evidence from China," *Review of Investment Studies*, vol. 32, no. 1, pp. 43–55, 2013.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 3112987, 13 pages
https://doi.org/10.1155/2022/3112987

Research Article

Promoting Variable Effect Consistency in Mixture Cure Model for Credit Scoring

Chenlu Zheng,[1,2] **Jianping Zhu**,[1,2] **Xinyan Fan**,[3] **Song Chen**,[4] **and Zhiyuan Zhang**[5]

[1]*School of Management, Xiamen University, Xiamen 361005, China*
[2]*Data Mining Research Center, Xiamen University, Xiamen 361005, China*
[3]*School of Statistics, Renmin University of China, Beijing 100872, China*
[4]*Taizhou University, Microfinance College, Taizhou 318000, China*
[5]*Science and Technology Development Department, Xiamen International Bank, Xiamen 361001, China*

Correspondence should be addressed to Jianping Zhu; xmzhujianping@163.com and Song Chen; chensong1978@tzc.edu.cn

Received 3 October 2021; Revised 10 December 2021; Accepted 31 December 2021; Published 21 February 2022

Academic Editor: Dehua Shen

Copyright © 2022 Chenlu Zheng 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.

Mixture cure models are widely adopted in credit scoring. Mixture cure models consist of two parts: an incident part which predicts the probability of default and a latency part which predicts when they are likely to default. The two model parts describe two quite relevant credit aspects. So, it is reasonable to expect that the two sets of the coefficients are somewhat related. Moreover, in practical cases, it is difficult to interpret the results when the two sets of the coefficients of the same variables have conflicting signs. Most existing works either ignore the interconnections of the two sets of coefficients or impose a strict constraint between them. We proposed a mixture cure model considering the variable effect consistency using a sign-based penalty. It is a more flexible model that allows the two sets of coefficients to be in different distributions and magnitudes. To accommodate high-dimensional credit data, a group lasso penalty is also imposed for variable selection. Simulation shows that the proposed method has competitive performance compared with alternative methods in terms of estimation and prediction. Furthermore, the empirical study illustrates that the proposed method outperforms the alternative method and can improve the interpretability of the results.

1. Introduction

Credit scoring is an effective and crucial approach for evaluating credit risk [1, 2]. A slight improvement in the prediction precision of credit scoring models can bring considerable benefits. Therefore, credit scoring has attracted increasing attention of scholars and practitioners. Many studies treat it as a classification problem to distinguish noncreditworthy customers from creditworthy ones [2, 3]. These studies focus on classification techniques including logistic regression, support vector machine, neural network, and random forest [4, 5]. For example, Zhou et al. [6] proposed a logistic regression method with clustering analysis for credit risk evaluation. Zhang et al. [7] proposed a cost-sensitive logistic regression model to assess the credit risk. Considering the high cost and time consumption of credit scoring, a credit granting process using three-way decisions is proposed to make efficient credit decisions [8]. Since the exposure to risk and the losses caused by default are strongly related to the time when they default [9], credit risk prediction overtime is of great significance for timely risk management.

Survival analysis, with its ability of predicting the probability of default over time, has been first applied in credit scoring in 1992 [10]. It can be more informative than the binary classification model. Subsequently, various survival analysis models are proposed to predict credit risk over time. For instance, Cox proportional hazards (PH) model is adopted to predict early repayment and time to default in personal loans and investigated the effect of different variables on time to default [11, 12]. In addition, macroeconomic factors and time-varying data are also incorporated in

survival analysis to improve the performance of prediction in credit scoring [13, 14]. And the models are further extended by a survival gradient boosting decision tree approach to enhance the prediction performance [15].

However, standard survival analysis assumes that the loan term is long enough and every customer will eventually default. In practice, a substantial proportion of customers will not default during the entire loan term. Since mixture cure models applied in medicine assume that some patients have been cured and will not die during the follow-up period, it is more appropriate in the credit market and was first introduced to credit scoring by Tong et al. [16].

Recently, the mixture cure model, an extension of the standard survival analysis, is widely adopted in credit scoring for its ability of predicting not only whether customers will default but also when they are likely to default. Results showed that the mixture cure model is more suitable for credit data compared with standard survival analysis models and the mixture cure model incorporating penalized spline has better performance in prediction [17]. Mixture cure models have been further developed by identifying different risk patterns of customers, considering the influence of competitive risk, and the relationship between the default times and the variables [18–20].

Mixture cure models consist of two parts: an incident part which predicts the probability of default and a latency part which predicts when they are likely to default. In the two model parts, the two sets of the coefficients indicate the two sets of the variable effects on the credit risk. The two model parts describe two quite relevant credit aspects. Nevertheless, most of the existing studies ignore the relations between the two sets of coefficients in two model parts. These works generally assume that there are no direct constraints between the two sets of coefficients, which may get conflicting results of variable effects. For example, Dirick et al. [21] propose a mixture cure model incorporating macroeconomic factors to predict credit risk. The results show that the customers' annual income has the opposite effect on whether and when to default. In other words, according to the results, customers with lower annual income have a lower probability of default, but they are more likely to default earlier. It is difficult to interpret the conflicting results and apply them in practice.

In fact, the two model parts describe two quite relevant credit aspects, namely, the probability of default and survival (nondefault) time. Customers with high default probability are more likely to default earlier. So, it is reasonable to expect that the two sets of the coefficients are somewhat related. Theoretical derivations [22] and empirical analysis [23] also suggest that relaxing the independence of two sets of coefficients can improve the model performance. The assumption has been relaxed by establishing a joint distribution of the defaulting predictor and the logarithm of the hazard rate in [23]. Note that the two model parts still describe two different aspects of default. The assumption of a joint distribution may be too strict. So, we consider a more flexible model that allows the two sets of coefficients to be in different distributions and magnitudes. Sign consistency penalty is proposed to promote the similarity in sign to get more interpretable results by Zhang et al. [24]. In this paper, we propose a variable effect consistency mixture cure model with a sign-based penalty. The proposed method can promote the similarity in the signs of variable effect in the two model parts to improve interpretability. To accommodate high-dimensional credit data, a group lasso penalty is also imposed for variable selection [25].

The contributions of this paper are as follows. First, we propose a variable effect consistency mixture cure model. The proposed method can lead to more interpretable results by promoting the similarity in the signs of coefficients in the two parts of the mixture cure model. Second, a group lasso penalty is imposed to select important variable subgroups and accommodate the high-dimensional data. Third, simulation and empirical analysis of credit data illustrate that the proposed method can improve the prediction accuracy as well as interpretability, which has important practical significance for applying the prediction results to the credit business.

The remainder of the paper is organized as follows. Section 2 introduces the variable effect consistency mixture cure model. Computational algorithm is presented in Section 3. Simulation is carried out in Section 4. Empirical study is presented in Section 5. Finally, conclusions are discussed in Section 6.

2. Methods

In this paper, we consider credit data with n customers and p variables. Denote U as the time to default and C as the time of censoring. Let Y be the unobservable binary variable with $Y = 0$ indicating that the customer is cured and will not default ($U = \infty$), whereas $Y = 1$ indicates an uncured customer and will eventually default.

Denote $\delta_i = \mathrm{I}(U_i < C_i)$ as the censoring indicator of customer i, where U_i and C_i are the time to default and censoring time of customer i, respectively. $\delta_i = 0$ if censored and $\delta_i = 1$ otherwise. Note that there are three possible credit states of customers. (a) $\delta_i = 0$ and $Y_i = 0$: censored, cured customers who will not default; (b) $\delta_i = 0$ and $Y_i = 1$: censored, uncured customers who will eventually default and have not been observed to default in censoring time C_i; (c) $\delta_i = 1$ and $Y_i = 1$: uncensored, uncured customers who have been observed to default.

Denote $t_i = \min(U_i, C_i)$. Note that, in many practical cases, variables can be naturally grouped. For instance, many categorical variables may have several levels and can be represented by subgroups of dummy variables [26]. The additive model with polynomial or nonparametric components can be expressed as groups of basis functions [27]. In addition, grouping structure can also be introduced by taking advantage of prior knowledge. For example, genes belonging to the same biological pathway can also be considered a group [28]. Let $\mathbf{x} = (\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_J^{\mathrm{T}})^{\mathrm{T}}$ be the variable vector with J subgroups. $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp_j})^{\mathrm{T}}$ is the j-th subgroup of variable vector, and $\sum_{j=1}^{J} p_j = p$. The observable data are $(t_i, \delta_i, \mathbf{x})$, $i = 1, \ldots, n$.

The incident part of the mixture cure model describes the probability of default, for which we adopt a logistic

regression model. Let π_i be the probability of cured (non-default) customer i.

$$\pi_i = \Pr(Y_i = 0|\mathbf{x}) = \frac{1}{1+\exp(\alpha_0 + \mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha})}, \tag{1}$$

where α_0 is the intercept, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_J^{\mathrm{T}})^{\mathrm{T}}$ is the vector of unknown regression coefficients, and the j-th subgroup of the coefficient vector is $\boldsymbol{\alpha}_j = (\alpha_{j1}, \ldots, \alpha_{jp_j})^{\mathrm{T}}$.

In the latency part, for uncured customers, we adopt an exponential model for survival. Note that the exponential model has been commonly adopted in mixture cure models [29, 30]. It is easy to capture the relations between the probability and the time to default for it includes only one parameter [23]. The survival function is

$$\begin{aligned} S_i(t|Y_i = 1, \mathbf{x}) &= \exp(-\lambda_i t), \\ \lambda_i &= \exp(\beta_o + \mathbf{x}^T\boldsymbol{\beta}), \end{aligned} \tag{2}$$

where λ_i is the hazard function of customer i, β_0 is the intercept, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_J^{\mathrm{T}})^T$ is the vector of unknown regression coefficients, and the j-th subgroup of the coefficient vector is $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp_j})^{\mathrm{T}}$. Survival function $S_i(t|Y_i = 1, \mathbf{x}) = P(U_i > t|Y_i = 1, \mathbf{x})$ indicates the survival probability of uncured customers in time t, that is, the probability of default in time t given the customer will default.

The mixture cure model can be given by

$$S_i(t|\mathbf{x}) = \pi_i + (1-\pi_i)S_i(t|Y_i = 1, \mathbf{x}), \tag{3}$$

where $S_i(t|\mathbf{x})$ is the survival probability of customer i in time t.

For observable data $(t_i, \delta_i, \mathbf{x})$, the objective function can be written as follows.

$$Q(\alpha_0, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta}) = -L(\alpha_0, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta}) + P(\boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{4}$$

where $L(\alpha_0, \alpha, \beta_0, \beta)$ is the log-likelihood function, which is

$$L(\alpha_0, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{n} \begin{pmatrix} \delta_i(\log(1-\pi_i) + \log(\lambda_i) - \lambda_i t_i) \\ +(1-\delta_i)\log((1-\pi_i)\exp(-\lambda_i t_i) + \pi_i) \end{pmatrix}, \tag{5}$$

with $\pi_i = 1/(1+\exp(\alpha_0 + \mathbf{x}^{\mathbf{T}}\boldsymbol{\alpha}))$, and $\lambda_i = \exp(\beta_o + \mathbf{x}^{\mathbf{T}}\boldsymbol{\beta})$. Here, $P(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the penalty function, which is

$$\begin{aligned} P(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mu_1 &\sum_{j=1}^{J} \sqrt{p_j}\left(\|\boldsymbol{\alpha}_j\| + \|\boldsymbol{\beta}_j\|\right) \\ &+ \frac{\mu_2}{2}\sum_{j=1}^{J}\sum_{k=1}^{p_j}\left(\operatorname{sign}(\alpha_{jk}) - \operatorname{sign}(\beta_{jk})\right)^2, \end{aligned} \tag{6}$$

where $\mu_1 > 0$ and $\mu_2 > 0$ are tuning parameters, $\|\cdot\|$ is the l_2 norm, and $\operatorname{sign}(\cdot)$ is the sign function. In many practical cases, grouping structure arises naturally. In addition, it is hard to interpret the results when coefficients corresponding to the same variables have conflicting signs. Therefore, we consider a flexible mixture cure model with sign consistency and group variable selection penalties. The first penalty is a group lasso penalty. It can conduct estimation and group variable selection by shrinking the coefficients of insignificant groups to 0. It considers grouping structures and has good prediction performance [26]. The second penalty is the sign consistency penalty. It promotes the sign consistency of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the two parts of the model, which can lead to more interpretable results [31].

3. Computational Algorithm

In this section, the Expectation Coordinate Descent (ECD) algorithm is developed to optimize the objective function. In E-step, a latent unobserved Y_i is introduced to obtain a complete log-likelihood function. In CD-step, group coordinate descent is adopted to iteratively update a single parameter with the remaining parameters fixed at their most recent values. Sign function $\operatorname{sign}(\cdot)$ is difficult to optimize for its discontinuity and nondifferentiability. Therefore, referring to [24, 32], we propose the approximation as follows:

$$\left(\operatorname{sign}(\alpha_{jk}) - \operatorname{sign}(\beta_{jk})\right)^2 \approx \left(\frac{\alpha_{jk}}{|\alpha_{jk}| + \tau} - \frac{\beta_{jk}}{|\beta_{jk}| + \tau}\right)^2, \tag{7}$$

where τ is a small positive constant (more discussions below).

The ECD algorithm updates $(\alpha_0, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta})$ in the m–th iteration as follows.

3.1. E-Step. Denote the observation of the latent Y_i as y_i and denote the complete data as $\{(t_i, \delta_i, y_i, \mathbf{x}), i = 1, \ldots, n\}$. The complete log-likelihood is

$$\begin{aligned} L^{[m]} &= \sum_{i=1}^{n}\left((1-y_i)\log(\pi_i^{[m]}) + y_i\log(1-\pi_i^{[m]})\right) \\ &\quad + \sum_{i=1}^{n}\left(\delta_i(\beta_0^{[m]} + \mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}^{[m]}) - y_i t_i \exp(\beta_0^{[m]} + \mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}^{[m]})\right) \\ &\coloneqq L_1^{[m]} + L_2^{[m]}, \end{aligned} \tag{8}$$

where

$$\pi_i^{[m]} = \frac{1}{1+\exp(\alpha_0^{[m]} + \mathbf{x}^{\mathbf{T}}\boldsymbol{\alpha}^{[m]})}. \tag{9}$$

The expectation of y_i is

$$E(y_i) = \begin{cases} 1 & \delta_i = 1 \\ \dfrac{(1-\pi_i^{[m]})\exp(-\lambda_i^{[m]} t_i)}{\pi_i^{[m]} + (1-\pi_i^{[m]})\exp(-\lambda_i^{[m]} t_i)} & \delta_i = 0 \end{cases}, \tag{10}$$

where

$$\lambda_i^{[m]} = \exp(\beta_0^{[m]} + \mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}^{[m]}). \tag{11}$$

When customer i is observed to default ($\delta_i = 1$), the unobserved $y_i = 1$, whereas the expectation of y_i is related to the probability of cured and the uncured but censored customers.

In E-step, we take the expectation of $L^{[m]}$ with respect to $E(y_i)$ given the complete data $\{(t_i, \delta_i, y_i, \mathbf{x}), \quad i = 1, \cdots, n\}$.

$$E\left(L^{[m]}\right) = E\left(L_1^{[m]}\right) + E\left(L_2^{[m]}\right), \tag{12}$$

where

$$\begin{aligned} l_1^{[m]} &= E\left(L_1^{[m]}\right) = \sum_{i=1}^{n}\left((1 - E(y_i))\log\left(\pi_i^{[m]}\right) + E(y_i)\log\left(1 - \pi_i^{[m]}\right)\right), \\ l_2^{[m]} &= E\left(L_2^{[m]}\right) = \sum_{i=1}^{n}\left(\delta_i\left(\beta_0^{[m]} + \mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}^{[m]}\right) - E(y_i)t_i \exp\left(\beta_0^{[m]} + \mathbf{x}^{\mathbf{T}}\boldsymbol{\beta}^{[m]}\right)\right). \end{aligned} \tag{13}$$

3.2. CD-Step. In CD-step, group coordinate descent is adopted to iteratively update $(\alpha_0, \boldsymbol{\alpha}, \beta_0, \boldsymbol{\beta})$. The intercept $\alpha_0^{[m+1]}$ is updated by

$$\alpha_0^{[m+1]} = \alpha_0^{[m]} - \left(\frac{\partial^2 l_1^{[m]}}{\partial\left(\alpha_0^{[m]}\right)^2}\right)^{-1} \frac{\partial l_1^{[m]}}{\partial \alpha_0^{[m]}}. \tag{14}$$

For $\boldsymbol{\alpha}_j^{[m+1]} \in R^{p_j}$, we adopted a fast unified algorithm, Groupwise Majorization Descent (GMD) proposed in [33] to solve the group lasso penalized objective function in (4). The upper bound of the objective function is as follows:

$$-\left(\boldsymbol{\alpha}_j^{[m+1]} - \boldsymbol{\alpha}_j^{[m]}\right)^{\mathrm{T}}\left(\frac{\alpha l_1^{[m]}}{\partial \boldsymbol{\alpha}_j^{[m]}} + \mu_2 V_{1j}^{[m]}\right) + \frac{M_{1j}^{[m]}}{2}\left(\boldsymbol{\alpha}_j^{[m+1]} - \boldsymbol{\alpha}_j^{[m]}\right)^{\mathrm{T}}\left(\boldsymbol{\alpha}_j^{[m+1]} - \boldsymbol{\alpha}_j^{[m]}\right)^{\mathrm{T}} + \mu_1\sqrt{p_j}\left\|\boldsymbol{\alpha}_j^{[m+1]}\right\|. \tag{15}$$

Here, $V_{1j}^{[m]}$ is p_j-length vector, and $M_{1j}^{[m]}$ is a constant as follows:

$$\begin{aligned} V_{1j}^{[m]} &= \left(\frac{1}{\left|\alpha_{jk}^{[m]}\right| + \tau}\left(\frac{\beta_{jk}^{[m]}}{\left|\beta_{jk}^{[m]}\right| + \tau} - \frac{\alpha_{jk}^{[m]}}{\left|\alpha_{jk}^{[m]}\right| + \tau}\right)\right)_{1 \le k \le p_j}, \\ M_{1j}^{[m]} &= \psi\left(-\left(\frac{\partial^2 l_1^{[m]}}{\partial \alpha_{jk_1}^{[m]} \partial \alpha_{jk2}^{[m]}}\right)_{1 \le k_1, k_2 \le p_j}\right) + \max_k\left(\mu_2\left(\frac{1}{\left|\alpha_{jk}^{[m]}\right| + \tau}\right)^2\right), \end{aligned} \tag{16}$$

where $\psi(\cdot)$ is the maximum eigenvalue function.

Similarly, the intercept $\beta_0^{[m+1]}$ is updated by

$$\beta_0^{[m+1]} = \beta_0^{[m]} - \left(\frac{\partial^2 l_2^{[m]}}{\partial\left(\beta_0^{[m]}\right)^2}\right)^{-1} \frac{\partial l_2^{[m]}}{\partial \beta_0^{[m]}}. \tag{17}$$

For $\boldsymbol{\beta}_j^{[m+1]} \in R^{p_j}$, consider the optimization function:

$$-\left(\boldsymbol{\beta}_j^{[m+1]} - \boldsymbol{\beta}_j^{[m]}\right)^{\mathrm{T}}\left(\frac{\partial l_2^{[m]}}{\partial \boldsymbol{\beta}_j^{[m]}} + \mu_2 V_{2j}^{[m]}\right) + \frac{M_{2j}^{[m]}}{2}\left(\boldsymbol{\beta}_j^{[m+1]} - \boldsymbol{\beta}_j^{[m]}\right)^{\mathrm{T}}\left(\boldsymbol{\beta}_j^{[m+1]} - \boldsymbol{\beta}_j^{[m]}\right) + \mu_1\sqrt{p_j}\left\|\boldsymbol{\beta}_j^{[m+1]}\right\|. \tag{18}$$

Here, $V_{2j}^{[m]}$ is p_j-length vector, and $M_{2j}^{[m]}$ is a constant as follows:

$$V_{2j}^{[m]} = \left(\frac{1}{\left|\beta_{jk}^{[m]}\right| + \tau} \left(\frac{\alpha_{jk}^{[m]}}{\left|\alpha_{jk}^{[m]}\right| + \tau} - \frac{\beta_{jk}^{[m]}}{\left|\beta_{jk}^{[m]}\right| + \tau} \right) \right)_{1 \le k \le p_j},$$
$$M_{2j}^{[m]} = \psi\left(-\left(\frac{\partial^2 l_2^{[m]}}{\partial \beta_{jk_1}^{[m]} \partial \beta_{jk_2}^{[m]}} \right)_{1 \le k_1, k_2 \le p_j} \right) + \max_k \left(\mu_2 \left(\frac{1}{\left|\beta_{jk}^{[m]}\right| + \tau} \right)^2 \right). \tag{19}$$

The tuning parameters, μ_1 and μ_2, are selected by 5-fold cross-validation. The parameter τ in the approximation of the sign function controls the degree of approximation [24]. A smaller τ leads to a better approximation but less stable estimation. The proposed method is valid as long as τ is not too large, and the parameters with different signs can be distinguished [34]. Therefore, as suggested in [31], we set $\tau = 0.1$, which leads to satisfactory results.

The ECD algorithm is summarized in Table 1.

4. Simulations

In this section, some experiment examples are given to illustrate the performance of the proposed method compared to alternative methods. The proposed method is a mixture cure model with group lasso and sign consistency (MCGS). Two alternative methods are the standard mixture cure model without variable selection and sign consistency penalty (Full) and the mixture cure model with group lasso penalty (MCG), respectively. For comparison, alternative methods both adopt the logistic regression in the incident part and the exponential model in the latency part.

Here, we set sample size $n = 1000$ and consider low-dimensional data with $p = 60$ and high-dimensional data with $p = 200$. The censoring time is generated from an exponential distribution with censoring rates $\eta = \{0.25, 5\}$. We consider three examples regarding different grouping structures of coefficients and different types of variables. The true values of coefficients are generated according to the following settings in three examples:

Example 1: for each subgroup, we set $p_j = 10$. Intragroup variables x_{jk_m} and x_{jk_n} are generated from a multivariate normal distribution with the correlation coefficient $\rho = 0.3^{|k_m - k_n|}$, whereas intergroup variables are independent. Denote the true coefficients as $\boldsymbol{\alpha}^{\text{true}}$ and $\boldsymbol{\beta}^{\text{true}}$. The coefficients of the two scenarios are shown as follows:

Scenario 1:

$$\alpha^{\text{true}} = (\underbrace{0.1, \ldots, 0.1}_{10}, \underbrace{-0.4, \ldots, -0.4}_{10}, \underbrace{0, \ldots, 0}_{p-20})^{\text{T}},$$
$$\beta^{\text{true}} = (\underbrace{0.4, \ldots, 0.4}_{10}, \underbrace{-0.3, \ldots, -0.3}_{10}, \underbrace{0, \ldots, 0}_{p-20})^{\text{T}}. \tag{20}$$

Scenario 2:

$$\alpha^{\text{true}} = (\underbrace{0.5, \ldots, 0.5}_{10}, \underbrace{0.2, \ldots, 0.2}_{8}, -0.2, -0.2, \underbrace{0, \ldots, 0}_{p-20})^{\text{T}},$$
$$\beta^{\text{true}} = (\underbrace{0.4, \ldots, 0.4}_{10}, \underbrace{-0.3, \ldots, -0.3}_{8}, \underbrace{0, \ldots, 0}_{p-20})^{\text{T}}. \tag{21}$$

Example 2: the settings are similar to Example 1 except for the subgroup settings. We set 15 variables in the first subgroup, 5 variables in the second subgroup, and 10 variables for the remaining subgroups. The coefficients are shown as follows:

$$\alpha^{\text{true}} = (\underbrace{0.5, \ldots, 0.5}_{15}, \underbrace{-0.4, \ldots, -0.4}_{5}, \underbrace{0, \ldots, 0}_{p-20})^{\text{T}},$$
$$\beta^{\text{true}} = (\underbrace{0.4, \ldots, 0.4}_{15}, \underbrace{-0.3, \ldots, -0.3}_{5}, \underbrace{0, \ldots, 0}_{p-20})^{\text{T}}. \tag{22}$$

Example 3: consider the case with some discrete variables. For each subgroup, we set $p_j = 10$. A latent variable Z_j is generated from a multivariate normal distribution with the intragroup correlation coefficient $\rho = 0.3^{|k_m - k_n|}$ and intergroup correlation independent. The coefficient setting is the same as Example 2. x_{jk} is defined as follows:

$$x_{jk} = \begin{cases} I(Z_j > 0) & j \le \dfrac{p}{2}, \\ Z_j & j > \dfrac{p}{2}. \end{cases} \tag{23}$$

The performance of each model is me asured by 5 measures. Denote $\theta \in \{\boldsymbol{\alpha}, \boldsymbol{\beta}, (\boldsymbol{\alpha}^{\text{T}}, \boldsymbol{\beta}^{\text{T}})^{\text{T}}\}$, $\widehat{\theta}$ as the estimation of θ, $\widehat{\pi}$ as the estimation of π, and $\widehat{\lambda}$ as the estimation of λ. The true positive rate (TPR), false positive rate (FPR), and mean square error (MSE) with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $(\boldsymbol{\alpha}^{\text{T}}, \boldsymbol{\beta}^{\text{T}})^{\text{T}}$ can be written as follows:

$$\text{TPR}(\theta) = \frac{\text{TP}}{\text{TP} + \text{FN}},$$
$$\text{FPR}(\theta) = \frac{\text{FP}}{\text{TN} + \text{FP}}, \tag{24}$$
$$\text{MSE}(\theta) = \frac{\sum_{j=1}^{p} \left(\theta_j^{\text{true}} - \widehat{\theta}_j\right)^2}{\sum_{j=1}^{p} \left(\theta_j^{\text{true}}\right)^2},$$

where

$$\begin{aligned} \mathrm{TP} &= \sum_{j=1}^{p} I\left(\theta_j^{\mathrm{true}} \neq 0 \cap \widehat{\theta}_j \neq 0\right), \\ \mathrm{TP} + \mathrm{FN} &= \sum_{j=1}^{p} I\left(\theta_j^{\mathrm{true}} \neq 0\right), \\ \mathrm{FP} &= \sum_{j=1}^{p} I\left(\theta_j^{\mathrm{true}} = 0 \cap \widehat{\theta}_j \neq 0\right), \\ \mathrm{TN} + \mathrm{FP} &= \sum_{j=1}^{p} I\left(\theta_j^{\mathrm{true}} = 0\right). \end{aligned} \tag{25}$$

The relative root mean square error of the cure rate estimation (RMSE_π) and the relative root mean square error of the hazard function estimation (RMSE_λ) are

$$\begin{aligned} \mathrm{RMSE}_\pi &= \frac{\sum_{i=1}^{n} \left(\widehat{\pi}_i - \pi_i\right)^2}{\sum_{i=1}^{n} \pi_i^2}, \\ \mathrm{RMSE}_\lambda &= \frac{\sum_{i=1}^{n} \left(\widehat{\lambda}_i - \lambda_i\right)^2}{\sum_{i=1}^{n} \lambda_i^2}. \end{aligned} \tag{26}$$

Tables 2–5 show the mean TPR, FPR, and MSE of the coefficients, as well as the standard deviations over the 100 replicates for each example.

As indicated in Tables 2–5, the two group selection methods (MCGS and MCG) perform significantly better than the Full method. This is expected since the group lasso can select important subgroups of variables. Comparing the two group methods, the proposed method has competitive performance compared with the MCG method. It indicates that promoting sign consistency improves the performance in terms of estimation. For instance, under Scenario 1 in Example 1 with $\eta = 5$ and $p = 200$ in Table 2, the mean MSEs of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ for the proposed method are 0.12, 0.04, and 0.09, respectively, compared to 1.00, 0.05, and 0.63 for the MCG method and 17.3, 2.38, and 11.56 for the Full method.

Tables 6–8 show the mean RMSE of π and λ, as well as the standard deviations over the 100 replicates for each example. The results illustrate the performance in terms of prediction of the probability of nondefault and survival.

As shown in Tables 6–8 the prediction performance of group selection methods is significantly better than that of the Full method, and the proposed method has competitive performance compared with the MCG method. For example, in Example 2 with $\eta = 0.25$ and $p = 200$ in Table 8, the mean RMSE_π and RMSE_λ for the proposed method are 0.01, and 0.07, respectively, compared to 0.04 and 0.12 for the MCG method and 0.27 and 10.01 for the Full method. In addition, compared with the results regarding low- and high-dimensional settings, the group selection methods have greater advantages in prediction performance when the dimensionality is higher.

Results of simulation reveal that the proposed variable effect consistency mixture cure model can improve the performance in terms of estimation and prediction compared with alternative methods.

5. Empirical Study

In this section, we applied our proposed method to real data on credit loans. The data come from the personal loan department of a city commercial bank in China, which contains 4796 personal loan samples from 2014 to 2019 after preprocessing. The data include mortgage loans and credit loans, covering consumer durables, personal housing decoration loans, and other personal consumption loans. Censoring time C_i is the interval between the loan value date and either default or the end of observation (June 1, 2019). Therefore, censoring times of customers vary from individual to individual. It has a mean of 1.93 years and a standard deviation of 0.8. Customers whose time to default Y_i is longer than the censoring time C_i are censored ($\delta_i = 0$). 47 out of 4796 customers are censored. By transforming the discrete variables into dummy variables, the data contain 27 variables. Table 9 provides a list of variables and their descriptions.

In this section, the alternative method is a mixture cure model with group lasso (MCG). Different from the simulation, the real values of parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown in real data. Referring to [30, 35], we adopt the (1) log-rank statistics and (2) negative log-likelihood to evaluate the performance of the models instead of $\mathrm{TPR}(\theta)$, $\mathrm{FPR}(\theta)$, $\mathrm{ME}(\theta)$, RMSE_π, and RMSE_λ in simulation.

Log-rank statistics is a commonly used indicator in survival analysis to test the null hypothesis that there is no significant difference in survival distribution between two or more independent groups. It is calculated by cross-validation. We sequentially take 1/10 samples as the validation set and the remaining as the training set. Apply the proposed and alternative methods to obtain the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and then calculate the $\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}$ and $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$ for the validation set. Results of $\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}$ and $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$ are based on 10 replicates. Divide the calculated $\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}$ into two groups at the median and calculate the log-rank statistics. Similarly, divide the calculated $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$ into two groups at the median and calculate the log-rank statistics. The mean log-rank of the proposed and the alternative methods is 5.6 and 4.3 respectively, indicating better performance of the proposed method.

Figure 1 shows the Kaplan–Meier curves stratified by different groups. Kaplan–Meier curves are commonly used to describe the change of the survival probability overtime in survival analysis [36, 37]. The probability of being cured is negatively related to $\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}$, and the survival time is negatively related to $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$. In Figure 1, a group is denoted by "low risk" with lower $\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}$ (a) or lower $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$ (b) whereas another group is denoted by "high risk." As indicated in Figure 1, there are clearly different trends in the curves in different groups. Customers with lower $\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}$ and $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$ have lower risk and are less likely to default.

To assess the performance of the model, the data are randomly divided into training set and test set by 2 : 1. The training set is used for fitting the model and the test set is used for evaluating the prediction performance of the fitted model. The tuning parameters are selected by 5-fold cross-validation. The mean (standard deviation) negative log-likelihood of the proposed method (MCGS) and the alternative method (MCG) is 106.04 (16.04) and 118.60

Table 1: ECD algorithm.

Expectation Coordinate Descent algorithm
Initialize $m = 0$, $\alpha_0^{[m]} = \beta_0^{[m]} = 0$, and $\alpha^{[m]} = \beta^{[m]} = 0_{J\times p}$. Compute $\pi_i^{[m]}$ and $\lambda_i^{[m]}$ from (9) and (11)
Repeat
E-step
Update $E(y_i)$ from (10)
CD-step
Update $\boldsymbol{\alpha}_0^{[m+1]}$ from (14)
For $j = 1, \ldots, p$, update $\boldsymbol{\alpha}_j^{[m+1]}$ from (15); update $\pi_i^{[m+1]}$ from (9)
Update $\boldsymbol{\beta}_0^{[m+1]}$ from (17);
For $j = 1, \ldots, p$, update $\boldsymbol{\beta}_j^{[m+1]}$ from (18); update $\lambda_i^{[m+1]}$ from (11)
$m = m + 1$
Until $\max\left\{\left\|\boldsymbol{\alpha}_j^{[m+1]} - \boldsymbol{\alpha}_j^{[m]}\right\|, \left\|\boldsymbol{\beta}_j^{[m+1]} - \boldsymbol{\beta}_j^{[m]}\right\|\right\} \le 10^{-3}$

Table 2: Results of the estimation performance in scenario 1, Example 1.

η	p		$\boldsymbol{\alpha}$			$\boldsymbol{\beta}$			$(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$		
			Full	MCG	MCGS	Full	MCG	MCGS	Full	MCG	MCGS
0.25	60	MSE	0.65 (0.23)	0.13 (0.04)	0.07 (0.02)	0.13 (0.03)	0.04 (0.01)	0.03 (0.01)	0.34 (0.10)	0.07 (0.02)	0.05 (0.01)
		TPR	—	0.92 (0.14)	1.00 (0.01)	—	1.00 (0.00)	1.00 (0.00)	—	0.96 (0.07)	1.00 (0.00)
		FPR	—	0.18 (0.18)	0.28 (0.19)	—	0.59 (0.16)	0.48 (0.18)	—	0.38 (0.13)	0.38 (0.17)
	200	MSE	29.04 (9.28)	0.17 (0.04)	0.08 (0.02)	1.05 (0.18)	0.04 (0.01)	0.04 (0.01)	12.38 (3.76)	0.09 (0.02)	0.06 (0.01)
		TPR	—	0.83 (0.20)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	0.91 (0.10)	1.00 (0.00)
		FPR	—	0.02 (0.03)	0.07 (0.05)	—	0.28 (0.09)	0.20 (0.09)	—	0.15 (0.05)	0.14 (0.07)
5	60	MSE	9.17 (3.02)	0.42 (0.21)	0.23 (0.04)	0.51 (0.14)	0.07 (0.02)	0.07 (0.02)	4.01 (1.23)	0.21 (0.09)	0.13 (0.02)
		TPR	—	0.55 (0.24)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	0.77 (0.12)	1.00 (0.00)
		FPR	—	0.03 (0.08)	0.04 (0.07)	—	0.67 (0.15)	0.06 (0.11)	—	0.35 (0.09)	0.05 (0.09)
	200	MSE	17.30 (2.12)	1.00 (0.00)	0.12 (0.02)	2.38 (0.48)	0.05 (0.01)	0.04 (0.01)	11.56 (1.26)	0.63 (0.00)	0.09 (0.02)
		TPR	—	0.00 (0.00)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	0.50 (0.00)	1.00 (0.00)
		FPR	—	0.00 (0.00)	0.02 (0.01)	—	0.35 (0.08)	0.02 (0.01)	—	0.17 (0.04)	0.02 (0.01)

Note. In each cell, there is mean (standard deviation).

(19.09), respectively. The result is based on 100 duplicates. It indicates that the proposed method performs better than the alternative in terms of model fit and prediction.

Table 10 reports the estimations of the MCGS method and the MCG method. A positive coefficient $\boldsymbol{\alpha}$ indicates that the variable is positively related to the probability of default, and a positive coefficient $\boldsymbol{\beta}$ indicates that the variable is negatively related to default time. Both probabilities of default and default time are two quite relevant credit aspects. Compared with the alternative method, the signs of the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of the proposed method are promoted to be more consistent, whereas the business type in the MCG model has an opposite effect on the probability and time to default. The results show that promoting variable effect consistency can improve prediction performance as well as interpretability.

The coefficient results of the proposed method reveal that interest rate, loan line, business type, gender, education, and employment status are important variables that affect the probability of default and the time to default. Loan term, age, medical insurance, entrusted payment, early repayment, annual household income, type of workplace, and housing status have no significant impact on credit. The impact of occupation and professional title on credit is not clear.

From the perspective of loan products, we find that interest rate has a negative impact. This is not surprising, as higher interest rates lead to higher costs, and the customers are more likely to default. The loan line has a positive effect. One possible explanation is that low-risk customers are more likely to obtain a higher loan line. Loan term, entrusted payment, and early repayment have no effect on the credit.

Table 3: Results of the estimation performance in scenario 2, Example 1.

η	p		$\boldsymbol{\alpha}$			$\boldsymbol{\beta}$			$(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$		
			Full	MCG	MCGS	Full	MCG	MCGS	Full	MCG	MCGS
0.25	60	MSE	0.51 (0.19)	0.11 (0.03)	0.11 (0.03)	0.13 (0.03)	0.04 (0.01)	0.04 (0.01)	0.33 (0.11)	0.08 (0.02)	0.08 (0.02)
		TPR	—	0.99 (0.03)	0.99 (0.03)	—	1.00 (0.00)	1.00 (0.00)	—	0.99 (0.01)	0.99 (0.01)
		FPR	—	0.33 (0.23)	0.33 (0.23)	—	0.71 (0.13)	0.71 (0.13)	—	0.52 (0.14)	0.52 (0.14)
	200	MSE	24.30 (6.60)	0.16 (0.03)	0.16 (0.03)	1.01 (0.19)	0.05 (0.01)	0.05 (0.01)	13.40 (3.50)	0.11 (0.02)	0.11 (0.02)
		TPR	—	0.98 (0.03)	0.98 (0.03)	—	1.00 (0.00)	1.00 (0.00)	—	0.99 (0.02)	0.99 (0.02)
		FPR	—	0.04 (0.05)	0.04 (0.05)	—	0.35 (0.08)	0.35 (0.08)	—	0.19 (0.06)	0.19 (0.06)
5	60	MSE	5.56 (1.83)	0.32 (0.10)	0.33 (0.07)	0.53 (0.15)	0.10 (0.03)	0.11 (0.03)	3.21 (0.97)	0.22 (0.06)	0.23 (0.04)
		TPR	—	0.63 (0.19)	0.87 (0.20)	—	1.00 (0.01)	1.00 (0.01)	—	0.81 (0.10)	0.93 (0.10)
		FPR	—	0.07 (0.13)	0.08 (0.11)	—	0.70 (0.15)	0.36 (0.33)	—	0.38 (0.11)	0.22 (0.19)
	200	MSE	24.46 (3.45)	0.60 (0.25)	0.39 (0.04)	4.52 (0.85)	0.11 (0.03)	0.13 (0.03)	15.13 (1.71)	0.37 (0.14)	0.27 (0.02)
		TPR	—	0.38 (0.21)	0.93 (0.17)	—	1.00 (0.00)	1.00 (0.02)	—	0.69 (0.11)	0.96 (0.08)
		FPR	—	0.00 (0.00)	0.02 (0.04)	—	0.32 (0.12)	0.08 (0.13)	—	0.16 (0.06)	0.05 (0.08)

Note. In each cell, there is mean (standard deviation).

Table 4: Results of the estimation performance in Example 2.

η	p		$\boldsymbol{\alpha}$			$\boldsymbol{\beta}$			$(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$		
			Full	MCG	MCGS	Full	MCG	MCGS	Full	MCG	MCGS
0.25	60	MSE	0.35 (0.13)	0.13 (0.03)	0.07 (0.02)	0.09 (0.02)	0.04 (0.01)	0.02 (0.01)	0.25 (0.08)	0.09 (0.02)	0.05 (0.01)
		TPR	—	1.00 (0.00)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)
		FPR	—	0.37 (0.24)	0.08 (0.06)	—	0.76 (0.09)	0.12 (0.07)	—	0.57 (0.13)	0.10 (0.06)
	200	MSE	20.24 (4.57)	0.23 (0.05)	0.09 (0.02)	0.68 (0.12)	0.04 (0.01)	0.02 (0.01)	12.70 (2.81)	0.16 (0.03)	0.06 (0.01)
		TPR	—	1.00 (0.00)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)
		FPR	—	0.02 (0.04)	0.04 (0.02)	—	0.37 (0.09)	0.06 (0.03)	—	0.20 (0.05)	0.05 (0.02)
5	60	MSE	4.72 (1.08)	0.98 (0.10)	0.10 (0.02)	0.30 (0.07)	0.04 (0.01)	0.03 (0.01)	3.02 (0.66)	0.62 (0.06)	0.07 (0.02)
		TPR	—	0.05 (0.22)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	0.53 (0.11)	1.00 (0.00)
		FPR	—	0.00 (0.00)	0.03 (0.03)	—	0.64 (0.17)	0.04 (0.03)	—	0.32 (0.09)	0.03 (0.03)
	200	MSE	17.30 (2.12)	1.00 (0.00)	0.12 (0.02)	2.38 (0.48)	0.05 (0.01)	0.04 (0.01)	11.56 (1.26)	0.63 (0.00)	0.09 (0.02)
		TPR	—	0.00 (0.00)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	0.50 (0.00)	1.00 (0.00)
		FPR	—	0.00 (0.00)	0.02 (0.01)	—	0.35 (0.08)	0.02 (0.01)	—	0.17 (0.04)	0.02 (0.01)

Note. In each cell, there is mean (standard deviation).

Table 5: Results of the estimation performance in Example 3.

η	p		$\boldsymbol{\alpha}$			$\boldsymbol{\beta}$			$(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$		
			Full	MCG	MCGS	Full	MCG	MCGS	Full	MCG	MCGS
0.25	60	MSE	0.33 (0.08)	0.19 (0.05)	0.10 (0.02)	0.04 (0.01)	0.03 (0.01)	0.02 (0.01)	0.22 (0.05)	0.13 (0.03)	0.07 (0.01)
		TPR	—	1.00 (0.01)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)
		FPR	—	0.53 (0.15)	0.29 (0.12)	—	0.66 (0.08)	0.29 (0.12)	—	0.60 (0.09)	0.29 (0.12)
	200	MSE	2.42 (0.57)	0.48 (0.14)	0.17 (0.02)	0.20 (0.03)	0.05 (0.01)	0.03 (0.01)	1.56 (0.35)	0.31 (0.08)	0.11 (0.01)
		TPR	—	0.95 (0.18)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	0.98 (0.09)	1.00 (0.00)
		FPR	—	0.05 (0.07)	0.14 (0.06)	—	0.57 (0.10)	0.16 (0.06)	—	0.31 (0.07)	0.15 (0.06)
5	60	MSE	1.31 (0.33)	0.99 (0.08)	0.43 (0.29)	0.08 (0.02)	0.03 (0.01)	0.03 (0.01)	0.83 (0.20)	0.62 (0.05)	0.27 (0.18)
		TPR	—	0.01 (0.10)	0.83 (0.38)	—	1.00 (0.00)	1.00 (0.00)	—	0.51 (0.05)	0.92 (0.19)
		FPR	—	0.00 (0.02)	0.14 (0.14)	—	0.51 (0.13)	0.36 (0.17)	—	0.25 (0.07)	0.25 (0.12)
	200	MSE	6.94 (1.12)	1.00 (0.00)	0.28 (0.06)	0.51 (0.07)	0.04 (0.01)	0.05 (0.01)	4.47 (0.68)	0.63 (0.00)	0.19 (0.04)
		TPR	—	0.00 (0.00)	1.00 (0.00)	—	1.00 (0.00)	1.00 (0.00)	—	0.50 (0.00)	1.00 (0.00)
		FPR	—	0.00 (0.00)	0.17 (0.06)	—	0.34 (0.08)	0.22 (0.06)	—	0.17 (0.04)	0.19 (0.06)

Note. In each cell, there is mean (standard deviation).

Table 6: Results of the prediction performance in Example 1.

η	p	RMSE_{π}			RMSE_{λ}		
		Full	MCG	MCGS	Full	MCG	MCGS
				Scenario 1			
0.25	60	0.06 (0.02)	0.02 (0.01)	0.01 (0.00)	0.57 (0.80)	0.10 (0.08)	0.12 (0.08)
	200	0.37 (0.04)	0.03 (0.01)	0.02 (0.01)	24.10 (57.28)	0.12 (0.08)	0.16 (0.10)
5	60	0.33 (0.05)	0.19 (0.18)	0.10 (0.03)	6.32 (20.47)	0.23 (0.28)	0.18 (0.09)
	20	0.41 (0.03)	0.96 (0.04)	0.02 (0.01)	>10^3 (>10^3)	0.15 (0.15)	0.10 (0.08)
				Scenario 2			
0.25	60	0.06 (0.01)	0.02 (0.01)	0.02 (0.01)	0.52 (0.71)	0.12 (0.07)	0.12 (0.07)
	200	0.31 (0.03)	0.03 (0.01)	0.03 (0.01)	72.58 (387.68)	0.15 (0.10)	0.15 (0.10)
5	60	0.30 (0.06)	0.08 (0.04)	0.14 (0.06)	5.72 (10.42)	0.21 (0.18)	0.27 (0.15)
	200	0.53 (0.04)	0.26 (0.24)	0.17 (0.04)	>10^3 (>10^3)	0.26 (0.17)	0.35 (0.17)

Note. In each cell, there is mean (standard deviation).

Different coefficient of business type reveals that, compared with other personal loans, consumer durables are more likely to default.

From the perspective of the influence of the variables of the customers, employed customers have a positive impact. Age, annual household income, housing status, and type of workplace have no significant effect on the credit. Compared with women, men are more likely to default. This is consistent with the results of [38] and the personality characteristics of men's risk preference [39]. Customers with higher education are less likely to default. Bachelor degree or above has a positive effect on credit. Generally, customers

Table 7: Results of the prediction performance in Example 2.

η	p	$RMSE_\pi$			$RMSE_\lambda$		
		Full	MCG	MCGS	Full	MCG	MCGS
0.25	60	0.05 (0.01)	0.02 (0.01)	0.01 (0.00)	0.36 (0.35)	0.10 (0.07)	0.06 (0.05)
	200	0.27 (0.03)	0.04 (0.01)	0.01 (0.00)	10.01 (33.80)	0.12 (0.17)	0.07 (0.06)
5	60	0.26 (0.03)	0.88 (0.16)	0.02 (0.01)	1.98 (4.56)	0.21 (0.30)	0.09 (0.08)
	200	0.41 (0.03)	0.96 (0.04)	0.02 (0.01)	>10^3 (>10^3)	0.15 (0.15)	0.10 (0.08)

Note. In each cell, there is mean (standard deviation).

Table 8: Results of the prediction performance in Example 3.

η	p	$RMSE_\pi$			$RMSE_\lambda$		
		Full	MCG	MCGS	Full	MCG	MCGS
0.25	60	0.21 (0.04)	0.11 (0.03)	0.06 (0.02)	0.04 (0.01)	0.03 (0.01)	0.02 (0.01)
	200	0.83 (0.10)	0.28 (0.08)	0.10 (0.02)	0.30 (0.07)	0.05 (0.01)	0.03 (0.01)
5	60	1.00 (0.19)	0.77 (0.09)	0.34 (0.21)	0.09 (0.03)	0.03 (0.01)	0.03 (0.01)
	200	2.62 (0.21)	0.72 (0.06)	0.21 (0.10)	0.99 (0.33)	0.04 (0.01)	0.04 (0.01)

Note. In each cell, there is mean (standard deviation).

Table 9: Description of variables.

Variables	Description
Business type	Consumer durables, personal housing decoration loans, and other personal consumption loans
Interest rate	$[0.029, 0.095]$
Loan line	$(0, +\infty)$
Loan term	$(0, +\infty)$
Early repayment	Yes, no
Entrusted payment	Yes, no
Age	$[18, 70]$
Gender	Male, female
Education	Master/doctor, bachelor, vocational education, high school, and below
Medical insurance	Yes, no
Housing status	Self-purchasing (with mortgage), self-purchasing (without mortgage), and others
Employment	Employed and others
Type of workplace	Government organization/institution, firm, and others
Occupation	Managers, commercial and service workers, and others
Professional title	Advanced, intermediate, primary, and no professional title
Annual household income (RMB)	$\leq 200,000$, $200,000 - 400,000$, $400,000 - 600,000$, $\geq 600,000$, and unknown

with higher education have a higher chance of getting decent jobs and income. They tend to maintain good credit records and are less likely to default. Compared with other employment groups such as self-employed, freelance, and unemployed ones, the employed group has a more stable income and is less likely to default.

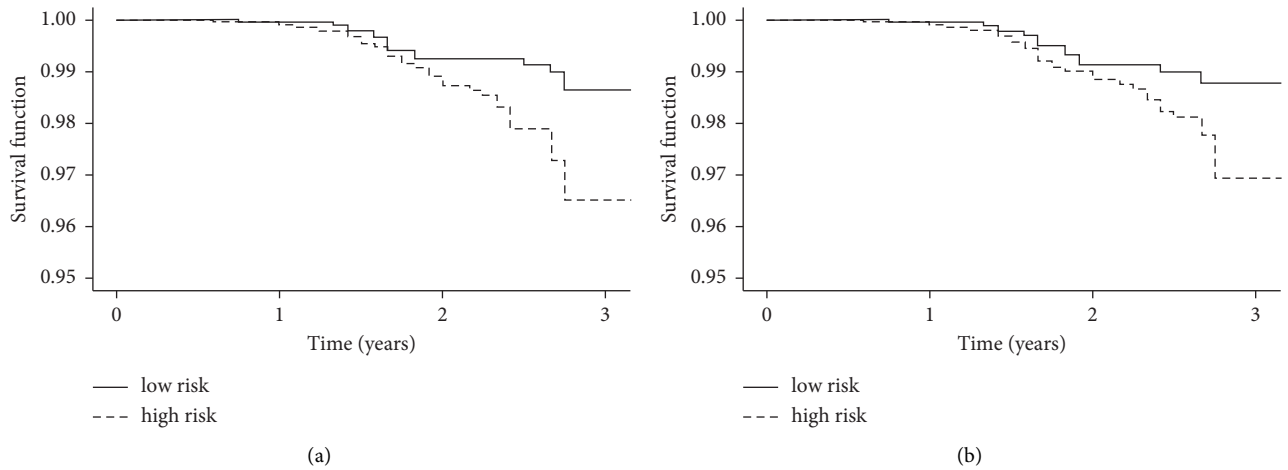


Figure 1: Kaplan–Meier curves stratified by different groups: (a) divided by the median of $\mathbf{x}^{\mathrm{T}}\boldsymbol{\alpha}$ and (b) divided by the median of $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$.

Table 10: Estimation of coefficients of the methods.

	MCG		MCGS	
	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$
$\boldsymbol{\alpha}_0/\boldsymbol{\beta}_0$	−1.17	−4.07	−1.43	−3.67
Business type				
Consumer durables	0.15	0.70	0.11	0.12
Personal housing decoration loans	−0.02	0.21	0	0
Interest rate	0.54	0.17	0.34	0.28
Loan line	−0.06	0	−0.03	−0.03
Loan term	0	0	0	0
Early repayment (yes)	0	−0.24	0	0
Entrusted payment (yes)	−0.13	0	0	0
Age	0	0	0	0
Gender (male)	0	0.82	0.25	0.30
Education				
Master/doctor	−0.03	−0.32	−0.02	−0.02
Bachelor	−0.04	−0.22	−0.06	−0.06
Vocational education	0.02	0.32	0.05	0.05
Medical insurance (yes)	0	0.34	0	0
Housing status				
Self-purchasing (with mortgage)	0	−0.01	0	0
Self-purchasing (without mortgage)	0	0.04	0	0
Employment (employed)	−0.08	−0.31	−0.24	−0.21
Type of workplace				
Government organization and institution	−0.14	−1.53	0	0
Firm	−0.14	−0.40	0	0
Occupation				
Managers	0.09	0.09	0.04	0.04
Commercial and service workers	−0.25	−0.29	−0.11	−0.11
Professional title				
Advanced	0.04	0.41	0.04	0.04
Intermediate	−0.02	−0.25	−0.04	−0.04
Primary	0	−0.03	0	0
Annual household income (RMB)				
≤200,000	0	0.05	0	0
200,000 – 400,000	0	0.38	0	0
400,000 – 600,000	0	−0.29	0	0
≥600,000	0	−0.19	0	0

6. Conclusions

The mixture cure model is widely adopted in credit scoring for its ability of predicting whether customers will default and when they are likely to default. However, most of the existing studies ignore the relations between the two sets of variable effects in the two model parts which may get conflicting results of variable effects. It can be difficult to interpret the results and apply them in practice.

In this paper, we propose a variable effect consistency mixture cure model, to promote the similarity of the sign of variables in the two model parts by imposing a sign consistency penalty. Meanwhile, to accommodate the high-dimensional credit data, we also impose a group lasso penalty to conduct variable selection and parameter estimation. Simulation shows that the proposed method has competitive performance compared with the MCG method and significantly outperforms the Full method in terms of estimation and prediction. Furthermore, the empirical study illustrates that the proposed method can improve prediction performance as well as interpretability. The results of the variable effect consistency mixture cure model also offer additional insights into the relationship between the variable effect before and after loan.

Data Availability

The raw/processed data used in the empirical study cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Office for Philosophy and Social Sciences of China under Grant no. 20&ZD137 and the National Bureau of Statistics of China under Grant no. 2020ZX20.

References

[1] D. M. B. Silva, G. H. A. Pereira, and T. M. Magalhães, "A class of categorization methods for credit scoring models," *European Journal of Operational Research*, vol. 296, no. 1, pp. 323–331, 2022.

[2] P. Pławiak, M. Abdar, J. Pławiak, V. Makarenkov, and U. R. Acharya, "A new deep genetic hierarchical network of learners for prediction of credit scoring," *Information Sciences*, vol. 516, pp. 401–418, 2020.

[3] R. Y. Goh, L. S. Lee, H.-V. Seow, and K. Gopal, "Hybrid harmony search–artificial intelligence models in credit scoring," *Entropy*, vol. 22, no. 9, https://doi.org/10.3390/e22090989, Article ID 989, 2020.

[4] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," *Journal of the Operational Research Society*, vol. 54, no. 6, pp. 627–635, 2003.

[5] S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, 2015.

[6] Q. Zhou, L. Wang, L. Juan, S. Zhou, and L. Li, "The study on credit risk warning of regional listed companies in China based on logistic model," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 6672146, 8 pages, 2021.

[7] L. Zhang, H. Ray, J. Priestley, and S. Tan, "A descriptive study of variable discretization and cost-sensitive logistic regression on imbalanced credit data," *Journal of Applied Statistics*, vol. 47, no. 3, pp. 568–581, 2020.

[8] S. Maldonado, G. Peters, and R. Weber, "Credit scoring using three-way decisions with probabilistic rough sets," *Information Sciences*, vol. 507, pp. 700–714, 2020.

[9] C. Lohmann and T. Ohliger, "Using accounting-based and loan-related information to estimate the cure probability of a defaulted company," *European Financial Management*, vol. 27, pp. 620–640, 2021, https://doi.org/10.1111/eufm.12279.

[10] B. Narain, "Survival Analysis and the Credit Granting Decision," in *Credit Scoring And Credit Control*, pp. 109–121, Oxford University Press, New York, NY, USA, 1992.

[11] J. Banasik, J. N. Crook, and L. C. Thomas, "Not if but when will borrowers default," *Journal of the Operational Research Society*, vol. 50, no. 12, pp. 1185–1190, 1999.

[12] M. Stepanova and L. Thomas, "Survival analysis methods for personal loan data," *Operations Research*, vol. 50, no. 2, pp. 277–289, 2002.

[13] T. Bellotti and J. Crook, "Credit scoring with macroeconomic variables using survival analysis," *Journal of the Operational Research Society*, vol. 60, no. 12, pp. 1699–1707, 2009.

[14] V. B. Djeundje and J. Crook, "Dynamic survival models with varying coefficients for credit risks," *European Journal of Operational Research*, vol. 275, no. 1, pp. 319–333, 2019.

[15] Y. Xia, L. He, Y. Li, Y. Fu, and Y. Xu, "A dynamic credit scoring model based on survival gradient boosting decision tree approach," *Technological and Economic Development of Economy*, vol. 27, no. 1, pp. 96–119, 2021, https://doi.org/10.3846/tede.2020.13997.

[16] E. N. C. Tong, C. Mues, and L. C. Thomas, "Mixture cure models in credit scoring: if and when borrowers default," *European Journal of Operational Research*, vol. 218, no. 1, pp. 132–139, 2012.

[17] L. Dirick, G. Claeskens, and B. Baesens, "Time to default in credit scoring using survival analysis: a benchmark study," *Journal of the Operational Research Society*, vol. 68, no. 6, pp. 652–665, 2017.

[18] B. C. Alves and J. G. Dias, "Survival mixture models in behavioral scoring," *Expert Systems with Applications*, vol. 42, no. 8, pp. 3902–3910, 2015.

[19] N. Zhang, Q. Yang, A. Kelleher, and W. Si, "A new mixture cure model under competing risks to score online consumer loans," *Quantitative Finance*, vol. 19, no. 7, pp. 1243–1253, 2019.

[20] C. Jiang, Z. Wang, and H. Zhao, "A prediction-driven mixture cure model and its application in credit scoring," *European Journal of Operational Research*, vol. 277, no. 1, pp. 20–31, 2019.

[21] L. Dirick, T. Bellotti, G. Claeskens, and B. Baesens, "Macroeconomic factors in credit risk calculations: including time-varying covariates in mixture cure models," *Journal of Business & Economic Statistics*, vol. 37, no. 1, pp. 40–53, 2019.

[22] C. Han and R. Kronmal, "Two-part models for analysis of Agatston scores with possible proportionality constraints," *Communications in Statistics-Theory and Methods*, vol. 35, no. 1, pp. 99–111, 2006.
[23] F. Liu, Z. Hua, and A. Lim, "Identifying future defaulters: a hierarchical Bayesian method," *European Journal of Operational Research*, vol. 241, no. 1, pp. 202–211, 2015.
[24] Q. Zhang, S. Ma, and Y. Huang, "Promote sign consistency in the joint estimation of precision matrices," *Computational Statistics & Data Analysis*, vol. 159, 2021 https://doi.org/10.1016/j.csda.2021.107210, Article ID 107210.
[25] Q. Zhang, S. Zhang, J. Liu, J. Huang, and S. Ma, "Penalized integrative analysis under the accelerated failure time model," *Statistica Sinica*, vol. 26, no. 2, pp. 492–508, 2016.
[26] J. Huang, P. Breheny, and S. Ma, "A selective review of group selection in high-dimensional models," *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, vol. 27, pp. 481–499, 2012.
[27] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B*, vol. 68, no. 1, pp. 49–67, 2006.
[28] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY, USA, 2009.
[29] M. E. Ghitany, R. A. Maller, and S. Zhou, "Exponential mixture models with long-term survivors and covariates," *Journal of Multivariate Analysis*, vol. 49, no. 2, pp. 218–241, 1994.
[30] X. Fan, M. Liu, K. Fang, Y. Huang, and S. Ma, "Promoting structural effects of covariates in the cure rate model with penalization," *Statistical Methods in Medical Research*, vol. 26, no. 5, pp. 2078–2092, 2017.
[31] C. Zheng, J. Zhu, and J. Zhu, "Promote sign consistency in cure rate model with Weibull lifetime," *AIMS Mathematics*, vol. 7, no. 2, pp. 3186–3202, 2022.
[32] X. Shi, S. Ma, and Y. Huang, "Promoting sign consistency in the cure model estimation and selection," *Statistical Methods In Medical Research*, vol. 29, no. 1, pp. 15–28, 2020.
[33] Y. Yang and H. Zou, "A fast unified algorithm for solving group-lasso penalize learning problems," *Statistics and Computing*, vol. 25, no. 6, pp. 1129–1141, 2015.
[34] K. Fang, X. Fan, Q. Zhang, and S. Ma, "Integrative sparse principal component analysis," *Journal of Multivariate Analysis*, vol. 166, pp. 1–16, 2018.
[35] J. Rodrigues, M. de Castro, V. G. Cancho, and N. Balakrishnan, "COM-Poisson cure rate survival models and an application to a cutaneous melanoma data," *Journal of Statistical Planning and Inference*, vol. 139, no. 10, pp. 3605–3611, 2009.
[36] T. Chen and P. Du, "Promotion time cure rate model with nonparametric form of covariate effects," *Statistics in Medicine*, vol. 37, no. 10, pp. 1625–1635, 2018.
[37] S. Pal and N. Balakrishnan, "Likelihood inference based on EM algorithm for the destructive length-biased Poisson cure rate model with Weibull lifetime," *Communications in Statistics - Simulation and Computation*, vol. 47, no. 3, pp. 644–660, 2018.
[38] Y. Li, Y. Li, and Y. Li, "What factors are influencing credit card customer's default behavior in China? A study based on survival analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 526, Article ID 120861, 2019.
[39] Y. Shu and Q. Y. Yang, "Research on auto loan default prediction based on large sample data model," *Management Review*, vol. 29, no. 9, pp. 59–71, 2017.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 9520486, 16 pages
https://doi.org/10.1155/2022/9520486

Research Article

On the Safe-Haven Ability of Bitcoin, Gold, and Commodities for International Stock Markets: Evidence from Spillover Index Analysis

Qian Wang, Yu Wei, Yao Wang, and Yuntong Liu

School of Finance, Yunnan University of Finance and Economics, Kunming, Yunnan 650221, China

Correspondence should be addressed to Yu Wei; weiyusy@126.com

Received 15 November 2021; Accepted 18 December 2021; Published 4 January 2022

Academic Editor: Dehua Shen

Copyright © 2022 Qian 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.

Stock market is susceptible to various external shocks for its tight dependence on economic fundamentals, financial speculation, and fragile emotions in massive traders, making it a very risky market for investors. In this paper, we aim to identify whether commonly recognized safe-haven assets, that is, bitcoin, gold, and commodities, can provide investors with effective hedging utility in international stock markets, especially during periods of extreme market turbulence. By using the spillover index method based on the TVP-VAR model, we find that firstly, bitcoin, gold, and commodities can only offer weak hedging effects on stock markets. Furthermore, their abilities to act as a safe haven are ranked as: commodities > gold > bitcoin. Secondly, in general, we have observed the increasing hedging ability of these safe-haven assets in times of extreme market turmoil. Thirdly, among international stock and safe-haven asset markets, the world and the developed stock markets act as the net spillover transmitters, while bitcoin, gold, and commodities are the net recipients. Lastly, the total spillover effects are time-varying and increase significantly after the outbreak of extreme events.

1. Introduction

"Black swan" events not only can impact the healthy development of the global economy but also can seriously pose a threat to the investors in the financial market. For instance, the outbreak of the European debt crisis deeply affects global stock markets [1, 2]. Moreover, the COVID-19 pandemic has made the global stock markets fluctuate frequently and led to a cross-market spillover of financial risks [3]. Specifically, the market spillover effects mean that when a "Black swan" event happens in a specific market, the risk would be spread to other markets or the transfer of information from one market to others could be witnessed. In fact, it is extremely difficult to ignore the existence of the market spillovers for investors in view of the returns on the investment portfolio, international investors, and portfolio managers, as the investors are always searching for proper assets or risk diversification. Therefore, it is significant to identify the safe-haven assets on global stock markets when the "Black swan" events occur.

Generally, gold, bitcoin, as well as commodities are considered three popular safe havens for stocks [4–6]. However, the conclusions on the hedging abilities of these assets have not reached a consensus on account of the time-varying properties of risk spillover, which makes it difficult for investors to compare the performances of gold, bitcoin, and commodities as a safe haven of the stock markets when extreme events occur [7].

Scholars have long debated on the performance of gold as a hedging instrument or a safe haven for stocks. Conventionally, as the saying of "gone with the gold," gold is regarded as an effective instrument protecting stock market investment from a decline [8–11]. For example, Triki and Ben Maatoug [12] analyzed the relationship between gold prices returns and the S&P500 stock index during the 1985 to 2018 period using the DCC MV-GARCH model to test the hedging role of gold and found that gold is a good diversifier and safe haven, particularly during the great tension. Additionally, He et al. [13] stated that gold does act as a good

diversifier in the US and the UK stock indices by applying the basic capital asset pricing model (CAPM). Conversely, recent studies have questioned the conclusions on gold as an effective hedging asset for stock markets [14, 15]. For instance, Disli et al. [16] examined the spillover effects between global equities and gold during the COVID-19 pandemic using a wavelet coherence model and proved gold lose hedging properties because of the nature of the underlying forces of the current pandemic.

The role of bitcoin as a safe-haven asset is also under discussion for years. Some analyses [17, 18] have posited that bitcoin can serve as an effective diversifier. For example, Kliber et al. [19] claimed that bitcoin is a safe-haven asset in Venezuela by applying the stochastic volatility model. Bouri et al. [20] used a dynamic conditional correlation model to explore whether bitcoin acts as a hedge and safe haven for major world stock indices (such as MSCI World, MSCI Europe, etc.); the results also verified the hedge and safe-haven properties of bitcoin. However, evidences rejecting the saying that bitcoin acts as a safe haven are not rare [21, 22]. In reaction to the COVID-19 bear market, Choi and Shin [23] utilized a vector autoregression (VAR) model to provide systematic evidence on the relationship among inflation, uncertainty, and bitcoin and gold prices during the COVID-19 pandemic and found that bitcoin prices increased significantly in response to a positive shock on the stock market, suggesting that bitcoin does not serve as a hedge for investment in stock markets.

Commodity indices investments are also debated assets for their safe-haven property. Some scholars believed that commodity indices are safe-haven assets to diversify risk in the stock market [24]. Bekiros et al. [25] proved that commodity indices in general reduced the risk in stock and increased the risk-adjusted performance from the full sample period (2002–2015). By contrast, other studies have asserted that commodities fail to provide a hedge. For example, Ali et al. [26] clarified the roles of commodities as a hedge for international stock markets using the cross-quantilogram approach from 2001 to 2018. Their conclusions proved that commodities were the safe-haven asset for the stock markets in emerging and frontier market groups.

Even though there are already some researches comparing the safe-haven abilities of bitcoin, gold, and commodity by couples. For example, the conditional diversification benefits offered by gold to equity investments in the G7 markets are comparatively much higher and more stable than those of bitcoin [27]. However, there is still a lack of empirical comparison among the safe-haven properties of bitcoin, gold, and commodities against world, developed, and emerging stock markets.

Several inadequacies exist in the research on the spillover among safe-haven instruments for stock markets during the market downturn. In the literature on which is proper safe-haven assets for stock markets, the relative importance of bitcoin, gold, and commodities seems not to be specific. Furthermore, the current study, which is to compare the role of bitcoin, gold, and commodities as a safe haven for stock markets during extreme events from a dynamic standpoint, has been ignored. Moreover, directional net spillover has long been ignored by most studies. For example, Shahzad et al. [6] analyzed the safe-haven property by the cross-quantilogram model. Bouri et al. [4] utilized a wavelet coherency approach to compare the safe-haven roles, which clearly ignored net directional spillover of safe-haven assets.

In order to identify the most effective safe-haven assets for stock markets in different periods, we use the TVP-VAR (time-varying parameter vector autoregressive) model by estimating the spillovers among bitcoin, gold, and commodities and stock markets following Diebold and Yilmaz [28]. The benefits of these methods are as follows. (1) Compared with the VAR (vector autoregressive) model, by using the TVP-VAR model, we can easily capture the potential time-varying features in the economy in a flexible and robust way. (2) The parameters in the model are assumed as a stochastic wandering process, which portray the spillover effects for multiple markets at different time points in a better way. (3) The window size is not necessary to be selected, which has the advantage of containing all the information without losing observations. (4) In the view of the net directional spillover analysis, the model applied is particularly useful in distinguishing the net transmitters and net receivers of a system. (5) The model can be more effective for identifying the attribute of a safe haven, as the rolling window analyses reveal evidence of time-variation in the safe-haven roles [6].

Therefore, this paper addresses the question of whether bitcoin, gold, and commodities exhibit a safe-haven property for the stock markets when facing the "black swan" events by adopting dynamic spillover method. This study makes contributions to the existing literature in three aspects: on one hand, this paper compares the performances of bitcoin, gold, and commodities as safe havens for world, developed, and emerging stock markets during the European debt crisis and COVID-19 pandemic from the perspectives of dynamic spillover to propose references for investors in searching for the best asset as hedge objectives during black swan events in the stock market. On the other hand, to the best of our knowledge, most studies on safe haven have only focused on the full sample phase, while we divide our sample into three phases according to the European debt crisis and COVID-19 pandemic to help investors alter their investment strategies during different periods. What is more, our research employs TVP-VAR based on the study of Diebold and Yilmaz [28], which is useful in recognizing the main transmitter and recipient, and provides information to the investors in forming strategies across markets.

The remaining part of the paper proceeds as follows. Section 2 is concerned with the methodology used for this study. Section 3 gives a brief overview of the data. The findings of the study are focused on Section 4. Section 5 displays the robustness of the empirical results by using a static spillover measurement and a dynamic framework on three different rolling windows. The conclusions are presented in Section 6.

2. Methodology

2.1. Time-Varying Parameter Vector Autoregressive Model. There are a number of instruments available for measuring the spillover among variables. The Bayesian time-varying parameter vector autoregressive (TVP-VAR) method is one of the practical ways of capturing the dynamic spillover between a variety of financial products.

Firstly, a structural vector autoregressive (SVAR) model is constructed as follows:

$$Ay_t = F_1 y_{t-1} n + \cdots + F_s y_{t-s} + \mu_t, \quad t = s+1, \ldots, n, \tag{1}$$

where y_t is the $k \times 1$ dimensional observed variable; in this paper, y_t is the observed variable of 6×1 dimensional consisting of bitcoin, gold, commodities index, world, developed, and emerging stocks. $A, F_1, \ldots, F_s$ are $k \times k$ dimensional coefficient matrices. The residuals μ_t is $k \times 1$ dimensional structural shock variables, and the contemporaneous relationship of the structural shocks is defined as a lower triangular matrix A: $A = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ a_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{k1} & \ldots & a_{k,k-1} & 1 \end{pmatrix}$.

On the basis of the above model, we rewrite equation (1) as a VAR model in reduced form as follows:

$$y_t = B_1 y_{t-1} + \cdots + B_s y_{t-s} + A^{-1} \sum \varepsilon_t, \tag{2}$$

where $\varepsilon_t \sim N(0, I_k)$, $B_i = A^{-1} F_i$, $i = 1, \ldots, s$, $\sum = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \sigma_k \end{pmatrix}$, and $\sigma_i (i = 1, \ldots, k)$ is the standard deviation of the structural shock. The elements in B is stacked by means of a row stack into a $k^2 s \times 1$ dimensional column vector β, defined by $X_t = I_k \otimes (y_{t-1}, \ldots, y_{t-s})$, where $\otimes$ denotes the Kronecker product. Then equation (2) can be expressed as follows:

$$y_t = X_t \beta + A^{-1} \sum \varepsilon_t. \tag{3}$$

Then, a time-varying vector autoregression model is defined as follows:

$$y_t = X_t \beta_t + A_t^{-1} \sum_t \varepsilon_t, \tag{4}$$

where coefficient β_t, parameters A_t, and covariance matrix of random fluctuations $\sum_t$ are both time-varying in nature. Let a_t is a stacked element in the lower triangular matrix A_t, then denoted a_t as $a_t = (a_{21}, a_{31}, a_{32}, a_{41}, \ldots, a_{k,k-1})$. At the same time, we suppose $h_t = (h_{1t}, \ldots, h_{kt})'$, where $h_{jt} = \log \sigma_{jt}^2$, $j = 1, \ldots, k$, $t = s+1, \ldots, n$.

Del Negro and Primiceri [29] proposed that even if the parameters are smooth and stable, the constructed model still has many parameters to be estimated. In this paper, we assume that all parameters in equation (4) obey the first-order random wandering process to reduce the number of parameters to be estimated, which finally results in: $\beta_{t+1} = \beta_t + \mu_{\beta t}$, $a_{t+1} = a_t + \mu_{at}$, and $h_{t+1} = h_t + \mu_{ht}$; in the meantime,

$$\begin{pmatrix} \varepsilon_t \\ \mu_{\beta t} \\ \mu_{at} \\ \mu_{ht} \end{pmatrix} \sim N\left(0, \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & \sum_{\beta} & 0 & 0 \\ 0 & 0 & \sum_{a} & 0 \\ 0 & 0 & 0 & \sum_{h} \end{pmatrix}\right), \tag{5}$$

where $t = s+1, \ldots, n$, $\beta_{s+1} \sim N(\mu_{\beta_0}, \sum_{\beta_0})$, $a_{s+1} \sim N(\mu_{a_0}, \Sigma_{a_0})$, and $h_{s+1} \sim N(\mu_{h_0}, \Sigma_{h_0})$. Assuming $\sum_{\beta}$, $\sum_{a}$, and $\sum_{h}$ are symmetry matrices, which accounts for time-varying parameters β_t, a_t, and h_t mutually irrelevant.

2.2. Spillover Indices Based on Diebold–Yilmaz Methods. In order to calculate the generalized impulse response functions (GIRF) and generalized forecast error variance decomposition (GFEVD), VAR can be transformed into an infinite-order vector moving average (VMA) process [30]:

$$y_t = \sum_{k=0}^{\infty} L' W_t^k L \varepsilon_{t-k}, y_t = \sum_{k=0}^{\infty} A_{kt} \varepsilon_{t-k}, \tag{6}$$

where $L = [I_N, \ldots, O_p]'$, $W = [\beta_t; I_{N(p-1)}, O_{N(p-1) \times N}]$, and A_{jt} represent, respectively, an $Np \times N$, $Np \times Np$, and $N \times N$ dimensional matrix.

The GIRF$_s$ reveals that to what extent the trajectory of a variable is due to the impact of itself or other variables in the system. The variance decomposition method measures the proportion of prediction error variance in any endogenous variable affected by different information shocks in the VAR system [31, 32]. The proportion explained by the variable k in the H-step-ahead prediction error variance of variable j, which can be calculated by

$$\begin{aligned} \mathrm{GIRF}_t\left(H, \delta_{k,t}, F_{t-1}\right) &= E\left(Y_{t+H} | \varepsilon_{k,t} = \delta_{k,t}, F_{t-1}\right) - E\left(Y_{t+H} | F_{t-1}\right), \\ \Psi_{k,t}^{g}(H) &= \frac{A_{H,t} S_t \varepsilon_{k,t}}{\sqrt{S_{kk,t}}} - \frac{\delta_{k,t}}{\sqrt{S_{kk,t}}}, \\ \Psi_{k,t}^{g}(H) &= \frac{1}{\sqrt{S_{kk,t}}} A_{H,t} S_t \varepsilon_{k,t}, \end{aligned} \tag{7}$$

where $\Psi_{k,t}^{g}(H)$ represents the GIRF$_s$ of variable k, $\delta_{k,t}$ represents selection vector with one on the k-th position and zero otherwise, and F_{t-1} the information set until $t-1$. In the next moment, GFEVD is $\widetilde{\phi}_{kj,t}^{g}(H)$, which can be explained the variance share one variance has on others.

$$\widetilde{\phi}_{jk,t}^{g}(H) = \frac{\sum_{t=1}^{H-1} \Psi_{jk,t}^{2,g}}{\sum_{k=1}^{N} \sum_{t=1}^{H-1} \Psi_{jk,t}^{2,g}}, \tag{8}$$

where $\sum_{k=1}^{N} \widetilde{\phi}_{jk,t}^{g}(H) = 1$ and $\sum_{j,k=1}^{N} \widetilde{\phi}_{jk,t}^{g}(H) = N$. According to the GFEVD, the total spillover index measures the overall

spillover level of bitcoin, gold, commodities, and stock markets systems, which reveals the influence proportion of information spillover contribution to system change.

$$C_t^g(H) = 100 * \frac{\sum_{j,k=1,j\neq k}^{N} \widetilde{\phi}_{jk,t}^g(H)}{\sum_{j,k=1}^{N} \widetilde{\phi}_{jk,t}^g(H)} = 100 * \frac{\sum_{j,k=1,j\neq k}^{N} \widetilde{\phi}_{jk,t}^g(H)}{N}. \tag{9}$$

First of all, we can measure the directional spillover variable j transmits its shock to all other variables k, called total directional spillover to others and defined as follows:

$$C_{j\longrightarrow k,t}^g(H) = 100 * \frac{\sum_{k=1,j\neq k}^{N} \widetilde{\phi}_{kj,t}^g(H)}{\sum_{k=1}^{N} \widetilde{\phi}_{kj,t}^g(H)}. \tag{10}$$

Then, the total directional spillover from others that variable j receives it from variables k in a similar pattern and defined as follows:

$$C_{j\leftarrow k,t}^g(H) = 100 * \frac{\sum_{k=1,j\neq k}^{N} \widetilde{\phi}_{jk,t}^g(H)}{\sum_{j=1}^{N} \widetilde{\phi}_{jk,t}^g(H)}. \tag{11}$$

At last, we figured the net directional spillover as follows:

$$C_{j,t}^g = C_{j\longrightarrow k,t}^g(H) - C_{j\leftarrow k,t}^g(H), \tag{12}$$

where $C_{j,t}^g$ helps identify the nature of the market that whether it is a net receiver or a net transmitter. Furthermore, the sign of the net total directional spillover illustrates if variable j is driving the network ($C_{j,t}^g > 0$) or driven by the network ($C_{j,t}^g < 0$). In other words, in the case where variable j is driven by the network ($C_{j,t}^g < 0$), variable j is viewed as a safe-haven asset.

3. Data

In this paper, our data set covered three potential safe-haven assets, namely, bitcoin, gold, commodities, and the stock market indices of world, developed, and emerging markets. A newly emerged digital asset, bitcoin's daily closing prices (00:00) are recorded from the website https://coinmarketcap.com/. For gold, the price used in this study is the Gold Fixing Price 10:30 a.m. (London time) in London Bullion Market, based in US dollars. The commodity index is the S&P Goldman Sachs Commodity Index (S&P GSCI) that represents the most widely used benchmark index in the commodity market [4]. For stock markets, the three stock indices (world, developed, and emerging markets) are represented by the corresponding Morgan Sanely Capital International (MSCI) indices.

We define the daily returns as the difference between the natural logarithm of two consecutive closing prices as $r_t = \ln(P_t) - \ln(P_{t-1})$. $\mathrm{Ln}(P_t)$ denotes the natural logarithm of the closing price at the time. The data spans from July 20, 2010, to June 23, 2021. The starting point of the sample phase is constrained by the availability of bitcoin data. Data with inconsistent trading time due to holidays or other factors are excluded; thereby, we have a total of 2,704 observations for each variable. The sample is further divided into three phases: phase I is ended by the phase of the European debt crisis, from July 20, 2010, to March 13, 2012, when the Greek parliament has approved the country's agreement with the European Union and the International Monetary Fund on a second bailout package, and the award of 130 billion euros ($171 billion) in new loans to Greece has gone a long way toward curbing the European debt crisis; phase II is economic calm, from March 14, 2012, to January 13, 2020, when the first cross-border COVID-19 transmission of infection was reported by the World Health Organization; and phase III is started with the outbreak of the COVID-19 pandemic, from January 14, 2020, to June 23, 2021.

In Table 1, we present the descriptive statistics of all the return series in our sample. It can be seen that all the means of returns are positive, except for commodities that have a negative mean return. Bitcoin is the riskiest asset, whereas emerging markets are the least risky asset. Broadly, the fat-tail is more prominent during phases I and III than phase II. Moreover, all of them are leptokurtic and not normally distributed according to the excess kurtosis and Jarque–Bera tests. Then they also have a high-degree of autocorrelations by the Ljung–Box Q statistics. The autocorrelation for gold is significantly stronger during phase III. Lastly, we check for the presence of unit roots using the augmented Dickey–Fuller (ADF) and Phillips–Perron (P-P) test statistics. The ADF and P-P show that all the return series are stable and satisfy the modeling conditions.

4. Empirical Results

4.1. Total Spillover. As can be seen from Figure 1, the dynamic total spillover effects (defined in equation (12)) among bitcoin, gold, commodities, and world, developed, emerging stock markets show significant time-varying characteristics, ranging from almost 20% to 80%, demonstrating a relatively strong information spillover among the bitcoin, gold, commodities, and three stock markets. That is, the risk of one market is likely to be transferred to other markets. More specifically, the dynamic total average spillover is 39.37% during phase II, while it is 50.84% and 48.43% in phases I and III, respectively. The six markets' dynamic total average spillover effects increase significantly during the European debt crisis and COVID-19 health pandemic, which means the contagion of risk is more sensitive when crisis events occur. Moreover, the most significant spike within the sample interval is observed at the beginning of 2020 when the COVID-19 occurred. With the government's control of the epidemic situation and people's understanding and attention to the epidemic situation, a downward trend is witnessed in spillover, decreasing from 75% to 40%.

To sum up, the above findings imply that the major "swan" events may enhance the information or risk spillover transmissions among bitcoin, gold, commodities, and three stock markets and increase the total spillover effects among the probable safe-haven assets and stock markets. With further deepening economic globalization, the rapid improvement of medical standards, and the continuous

Table 1: Descriptive statistics of all return series.

	Bitcoin	Gold	Commodities	World	Developed	Emerging
Entire phase: July 20, 2010–June 23, 2021						
Observations	2704	2704	2704	2704	2704	2704
Mean	0.0047	0.0001	−0.0002	0.0003	0.0003	0.0001
Standard deviation	0.0768	0.0099	0.0132	0.0093	0.0096	0.0102
Minimum	−0.8488	−0.0970	−0.1252	−0.0999	−0.1044	−0.0694
Maximum	1.4742	0.0493	0.0762	0.0806	0.0841	0.0557
Skewness	2.5801	−0.6755	−0.7964	−1.1364	−1.1081	−0.5618
Excess kurtosis	68.7331	9.6108	11.9570	19.2062	19.9086	7.6602
Jarque–Bera	489814.3***	5129.4***	9324.8***	30172.9***	32764.9***	2589.1***
$Q(5)$	22.3930***	3.5407	16.5090***	40.1330***	37.9040***	84.6500***
$Q(10)$	29.5070***	5.0090	20.7670**	115.4700***	128.2500***	93.6780***
$Q(20)$	45.7420***	14.8400	33.8330**	152.7000***	172.0400***	101.1500***
ADF	−53.6237***	−51.2561***	−53.9454***	−16.9024***	−16.9627***	−31.9317***
P-P	−53.6050***	−51.2512***	−54.0550***	−50.0002***	−51.9171***	−44.5217***
Phase I: July 20, 2010–March 13, 2012						
Observations	379	379	379	379	379	379
Mean (%)	0.0105	0.000916	0.0007	0.000462	0.000484	0.000319
Standard deviation	0.1263	0.012317	0.013933	0.011993	0.01234	0.012693
Minimum	−0.4308	−0.05679	−0.06744	−0.05223	−0.05256	−0.06524
Maximum	0.6931	0.037309	0.04661	0.043098	0.046433	0.048096
Skewness	0.7216	−0.56693	−0.61513	−0.42635	−0.37636	−0.56629
Excess kurtosis	8.9885	5.463482	5.345873	5.57309	5.553498	5.407944
Jarque–Bera	599.2170***	116.1377***	110.8051***	116.0354***	111.9143***	111.8194***
$Q(5)$	3.2423	0.6646	3.0721	13.8870**	13.115**	25.029***
$Q(10)$	8.3026	10.1590	13.5230	18.3170**	17.448***	27.240***
$Q(20)$	21.4110	17.8200	19.1980	27.9300	27.198**	33.366**
ADF	−20.5191***	−18.9369***	−19.9059***	−17.7343***	−18.5337***	−15.32656***
P-P	−20.4961***	−18.9299***	−19.9029***	−17.6561***	−18.5378***	−15.25227***
Phase II: March 3, 2012–January 13, 2020						
Observations	1961	1961	1961	1961	1961	1961
Mean (%)	0.00374	$-3.87E-05$	−0.00038	0.000276	0.000307	$3.58E-05$
Standard deviation	0.067284	0.009058	0.011619	0.006921	0.007041	0.008796
Minimum	−0.84883	−0.09704	−0.06586	−0.04881	−0.05029	−0.05128
Maximum	1.47418	0.045897	0.076166	0.029864	0.029383	0.043654
Skewness	4.461687	−0.69559	0.006042	−0.61131	−0.62609	−0.21999
Excess kurtosis	137.811	12.17903	6.054368	6.624474	6.669165	5.055585
Jarque–Bera	1491474***	7042.443***	762.2823***	1195.523***	1228.135***	361.0696***
$Q(5)$	50.598***	1.8772	21.817***	79.176***	59.232***	88.070***
$Q(10)$	65.767***	8.9200	26.118***	85.493***	64.145***	99.324***
$Q(20)$	74.945***	22.666	32.858**	101.82***	82.003***	107.44***
ADF	−22.07981***	−44.52072***	−47.35964	−36.87325***	−38.06258***	−36.33821***
P-P	−44.39223***	−44.52230***	−47.25237	−36.39581***	−37.67187***	−36.32432***
Phase III: January 14, 2020–June 23, 2021						
Observations	364	364	364	364	364	364
Mean (%)	0.003911	0.000388	$3.04E-05$	0.000607	0.000625	0.000476
Standard deviation	0.05345	0.011364	0.019062	0.01521	0.015976	0.013915
Minimum	−0.49728	−0.052272	−0.12522	−0.09997	−0.10441	−0.06943
Maximum	0.19182	0.04934	0.071175	0.080587	0.084062	0.055737
Skewness	−2.27269	−0.817685	−1.76762	−1.4396	−1.3505	−0.9786
Excess kurtosis	24.37342	7.096871	14.46007	16.09758	15.77715	8.544046
Jarque–Bera	7241.834***	295.125***	2181.44***	2727.518***	2586.688***	524.2676***
$Q(5)$	14.529**	8.8985	4.8698	42.706***	46.605***	15.504***
$Q(10)$	22.172**	16.578***	9.4020	125.96***	141.43***	25.195***
$Q(20)$	37.942***	28.859***	20.095	157.10***	177.07***	31.924**
ADF	−21.98285***	−17.94773***	−18.70611***	−6.224532***	−5.147020***	−11.19319***
P-P	−21.80890***	−17.97224***	−18.89272***	−21.87236***	−22.44193***	−19.26622***

Note. The Jarque–Bera statistic tests for the null hypothesis of normality in sample return distribution. $Q(n)$ is the Ljung–Box statistics of the return series for up to n-th order serial correlation. ADF and P-P are statistics of augmented Dickey–Fuller and Phillips–Perron unit root tests, respectively, based on the least AIC criterion. The symbols ***, **, and * indicate rejection at the 1%, 5%, and 10% significance level. The whole sample phase spans from July 20, 2010, to June 23, 2021.

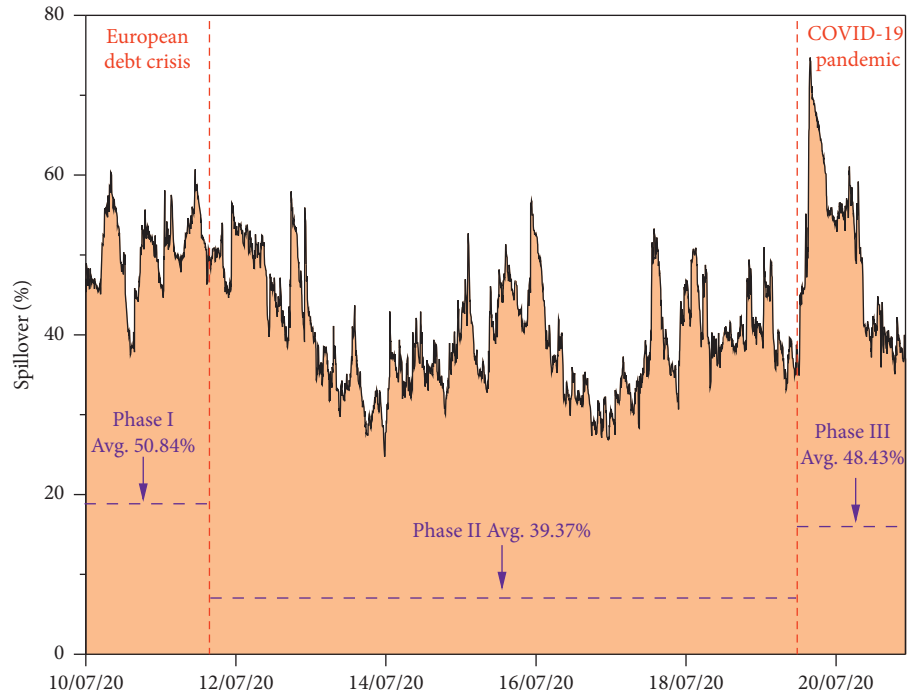


Figure 1: Dynamic return total spillover for TVP-VAR. The orange areas of accumulation define the dynamic return total spillover. The return total spillover plotted here is computed by equation (9). All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging). The sample period for returns is from July 20, 2020, to June 23, 2021.

innovation of information technology, information or risk spillovers between all kinds of markets in the world are getting stronger. Once an extreme risk event occurs in one market, the negative impacts will quickly transmit to other markets through many channels, such as trade, investment, and speculation, thus further enhancing the spillovers among the probable safe-havens assets and stock markets.

4.2. Net Directional Spillover. The time-varying characteristics of net total directional return spillover from each market to all other markets are shown in Figure 2.

We can find that world and developed stock markets act as the net spillover transmitters, while bitcoin, gold, commodities, and emerging stock market are the net recipients. Further analysis shows that the following:

First of all, as for stock markets, we find that the world stock index makes the largest net spillover transmitter with values of 2.26%, 2.56%, and 3.15% (1.70%, 1.72%, and 2.40% for the developed stock market), respectively, during the three periods, which demonstrates that the world stock market still occupies the leading status in the whole stock markets. In addition, the emerging stock market is the small net spillover in the stock markets with values of −0.24%, −1.49%, and −1.31% in phases I, II, and III, respectively. This result is expected considering the slow development and low openness in the stock markets; the emerging stock market has little advantage in the global stock markets; and its fluctuations will have a slender impact on other markets and the global economy.

Interestingly, with regard to the three conceivable safe-haven assets, it is obvious that the net total directional spillover effects of bitcoin, gold, and commodities are almost always negative during all sample phases. Furthermore, the negative net total directional spillover means not only the role of the market as a recipient in information spillover but also the role as a hedge asset hedges risk from the investor's perspective. The results are likely to be related to the transfer of information or risk. Therefore, it is fair to claim that these three assets can be considered safe havens for the stock markets. In addition, in phase II, the total net spillover effects of bitcoin, gold, and commodities are −0.55%, −0.99%, and −1.25%, respectively, and similar results are obtained in phase III (−0.99, −1.55%, and −1.70%, respectively). Bitcoin gets the smallest net spillover with a value of −0.81% in phase I. However, there is a difference in phases II and III: the net spillover of gold is −2.06% much higher than the commodities (−0.84%). It seems possible that the above results are due to the event that Spain's stock market plunged 6% as the debt crisis rattled markets on February 5, 2010, and was warned by Moody's Investors Service on February 26 that if the country does not meet targets in its deficit-reduction plan, Greece's sovereign debt rating would be downgraded within months. It was a time of heightened risk aversion and a sharp sell-off in the euro as investors rushed to get out of a weaker currency. Capital turned to gold as a safe-haven asset, making gold's monetary properties emerge, and prices began to recover.

Last but not least, the commodities index is tightly connected to the stock markets during the negative extreme

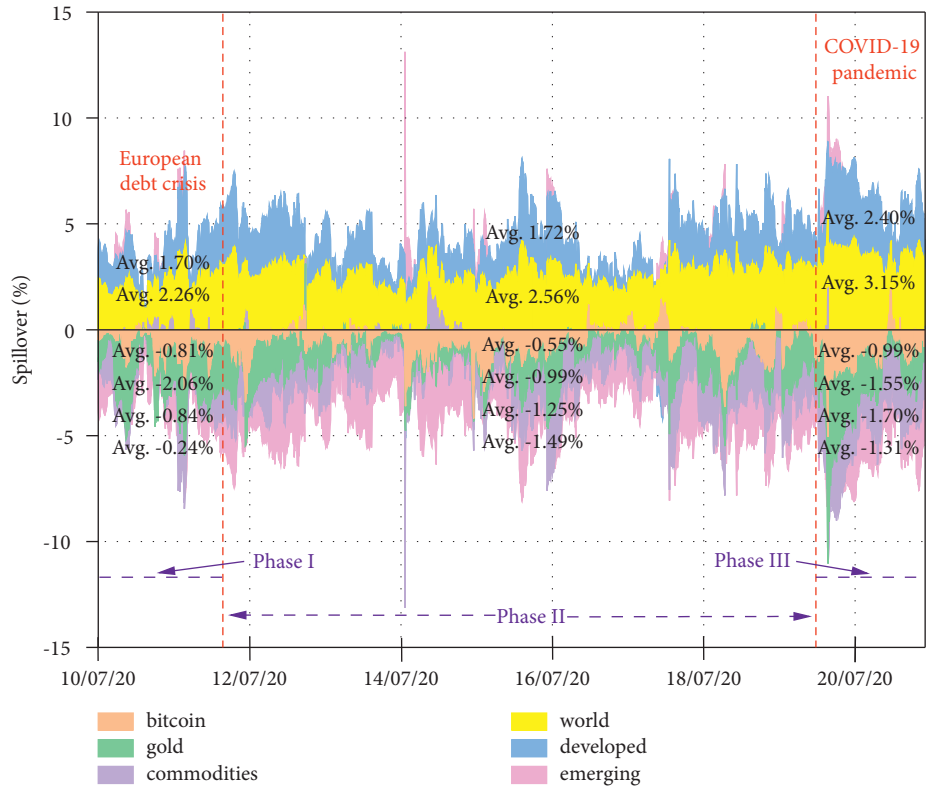


Figure 2: Dynamic return net total directional spillover for TVP-VAR. The different color areas define the accumulation of net total directional spillover. The net spillover of returns plotted here is computed by equation (12), which is defined as the difference in total directional connectedness to others and from others. All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging). The sample period for returns is from July 20, 2020, to June 23, 2021.

events on the sample markets. This result may be explained by that global investors prefer the commodity markets with the financialization of commodity assets for asset allocations.

In summary, the bitcoin, gold, and commodities indices have weak hedging effects on stock markets, and their abilities to act as safe-haven properties could be ranked as: commodities > gold > bitcoin.

4.3. Directional Spillover. Figures 3 and 4 show the return dynamic total directional "to" and "from" spillover, respectively (defined in equations (10) and (11)). The following conclusions are obtained.

Firstly, the spillover effects change over time. To be more specific, the spillover effects of the stock market declined when the European debt crisis ended and is less fluctuated when the economy is stable, but the spillover effects of the system increased rapidly during the outbreak of COVID-19. In other words, the spillover in phase II is smaller than in phases I and III in all markets.

Secondly, "to" and "from" represent the opposite relationship. The directional spillover almost changes after the outbreak of the extreme events. The net directional spillover is figured by the difference between the "to" and "from" (defined in equation (12)). In Figures 3 and 4, there is a clear trend that the spillover of "from" is invariably greater than "to" for bitcoin, gold, and commodities markets in all phases. This result further supports that the idea that bitcoin, gold, and commodities have weak hedging effects on stock markets, and it is in line with the results in the section above.

Thirdly, the row "to" (defined in equation (10)) indicates the total spillover effects transmitted by market j to all other markets. From Figure 3, since that the "to" spillovers of bitcoin, gold, and commodities markets are less than the "from" spillovers, these three markets are defined as the net spillover recipients. Furthermore, a downward trend of the total directional spillover of commodities from other markets is witnessed after the European debt crisis, ranging from 8.54% to 3.91%. However, when the COVID-19 erupted in January 2020, the spillover increased rapidly from 3.91% to 5.12%. The European debt crisis and COVID-19 pandemic eruption significantly increased the total directional spillover from other markets to commodities. Another similar example is gold, after March 2012 and COVID-19 pandemic; the directional spillover from other markets to gold declined rapidly from 4.17% to 2.27%, then to 3.07%. Bitcoin also has a similar tendency; particularly, the total directional spillover from other markets to bitcoin increases at a rapid rate from 0.99% to 4.60% when the COVID-19 pandemic occurs.

Finally, the column "from" (defined in equation (11)) indicates the total spillovers received by market j from all other markets. As is shown in Figure 4, world and developed stock

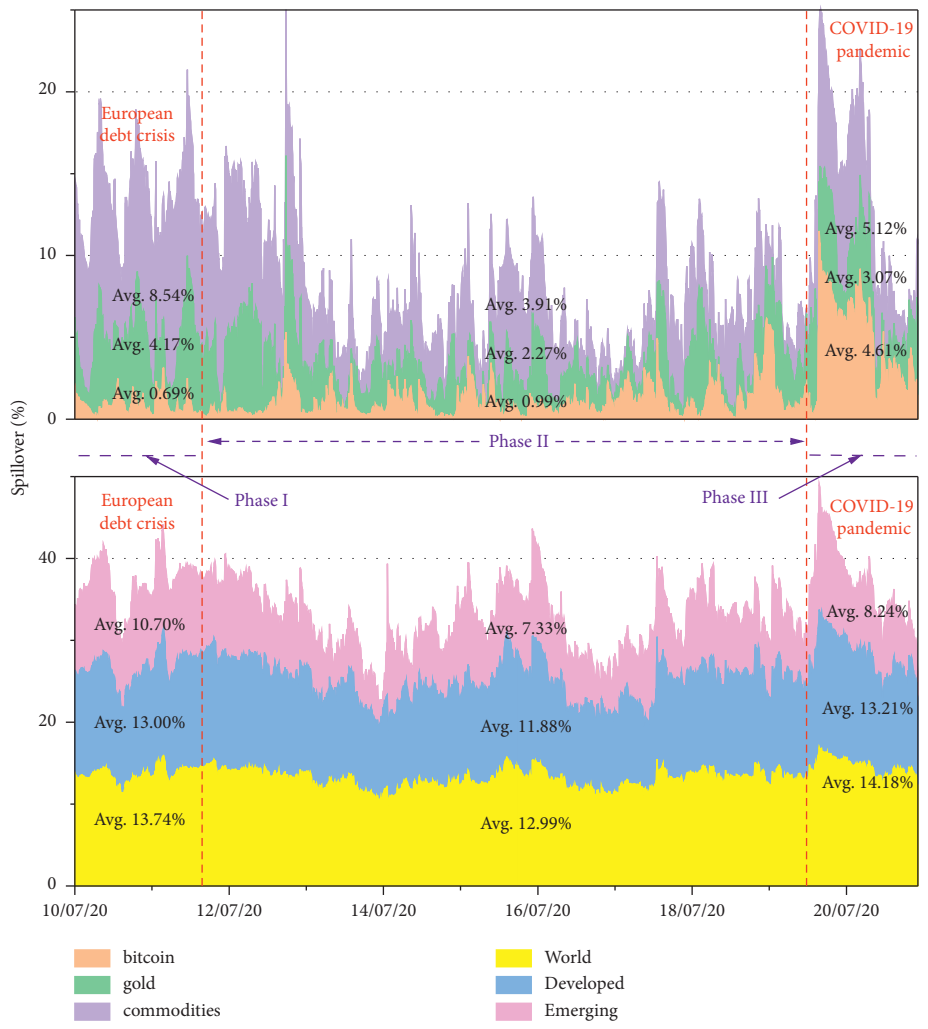


Figure 3: Dynamic total directional spillover to others for TVP-VAR. The different color areas define the accumulation of total directional spillover to other markets. The total directional spillover to others of returns plotted here is computed by equation (10), which is defined as the difference in total directional connectedness to others. All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging). The sample period for returns is from July 20, 2020, to June 23, 2021.

markets act as the net spillover transmitters in the sample for that these markets' "to" spillovers are not lower than "from" spillovers. In addition, the tendency of directional spillover effects of other markets from the world and developed markets are similar. The value of the directional spillover from developed stock markets to other markets declines from 11.30% to 10.16% after the European debt crisis in phase II to 10.81% in January 2020. However, the directional spillover of other markets from the emerging stock market (10.94%, 8.82%, and 9.55%, respectively) are greater than "to" (10.70%, 7.33%, and 8.24% successively) in each stage. Hence, the most striking observation from the data is that the emerging stock market acts as the net spillover recipient. A possible explanation for this may be the absence of adequate information technology and core competitiveness for the emerging stock market to develop.

5. Robustness Test

In this section, we further conduct robustness checks of our major empirical results made in Section 4 by using a mixed methodology model based on Diebold and Yilmaz [28]. What is more, it is made up of a static spillover measurement (on time domain) and a dynamic framework on three rolling windows (i.e., 80, 100, and 120 lengths).

5.1. Static Spillover on the Time Domain. From the perspective of the static spillover measurement [28], we mainly study the correlation among bitcoin, gold, commodities, and three stock markets. Table 2 displays the static spillover measures between each market on the time domain during the entire phase and three sample phases.

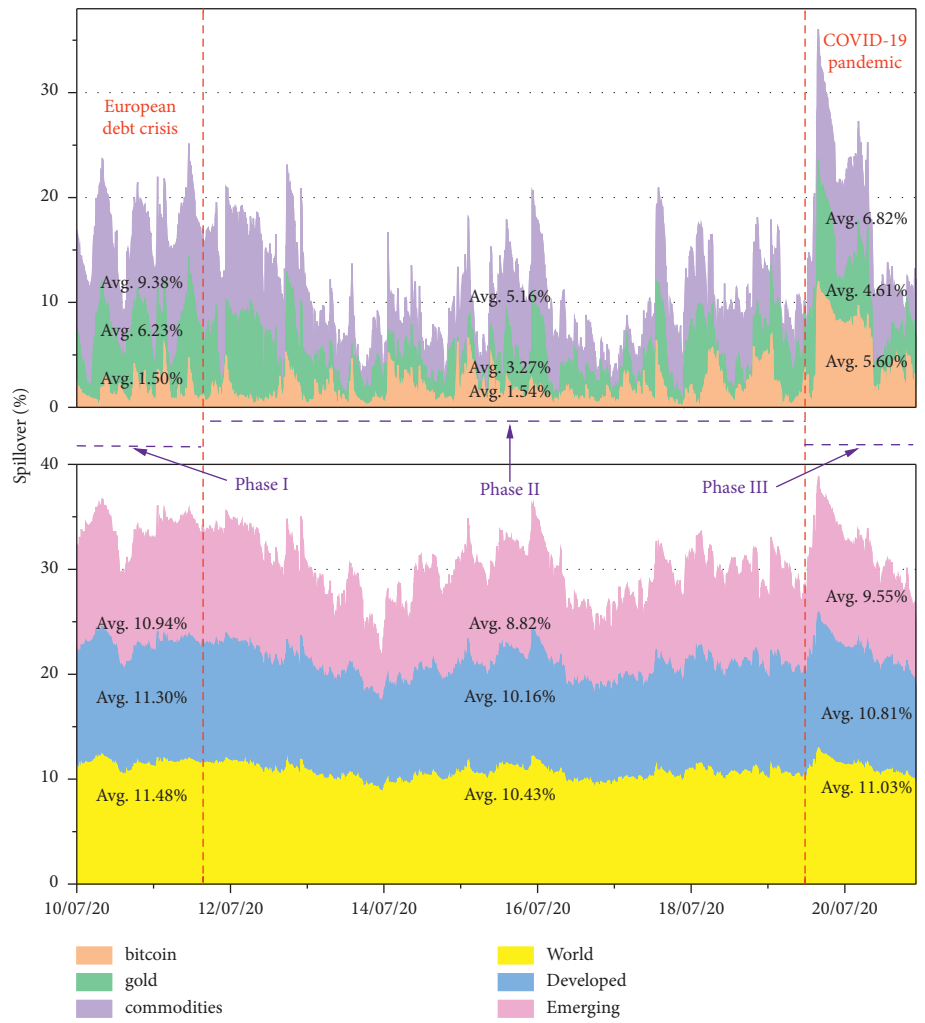


Figure 4: Total directional spillover from others for TVP-VAR. The different color areas define the accumulation of total directional spillover from other markets. The total directional spillover from others of returns plotted here is computed by equation (11), which is defined as the difference in total directional connectedness from others. All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging). The sample period for returns is from July 20, 2020, to June 23, 2021.

First of all, it is noticeable that the total connectedness in phase II is 35.11%, while in phase I and phase III it is 47.15% and 49.96%, respectively, which are higher than the total connectedness in phase II, demonstrating a relatively strong information spillover after the European debt crisis and COVID-19 outbreak. This result is in accordance with Section 4.1.

Secondly, from the net values of the entire phase in Table 2, we can see that the spillovers of bitcoin, gold, commodities, and emerging stock markets are negative, illustrating these markets play the role of net spillover recipients. Nevertheless, the world and developed stock markets are net spillover transmitters since the spillover effects are 2.32% and 1.66%, respectively. Furthermore, the results of the net spillover recipient prove that bitcoin, gold, and commodities can be seen as safe-haven assets, and the ranking of their abilities is: commodities (−1.86%) > gold (−0.47%) > bitcoin (−0.16%). These results are in consistent with those obtained by the former results.

What is more, there is a clear trend that the spillover of "from" is invariably greater than "to" for bitcoin, gold, commodities, and emerging markets while the spillovers of world and developed stock markets are opposite in all phases. Contrary to expectations, one unanticipated finding is the difference between the "to" and "from" of gold and commodities at phase I, where the value difference for gold (−1.83%) is to be outbidding about commodities (−0.81%).

5.2. Dynamic Test with Different Rolling Windows. In estimating the robustness of the above empirical results, we set the 3 dynamic rolling window sizes of 80,100, and 120 equivalent to left 2,625, 2,605, and 2,585 out-of-sample observations, respectively. By the way of orientation, we also compute the total, net, "to," and "from" spillover spanning from the 3 rolling windows.

From Figure 5, we can see the six markets total spillover based on three different rolling window sizes (i.e., 80, 100, and

Table 2: Full-sample return spillover measures in the static time domain.

	Bitcoin	Gold	Commodities	World	Developed	Emerging	From
Entire phase							
Bitcoin	98.14	0.30	0.16	0.63	0.71	0.06	0.31
Gold	0.23	92.60	3.73	1.04	0.91	1.49	1.23
Commodities	0.11	2.48	61.60	13.21	12.70	9.90	6.40
World	0.23	0.41	7.78	35.74	35.45	20.39	10.71
Developed	0.27	0.36	7.70	36.74	37.04	17.89	10.49
Emerging	0.06	1.08	7.86	26.55	23.17	41.27	9.79
To	0.15	0.77	4.54	13.03	12.16	8.29	**38.94**
Net	−0.16	−0.47	−1.86	2.32	1.66	−1.50	
Phase I: July 20, 2010–March 13, 2012							
Bitcoin	96.98	0.78	0.48	0.67	0.86	0.22	0.50
Gold	0.27	76.89	12.6	2.92	2.52	4.79	3.85
Commodities	0.07	7.12	43.72	17.34	16.66	15.09	9.38
World	0.18	1.07	12.81	32.16	31.96	21.82	11.31
Developed	0.25	0.94	12.68	32.96	33.19	19.97	11.13
Emerging	0.05	2.26	12.9	26.51	24.16	34.13	10.98
To	0.14	2.03	8.58	13.40	12.69	10.32	**47.15**
Net	−0.37	−1.83	−0.80	2.10	1.56	−0.66	
Phase II: March 14, 2012–January 13, 2020							
Bitcoin	99.62	0.26	0.01	0.03	0.04	0.03	0.06
Gold	0.09	97.04	2.39	0.06	0.07	0.35	0.49
Commodities	0.03	1.73	71.25	10.22	9.81	6.96	4.79
World	0.09	0.03	5.49	37.88	37.55	18.95	10.35
Developed	0.1	0.01	5.45	39.04	39.49	15.9	10.08
Emerging	0.06	0.49	5.39	27.07	22.97	44.03	9.33
To	0.06	0.42	3.12	12.74	11.74	7.03	**35.11**
Net	−0.00	−0.07	−1.67	2.39	1.65	−2.30	
Phase III: January 14, 2020–June 23, 2021							
Bitcoin	64.36	4.92	3.22	10.92	11.04	5.53	5.94
Gold	5.93	77.88	2.09	5.35	5.29	3.46	3.69
Commodities	3.02	1.47	54.66	14.93	14.5	11.42	7.56
World	6.09	2.16	8.44	32.03	31.83	19.46	11.33
Developed	6.51	2.15	8.31	32.69	32.87	17.46	11.19
Emerging	3.45	2.59	9.35	24.49	21.71	38.41	10.27
To	4.16	2.22	5.24	14.73	14.06	9.56	**49.96**
Net	−1.77	−1.47	−2.32	3.40	2.87	−0.71	

Note. (1) The bold value in the bottom right corner is the total connectedness computed in the system. (2) Values in the last row at each phase denote the net spillover of six markets that are mainly compared with Section 4.1. (3) The values of "to" or "from" represent the spillover that an individual market contributes or receives totally in the whole sample system. (4) All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging). (5) The entire phase for returns is from July 20, 2020, to June 23, 2021.

120 lengths) and the dynamic total spillover based on TVP-VAR, and the estimated total spillovers remain the similar tendency. Total spillovers fell after the European debt crisis and then rose sharply when the outbreak of the COVID-19 pandemic by the selection of rolling window sizes, implying that the dynamic total spillover among the six markets has significant time-varying characteristics, and the total spillover effects increase significantly after extreme events occur. Consequently, the dynamic total spillover with different rolling windows confirms the robustness of the above empirical analysis results of Section 4.1 and the static test in Section 5.1.

Figure 6 exhibits the net directional spillover measures based on three different rolling window sizes (i.e., 80, 100, and 120 lengths); the results illustrate that exactly similar time-varying characteristics. Table 3 is organized by the average value of net directional spillover on the basis of Figure 6, illustrating that the dynamic net spillovers of bitcoin, gold, commodities, and emerging markets are not positive, while the world and developed stock markets are non-negative with the three different rolling windows in all phases. Tables 4 and 5 provide, respectively, the mean dynamic "from" and "to" spillovers with different rolling windows in line with Figures 7 and 8. These results show that developed and world stock markets act as the net spillover transmitters, while the emerging stock market and three possible safe-haven markets are the net recipients in all phases, similar to the results in Section 4.2. What calls for special attention is that the ranking of performance as a safe-haven property seems to be inconsistent with former results. This difference can be explained by the loss of valid sample dates by rolling window lengths; thus, the net spillover effects of safe-haven assets are seriously affected. In summary, by setting different rolling window sizes, we further confirm the robustness of the conclusions obtained in Section 4.3.

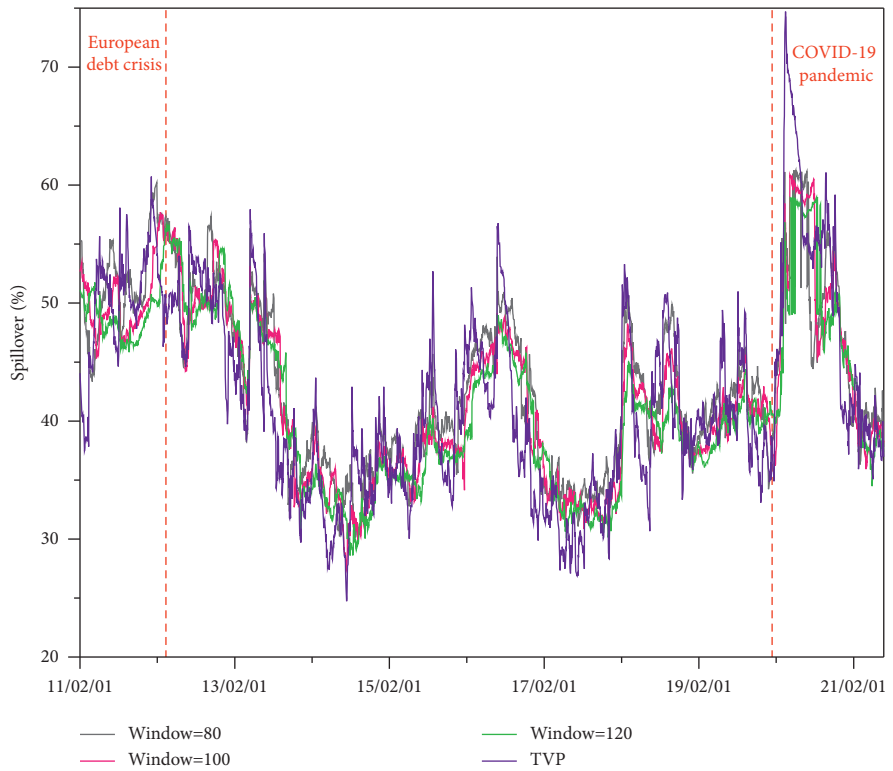


Figure 5: Dynamic return total spillover with different rolling windows and TVP-VAR. The blue line is the dynamic return total spillover for TVP-VAR. The red dotted line defines the time points that distinguish three sample phases. The three window sizes plot the dynamic return total spillover measures on the 80, 100, and 120. All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging). The sample period for returns is from February 2, 2011, to June 23, 2021.

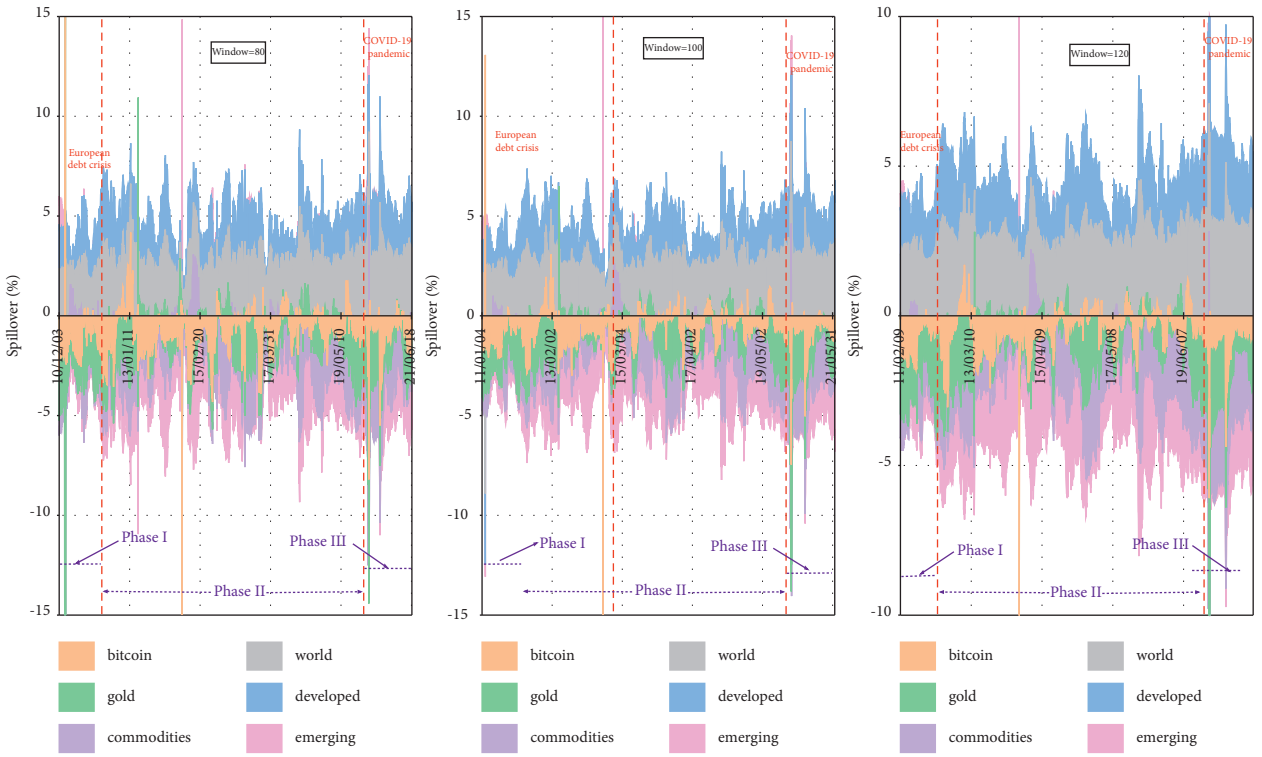


Figure 6: Dynamic net spillover with different rolling windows. The red dotted line defines the time points that distinguish three sample phases. The three window sizes plot the dynamic return total spillover measures on the 80 (the sample period for returns is from December 8, 2010, to June 23, 2021), 100 (the sample period for returns is from January 4, 2011, to June 23, 2021), and 120 (the sample period for returns is from February 9, 2011, to June 23, 2021.). All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging).

Table 3: Dynamic net spillover with different rolling windows.

	Window = 80	Window = 100	Window = 120
Phase I:	December 2, 2010–March 13, 2012	December 31, 2010–March 13, 2012	February 1, 2011–March 13, 2012
Bitcoin	−1.100	−0.926	−0.832
Gold	−1.829	−1.759	−1.700
Commodities	−0.549	−0.596	−0.620
World	2.134	2.122	2.105
Developed	1.629	1.582	1.550
Emerging	−0.283	−0.421	−0.501
Phase II: March 14, 2012–January 13, 2020			
Bitcoin	−0.734	−0.547	−0.444
Gold	−0.841	−0.799	−0.773
Commodities	−1.301	−1.289	−1.274
World	2.646	2.599	2.573
Developed	1.919	1.865	1.838
Emerging	−1.689	−1.828	−1.919
Phase III: January 14, 2020–June 23, 2021			
Bitcoin	−0.937	−0.818	−0.829
Gold	−1.755	−1.741	−1.665
Commodities	−1.848	−1.989	−2.078
World	3.283	3.264	3.252
Developed	2.485	2.503	2.529
Emerging	−1.229	−1.219	−1.209

Note. The values are produced by the average value of net directional spillover on the basis of Figure 6. All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging). The sample period for returns is from July 20, 2020, to June 23, 2021.

Table 4: Dynamic from spillover with different rolling windows.

	Window = 80	Window = 100	Window = 120
Phase I:	December 2, 2010–March 13, 2012	December 31, 2010–March 13, 2012	February 1, 2011–March 13, 2012
Bitcoin	2.028	1.747	1.451
Gold	6.330	5.519	4.741
Commodities	9.444	9.522	9.349
World	11.485	11.423	11.331
Developed	11.318	11.247	11.149
Emerging	10.998	11.015	10.917
Phase II: March 14, 2012–January 13, 2020			
Bitcoin	2.148	1.683	1.423
Gold	3.045	3.196	3.005
Commodities	4.929	5.283	5.199
World	10.470	10.546	10.520
Developed	10.209	10.293	10.264
Emerging	9.265	9.421	9.404
Phase III: January 14, 2020–June 23, 2021			
Bitcoin	5.618	5.497	5.469
Gold	5.141	4.763	4.518
Commodities	6.767	6.672	6.652
World	11.063	11.048	11.063
Developed	10.833	10.828	10.853
Emerging	9.598	9.629	9.690

Note. The values are produced by the average value of total directional spillover from others on the basis of Figure 7. All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging).

Table 5: Dynamic to spillover with different rolling windows.

	Window = 80	Window = 100	Window = 120
Phase I:	December 2, 2010–March 13, 2012	December 31, 2010–March 13, 2012	February 1, 2011–March 13, 2012
Bitcoin	1.131	0.821	0.619
Gold	4.496	3.759	3.041

Table 5: Continued.

	Window = 80	Window = 100	Window = 120
Commodities	9.126	8.925	8.729
World	13.665	13.546	13.436
Developed	12.993	12.829	12.699
Emerging	10.856	10.594	10.414
Phase II: March 14, 2012–January 13, 2020			
Bitcoin	1.368	1.135	0.979
Gold	2.617	2.397	2.232
Commodities	4.159	3.993	3.925
World	13.236	13.146	13.093
Developed	12.259	12.158	12.102
Emerging	7.772	7.593	7.485
Phase III: January 14, 2020–June 23, 2021			
Bitcoin	4.681	4.678	4.639
Gold	3.386	3.022	2.853
Commodities	4.919	4.683	4.574
World	14.347	14.312	14.315
Developed	13.318	13.330	13.382
Emerging	8.369	8.409	8.481

Note. The values are produced by the average value of total directional spillover to others on the basis of Figure 8. All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging).

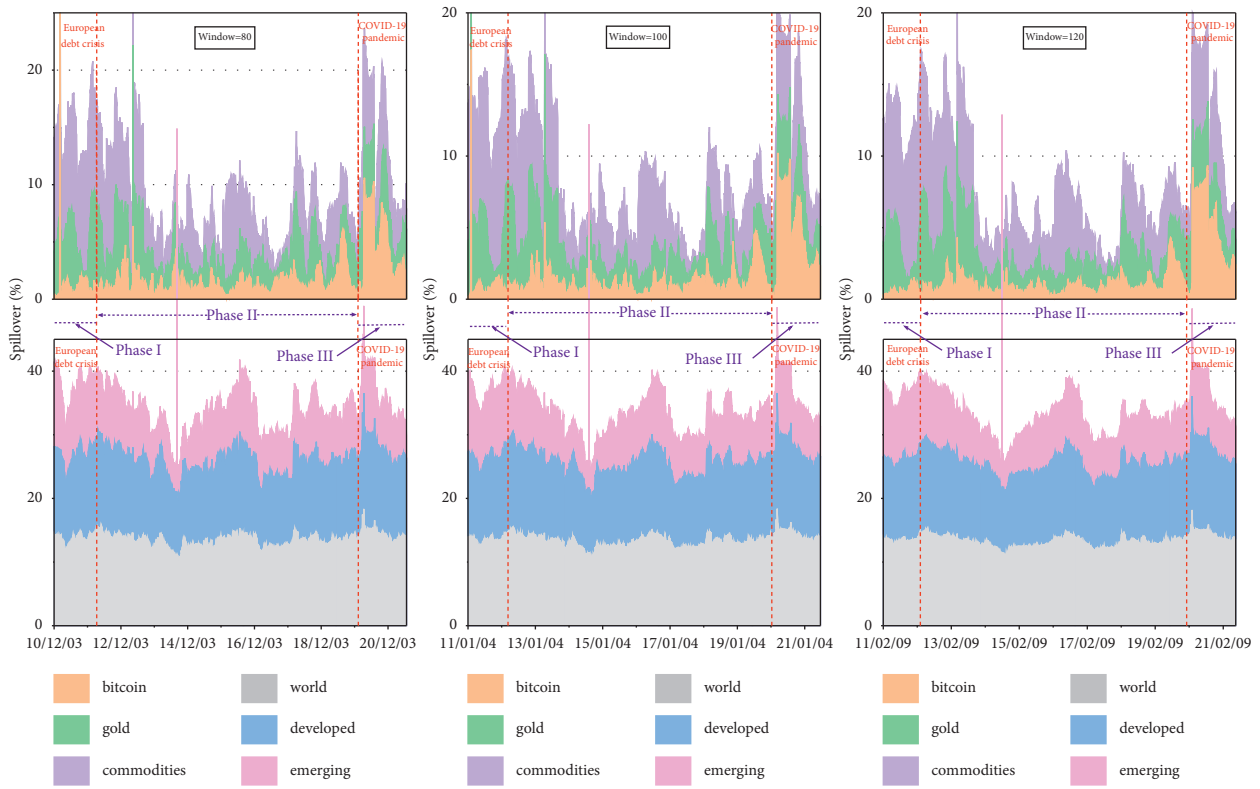


Figure 7: Dynamic net spillover to others with different rolling windows. The red dotted line defines the time points that distinguish three sample phases. The three window sizes plot the dynamic return total spillover measures on the 80 (the sample period for returns is from December 8, 2010, to June 23, 2021), 100 (the sample period for returns is from January 4, 2011, to June 23, 2021), and 120 (the sample period for returns is from February 9, 2011, to June 23, 2021). All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging).

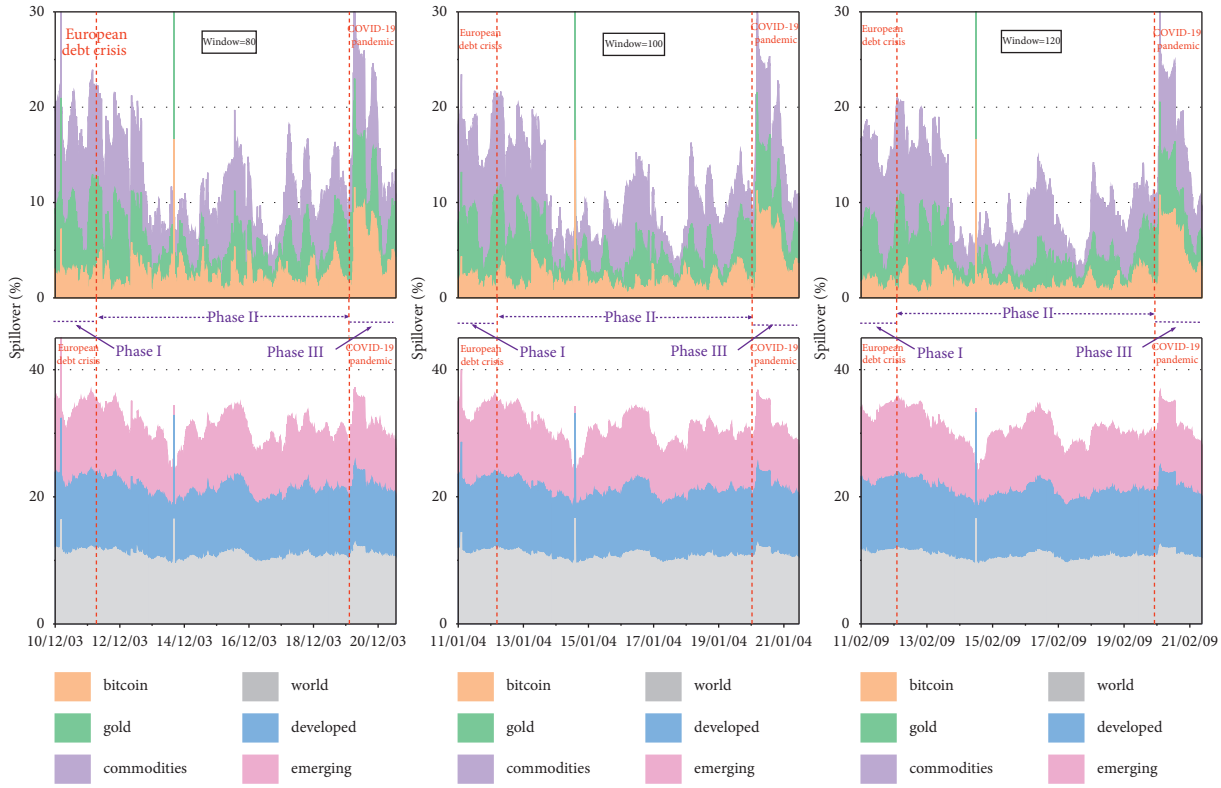


Figure 8: Dynamic net spillover from others with different rolling windows. The red dotted line defines the time points that distinguish three sample phases. The three window sizes plot the dynamic return total spillover measures on the 80 (the sample period for returns is from December 8, 2010, to June 23, 2021), 100 (the sample period for returns is from January 4, 2011, to June 23, 2021), and 120 (the sample period for returns is from February 9, 2011, to June 23, 2021). All abbreviations are as follows: bitcoin, gold, and commodities indices (commodities); the world stock index (world); the developed stock index (developed); and the emerging stock index (emerging).

6. Conclusion

The outbreak of extreme events has brought strong shocks to a range of stock markets. The purpose of the research is to examine whether bitcoin, gold, or commodities can provide investors with effective hedging instruments for international stock markets during the European debt crisis and COVID-19 pandemic.

Firstly, both the dynamic net directional spillover analysis and robustness test demonstrate that bitcoin, gold, and commodities all have weak hedging effects on stock markets, and their abilities to act as a safe-haven property are ranked as: commodities > gold > bitcoin. Specifically, in the dynamic net directional spillover analysis, the commodities market is identified to have the largest safe-haven property with the values of −0.84%, −1.25%, and −1.70% in turn during the three periods, respectively. In addition, the bitcoin market is the smallest net spillover receiver with values of −0.81%, −0.55%, and −0.99% in phases I, II, and III order, respectively. In the static time domain of the robustness test, the spillover effects of bitcoin, gold, and commodities are −0.16%, −0.47%, and −1.86%, respectively. What is more, the robustness check results further validate the empirical findings. Secondly, the dynamic return net directional spillover analysis directly shows that world (2.65%) and developed (1.94%) stock markets act as the net spillover transmitters, while bitcoin, gold, and commodities are the net recipients with a smaller "to" than "from" spillover. Thirdly, the dynamic total average spillovers increase significantly after extreme risk event occurs; to be more specific, the spillover effects of the market declined from 50.84% to 39.37% when the European debt crisis ended, but the spillover effects of the system increased rapidly to 48.43% when the outbreak of COVID-19, which means the contagion of risk is more sensitive when crises occur. Finally, the dynamic total spillover measurements reveal that the six markets' total spillover shows significant time-varying characteristics, ranging from almost 20% to 80%, demonstrating the relatively strong information spillover among the bitcoin, gold, commodities, and three stock markets.

Given the fact of strong time-varying connectedness among bitcoin, gold, commodities, and stock markets, policymakers and international investors need to pay close attention to the spillover effects among them from time to time when making regulatory and investment decisions. In addition, the significant effects of the European debt crisis and COVID-19 pandemic on the dynamic spillover among the safe-haven assets and three major stock markets imply that policymakers and global investors should keep an eye on the factor of "black swan," which can alter assets' spillover relationship. Moreover, since the dynamic return net directional spillover analysis is quite significant, various

economic supervision agents and international traders have to adjust their expectations and spillovers transmitted from other markets when making important decisions. They have to alter their expectations in time to avoid the negative impacts of shocks when the market policy changes. Last but not least, investors cannot ignore the importance of bitcoin, gold, and commodities in selecting preferable safe-haven assets when searching for more profitable portfolio policies and management strategies.

Data Availability

The bitcoin price was collected from https://coinmarketcap.com/. The gold price, commodity index, and three stock indices were subscribed from the Wind database.

Conflicts of Interest

The authors declare that there are no conflicts of interest in this paper.

Authors' Contributions

Qian Wang provided the original idea and wrote the draft of this paper. Yu Wei supervised the writing process of this paper. Yao Wang made the data collection. Yuntong Liu made the estimations of empirical models. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (71671145 and 71971191), Science and Technology Innovation Team of Yunnan Provincial Universities (2019014), and Yunnan Fundamental Research Projects (202001AS070018).

References

[1] B. Keddad and C. Schalck, "Evaluating sovereign risk spillovers on domestic banks during the European debt crisis," *Economic Modelling*, vol. 88, pp. 356–375, 2020.

[2] V. G. Papavassiliou and H. Kinateder, "Information shares and market quality before and during the European sovereign debt crisis," *Journal of International Financial Markets, Institutions and Money*, vol. 72, 2021.

[3] L. Bai, Y. Wei, G. Wei, X. Li, and S. Zhang, "Infectious disease pandemic and permanent volatility of international stock markets: a long-term perspective," *Finance Research Letters*, vol. 40, Article ID 101709, 2021.

[4] E. Bouri, S. J. H. Shahzad, D. Roubaud, L. Kristoufek, and B. Lucey, "Bitcoin, gold, and commodities as safe havens for stocks: new insight through wavelet analysis," *The Quarterly Review of Economics and Finance*, vol. 77, pp. 156–164, 2020.

[5] X. Li and Y. Wei, "The dependence and risk spillover between crude oil market and China stock market: new evidence from a variational mode decomposition-based copula method," *Energy Economics*, vol. 74, pp. 565–581, 2018.

[6] S. J. H. Shahzad, E. Bouri, D. Roubaud, L. Kristoufek, and B. Lucey, "Is Bitcoin a better safe-haven investment than gold and commodities?" *International Review of Financial Analysis*, vol. 63, pp. 322–330, 2019.

[7] Y. Wei, S. Qin, X. Li, S. Zhu, and G. Wei, "Oil price fluctuation, stock market and macroeconomic fundamentals: evidence from China before and after the financial crisis," *Finance Research Letters*, vol. 30, pp. 23–29, 2019.

[8] M. Akhtaruzzaman, S. Boubaker, B. M. Lucey, and A. Sensoy, "Is gold a hedge or a safe-haven asset in the COVID–19 crisis?" *Economic Modelling*, vol. 102, 2021.

[9] H. Boubaker, J. Cunado, L. A. Gil-Alana, and R. Gupta, "Global crises and gold as a safe haven: evidence from over seven and a half centuries of data," *Physica A: Statistical Mechanics and Its Applications*, vol. 540, 2020.

[10] J. Jin, J. Yu, Y. Hu, and Y. Shang, "Which one is more informative in determining price movements of hedging assets? Evidence from Bitcoin, gold and crude oil markets," *Physica A: Statistical Mechanics and Its Applications*, vol. 527, 2019.

[11] Y. Wei, C. Liang, Y. Li, X. Zhang, and G. Wei, "Can CBOE gold and silver implied volatility help to forecast gold futures volatility in China? Evidence based on HAR and Ridge regression models," *Finance Research Letters*, vol. 35, 2020.

[12] M. B. Triki and A. Ben Maatoug, "The GOLD market as a safe haven against the stock market uncertainty: evidence from geopolitical risk," *Resources Policy*, vol. 70, 2021.

[13] Z. He, F. O'Connor, and J. Thijssen, "Is gold a sometime safe haven or an always hedge for equity investors? A markov-switching CAPM approach for US and UK stock indices," *International Review of Financial Analysis*, vol. 60, pp. 30–37, 2018.

[14] Q. Ji, D. Zhang, and Y. Zhao, "Searching for safe-haven assets during the COVID-19 pandemic," *International Review of Financial Analysis*, vol. 71, 2020.

[15] K.-M. Wang and Y.-M. Lee, "Is gold a safe haven for exchange rate risks? An empirical study of major currency countries," *Journal of Multinational Financial Management*, Article ID 100705, 2021.

[16] M. Disli, R. Nagayev, K. Salim, S. K. Rizkiah, and A. F. Aysan, "In search of safe haven assets during COVID-19 pandemic: an empirical analysis of different investor types," *Research in International Business and Finance*, vol. 58, 2021.

[17] C. Dwita Mariana, I. A. Ekaputra, and Z. A. Husodo, "Are Bitcoin and Ethereum safe-havens for stocks during the COVID-19 pandemic?" *Finance Research Letters*, vol. 38, Article ID 101798, 2021.

[18] A. Stensås, M. F. Nygaard, K. Kyaw, S. Treepongkaruna, and Z. Yang, "Can Bitcoin be a diversifier, hedge or safe haven tool?" *Cogent Economics & Finance*, vol. 7, 2019.

[19] A. Kliber, P. Marszałek, I. Musiałkowska, and K. Świerczyńska, "Bitcoin: safe haven, hedge or diversifier? Perception of bitcoin in the context of a country's economic situation-a stochastic volatility approach," *Physica A: Statistical Mechanics and Its Applications*, vol. 524, pp. 246–257, 2019.

[20] E. Bouri, P. Molnár, G. Azzi, D. Roubaud, and L. I. Hagfors, "On the hedge and safe haven properties of Bitcoin: is it really more than a diversifier?" *Finance Research Letters*, vol. 20, pp. 192–198, 2017.

[21] T. Conlon and R. McGee, "Safe haven or risky hazard? Bitcoin during the Covid-19 bear market," *Finance Research Letters*, vol. 35, Article ID 101607, 2020.

[22] L. A. Smales, "Bitcoin as a safe haven: is it even worth considering?" *Finance Research Letters*, vol. 30, pp. 385–393, 2019.

[23] S. Choi and J. Shin, "Bitcoin: an inflation hedge but not a safe haven," *Finance Research Letters*, Article ID 102379, 2021.

[24] T. E. Sønsteng Henriksen, "Properties of long/short commodity indices in stock and bond portfolios," *Journal of Alternative Investments*, vol. 20, pp. 51–68, 2018.
[25] S. Bekiros, S. Boubaker, D. K. Nguyen, and G. S. Uddin, "Black swan events and safe havens: the role of gold in globally integrated emerging markets," *Journal of International Money and Finance*, vol. 73, pp. 317–334, 2017.
[26] S. Ali, E. Bouri, R. L. Czudaj, and S. J. H. Shahzad, "Revisiting the valuable roles of commodities for international stock markets," *Resources Policy*, vol. 66, 2020.
[27] S. J. Hussain Shahzad, E. Bouri, D. Roubaud, and L. Kristoufek, "Safe haven, hedge and diversification for G7 stock markets: gold versus bitcoin," *Economic Modelling*, vol. 87, pp. 212–224, 2020.
[28] F. X. Diebold and K. Yilmaz, "Better to give than to receive: predictive directional measurement of volatility spillovers," *International Journal of Forecasting*, vol. 28, no. 1, pp. 57–66, 2012.
[29] M. Del Negro and G. E. Primiceri, "Time varying structural vector autoregressions and monetary policy: a corrigendum," *The Review of Economic Studies*, vol. 82, no. 4, pp. 1342–1345, 2015.
[30] H. H. Pesaran and Y. Shin, "Generalized impulse response analysis in linear multivariate models," *Economics Letters*, vol. 58, no. 1, pp. 17–29, 1998.
[31] F. X. Diebold and K. Yilmaz, "Measuring financial asset return and volatility spillovers, with application to global equity markets," *The Economic Journal*, vol. 119, no. 534, pp. 158–171, 2009.
[32] F. X. Diebold and K. Yilmaz, "On the network topology of variance decompositions: measuring the connectedness of financial firms," *Journal of Econometrics*, vol. 182, pp. 119–134, 2011.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 2048811, 11 pages
https://doi.org/10.1155/2021/2048811

Research Article

Research on a Dynamic Evaluation of Financial Risks of Intellectual Property Pledge Financing of High-Tech Enterprises

Xiangyu Ge,[1] **Ying Jin**,[1,2] **Qing Li**,[1] **Xiaofang Yao**,[1] **and Shican Liu**[1]

[1]*School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430 073, China*
[2]*School of Statistics and Big Data, Henan University of Economics and Law, Zhengzhou 450 046, China*

Correspondence should be addressed to Shican Liu; canbulajiji@163.com

Received 17 October 2021; Accepted 2 December 2021; Published 26 December 2021

Academic Editor: Dehua Shen

Copyright © 2021 Xiangyu Ge 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.

Intellectual property pledge financing is effective in alleviating the financing problems of scientific and technological enterprises to a certain extent. However, compared with traditional loans, intellectual property pledge financing is at greater risk with an eye to the particularity of intellectual property pledges. Therefore, it is very important to evaluate the risk of intellectual property pledge financing. This article first outlines the information entropy method, functionalizing indicator data and weight data, and applies the combination of the weight function and indicator function in dynamic evaluation function, to get the final evaluation result. Second, taking four high-tech listed companies on the growth enterprises market (GEM) as example, an evaluation indicator system has been constructed with 16 relevant financial indicators from 2015 to 2019. Then, based on the dynamic function method, we can construct the evaluation function to describe the change trend of the financial indicators of enterprises with a comparison of the advantages and disadvantages of the financial indicators of enterprises. Finally, this article has made a comprehensively comparative analysis on the dynamic evaluation, and the financial risk of intellectual property pledge financing of high-tech enterprises has been done with the combination of the dynamic evaluation curve with the optimal time weight determined by the maximum entropy method.

1. Introduction

The key of economic and social development to promoting high-quality development is innovation, particularly scientific and technological innovation. Therefore, it is encouraged to carry out scientific and technological innovation for enterprises to implement the strategy of innovation-driven development, which is the main task of economic and social development in China. Nowadays, high-tech enterprises have become the backbone of economic development and innovative growth and play an important role in gathering scientific and technological resources, creating intellectual property rights, and driving economic transformation. The core competency of high-tech enterprises is to accelerate the transformation rate of scientific and technological research. However, most high-tech enterprises are new on a small scale, which are not available with appropriate fixed assets as collateral and fail to get funds from banks and other financial institutions, facing the predicament of lacking funds. Financing difficulties limit the development of most high-tech enterprises to some extent.

According to the data released by the Chinese State Intellectual Property Office from 2007 to 2020, the number of valid invention patents in China (excluding Hong Kong, Macao, and Taiwan) has increased from 84,000 to 2.213 million, with the invention of patents per 10,000 population reaching 15.8. With the significantly increasing ownership of China's intellectual property, it is evident that China is a real intellectual property power. Now, many high-tech enterprises in China are fitted with advanced patent technologies, which are valuable in the market. At the same time, financing a loan of intellectual property pledge effectively offers a loan to solve the financing difficulties to promote the technical innovation of these small- and medium-sized high-tech enterprises with advanced patents. As for the current state and government departments, some relevant policies

have been issued by governments to promote the development of local intellectual property pledge financing businesses. Unfortunately, in reality, intellectual property pledge financing has a long way to go. The key point is to evaluate and control the financial risk of high-tech financing enterprises. Therefore, the assessment and monitoring of financial risk of the enterprise are obvious to be particularly important, which has always been one of the main reasons affecting the development of intellectual property pledge financing business. It is of great practical significance to discuss the dynamic evaluation of financial risk in intellectual property pledge financing of high-tech enterprises.

In recent years, attention has been paid to the financing of small and medium enterprises (SMEs) in the field of finance. Significant analysis has been presented by Banerjee et al. [1], who believed that small and medium financial institutions have advantages in terms of information and cost to provide services over SMEs. Angelini et al. [2] found that relational lending featured with long term and concentration of trading objects are supposed to improve the availability and terms of loans for SMEs. The influence of government policies and national financial structure on the credit availability of SMEs has been reported by Berger and Udell [3]. Another research on a new financing model for SMEs based on guarantees has been studied by Zhang and Zhang [4]. Most researches have analyzed the relationship between bank credit and SMEs financing (see Lin and Li [5] and Xu and Zou [6]), while the interest rate and guarantee of SMEs loans have been discussed by Xiong et al. [7, 8]. Meanwhile, intellectual property pledge financing has been studied by many scholars in many aspects. The research of Jarboe and Ellis [9] focused on the innovative financing of intangible assets in the United States, including improvement of the value evaluation, trading mechanism of intellectual property, characteristics of intellectual property, and the disclosure of intellectual property information with policy suggestions proposed. Crawford and Strasser [10] proposed a series of measures to punish intellectual property infringement. However, Chessy and Toivanen [11] found that financing risk is not significantly correlated with the value of pledges. The study by Niinimaki [12] also believed that risk types cannot be identified simply by the value of objects. The financing behavior of intellectual property pledges and their legal uncertainty has been reported by Davies [13]. Detailed information of the risks in different stages of intellectual property has been studied from formation to pledge by Pennington and Sanchez [14].

The emphasis of existing literature is mainly placed on the risk sources and formation mechanism of intellectual property pledges. As the subject of the intellectual property pledge loan is still the enterprise itself, the enterprise's operation income is regarded as the first repayment source of the intellectual property pledge loan. When financial problems happen to a pledge financing enterprise happens, there will definitely be a risk of default. If a pledge financing enterprise is allowed to dispose of intellectual property rights, the risk is greater than the mortgage of tangible assets. Therefore, it is necessary to comprehensively evaluate the financial situation of pledge financing enterprises. It is more likely to minimize the cost of default risk when the problem is found early; the sooner, the better. The method commonly used in science and technology evaluation consists of the AHP analysis method by Schenkerman [15], fuzzy comprehensive evaluation method by Chen and Hwang [16], DEA evaluation method by Ding et al. [17], and VIKOR method by Liu et al. [18].

In the era of big data, traditional data analysis technology is faced with many limitations in information data mining. Functional data analysis is a discrete realization regarding observation data as a random process to analyze the changing pattern of data from the perspective of the whole function, which not only relax the structural constraints and distribution assumptions of data collection but also dig down inside the deep dynamic change patterns of data. Overall, it is of great significance to enrich information mining technology in the era of big data. The implementation of the traditional dynamic comprehensive evaluation method is mainly based on the cross-sectional data and time series data, which turns out to be discrete data in the process of evaluation, including original data, weight, and final evaluation. With the fusion of the cross-sectional data and the time series data, the original data have been changed to the functional data, the concept of which has been proposed by Ramsay [19], formed by the function feature of time as the independent variable. Later, the improvement of the analysis framework system has been further studied by several world-renowned statisticians such as Ramsay and Dalzell [20], Silverman [21], Lu [22], Delagle and Hall [23], Petersen and Muller [24], and Kokoszka and Reimherr [25] with the result of relaxing the structural constraints and distribution setting of data collection and deeply digging the potential dynamic information of data from the perspective of the interactive derivative function. Many successful explorations have been made on the expansion of classical statistical methods in functional data, for example, the principal component analysis [21], discriminant analysis [26], linear model [27], principal differential analysis [28], panel data model [29], and cluster analysis [30]. Based on the process of functional data, it is easy to carry out the dynamic comprehensive evaluation function curve for a period of time, which is more useful and effective for evaluation analysis.

In previous literature, the comprehensive evaluation of the risk of intellectual property pledge financing focuses on static analysis [31, 32]. Meanwhile, the comprehensive evaluation index for the small commodity market has been constructed by the functional data analysis, with making the dynamic evaluation of the quality of life [33, 34]. This is the first time to evaluate the financial risk of intellectual property pledge financing enterprises by functional data analysis in the article.

In this article, the function analysis method has been used to evaluate the financial risk of intellectual property pledge financing enterprises. Compared with the risk evaluation with annual data, quarter data with shorter time intervals are better to reflect the function characteristics of data in the case of time as an independent variable, which is more accurate and effective in describing the financial status

of enterprises in time. Whether at a specific moment or the overall time interval, the objectivity and comparability of the evaluation results are obvious.

The rest of the article is organized as follows. In Section 2, we will establish theoretical models and fractionalize the indicator data and the weight data to obtain the dynamic evaluation function with a combination of the weight function and the indicator function. In Section 3, the system of intellectual property pledge financing risk evaluation indicator will be constructed. In Section 4, we will make a dynamic evaluation with the functional curve for four high-tech listed enterprises, with a comprehensive comparison of the intellectual property pledge financing risk. A conclusion of this article is given in Section 5.

2. The Risks Evaluation Method of Intellectual Property Pledge Financing

Intellectual property pledge financing is indeed risky with great uncertainty. If the default happens to enterprises, debt assets are difficult to realize with creating a chain reaction, which requires financial institutions to evaluate the financial status of enterprises' pledge financing. Based on the quarterly financial data to be evaluated, the process of the generation of sample data is treated as a functional process. Regarding the observed data as a whole, the functional data analysis method is used to fit the overall function process of the observed discrete data to get the overall curve of financial data and reflect the dynamic changes of financial data. According to the weight function, the comprehensive evaluation value of finance can be obtained to judge the financial risk of enterprises. The specific theoretical model is as follows.

2.1. Functional Data Analysis Methods. The basic idea of functional data analysis (FDA) is to treat each sample observation as a whole rather than a sequence of individual observations. Due to errors or noise in the original data, the relationship between the actual observed data and the underlying function is as follows:

$$y_h = x(t_h) + \varepsilon_h, \quad h = 1, 2, \ldots, T, \tag{1}$$

where ε_h represents the "roughness" of the original data due to disturbances, errors, deviations, and other exogenous factors. The essence of transforming observed discrete data into functional data is to recover the underlying data from the original data. It is assumed that generating data are a smooth process, and the process of transforming the original discrete data into smooth function data has been realized based on the linear combination of basis functions, including Fourier basis functions, B-spine basis functions, polynomial basis functions, and wavelet basis functions, in which the Fourier base function and B-spine base function are more widespread. For periodic data, the Fourier base function is suitable. However, for aperiodic data, the B-spine base function is suitable. For a set of basis functions ϕ_k, $k = 1, 2, \ldots, K$, estimation of $x(t)$ can be obtained by the linear combination of the basis functions as follows:

$$\widehat{x}(t_h) = \sum_{k=1}^{K} c_k \phi_k(t_h). \tag{2}$$

It is known that the choice of basis function is also important to estimate the following derivative:

$$D^m \widehat{x}(t_h) = \sum_{k=1}^{K} c_k D^m \phi_k(t_h), \tag{3}$$

where $D^m \widehat{x}(t_h)$ means the m-order derivative of the function, usually $m = 2$. Related to the choice of a basis function, the estimation of the function is required to be considered, along with the estimation of the first derivative and the second derivative.

When a linear combination of basis functions is used to smooth the data, the coefficients c_k, $k = 1, 2, \ldots, K$, are needed to determine, by the least-squares criterion, i.e.,

$$\min \sum_{h=1}^{T} \left[x(t_h) - \widehat{x}(t_h)\right]^2, \tag{4}$$

or

$$\min \sum_{h=1}^{T} \left[x(t_h) - \sum_{k=1}^{K} c_k \phi_k(t_h)\right]^2. \tag{5}$$

Assuming that the value of the basis function $\phi_k(t_h)$ $(k = 1, 2, \ldots, K)$ at the observation point t_h $(h = 1, 2, \ldots, n)$ with $n \times K$-order matrix as $\Phi = \phi_k(t_h)$, where Φ is full rank, then the minimized solution is as follows:

$$c = \left(\Phi'\Phi\right)^{-1}\Phi' x. \tag{6}$$

However, the control of the expansion is not flexible enough in the degree of uniform smoothness without continuous control ability. Therefore, in the practical analysis, the main concern is the degree of fitting and smoothness of the fitting function. In general, the higher the degree of the fitting is, the greater the volatility of the fitting function is. To ensure the smoothness of the fitting, a penalty term is introduced in the object function:

$$\begin{aligned} \text{PENSSE} &= \min \sum_{h=1}^{T} \left[x(t_h) - \widehat{x}(t_h)\right]^2 + \lambda \int_T \left\{D^2 x(t)\right\} \mathrm{d}t \\ &= \min \sum_{h=1}^{T} \left[x(t_h) - \sum_{k=1}^{K} c_k \phi_k(t_h)\right]^2 + \lambda \int_T \left\{D^2 x(t)\right\} \mathrm{d}t. \end{aligned} \tag{7}$$

The first term is the sum of squares of residuals, which indicates the degree of fitting; and the penalty term is added to indicate the degree of fluctuation. Specifically, λ means a measure between the degree that the function of $x(t)$ fits the data and the volatility of the function itself, which is called a smoothing parameter. When $\lambda \longrightarrow 0$, the smooth fitting curve $x(t_h)$ is the interpolation function of the data, with no arbitrary fluctuation interpolation curves, which are the smoothest second-order differentiable function of the data curve accurately fitted. When $\lambda \longrightarrow \infty$, $\int_T \{D^2 x(t)\} \mathrm{d}t$, the

smooth fitting curve of $x(t_h)$ is the standard linear regression of the data. More attention has been paid to the smoothness of the function as λ gets bigger, with an estimated value as follows:

$$c = \left(\Phi'\Phi + \lambda \int_T \left\{D^2 x(s)\right\}\left\{D^2 x'(s)\right\} \mathrm{d}s\right)^{-1} \Phi' x. \tag{8}$$

By comparing equations (6) and (8), it can be found that they are equal when $\lambda = 0$. The data smoothing degree can be varied with the change of λ, a vector including the elements of node number of B-spline base. It is easy to estimate λ using the generalized cross-validation (GCV) method.

2.2. Comprehensive Evaluation Method of Pledge Financing Risks. In this subsection, we will first introduce the dimensionless processing method of the data, followed by using the information entropy method to obtain the objective weight of the indicators and functionalizing the indicator data and the weight data to obtain a function curve of the comprehensive evaluation, resulting in determining the weight of the time measurement by the maximum entropy method. Combining the evaluation function to obtain the final evaluation result is expected. Before we go into the details, we have to introduce some notations. $S = \{s_1, s_2, \ldots, s_n\}$ means the number of evaluated objects to n, $X = \{x_1, x_2, \ldots, x_m\}$ shows the number of evaluation indicators to m, and $H = \{t_1, t_2, \ldots, t_T\}$ are the time of T. $x_{ij}(t_h)$ represents the observed value of indicator x_j $(j = 1, 2, \ldots m)$ of the evaluated object s_i $(i = 1, 2, \ldots n)$ at time t_h $(h = 1, 2, \ldots T)$.

2.2.1. Dimensionless Processing of Data. For an interval-type indicator, the optimal value is in a certain interval, which is transformed into a positive indicator:

$$x_{ij}^*(t_h) = \begin{cases} 1 - \dfrac{a - x_{ij}(t_h)}{M}, & x_{ij}(t_h) < a, \\ 1, & a \le x_{ij}(t_h) \le b, \\ 1 - \dfrac{x_{ij}(t_h) - b}{M}, & x_{ij}(t_h) > b, \end{cases} \tag{9}$$

where $M = \max\left\{a - \min\left\{x_{ij}(t_h)\right\}, \max\left\{x_{ij}(t_h)\right\} - b\right\}$.

For dimensionless data, the complete process of sequence extremum processing is conducted using the following formulas reducing the dimension, which formula the forward type indicator/backward type indicator. Forward type indicator is shown as follows:

$$x_{ij}^*(t_k) = \frac{x_{ij}(t_k) - \min_{i,k}\left\{x_{ij}(t_k)\right\}}{\max_{i,k}\left\{x_{ij}(t_k)\right\} - \min_{i,k}\left\{x_{ij}(t_k)\right\}} + 1. \tag{10}$$

Backward type indicator is shown as follows:

$$x_{ij}^*(t_h) = \frac{\min_{i,h}\left\{x_{ij}(t_h)\right\} - x_{ij}(t_h)}{\max_{i,h}\left\{x_{ij}(t_h)\right\} - \min_{i,h}\left\{x_{ij}(t_h)\right\}} + 1. \tag{11}$$

2.2.2. Information Entropy Method for Calculating Objective Weight of Indicators. The entropy method is to determine the weight of an indicator by comparing the amount of information of the indicator. The concept of entropy is introduced from thermodynamics. In information theory, entropy is used to measure the degree of uncertainty, also known as the average amount of information. If the amount of information is more, the understanding of the objective function is accordingly more comprehensive. However, if the uncertainty is smaller, the impact on the system is smaller, and the corresponding target entropy is smaller. The reverse is also true.

The information entropy method of calculating the objective weight of indicators contains the following four steps:

Firstly, the characteristic proportion of each indicator is calculated. The characteristic proportion of the evaluated object of i under indicator of j in year of t_k is as follows:

$$p_{ij}(t_h) = \frac{x_{ij}^*(t_h)}{\sum_{i=1}^n x_{ij}^*(t_h)}, \tag{12}$$

where $x_{ij}^*(t_h) > 0$ and $\sum_{i=1}^n x_{ij}^*(t_h) > 0$.

Secondly, the entropy of each indicator is calculated. The entropy value of indicator j in year of t_h is as follows:

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^n p_{ij}(t_h) \ln\, p_{ij}(t_h). \tag{13}$$

Next, we calculate the difference coefficient ratio of indicators:

$$g_j = 1 - e_j, \quad j = 1, 2, \ldots, m. \tag{14}$$

Lastly, we determine the weight coefficient based on the difference coefficient:

$$w_j = \frac{g_j}{\sum_{j=1}^m g_j}, \quad j = 1, 2, \ldots, m. \tag{15}$$

2.2.3. Function Curve of Comprehensive Evaluation. The basis function method is used to express the data as smooth curves. Generally speaking, the estimation of λ is obtained by the GCV method.

$$\mathrm{GCV}(\lambda) = \frac{\mathrm{SSE}}{\left[\mathrm{trace}\left(I - S_{\phi,\lambda}\right)\right]^2}, \tag{16}$$

where

$$\begin{aligned} \mathrm{SSE} &= \int_T \sum_{i=1}^n \sum_{j=1}^m \left(y_{ij}(t) - x_{ij}(t)\right)^2 \mathrm{d}t, \\ S_{\phi,\lambda} &= \Phi M(\lambda)^{-1} \Phi^T W, \\ M(\lambda)^{-1} &= \Phi^T W \Phi + \lambda R, \\ R &= \int_T \left\{D^2\Phi(t)\right\}\left\{D^2\Phi'(t)\right\} \mathrm{d}t. \end{aligned} \tag{17}$$

After transforming the discrete data into functional data, the fitting curve $x_{ij}(t)$ of m indicator data of each evaluated object and the fitting curve $w_{ij}(t)$ $(j = 1, 2, \ldots, m)$ of weight function are found. The evaluation function curve is described by substituting the fitting curve into the evaluation model (18).

$$y_i(t) = w_1(t)x_{i1}(t) + w_2(t)x_{i2}(t) + \cdots + w_m(t)x_{im}(t). \tag{18}$$

2.2.4. Determining the Weight of Time Measurement. During the observation period, the influence of different time data on the total evaluation function is often different. Therefore, the weight of time measurement needs to be considered. Compared to the present impact, the past impact is relatively small. Thereupon, the evaluation of an observation object can be estimated by the following formula:

$$\overline{y}_i = \frac{1}{T}\int_T w_a(t)y_i \mathrm{d}t, \quad i = 1, 2, \ldots, n, \tag{19}$$

where $w_a(t)$ denotes the time weight function. In particular, if it is considered to have the same influence on any time, the time weight function of $w_a(t)$ represents a straight line parallel to the X-axis. In this case, equation (19) can be abbreviated as follows:

$$\overline{y}_i = \frac{1}{T}\int_T y_i(t)\mathrm{d}t. \tag{20}$$

Among them, the time weight is determined using the maximum entropy nonlinear programming method, which reveals the sample information as much as possible and takes into account the difference in the time sequence of the evaluated objects.

If the time weight vector is designed as $W_a = (W_{a1}, W_{a2}, \ldots, W_{aT})^T$, the time degree can be taken into consideration as follows [35]:

$$\alpha = \sum_{h=1}^{T}\frac{T-k}{T-1}w_{ah}, \tag{21}$$

where α reflects the importance degree of each period. When the value of α is closer to 0, it means that the comprehensive evaluation pays more attention to the data at the recent period; when the value of α is closer to 1, it means that the comprehensive evaluation pays more attention to the data at the earlier period; but when $\alpha = 0.5$, it means that the comprehensive evaluation is paid equal attention to the data at each period. Therefore, the weight of time measurement can be solved by the following nonlinear programming method:

$$\max\left\{-\sum_{h=1}^{T} w_{ah}\ln w_{ah}\right\},$$

$$\text{s.t.}\begin{cases}\alpha = \displaystyle\sum_{h=1}^{T}\frac{T-k}{T-1}w_{ah}, \\ \displaystyle\sum_{h=1}^{T} w_{ah} = 1, \ (w_{ah} \in [0,1]), \ h = 1, 2, \ldots, T.\end{cases} \tag{22}$$

In this way, a more objective and reasonable measurement of the weight of time can be obtained.

3. Risks Evaluation Indicator System of Intellectual Property Pledge Financing

As the enterprise has a role as the subject of intellectual property pledge loans, it is general to know that the first repayment source of intellectual property pledge financing is the cash flow earned from the enterprise's investment or other activities. The financial operation situation provides a comprehensive performance of enterprise development, which is the most direct embodiment of the current development situation and development potential of the enterprise. Moreover, the financial operation of the enterprise has a very important impact on both investors and other external stakeholders, as well as the development of the enterprise itself. When it comes to deterioration, not only will it affect the confidence and decision-making of external investors, but also invisible pressure will be put on the internal personnel of the enterprise and enthusiasm of research and development of scientific researchers, furthermore leading to the decrease of realization of intellectual property value. Therefore, it is a direct concern with the future development of the enterprise. The better the enterprise is, the more the funds will be invested in intellectual property. Even though the scale of the enterprises is limited, it is difficult for small- and medium-sized businesses to face more financial risks.

Based on previous research methods and the principles of scientifically, operability, independence, and comparability, 16 financial indicators have been selected from the six aspects, debt paying ability, profitability, operating ability, growth ability, cash flow indicators, and innovation ability, to construct an evaluation indicator system as shown in Table 1.

The debt paying ability is directly related to the survival of the enterprise, which is the basic premise of the healthy development of the enterprise. As a reflection of its continuous operating ability, the debt paying ability of an enterprise is applied to predict its future earnings and analyze its financial risks, which is measured by current ratio, asset-liability ratio, and current liability ratio.

Table 1: Risks evaluation indicator system.

First-level indicator	Second level indicator
Debt paying ability	Current ratio, asset-liability ratio, current liability ratio
Profit ability	Operating profit ratio, return on total assets ratio, return on equity
Operating ability	Inventory turnover ratio, total assets turnover, fixed assets turnover
Growth ability	Net profit growth rate, total assets growth rate, Increase rate of main business revenue
Cash flow	Net cash flow from operating activities per share, Cash ratio of operating income, ratio of cash to current liability
Innovation ability	Intangible asset ratio

Profit ability refers to earning profits to ensure the continuous operation of the enterprise. Profit is regarded as the central concern for business operators, investors, and creditors. Improving profitability will strengthen the growth in the future and, in turn, will become more resilient to risks, which is represented by the income from patented technologies or products revenue, the characteristics of high-tech enterprises. As a result, the profitability of enterprises is measured by operating profit rate, return on total assets ratio, and return on equity in the article.

The ability to make a profit using various assets is defined as the operating ability of enterprises, measured by inventory turnover ratio, total assets turnover, and fixed assets turnover to reflect the efficiency of economic resource management of high-tech enterprises in this article.

In order to better reflect the development potential of the enterprise in the future, the growth ability can be pictured by net profit growth rate, total assets growth rate, and increased rate of main business revenue.

Good enterprise cash flow is an important guarantee of enterprise financial stability and efficient operation, which can be illustrated by net cash flow from operating activities per share, cash ratio of operating income, and the ratio of cash to current liability.

The innovation ability of enterprises has some impacts on the financial risk of intellectual property pledge financing of high-tech enterprises. Due to the limitation of the quarterly data selection, the intangible asset ratio is provided to reflect the innovation ability.

4. Dynamic Evaluation and Analysis of the Risks of Intellectual Property Pledge Financing

Based on the time series data of the financial indicators of the financing enterprise, combined with the data related to the enterprise patent research and development and application, by the entropy method to make a dynamic comprehensive evaluation, obtaining the comprehensive weight of the distribution of the pros and cons of the evaluation data relative to the overall development level at different moments. Then, the comprehensive weighted dynamic evaluation method is used to better measure the dynamic risk of pledge financing enterprises. In this article, we consider high-tech listed companies as research objects, Wangsu Technology (Wangsu Keji), Shuzhi Technology (Shuzhi Keji), Hekang New Energy (Hekang Xinneng), Sanju Environmental Protection (Sanju Huanbao), the quarter data from 2015 to 2019 of the CSMAR database, and some other data from the annual financial reports of listed enterprises.

The objective weight of risk evaluation indicators is shown in Table 2.

The data are fitted by adding penalties. Since the indicator of all samples is nonperiodic, the data can be represented as smooth curves by the basis function method. In general, the process of fitting is realized by the fourth-order spline basis with the same basic function and different smoothing parameters. When the smoothing parameter of λ is large, the smoothness of the fitting curve is good. When it comes to being small, the volatility of the fitting curve is large. GCV is used to estimate the value of λ with the lowest indicator GCV (as shown in Figure 1) to fit the original data to obtain the best fitting curve.

Moreover, the weight function is integrated to obtain the objective weight of each indicator. From Table 2, the weights of each first-level indicator are ranked as follows:

$$\begin{aligned}&\text{debt paying ability} > \text{cash flow} > \text{operating ability}\\&> \text{innovation ability} > \text{profit ability} > \text{growth ability}.\end{aligned} \tag{23}$$

In Figure 2, even though there are some changes in the dynamic evaluation curve of Wangshu Keji from 2015 to 2019, the overall trend of the evaluation curve remains the same. On the other hand, the dynamic evaluation curve of Shuzhi Keji from 2015 to 2019 is a downward trend, with a short-term rebound. In particular, from 2018 to 2019, the evaluation curve of Shuzhi Keji is far lower than that of the other three enterprises. For Hekang Xinneng, there is a strong upward trend of the evaluation curve with the ups and downs from 2015 to 2019. For Sanju Huanbao, the dynamic evaluation curve from 2015 to 2019 has continued its steady upward trend, which is obviously higher than the other three curves. The maximum evaluation curve is in the second quarter of Wangshu Keji in 2015, whereas the minimum evaluation curve is in the first quarter of Shuzhi Keji in 2018.

As for Wangsu Keji, a rising trend in the evaluation curve is observed from the first quarter to the second quarter in 2015. In Figure 2, it has gradually slowed down to reach the peak value at the end of the second quarter. From the third quarter of 2015 to the third quarter of 2018, the evaluation curve has experienced an inverted V-shape, which reaches its bottom in the first quarter of 2017, and begins to rise. From Figures 3 and 4, great changes in the

Table 2: Objective weight of risk evaluation indicators.

First-level indicator	Second level indicator	Weight 1	Weight 2
Debt paying ability	Current ratio	0.041	0.306
	Asset-liability ratio	0.199	
	Current liability ratio	0.066	
Profitability	Operating profit ratio	0044	0.107
	Return on total assets ratio	0.032	
	Return on equity	0.031	
Operating ability	Inventory turnover ratio	0.051	0.162
	Total assets turnover	0.051	
	Fixed assets turnover	0.06	
Growth ability	Net profit growth rate	0.01	0.103
	Total assets growth rate	0.066	
	Increase rate of main business revenue	0.027	
Cash flow	Net cash flow from operating activities per share	0.044	0.178
	Cash ratio of operating income	0.076	
	Net cash flow from operating activities per share	0.058	
Innovation ability	Intangible asset ratio	0.144	0.144

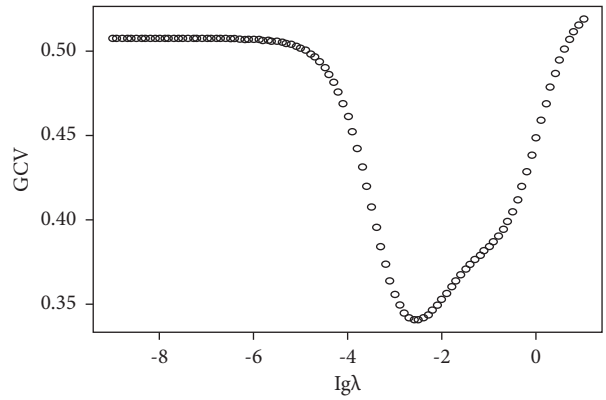


Figure 1: GCV curve.

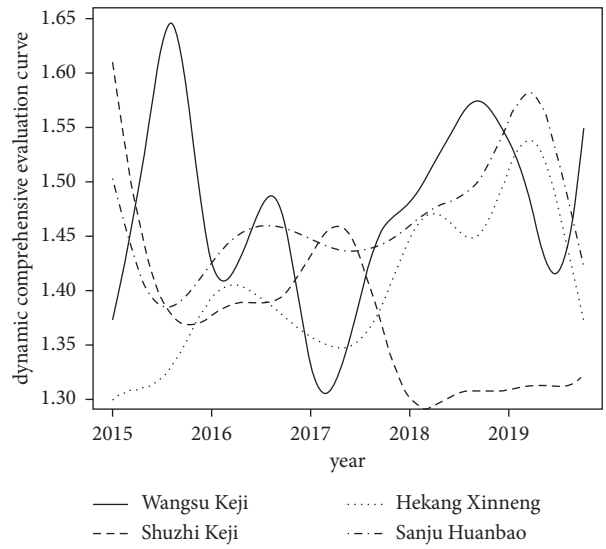


Figure 2: Dynamic comprehensive evaluation curve.

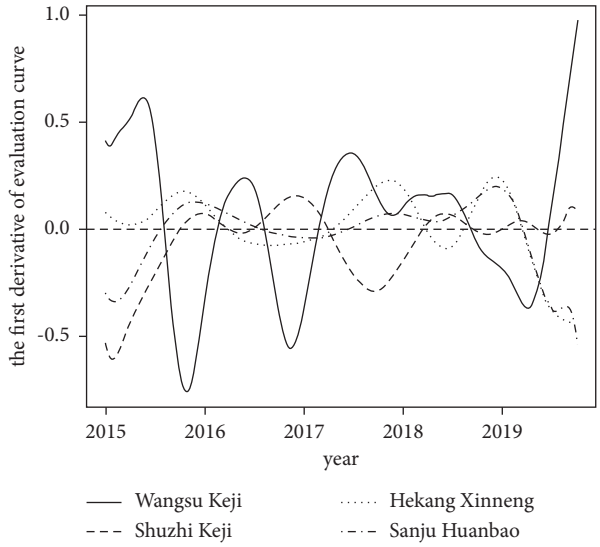


Figure 3: Graph of the first derivative of comprehensive evaluation.

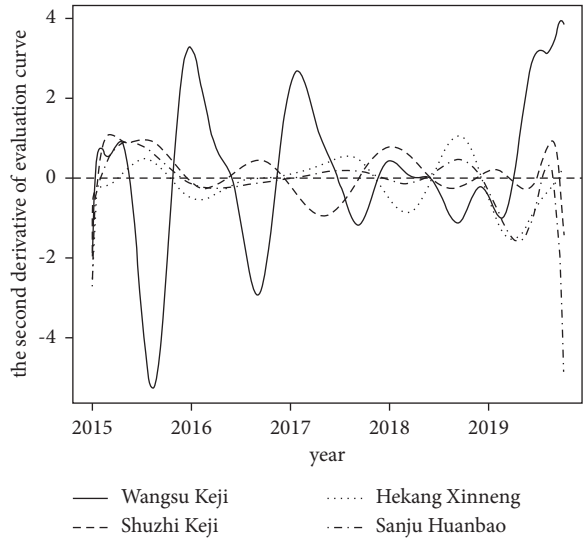


Figure 4: Graph of the second derivative of comprehensive evaluation.

overall financial condition can be seen. The rate of growth grows fast and slowly changes to drop. A decline has occurred in the fourth quarter of 2018, with a strong rebound in the third quarter of 2019. The first derivative chart shows that its growth rate greatly outpaces that of other enterprises. The fastest growing curve is the second derivative (acceleration), which hits the peak at the end of 2019. Rapid optimizing development has been preserved by the financial situation of Wangsu Keji in the second half of 2019.

From Figure 2, we find that the evaluation curve of Shuzhi Keji declines significantly in the first three quarters of 2015, and from the second quarter to the fourth quarter of 2017, it reaches its lowest point quarter of 2018. Although it started to rise in the fourth quarter in 2015, the fluctuation of speed is between 0 and 0.1 in Figures 3 and 4, with the slowing acceleration and upward trend. From the second quarter of 2018 to the end of the fourth quarter of 2019, although there is an upward trend, speed fluctuation is controlled with a slowing upward trend around 0.

The dynamic evaluation curve of Hekang Xinneng falls to its lowest point in the first quarter of 2015, followed by a rising strong upward trend in 2015, the second, third, and fourth quarters of 2017, and especially 2018, as shown in Figure 2. As shown in Figures 3 and 4, it has reached a peak of accelerated rising speed in late 2018 and early 2019. There is a relatively flat decline in the second, third, and fourth quarters of 2016 and also a relatively flat decline in the second quarter of 2018, with a sharp downward trend in 2019.

From Figure 2, we can also see that although the overall dynamic evaluation curve of Sanju Environmental Protection is showing an upward trend, there are also two strong decline stages, namely, in the first and second quarters of 2015 and the second, third, and fourth quarters of 2019.

Table 3: Comprehensive dynamic's evaluation value and ranking.

Enterprise	Wangsu Keji	Shuzhi Keji	Hekang Xinneng	Sanju Huanbao
Evaluation value	0.015 431 27	0.013 618 33	0.015 122 52	0.015 580 47
Ranking	2	4	3	1

Table 4: Evaluation value (EV) and ranking (R) of first-level indicators.

Enterprise	R	Debt paying ability		Profitability		Operating ability		Growth ability		Cash flow		Innovation ability	
		EV	R	EV	R	EV	R	Ev	R	Ev	R	Ev	R
Wangsu Keji	2	0.004 44	4	0.001 69	2	0.003 80	1	0.001 03	4	0.002 72	3	0.002 75	1
Shuzhi Keji	4	0.004 47	3	0.001 70	1	0.002 78	2	0.001 21	1	0.001 75	2	0.001 72	4
Hekang Xinneng	3	0.006 23	2	0.001 41	4	0.002 55	4	0.001 05	4	0.001 57	4	0.002 32	2
Sanju Huanbao	1	0.006 24	1	0.001 51	3	0.002 74	3	0.001 05	3	0.001 77	1	0.002 27	3

However, there is a smaller decline in the third and fourth quarters of 2016 and the first and second quarters of 2017. In Figure 3, it approximates to M type in the overall trend of the first derivative of the evaluation curve. Most speeds in the intermediate stage are greater than 0. Especially in Figure 4, the second derivative is greater than that in the third quarter of 2015 and the second quarter to fourth quarter of 2018, while an increasing trend in the first derivative illustrates the rapid development of the enterprise financial conditions during the two stages.

The function data of the comprehensive evaluation function curves have been obtained through the above empirical analysis. If further comprehensive comparative analysis of the financing risks of the four companies is required, the evaluation curve needs to be integrated to obtain the corresponding evaluation results. At the same time, the evaluation results change over different times during the observation period. In practical applications, dynamic analysis tends to pay more attention to recent results, so the weight of the recent period is larger, and the weight of the earlier period is relatively small. Based on the experience of all aspects, the recent data have the greatest impact on the comprehensive evaluation. Therefore, $\alpha = 0.1$ is taken in this article by referring to the relevant literature.

The comprehensive dynamic evaluation's values of different enterprises can be obtained by using formula (19) and the result is shown in Table 3:

According to Table 3, the financial risk ranking of each enterprise is

$$\text{Sanju Huanbao} > \text{Wangsu Keji} > \text{Hekang Xinneng} > \text{Shuzhi Keji}. \tag{24}$$

Shuzhi Keji has the largest financial risk, whereas Sanju Huanbao has the smallest financial risk. At the same time, the dynamic comprehensive evaluation value and ranking of the first-level indicators of each enterprise are calculated as shown in Table 4.

The results in Table 4 show that the two enterprises ranks No. 1 and No. 2 point to the first ranking of first-level indicators evaluation values. At the same time, the dispersion degrees of the evaluation value of the six first-level indicators evaluation values are different among the companies. There are differences in the impact on the total value of the dynamic comprehensive evaluation. Therefore, it is necessary to calculate the coefficient of variation of the evaluation value of each first-level index and draw the radar chart, as shown in Figure 5.

In Figure 5, first of all, debt paying ability shows the largest difference in the first-level indicator. Due to the first ranking of first-level evaluation value, the first ranking of the dynamic comprehensive evaluation value is Sanju Huanbao. Secondly, operating ability and innovation ability are the first-level indicators with a large difference. It can be seen that the evaluation value of Wangsu Keji ranks first in these two first-level indicators, leading to the second ranking of the dynamic comprehensive evaluation value. In addition, the first-level indicator (the indicator of debt paying ability and innovation ability) with the prominent coefficient of variation will significantly improve the ranking of the dynamic comprehensive evaluation. In contrast, the first-level indicator with the insignificant coefficient of variation (the indicator of profit and growth ability) has a weak impact on the dynamic comprehensive evaluation results. For example, Shuzhi Keji ranks first in terms of profitability and growth ability, with the rank of third in debt paying ability and fourth in innovation ability. As a result, Shuzhi Keji gets fourth place in the dynamic comprehensive evaluation result. Finally, although Hekang Xinneng does not rank first with the first-level index, it ranks second in two first-level indicators of solvency and innovation ability, the dynamic comprehensive evaluation result of Hekang Xinneng exceeds Shuzhi Keji to rank third.

In order to further analyze the change of the financial risk level, during the observation period from 2015 to 2017 and 2018 to 2019, the above method is used to calculate the dynamic comprehensive evaluation value of each subperiod, as shown in Table 5.

In the first stage, there is a difference in the rank of dynamic comprehensive evaluation value from 2015 to 2017 and the five-year dynamic comprehensive evaluation. In the second stage, the rank of dynamic comprehensive evaluation value during 2018 to 2019 is basically consistent with the results of the five-year dynamic comprehensive evaluation,

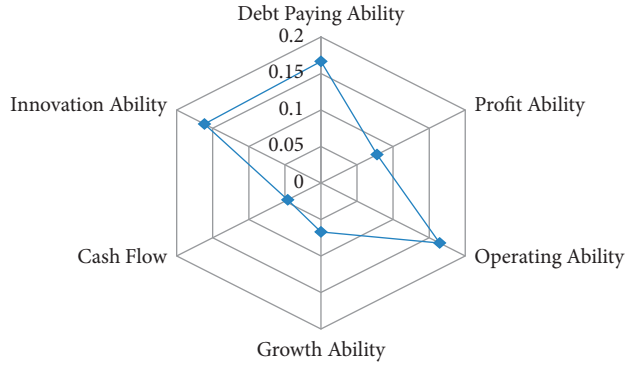


Figure 5: Coefficient of the variation radar chart.

Table 5: Dynamic comprehensive evaluation value and ranking for two time periods.

Enterprise	From 2015 to 2017		From 2018 to 2019	
	Evaluation value	Ranking	Evaluation value	Ranking
Wangsu Keji	0.022 141 74	3	0.032 856 94	2
Shuzhi Keji	0.022 804 46	2	0.029 073 65	4
Hekang Xinneng	0.021 503 32	4	0.031 889 28	3
Sanju Huanbao	0.023 598 25	1	0.033 052 03	1

which indicates that the development of each enterprise has a great change in the two periods. It can be seen from the ranking results that the dynamic comprehensive evaluation values of Wangsu Keji and Hekang Xinneng have a rising trend, while that of Shuzhi Keji is has a decreasing trend.

5. Conclusion

In this article, a new dynamic comprehensive evaluation method is constructed using the functional evaluation method, while the maximum information entropy method describes the dynamic financial risks of an intellectual property pledge financing of listed enterprises. The main conclusions of this article are as follows: first, with the transformation of the quarterly financial observation data of all enterprises into a function curve, the dynamic evaluation curve of each enterprise is obtained, accompanied by the first and second derivatives of the evaluation function curve, to analyze the changing trend of the evaluation curve. At the same time, with the combination of the dynamic evaluation curve with the optimal time weight curve, it is more accurate to calculate the total period value of financing risk. Second, the financial risk of each enterprise is mainly affected by the first-level indicators with outstanding advantages of coefficient of variation, while enterprises with small risk may also have weak influence. The empirical analysis shows that under the guidance of innovative development concepts, the advantages and development characteristics of the enterprise are highlighted to foster strengths and avoid weaknesses, making it necessary to strengthen financial management ability and innovation ability and reduce the risk of enterprise pledge financing. More importantly, the FDA method is applied to analyze the financial risk of enterprise pledge financing dynamically, in favor of objectively evaluating the financial risk of enterprise, which is subject to provide a new evaluation method for the financial risk analysis of high-tech enterprise pledge financing.

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 manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant nos. 71 901 222 and 71 974 204. This work was supported in part by the Humanities and Social Science Foundation of the Ministry of Education of China under Grant no. 21YJC790065, "the Fundamental Research Funds for the Central Universities", and Zhongnan University of Economics and Law under Grant nos. 2722020JCG062 and 2722020JX005.

References

[1] A. V. Banerjee, B. Timothy, and T. W. Guinnane, "The neighbor's keeper: the design of a credit cooperative with theory and a test," *Quarterly Journal of Economics*, vol. 109, no. 2, pp. 491–515, 1994.

[2] P. Angelini, R. D. Salvo, and G. Ferri, "Availability and cost for small businesses: customer relationships and credit cooperative," *Journal of Banking and Finance*, vol. 22, pp. 925–954, 1998.

[3] A. N. Berger and G. F. Udell, "A more complete conceptual framework for SME finance," *Journal of Banking and Finance*, vol. 30, no. 11, pp. 2945–2966, 2006.

[4] W. Zhang and X. D. Zhang, "New Financing Mode for small and medium enterprises coordinating the interests among the guarantee institution dominated participants," *The Theory and Practice of Finance and Economics*, vol. 34, no. 1, pp. 2–6, 2013, in Chinese.

[5] Y. F. Lin and Y. J. Li, "Promoting the growth of medium and small-sized enterprises through the development of medium and small-sized financial institutions," *Economic Research Journal*, vol. 1, pp. 10–18, 2001, in Chinese.

[6] Z. Xu and C. W. Zou, "Allocation and incentive Mechanism design of loan approval right in banks under the framework of hard and soft information: enlightenment on financing problems of small and medium-sized enterprises," *Journal of Financial Research*, vol. 8, pp. 1–15, 2010, in Chinese.

[7] X. Xiong, C. Guo, W. Zhang, and Y. J. Zhang, "Loan rate pricing of SME financing based on agent-based computational finance approach," *Systems Engineering-Theory & Practice*, vol. 29, no. 12, pp. 9–14, 2009, in Chinese.

[8] X. Xiong, J. M. Tan, W. Zhang, Y. J. Zhang, and Y. Yang, "Study on counter guarantee and proportional guarantee mechanism of SME financing," *Soft Science*, vol. 25, no. 6, pp. 80–85, 2011, in Chinese.

[9] K. P. Jarboe and I. Ellis, "Intangible assets: innovative financing for innovation," *Issues in Science and Technology*, vol. 26, no. 2, pp. 75–80, 2010.

[10] J. Crawford and R. Strasser, "Management of infringement risk of intellectual property assets," *Intellectual Property and Technology Law Journal*, vol. 20, no. 12, pp. 7–10, 2008.

[11] R. Chessy and O. Toivanen, "Is there adverse selection in the credit market?" *Venture Capital: An International Journal of Entrepreneurial Finance*, vol. 3, no. 3, pp. 215–238, 2001.

[12] J. P. Niinimaki, "Nominal and true cost of loan collateral," *Journal of Banking and Finance*, vol. 35, no. 10, pp. 2782–2790, 2011.

[13] I. Davies, "Secured financing of intellectual property assets and the reform of English personal property security law," *Oxford Journal of Legal Studies*, vol. 26, no. 3, pp. 559–583, 2006.

[14] R. Pennington and C. Sanchez, "Negotiating liability allocation terms: risk, indemnity, and intellectual property," *Contract Management*, vol. 47, pp. 42–53, 2007.

[15] S. Schenkerman, "Avoiding rank reversal in AHP decision-support models," *European Journal of Operational Research*, vol. 74, no. 3, pp. 407–419, 1994.

[16] S. J. Chen and C. L. Hwang, *Fuzzy Multiple Attribute Decision Making*, Springer, Berlin, Germany, 1992.

[17] T. Ding, J. Yang, H. Wu, W. Yao, C. Tan, and L. Liang, "Research performance evaluation of Chinese university: a non-homogeneous network DEA approach," *Journal of Management Science and Engineering*, 2020.

[18] A. Liu, T. Liu, J. Mou, and R. Wang, "A supplier evaluation model based on customer demand in block chain tracing anti-counterfeiting platform project management," *Journal of Management Science and Engineering*, vol. 5, no. 3, pp. 172–194, 2020.

[19] J. O. Ramsay, "When the data are functions," *Psychometrika*, vol. 47, no. 4, pp. 379–396, 1982.

[20] J. Q. Ramsay and C. J. Dalzell, "Some tools for functional data analysis," *Journal of the Royal Statistical Society: Series B*, vol. 53, no. 3, pp. 539–572, 1991.

[21] B. W. Silverman, "Smoothed functional principal components analysis by choice of norm," *Annals of Statistics*, vol. 24, no. 1, pp. 1–24, 1996.

[22] Z. Lu, "Nonparametric functional data analysis: theory and practice," *Technometrics*, vol. 49, no. 2, pp. 225-226, 2007.

[23] A. Delagle and P. Hall, "Defining probability density for a distribution of random function," *Annals of Statistics*, vol. 38, no. 2, pp. 1171–1193, 2010.

[24] A. Petersen and H. G. Muller, "Functional data analysis for density functions by transformation to a Hilbert space," *Annals of Statistics*, vol. 44, no. 1, pp. 183–218, 2016.

[25] P. Kokoszka and M. Reimherr, *Introduction to Functional Data Analysis*, Chapman Hall/CRC, FL, USA, 2017.

[26] T. Hastie, A. Buja, and R. Tibshirani, "Penalized disciminant analysis," *Annals of Statistics*, vol. 23, no. 1, pp. 73–102, 1995.

[27] G. M. James, "Generalized linear models with functional predictors," *Journal of the Royal Statistical Society: Series B*, vol. 64, pp. 422–432, 2002.

[28] J. O. Ramsay, G. Hooker, D. Campbell, and J. Gao, "Parameter estimation for differential equations: a generalized smoothing approach," *Journal of the Royal Statistical Society: Series B*, vol. 16, pp. 741–796, 2007.

[29] F. Yao, H. G. Muller, and J. L. Wang, "Functional data analysis for longitudinal data," *Annals of Statistics*, vol. 23, pp. 2973–2903, 2005.

[30] N. Serban and L. Wasserman, "Cats: clustering after transformation and smoothing," *Journal of the American Statistical Association*, vol. 100, no. 1, pp. 990–999, 2005.

[31] J. Zhang, "Risk management on intellectual property pledge financing by small and medium S & T enterprises: from the aspect of bank," *Scientific Management Research*, vol. 31, no. 2, pp. 98–101, 2013.

[32] H. Li, Z. Yuan, and S. Li, "Innovative enterprises intellectual property pledge loan risk assessment," *Studies in Science of Science*, vol. 35, no. 8, pp. 1253–1263, 2017.

[33] W. H. Su, L. R. Sun, and F. Cui, "One comprehensive evaluation method study based on functional data," *Statistical Research*, vol. 2, pp. 88–94, 2013, in Chinese.

[34] M. Y. Yan, "The evaluation of quality of life a method based on functional features of data," *Journal of Statistics and Information*, vol. 2, pp. 13–17, 2007, in Chinese.

[35] Y. J. Guo, Y. Yao, and P. T. Yi, "A dynamic comprehensive evaluation method and its application," *Systems Engineering–Theory & Practice*, vol. 10, pp. 154–158, 2007, in Chinese.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 9592906, 16 pages
https://doi.org/10.1155/2021/9592906

Research Article

Taming the Factor Zoo: New Evidence from China

Dongxu Chen, Xieyang Shen, and Tao Liu

School of Finance, Renmin University of China, Beijing, China

Correspondence should be addressed to Xieyang Shen; xieyang.shen@ruc.edu.cn

Received 25 July 2021; Revised 10 October 2021; Accepted 11 October 2021; Published 29 November 2021

Academic Editor: Dehua Shen

Copyright © 2021 Dongxu Chen 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.

We address the well-known "factor zoo" problem in the Chinese stock market. By replicating a generation of pricing factors, we verify the Liu–Stambaugh–Yuan four-factor model which subsumes other counterparts in the Chinese A-share market. We further construct a characteristic library and apply the double-selection LASSO approach to explore whether significant anomalies contribute to current pricing factors. We find that some anomalies indeed play a significant role in pricing cross-sectional returns, but the improvement to the Liu–Stambaugh–Yuan four-factor model is limited.

1. Introduction

Recent studies have proposed a series of factor models to price cross-sectional expected returns. As indicated by Cochrane [1], there exists a zoo of factors which can potentially explain expected stock returns. Searching for truly effective factors is a hot issue in asset pricing field. Once a new factor is constructed, researchers will evaluate whether it can provide incremental information by comparing with other mainstream factor models; see, for instance, works of Hou et al. [2], Racicot and Rentz [3], Racicot et al. [4], and Racicot et al. [5]. Hou et al. [2] show that the q-factor dominates all other pricing factors by various spanning tests. Racicot and Rentz [3]; Racicot et al. [4]; Racicot et al. [5] study some variants of the new Fama and French [6] model by using a new generalized method of moments estimator (GMM_d) and find that only the market risk factor is the consistently significant factor. Furthermore, these papers also find that adding to the new Fama and French [6] model with the illiquidity factor of Pástor and Stambaugh [7] does not provide more explanatory power. Certainly, other researchers may put forward different opinions in the future, but studies in this field have one thing in common: they all depend on WRDS, a large and complete data library in the US market. However, it still remains a puzzle whether these powerful factors take effects in other emerging markets, such as the Chinese stock market.

As the second largest economy of the world, China has developed its capital market vigorously in recent years and the total market value exceeded 8 trillion dollars in the end of 2019. Even though Chinese stocks are getting more appealing to global investors, limited studies focus on asset pricing problems in the A-share market, especially about which factors drive cross-sectional Chinese stock returns. The political and economic environments are quite different from those in the US. Besides, China limits foreign investors to participating in its domestic stock market directly. Hence, the factor models in the US may not all work in China. We hope to determine the most effective factor models for A-share market in recent 20 years and explore whether there exist undiscovered latent pricing factors.

Using Chinese A-share data from July 2000 to December 2019, we start by replicating a generation of pricing factor models emerging in recent literature, including Fama–French six-factor factor model (FF6, henceforth), Hou–Xue–Zhang q5-factor model (q5, henceforth), Stambaugh–Yuan four-factor model (SY4, henceforth), Daniel–Hirshleifer–Sun three-factor model (DHS-3, henceforth), and Liu–Stambaugh–Yuan four-factor model (LSY4, henceforth). Our paper independently constructs a database based on the work of Qiao [8] and Hou et al. [9]. We find LSY4 factors can generate a maximum Sharpe ratio of 2.16, significantly higher than other models. Differing from the US market, the investment factors hardly have any

risk premiums for A-share stocks, regardless of whether we use the q-factor proposed by Hou et al. [2] or CMA factor proposed by Fama and French [10]. These two factors only generate the average return of 0.07% and −0.10% across 20 years, respectively. Profitability factors, however, indeed show some significant risk premiums. The ROE factor of q5 is better than Fama's RMW, since the former is constructed with quarterly updated ROE rather than annually updated ROE. Consisting with this result, we find for SY4 and DHS-3, which is based on mispricing and investor behaviors, respectively; only the parts related to profitability measures have significant risk premiums.

Next, we test the pricing power of these pricing models. Following Hou et al. [9], we construct 104 anomalies and obtain their value-weighted long-short return spreads. By calculating the alphas with respect to various pricing models, we show that LSY4 can explain up to 98 anomalies, far more than other models. Regardless of average alphas, average t-statistics, or Gibbons–Ross–Shanken statistics, LSY4 always achieves the lowest value, indicating its outstanding performance to price A-share stocks. When we turn to spanning tests, it is not surprising that LSY4 can explain all factors except for the postearnings-announcement drift (PEAD) factor. For example, the powerful q-factor recommended by Hou et al. [2] is totally subsumed and has a tiny alpha of −0.06% ($t = -0.38$). However, PEAD survives from LSY4 with a significant alpha of 0.72% ($t = 3.35$).

Finally, we explore whether other potential risk factors can be added into LSY4 in future research. In contrast to the previous risk premium perspective, we focus on SDF loadings in this section. We mainly adopt the double-selection LASSO approach proposed by Feng et al. [11] to evaluate the marginal contribution of other latent risk factors. This approach accounts for model selection mistakes in finite sample and tests the importance of a given factor beyond other benchmark factors. The factors we test here include significant anomalies in the A-share market found by Qiao [8] as well as anomalies with significant LHY4 alphas in the previous section. We choose the benchmark factors as the union of aforementioned pricing factors. The empirical result shows 8 out of 18 anomalies can provide incremental information, most of which describe trading frictions, such as idiosyncratic volatility per the FF3 model, daily trading volume, and volatility of turnover. As a robustness check, we also extend the number of benchmark factors and replace test portfolios, but the conclusion changes a little. However, we notice a slight improvement when we add these newly found factors into the LSY4 model. It indicates that current factor zoo is limited to price A-share stocks better than LSY4. We suppose future research should focus on constructing new characteristics to come up with new models.

The contribution of our paper is twofold. First, we fill the gap in asset pricing studies related to Chinese market. As we discussed before, few studies apply the forefront asset pricing models to Chinese market. Although the data library of A-share market is not comparable to WRDS, we suppose it is enough for us to conduct a thorough research. Unfortunately, most of researchers still focus on testing Fama–French factors, ignoring the rapid development of literature in this field. On the contrary, these studies cannot even reach a consensus due to various samples. Examples of such studies include Hu et al. [12], Jiang et al. [13], and Cheung et al. [14]; some papers emphasize the importance of the value factor but others hold the opposite opinion. However, a recent research by Liu et al. [15] is remarkable, which can be regarded as a pioneer to explore factor models suitable for China. They construct a three-factor model using market return, size, and E/P ratio and find their model plus a turnover factor can explain 10 significant anomalies in the Chinese market after excluding the 30% smallest stocks. Nonetheless, this model still needs time to be universally accepted since it is newly published. Our paper inherits the work of Liu et al. [15] and will provide more robust evidences to support it.

Second, we conform to the trend in recent asset pricing literature that using machine learning techniques to search for latent pricing factors. As more and more factors are discovered, researchers are challenged by "multidimensionality" problem. Thus, a strand of literature takes advantage of variable selection approaches to screen effective factors. Representative works include Han et al. [16], Freyberger et al. [17], and Feng et al. [11]; all of which use LASSO regressions to exclude redundant factors. Inspired by their work, our paper applies LASSO to estimate SDF loadings based on a data library constructed by Qiao [8]. This author followed Hou et al. [9] and replicated 231 anomalies with Chinese stocks. Given some unique characteristics in the A-share market, the category and number of significant anomalies are really different from the US market. Thus, we pick out these significant anomalies and test whether they have contributions beyond the existing pricing factors using a double-selection LASSO approach. We hope our findings will inspire future asset pricing studies.

The rest of paper is organized as follows. Section 2 describes data and construction details of pricing factors. Section 3 reports the main empirical results of comparing asset pricing models, including explaining anomalies, spanning regression, and GRS test. Section 4 introduces the double-selection LASSO approach and shows its application on evaluating other potential risk factors. Section 5 is the conclusion.

2. Data

2.1. Data Source and Filters. We collect trading data and financial data from China Stock Market & Accounting Research Database (CSMAR). Our sample period is from July 2000 to December 2019. We do not use data before June 2000 since the cash flow statements of A-share stocks begin at the end of 1997, and the calculation of some factors needs the data from previous consecutive quarters. In A-share market, income statement and cash flow statement only report flow data between two recent quarters. Thus, we recover stock values of accounting data by subtracting values from the prior quarter. Note that Chinese listed firms are required to release quarterly reports after 2002; we subtract flow data semiannually before 2002.

Our sample includes all A-shares, and we impose several filters to exclude stocks (1) of financial industry, (2) listed less than six months, (3) having less than 15 trading days last month, and (4) having less than 120 trading days in the past 12 months. In addition, referring to the views of Liu et al. [15], the smallest 30% stocks in the A-share market are possibly shell targets in reverse mergers, which contaminate stock returns. Thus, we also exclude the smallest 30% stocks each month to avoid the shell value contamination.

2.2. Factor Pricing Models. This section briefly introduces the factor pricing models we are devoted to compare in our paper, including Fama–French six-factor factor model (FF6), Hou–Xue–Zhang q5-factor model (q5), Stambaugh–Yuan four-factor model (SY4), Daniel–Hirshleifer–Sun three-factor model (DHS-3), and Liu–Stambaugh–Yuan four-factor model (LSY4). The details to construct each factor are provided in the Supplementary Appendix (available (here)).

2.2.1. FF6 Model. Fama and French [6] propose their five-factor model by adding two additional factors, CMA and RMW, into the well-known 3-factor model [18]. CMA is the average returns on stocks with low investment minus the average returns on stocks with high investment. RMW is the average returns on stocks with high operating profitability minus the average returns on stocks with low operating profitability. Fama and French [10] further add the sixth factor, UMD, to control the momentum effect found by Jegadeesh and Titman [19]. UMD is the difference of returns on winner and loser portfolios.

2.2.2. q5 Model. Hou et al. [20] propose the famous q-factor model and Hou et al. [2] augment it to the q5-factor model, whose factors include MKT, R_{me}, R_{roe}, R_{ia}, and R_{eg}. In accordance with Fama–French factors, R_{me}, R_{roe}, and R_{ia} correspond to SMB, RMW, and CMA, respectively. However, they are derived from the q-theory instead of cashflow discounting. The fifth factor, R_{eg}, describes expected investment growth.

2.2.3. SY4 Model. Stambaugh and Yuan [21] do not construct their pricing factor by sorting stocks based on a specific characteristic; instead, they combine rankings of stocks with respect to 11 documented anomalies. The two new factors in addition to size and market factors, MGMT and PERF, are return spreads between underpriced stocks and overpriced stocks.

2.2.4. DHS-3 Model. Daniel et al. [22] propose a model that supplements CAPM with another two behavior factors. The financing factor FIN captures the long-horizon mispricing, which is induced by investors' overconfidence and exploited by managers' decisions to issue or repurchase equity. The postearnings' announcement drift factor PEAD captures the short-horizon mispricing, which is induced by investors' limited attention and underreaction to earnings' surprises.

2.2.5. LSY4 Model. Liu et al. [15] re-examine the size and value effect in China and construct new factors suitable for the A-share market. Compared to Fama and French [18], they replace the book-to-market ratio with earnings-to-price ratio and propose a new factor named VMG. Moreover, to subsume more anomalies, they add the fourth factor PMO based on one-month abnormal turnover. PMO represents the return spread between stocks about which investors are relatively pessimistic and stocks about which investors are relatively optimistic.

2.3. Test Anomalies. Hou et al. [9] have listed totally 447 anomalies published in top finance and accounting journals. However, one characteristic can construct several anomalies using different holding periods. Thus, their 447 anomalies are derived from 190 characteristics, which can be classified into 6 categories: *momentum, value versus growth, investment, profitability, intangibles,* and *trading frictions*. According to details disclosed in the online appendix of Hou et al. [9], our paper replicates 104 characteristics that are available and have fewest missing values in the A-share market. Qiao [8] also performs similar work but our characteristics accommodate theirs are more diverse in each category. We show their acronyms and definitions in Table 1.

One problem is how to combine characteristics with different frequencies. We need to merge the values of low-frequency characteristics with monthly stock returns. There are three possible cases: (1) for annually updated characteristics, returns from July of year t to June of year $t+1$ are matched with characteristics constructed in the December of year $t-1$; (2) for quarterly updated characteristics, returns in current month are matched with latest available characteristics, which are constructed with data published in the newest financial report; (3) for monthly updated characteristics, returns in current month are matched with characteristics constructed at the end of last month. Then, we sort each characteristic into 10 deciles and calculate the value-weighted decile returns for the next month. The deciles are rebalanced monthly. The anomaly returns are earned by holding a decile portfolio with respect to high expected returns and selling short a decile portfolio with respect to low expected returns.

3. Comparing Factor Models

3.1. Summary Statistics. We start by reporting the in-sample performance of individual factors. Table 2 displays some summary statistics for each factor model. The sample period is from July 2000 to December 2019. We find no Fama–French factors are significant and its maximum Sharpe ratio is only 1.25, which is the annualized Sharpe ratio of the tangency portfolio spanned by these factors. Specifically, the value effect measured by HML only derives the average return of 0.37% ($t = 1.43$) per month. Other factors SMB, RMW, and UMD are even worse, whose returns are 0.35% ($t = 1.40$), 0.24% ($t = 1.11$), and 0.23% ($t = 0.86$), respectively.

Table 1: List of characteristics. This table lists the acronyms, names, categories, and updating frequencies of 104 characteristics we consider in our empirical analysis. We build these characteristics using both trading and financial data from China Stock Market & Accounting Research Database (CSMAR), accessible at Wharton Research Data Service (WRDS). The inclusion of characteristics, construction details, and classification follow Hou et al. [9], recognizing the discrepancies between accounting standards in China and US GAAP.

Var	Definition	Category	Frequency
abr	Cumulative abnormal returns around earnings announcement dates	Momentum	Quarterly
abtur	Abnormal turnover	Trading frictions	Monthly
age	Firm age	Intangible	Monthly
alaq	Asset liquidity scaled by 1-quarter-lagged total assets	Intangible	Quarterly
almq	Asset liquidity scaled by 1-quarter-lagged market value of assets	Intangible	Quarterly
amq	Assets-to-market	Value versus growth	Quarterly
ami	Absolute return-to-volume	Trading frictions	Monthly
atoq	Return on quarterly assets turnover	Profitability	Quarterly
beta	Market beta	Trading frictions	Monthly
beta_	Downside beta	Trading frictions	Monthly
betad	The dimson beta	Trading frictions	Monthly
betafp	The Frazzini–Pedersen beta	Trading frictions	Monthly
blq	Quarterly book leverage	Profitability	Quarterly
bm	Book-to-market equity	Value versus growth	Annually
bmj	Book-to-june-end market equity	Value versus growth	Annually
bmq	Quarterly book-to-market equity	Value versus growth	Quarterly
cdi	Composite debt issuance	Investment	Annually
cei	Composite debt issuance	Investment	Annually
claq	Quarterly cash-based operating profits-to-lagged assets	Profitability	Quarterly
cop	Cash-based operating profitability	Profitability	Annually
cpq	Cash flow to price	Value versus growth	Quarterly
cs	Coskewness	Trading frictions	Monthly
cta	Cash to assets	Intangible	Annually
ctoq	Capital turnover	Profitability	Quarterly
cvd	Coefficient of variation of trading volume	Trading frictions	Monthly
cvt	Coefficient of variation of share turnover	Trading frictions	Monthly
dcoa	Changes in current operating assets	Investment	Annually
dcol	Changes in current operating liabilities	Investment	Annually
dgs	Change in gross margin minus change in sales	Intangible	Annually
dmq	Debt to market	Value versus growth	Annually
dnca	Changes in noncurrent assets	Investment	Annually
dnco	Changes in net noncurrent operating assets	Investment	Annually
dnoa	Changes in net operating assets	Investment	Annually
dpia	Changes in ppe and inventory to assets	Investment	Annually
droa	4-quarter change in return on assets	Profitability	Quarterly
droe	4-quarter change in return on equity	Profitability	Quarterly
dsa	Change in sales minus change in accounts receivable	Intangible	Annually
dsi	change in sales minus change in inventory	Intangible	Annually
dss	Change in sales minus change in sg&a	Intangible	Annually
dtv	Trading volume	Trading frictions	Monthly
dwc	Changes in net noncash working capital	Investment	Annually
e6	6-month residual momentum	Momentum	Monthly
e11	11-month residual momentum	Momentum	Monthly
ebpq	Enterprise book-to-price	Value versus growth	Annually
emq	Enterprise multiple	Value versus growth	Annually
epq	Earnings to price	Value versus growth	Quarterly
etr	Effective tax rate	Intangible	Annually
fpq	Failure probability	Profitability	Quarterly
glaq	Gross profits-to-lagged assets	Profitability	Annually
gpa	Gross profits to assets	Profitability	Annually
ia	Investment to assets	Investment	Annually
iaq	Quarterly investment to assets	Investment	Quarterly
ir	Intangible return	Value versus growth	Annually
isc	Idiosyncratic skewness per the CAPM	Trading frictions	Monthly
isff	Idiosyncratic skewness per the FF3 model	Trading frictions	Monthly
isq	Idiosyncratic skewness per the q-factor model	Trading frictions	Monthly
iv	Idiosyncratic volatility	Trading frictions	Monthly

Table 1: Continued.

Var	Definition	Category	Frequency
ivq	Idiosyncratic volatility per the q-factor model	Trading frictions	Monthly
ivff	Idiosyncratic volatility per the FF3 model	Trading frictions	Monthly
ivg	Inventory growth	Investment	Annually
kzq	Financial constraints (the Kaplan–Zingales index)	Intangible	Quarterly
mdr	Maximum daily return	Trading frictions	Monthly
me	Market equity	Trading frictions	Annually
ndpq	Quarterly net debt to price	Value versus growth	Quarterly
noa	Net operating assets	Investment	Annually
nopq	Quarterly net payout yield	Value versus growth	Quarterly
nsi	Net stock issues	Investment	Annually
ocpq	Operating cash flow to price	Value versus growth	Quarterly
olaq	Operating profits-to-lagged assets	Profitability	Quarterly
oleq	Operating profits-to-lagged equity	Profitability	Quarterly
olq	Quarterly operating leverage	Intangible	Quarterly
opa	Operating profits to assets	Profitability	Annually
ope	Operating profits to equity	Profitability	Annually
opq	Quarterly payout yield	Value versus growth	Quarterly
o	Ohlson's *o*-score	Profitability	Quarterly
pmq	Quarterly return on profit margin	Profitability	Quarterly
poa	Percent operating accruals	Investment	Annually
pps	Share price	Trading frictions	Monthly
ra25	Seasonality (average return over months t-24, t-36, t-48, and t-60)	Intangible	Monthly
rn25	Seasonality (average return from month t-60 to t-13)	Intangible	Monthly
r1a	Seasonality (return in month t-12)	Intangible	Monthly
r1n	Seasonality (average return from month t-11 to t-1)	Intangible	Monthly
r6	Prior 6-month returns	Momentum	Monthly
r11	Prior 11-month returns	Momentum	Monthly
rev	Reversal	Value versus growth	Monthly
rnaq	Return on net operating assets	Profitability	Annually
roa	Return on assets	Profitability	Quarterly
roe	Return on equity	Profitability	Quarterly
rs	Revenue surprises	Momentum	Quarterly
sgq	Sales growth	Value versus growth	Annually
size	Firm size	Trading frictions	Monthly
spq	Sales to price	Value versus growth	Quarterly
sr	5-year sales growth rank	Value versus growth	Annually
srev	Short-term reversal	Trading frictions	Monthly
sue	Standardized unexpected earnings	Momentum	Quarterly
tail	Tail risk	Trading frictions	Monthly
tanq	Tangibility	Intangible	Annually
tbiq	Taxable income-to-book income	Profitability	Quarterly
tes	Tax expense surprises	Momentum	Quarterly
ts	Total skewness	Trading frictions	Monthly
tur	Share turnover	Trading frictions	Monthly
tv	Total volatility	Trading frictions	Monthly
wwq	The whited-Wu index	Intangible	Quarterly
zq	Altman's *z*-score	Profitability	Quarterly

Consistent with Liu et al. [15], we see the investment factor CMA is very weak in the A-share market, since its mean return is even negative. As a result, the q-factors showing strong explanative power in the US market does not have enough risk premiums as we expected. Similar to Fama's CMA, the investment factor R_{ia} also has a mean return approaching to zero, about 0.07% ($t = 0.49$) per month. And the addition of expected growth factor R_{eg} is likewise useless; it only yields a tiny premium of 0.18% ($t = 0.85$). However, we find the profitability factor R_{roe} is much stronger than RMW, with an average return of 0.87% ($t = 4.11$) versus 0.24% ($t = 1.11$). Although both of them are return spreads sorted by firms' ROE, R_{roe} uses quarterly data, and its value is updated every three months. That indicates stocks in the A-share market are more sensitive to recent changes of earnings, and firms with high profitability indeed generate substantially higher stock returns, which is exactly the main conclusion of Jiang et al. [13].

This fact can also be seen from anomaly-based factors. We find the maximum Sharpe ratio for Stambaugh and Yuan's mispricing factors is only about 0.73 annually, the lowest in Table 2. As we have listed in the Supplementary Appendix, MGMT includes various investment measures, whereas PERF includes various profitability measures. Thus,

Table 2: Summary statistics of pricing factors. This table shows means, standard deviations, t-statistics, skewness, kurtosis, and Jarque–Bera statistic of each pricing factor and the maximum Sharpe ratio achievable from each factor model. The sample period is from July 2000 to December 2019. The first to last row correspond to the Fama–French 6-factor model [10], the Hou–Xue–Zhang q5-factor model [2], the Stambaugh–Yuan 4-factor model [21], the Daniel–Hirshleifer–Sun 3-factor model [22], and the Liu–Stambaugh–Yuan 4-factor model [15].

Pricing factors		Mean (%)	Std. dev (%)	t value	Skewness	Kurtosis	J-B stat.	Sharpe ratio
FF6	MKT	0.50	8.10	0.94	−0.23	4.06	19.93	1.25
	SMB	0.35	3.78	1.40	−0.29	4.60	28.84	
	HML	0.37	4.01	1.43	−0.18	7.02	163.34	
	RMW	0.24	3.28	1.11	0.25	5.85	83.45	
	CMA	−0.10	2.47	−0.61	0.19	5.48	62.98	
	UMD	0.23	4.05	0.86	−0.47	4.38	27.71	
Q-5	MKT	0.50	8.10	0.94	−0.23	4.06	19.93	1.72
	R_{me}	0.55	3.39	2.47	−0.21	5.38	58.31	
	R_{ia}	0.07	2.05	0.49	0.22	3.62	5.83	
	R_{roe}	0.87	3.25	4.11	0.20	4.34	19.68	
	R_{eg}	0.18	3.32	0.85	0.07	6.78	143.20	
SY-5	MKT	0.50	8.10	0.94	−0.23	4.06	19.93	0.73
	SMB	0.44	4.50	1.50	−0.62	9.35	418.85	
	MGMT	0.01	3.06	0.07	0.32	9.89	479.14	
	PERF	0.63	4.53	2.12	0.36	4.36	23.72	
DHS-3	MKT	0.50	8.10	0.94	−0.23	4.06	19.93	1.23
	PEAD	0.98	3.01	4.97	−0.13	3.46	2.80	
	FIN	0.12	2.17	0.84	−0.48	9.42	420.95	
LSY4	MKT	0.50	8.10	0.94	−0.23	4.06	19.93	2.16
	SMB	0.52	4.01	2.00	0.06	4.77	20.31	
	VMG	1.09	3.29	5.04	0.12	4.45	7.04	
	PMO	0.81	3.35	3.68	−0.82	9.78	401.52	

it is not surprising that the average return of MGMT is nearly zero (0.01% and $t = 0.07$), but the average return of PERF is significantly positive (0.63% and $t = 2.12$). Likewise, we see for the DHS-3, the postearnings-announcement drift (PEAD) factor, which is constructed by anomalies reflecting earnings' surprises, earns on average 0.98% ($t = 4.97$) per month. However, the financing factor FIN only delivers a premium of 0.12% ($t = 0.84$). At last, LSY-4 factors show the most significant risk premiums, with a maximum Sharpe ratio of 2.16. VMG effectively reflects the value effect; it earns an average return of 1.09% ($t = 5.04$), the highest among all factors. The sentiment factor PMO is also profitable and yields on average 0.81% ($t = 3.68$).

Interestingly, Table 2 also shows each pricing factor and illustrates the characteristic of leptokurtosis and fat tail, since all values of kurtosis are greater than 3. This fact is consistent with Jarque–Bara statistics, since we find the null hypotheses of the time series following normal distribution are rejected for all factors at the 1% significance level, except for R_{ia} and PEAD.

Table 3 reports the correlation matrix of all pricing factors. To make it clear, SMB in Table 3 corresponds to the size factor in LSY4. We are mainly interested about how LSY4 factors are related to the others. Not surprisingly, SMB is nearly equivalent to the size factor in the q-factor model, with a correlation of 0.92. The value factor VMG is highly correlated with profitability factors. It has a correlation of 0.68 with RMW and 0.81 with R_{roe}. It is because VMG is derived from earnings-to-price ratio, which is a comprehensive characteristic reflecting the value and expected earnings' growth of a stock. The conventional value factor HML is not closely related to VMG. For sentiment factor PMO, we find few factors are highly correlated with it. The closest factor is HML, only with a correlation of −0.27. That indicates other factors cannot subsume the premium of turnover effect, and thus, it is reasonable to add it into the model to price A-share stocks.

3.2. Explaining Anomalies. In Table 4, we regress 104 anomaly returns on various factor models, to investigate the performance of factor models in explaining anomalies. Columns 1–3 report the mean return, standard deviation, and t-statistic for each anomaly over the period from July 2000 to December 2019. For these unadjusted raw returns, we find 42 of 104 anomalies are significant (shown in bold). That indicates mispricing is a little severe in the A-share market, and strategies based on mispricing can earn significant profits. Next, columns 4–8 correspond to t-statistics of alphas controlling for various pricing factors, given Newey–West adjustments over 12-month lag. Surprisingly, it shows for models proposed from the US market, which are displayed in columns 4–7, are nearly useless: there are still 49, 43, 55, and 39 alphas with t-statistics higher than 1.96, respectively. In other words, significant anomalies can hardly be eliminated by these four models. In comparison, LSY4 has great effect on explaining these anomalies, with only six of 104 significant alphas, including Dimson beta (**betad**), one measure of seasonality (**rn25**), two measures of idiosyncratic skewness per different models (**isff** and **isq**), cumulative abnormal returns around earnings announcement dates (**abr**), and 6-month residual momentum (**e6**).

Table 3: Correlation matrix of pricing factors. This table displays the correlation matrix of mainstream pricing factors. The sample period is from July 2000 to December 2019. MKT, SMB, VMG, and PMO are market, size, value, and turnover factors in the Liu–Stambaugh–Yuan 4-factor model [15]. HML, RMW, CMA, and UMD are value, profitability, investment, and momentum factors in the Fama–French 6-factor model [10]. R_{me}, R_{ia}, R_{roe}, and R_{eg} are size, investment, profitability, and expected growth factor in the Hou–Xue–Zhang q5-factor model [2]. MGMT and PERF are two mispricing factors in the Stambaugh–Yuan 4-factor model [21]. PEAD and FIN are two long-short horizon factors in the Daniel–Hirshleifer–Sun 3-factor model [22].

	SMB	MKT	VMG	HML	RMW	CMA	UMD	PMO	R_{me}	R_{ia}	R_{roe}	R_{eg}	MGMT	PERF	FIN	PEAD
SMB	1.00															
MKT	0.23	1.00														
VMG	−0.63	−0.31	1.00													
HML	−0.41	0.04	0.38	1.00												
RMW	−0.62	−0.35	0.68	−0.10	1.00											
CMA	0.21	0.01	−0.12	0.54	−0.59	1.00										
UMD	−0.37	−0.19	0.28	−0.18	0.49	−0.30	1.00									
PMO	0.04	−0.23	−0.01	−0.27	0.11	−0.18	0.08	1.00								
R_{me}	0.92	0.21	−0.62	−0.52	−0.49	0.05	−0.15	0.08	1.00							
R_{ia}	0.03	−0.07	0.12	0.55	−0.37	0.89	−0.15	−0.09	−0.08	1.00						
R_{roe}	−0.54	−0.32	0.81	−0.01	0.79	−0.34	0.57	0.08	−0.40	−0.05	1.00					
R_{eg}	−0.33	0.00	0.34	0.85	−0.18	0.58	−0.24	−0.21	−0.48	0.60	−0.03	1.00				
MGMT	0.17	−0.01	−0.08	0.34	−0.40	0.71	−0.20	−0.03	0.08	0.64	−0.19	0.43	1.00			
PERF	−0.31	−0.31	0.46	−0.34	0.69	−0.43	0.73	0.17	−0.11	−0.24	0.75	−0.38	−0.19	1.00		
FIN	−0.11	0.05	0.13	0.32	−0.02	0.29	−0.05	−0.10	−0.14	0.36	0.07	0.32	0.34	−0.06	1.00	
PEAD	−0.26	−0.24	0.37	−0.05	0.39	−0.12	0.49	0.02	−0.12	0.04	0.52	−0.03	−0.03	0.50	0.08	1.00

Table 5 further reports other statistics reflecting explaining power of pricing models. Again, we use these 104 anomalies as test assets. Average alpha is the absolute mean of 104 alphas with respect to each model, and average t is the absolute mean of corresponding t-statistics. Following Shanken [23], pricing errors are computed as

$$\Delta = \alpha\prime\Sigma^{-1}\alpha, \tag{1}$$

where α is the vector of alpha and Σ is the covariance matrix of residuals for each factor model. We also perform the well-known GRS test proposed by Shanken et al. [24] to show whether alphas of 104 anomalies are jointly zero. Generally speaking, it seems that LSY4 has great advantage over other counterparts. It generates the smallest average absolute alpha of 0.31% as well as average absolute t-statistic of only 1.00. Its pricing error is the second lowest with the value of 3.65, slightly higher than the q-factor model (3.61). As for GRS test, we find all models reject the null hypothesis that alphas are jointly zero. It is not surprising since, as Table 4 shows, no factor model can accommodate all anomalies. However, LSY4 derives the smallest GRS statistic.

3.3. Spanning Regressions. So far, we have shown the strong pricing power of the LSY4 model. Another key question is whether LSY4 can price factors of other models, which matters for model comparison. In this section, we regress FF6 factors, q5 factors, SY4 factors, and DHS-3 factors on LSY4 factors in turn, to show whether LSY4 factors can explain their risk premiums. Table 6 displays the results. From the Durbin–Watson statistics in the last column, we realize we cannot reject the hypothesis that residuals may suffer from autocorrelations. Thus, we adjust standard errors of regression coefficients using the Newey and West [25] correction with three lags. We find, except for PEAD and R_{me}, other alphas are insignificant with the specification of the LSY4 model, indicating corresponding factors are accommodated by LSY4 factors. For example, the alpha of powerful q-factor R_{roe} is only 0.07% ($t = 0.48$), and the investment premium R_{ia} is further reduced to −0.06% ($t = -0.41$). On the contrary, PEAD survives from the LSY4 model with a highly significant alpha of 0.71% ($t = 3.35$). This fact consists with Table 4, since, as we have illustrated previously, PEAD is constructed with four-day cumulative abnormal return (**abr**), which cannot be explained by LSY4 factors. We also report two GRS statistics. We can hardly reject the null that all alphas are jointly equal to zero (GRS = 2.23), but once excluding PEAD, the GRS statistic declines to 1.47 immediately.

4. Exploring New Factors

In this section, we will explore whether there any significant anomalies can serve as latent pricing factors which can be added into the current LSY4 model. We mainly depend on the double-selection LASSO approach proposed by Feng et al. [11] and evaluate marginal contributions of a series of new factors beyond existing factors. We start with a brief introduction of this methodology and then report our findings.

4.1. Double-Selection LASSO. In asset pricing theory, the ability of a given factor to explain asset prices is reflected in its stochastic discount factor (SDF) loadings, as discussed by Cochrane [26] and Ferson [27]. Typically, we assume the stochastic discount factor has the following linear specification:

$$m_t := \gamma_0^{-1} - \gamma_0^{-1}\lambda_v' v_t := \gamma_0^{-1}\left(1 - \lambda_g' g_t - \lambda_h' h_t\right), \tag{2}$$

Table 4: Anomalies by univariate characteristic sorting. This table displays the return properties of anomalies constructed by sorting on each of the 104 characteristics over the whole sample of July 2000 to December 2019. With monthly rebalancing, we sort all stocks by individual characteristics into 10 deciles, form value-weighted portfolio within each decile, and build a long-short spread portfolio using decile ten and one. For each portfolio, we report the monthly return means, standard deviations, and *t*-statistics in columns labeled "Mean," "Std. dev," and "t-Raw." We also use various factor models to explain these anomalies and report corresponding *t*-statistics of alphas with Newey–West adjustment over 12 monthly lags, including the Fama–French 6-factor model [10], the Hou–Xue–Zhang q5-factor model [2], the Stambaugh–Yuan 4-factor model [21], the Daniel–Hirshleifer–Sun 3-factor model [22], and the Liu–Stambaugh–Yuan 4-factor model [15], in columns "t-FF6," "t-Q5," "t-SY4," "t-DHS3," and "t-LSY4," respectively. Bold types denote values that are statistically significant at 95% confidence.

Anomaly	(1) Mean (%)	(2) Std. dev (%)	(3) t-Raw	(4) t-FF6	(5) t-Q5	(6) t-SY4	(7) t-DHS3	(8) t-LSY4
abr	0.85	4.74	**2.92**	**2.75**	1.89	**2.97**	0.09	**2.41**
abtur	1.12	5.91	**3.92**	**3.98**	**3.5**	**3.15**	**5.35**	0.16
age	0.01	4.07	0.05	0.77	0.52	0.8	0.92	0.5
alaq	0.86	4.21	**3.34**	**3.59**	1.71	**2.2**	**2.55**	1.33
almq	0.71	6.40	**2**	**2.78**	**5.41**	**5.74**	**2.96**	1.24
ami	0.81	6.63	1.62	**3.89**	**1.96**	1.27	**2.5**	0.59
amq	0.52	7.34	1.23	1.46	**4.17**	**5.1**	**3.28**	0.4
atoq	0.90	3.78	**2.96**	**4.03**	**2.24**	**2.66**	**2.4**	1.69
beta	0.53	6.39	1.46	1.91	0.92	1.82	0.27	0.14
beta_	0.26	6.57	0.67	1.86	0.98	1.78	0.31	1.27
betad	1.05	5.91	**2.61**	**2.19**	**2.6**	**3.05**	**3.09**	**2.57**
betafp	0.64	6.98	1.68	**2.28**	0.95	1.69	0.47	0
blq	0.05	5.11	0.14	0.53	0.82	**2.13**	0.35	0.32
bmj	0.21	4.87	0.62	0.44	1.6	**2.78**	1.07	0.32
bmq	0.59	7.22	1.43	**2.22**	**4.7**	**5.21**	**3.43**	0.36
cdi	0.04	3.44	0.18	0.09	1.15	0.91	0.64	0.69
cei	0.22	4.16	0.89	1.24	0.31	**2.2**	0.24	0.27
claq	0.15	4.51	0.56	0.28	0.09	1.17	0.72	0.29
cop	0.00	4.77	0	1.14	0.47	0.56	0.31	0.76
cpq	0.66	4.00	**2.32**	**2.34**	**2.61**	**3.51**	**3.35**	0.45
cs	0.15	5.30	0.5	0.17	0.27	1.54	0.09	1.11
cta	0.54	5.07	**2**	**2.29**	0.5	0.62	0.66	0.03
ctoq	0.63	4.08	**2.64**	**3.31**	1.52	**2.27**	1.74	1.08
cvd	0.95	4.95	**2.95**	**3.68**	1.82	**2.25**	1.42	1.16
cvt	0.65	4.98	1.8	**2.23**	1.28	1.02	0.52	0.94
dcoa	0.12	3.72	0.57	1.06	0.52	1	1.58	0.17
dcol	0.02	2.96	0.1	0.07	0.42	0.53	0.45	0.27
dgs	0.02	3.11	0.14	1.65	1.57	0.29	1.35	1.14
dmq	0.24	5.98	0.61	0.19	**2.43**	**3.3**	1.79	0.16
dnca	0.26	3.06	1.32	1.44	0.58	1.87	**2.08**	0.13
dnco	0.14	2.70	0.77	0.35	0.64	1.3	1.37	0.45
dnoa	0.23	3.35	1.04	0.96	0.72	1.48	1.17	0.17
dpia	0.07	3.37	0.36	0.29	0.7	0.25	1.22	1.22
droa	0.98	4.24	**3.96**	**5.66**	1.18	**3.39**	**2.48**	0.83
droe	1.10	4.15	**4.98**	**6.32**	1.34	**4.46**	**3.11**	1.12
dsa	0.37	2.62	**2.19**	**2.26**	1.58	1.83	1.58	1.25
dsi	0.16	2.41	1.02	0.63	0.33	0.76	0.69	0.28
dss	0.17	2.71	0.88	1.81	**2**	0.84	1.08	1.86
dtv	0.84	6.12	1.79	**3.61**	1.39	1.55	**2.6**	0.08
dwc	0.09	3.03	0.55	0.43	0.51	0.73	0.11	0.33
e11	0.40	4.72	1.07	0.84	0.2	0.68	0.28	1.37
e6	0.47	4.80	1.52	1.52	0.92	1.23	0.38	**2.48**
ebpq	1.15	7.53	**2.88**	**3.22**	**5.4**	**5**	**3.32**	1.8
emq	1.34	6.21	**3.5**	**5.02**	**3.17**	**5.41**	**2.41**	0.5
epq	1.27	6.07	**3.53**	**4.88**	**2.84**	**5.22**	**2.48**	0.32
etr	0.28	3.21	1.43	0.9	0.6	0.95	0.42	0.23
fpq	0.48	6.11	1.05	1.41	1.22	0.39	0.94	0.35
glaq	1.13	5.63	**2.86**	**4.18**	1.75	**3.68**	1.68	0.64
gpa	0.60	5.81	1.69	**3.01**	1.7	1.61	0.63	1.37
ia	0.12	3.77	0.62	0.8	0.47	1.26	1.23	0.38
iaq	0.31	4.20	1.37	**2.82**	0.91	0.27	0.02	1.37

Table 4: Continued.

Anomaly	(1) Mean (%)	(2) Std. dev (%)	(3) t-Raw	(4) t-FF6	(5) t-Q5	(6) t-SY4	(7) t-DHS3	(8) t-LSY4
ir	0.23	5.94	0.69	0.23	**3.23**	**2.14**	**2.49**	1.89
isc	0.31	3.68	1.38	1.06	**2.07**	1.02	0.95	1.91
isff	0.25	3.43	1.06	1.13	1.49	1.56	0.92	**2.56**
isq	0.27	3.60	1.2	1.43	**2.18**	**2.24**	0.91	**2.51**
iv	0.74	6.22	**2.01**	**3.36**	**3.28**	**5.17**	**2.1**	1.56
ivc	1.06	5.86	**2.97**	**3.23**	**3.22**	**4.28**	**2.91**	0.08
ivff	1.19	5.70	**3.17**	**3.7**	**4.05**	**4.47**	**2.94**	0.56
ivg	0.09	2.99	0.62	0.9	0.42	1.38	1.29	0.82
ivq	1.15	5.79	**2.92**	**3.22**	**3.59**	**4.14**	**2.7**	0.61
kzq	0.06	4.41	0.22	0.64	0.18	0.07	0.4	0.31
mdr	0.73	5.51	**2.32**	**2.83**	**2.24**	**2.93**	**2.08**	1.44
me	0.30	6.56	0.66	0.8	**2.69**	0.81	1.23	0.24
ndpq	0.42	6.25	1.12	1.13	**3.13**	**3.04**	1.52	1.59
noa	0.50	3.68	**2.72**	**3.89**	**2.43**	**2.56**	1.92	0.51
nopq	0.40	4.93	1.36	1.87	**2.4**	**4.41**	**2.03**	0.01
nsi	0.17	3.31	0.94	0.35	0.02	1.4	1.91	0.11
ocpq	0.76	5.27	**3.02**	**3.4**	**4.08**	**5.42**	**3.04**	0.19
olaq	0.75	6.27	1.71	**3.15**	0.5	**3.33**	0.59	0.19
oleq	0.89	6.31	**2.12**	**4.01**	0.84	**3.51**	0.78	0.13
olq	0.32	3.64	1.38	1.35	0.83	0.85	1.21	0.2
opa	0.11	5.84	0.23	0.26	0.62	0.75	0.74	0.29
ope	0.25	6.00	0.61	0.86	1.3	1.81	0.43	1.13
opq	0.36	5.64	1.16	1.31	**2.53**	**4.06**	1.66	0.24
oq	0.68	4.73	**1.99**	**4.49**	**2.67**	**2.69**	1.2	1.28
pmq	0.87	5.19	**2.58**	**4.64**	1.3	**4.11**	1.43	1.29
poa	0.12	3.61	0.81	0.69	0.32	1.39	0	0.44
pps	0.77	7.20	1.74	1.56	**3.7**	**5.11**	**4.64**	0.73
Ra25	0.66	4.19	**2.69**	**3.03**	**2.7**	**2.9**	**2.58**	**2.77**
Rn25	0.77	6.61	**2.21**	1.25	**2.16**	**3.48**	**2.89**	1.85
r11	0.68	6.64	1.39	1.41	1.08	0.28	0.97	0.88
r1a	0.68	4.64	**2.19**	**2.81**	1.19	1.19	1.38	1.11
r1n	0.07	6.76	0.14	1.17	**2.25**	1.28	**2.21**	0.96
r6	0.25	6.49	0.56	0.1	0.74	0.32	1.67	1.91
rev	0.40	5.38	1.44	0.97	**2.04**	**2.29**	**2.4**	1.8
rnaq	1.28	5.51	**3.89**	**5.04**	1.56	**5.83**	**2.2**	0.62
roa	1.34	5.65	**3.5**	**4.97**	**2.38**	**5.51**	**1.98**	1.4
roe	1.51	5.71	**4.41**	**5.62**	**2.29**	**6.29**	**2.35**	0.98
rs	0.57	3.61	**2.05**	1.46	0.18	1.53	0.39	0.3
sgq	0.96	3.70	**4.28**	**5.2**	1.74	**3.26**	**2.64**	1.33
size	0.66	7.06	1.28	**3.41**	1.79	0.9	**2.39**	1.51
spq	1.16	6.55	**2.59**	**3.12**	**3.1**	**5.3**	**3.11**	0.52
sr	0.11	4.68	0.33	0.56	0.4	0.34	0.12	0.1
srev	0.91	6.29	**2.68**	**3.14**	**2.81**	**2.35**	**3.4**	0.5
sue	1.28	3.81	**5.5**	**6.09**	**2.05**	**4.95**	**3.16**	1.64
tail	0.33	3.75	1.45	1.88	0.05	0.89	0.84	0.7
tanq	0.23	3.89	1.01	1.91	0.72	0.41	0.39	0.3
tbiq	0.30	3.14	1.28	1.56	1.54	**2.04**	1.21	1.1
tes	1.00	3.54	**4.77**	**6.03**	**2.27**	**4.07**	**2.73**	1.7
ts	0.64	3.90	**3.2**	**2.32**	**2.07**	**2.45**	**3.23**	0.74
tur	0.61	6.88	1.64	**3.9**	**3.6**	**5.04**	1.49	1.52
tv	0.80	6.41	**2.06**	**2.52**	**2.15**	**3.08**	1.54	0.33
wwq	0.07	7.12	0.16	0.45	**2.77**	**2.79**	0.21	0.08
zq	0.05	6.47	0.13	0.7	1.51	**2.79**	0.98	0.12

where γ_0 is the zero-beta rate, h_t is a $p \times 1$ vector of benchmark factors, g_t is a $q \times 1$ vector of new factors needed to be tested, and λ_g and λ_h are so-called SDF loadings. Given a vector of test asset returns r_t, we may want to evaluate the explanatory power of g_t while controlling for other benchmark factors h_t. According to definition of SDF, formula (2) can be expressed as the relationship between expected returns and covariances with factors h_t and g_t:

Table 5: Comparison of models in explaining anomalies. This table reports some statistics derived from using various factor models to explain 104 anomalies. We display results of the Fama–French 6-factor model [10], the Hou–Xue–Zhang q5-factor model [2], the Stambaugh–Yuan 4-factor model [21], the Daniel–Hirshleifer–Sun 3-factor model [22], and the Liu–Stambaugh–Yuan 4-factor model [15], in columns "FF6," "Q5," "SY4," "DHS3," and "LSY4," respectively. Average alpha and Average t are the absolute mean of alphas and corresponding t-statistics over 104 anomalies. Δ is the cross-sectional pricing error with respect to each model, which is computed as formula (1). GRS is the statistic of well-known GRS test proposed by Shanken et al. [24].

	FF6	Q5	SY4	DHS3	LSY4
Average alpha (%)	0.55	0.46	0.62	0.52	0.31
Average t	2.21	1.64	2.29	1.55	1
Δ	3.91	3.61	4.29	3.93	3.65
GRS	4.26	3.59	5.11	4.33	3.28

Table 6: Spanning regression. This table reports results of regressing FF6 factors, q5 factors, SY4 factors, and DHS-3 factors on LSY4 factors. MKT, SMB, VMG, and PMO are market, size, value, and turnover factors in the Liu–Stambaugh–Yuan 4-factor model [15]. HML, RMW, CMA, and UMD are value, profitability, investment, and momentum factors in the Fama–French 6-factor model [10]. R_{me}, R_{ia}, R_{roe}, and R_{eg} are size, investment, profitability, and expected growth factor in the Hou–Xue–Zhang q5-factor model [2]. MGMT and PERF are two mispricing factors in the Stambaugh–Yuan 4-factor model [21]. PEAD and FIN are two long-short horizon factors in the Daniel–Hirshleifer–Sun 3-factor model [22]. For each factor, we report its coefficient estimate and corresponding t-statistic (given in parentheses). R^2 is adjusted R-squared of each regression, and D-W Stat. is corresponding Durbin–Watson statistic of each regression.

	Alpha	Mktrf	SMB	VMG	PMO	R^2	D-W stat.
HML	0.39	0.06**	−0.27***	0.29***	−0.28***	0.26	2
	−1.48	−2	(−3.79)	−3.25	(−3.98)		
RMW	−0.14	−0.05**	−0.27***	0.43***	0.10**	0.55	1.85
	(−0.85)	(−2.52)	(−5.74)	−7.55	−2.2		
CMA	−0.03	−0.03	0.14***	−0.01	−0.16***	0.07	1.85
	(−0.16)	(−1.30)	−2.85	(−0.07)	(−3.26)		
UMD	0.28	−0.04	−0.33***	0.06	0.09	0.14	1.93
	−0.99	(−1.17)	(−4.22)	−0.58	−1.23		
R_{me}	0.19*	0	0.75***	−0.06*	0.04	0.85	2.11
	−1.9	(−0.05)	−27.25	(−1.85)	−1.37		
R_{ia}	−0.06	−0.02	0.10**	0.13**	−0.08*	0.03	2.13
	(−0.41)	(−1.13)	−2.24	−2.48	(−1.79)		
R_{roe}	0.07	−0.02	−0.05	0.74***	0.07*	0.66	1.74**
	−0.48	(−1.43)	(−1.24)	−14.87	−1.88		
R_{eg}	0.1	0.03	−0.16**	0.25***	−0.18***	0.17	1.88
	−0.42	−1.27	(−2.42)	−3.11	(−2.93)		
MGMT	−0.11	−0.02	0.16**	0.03	−0.05	0.02	2.08
	(−0.47)	(−0.85)	−2.52	−0.43	(−0.78)		
PERF	−0.03	−0.09**	−0.05	0.52***	0.18**	0.25	1.98
	(−0.11)	(−2.51)	(−0.59)	−5.1	−2.32		
PEAD	0.72***	−0.05**	0.03	0.28***	0	0.14	1.96
	−3.35	(−2.09)	(−0.45)	−3.86	(−0.04)		
FIN	0.08	0.02	−0.03	0.08	−0.05	0.02	1.95
	−0.51	−1.14	(−0.66)	−1.35	(−1.20)		

$$E(r_t) = \iota_n \gamma_0 + C_v \lambda_v = \iota_n \gamma_0 + C_g \lambda_g + C_h \lambda_h, \tag{3}$$

where $E(r_t)$ is a $n \times 1$ vector of expected returns, ι_n is a $n \times 1$ vector of ones, C_g is the covariance matrix of r_t and g_t, C_h is the covariance matrix of r_t and h_t. Intuitively, SDF loadings λ_g and λ_h measure how expected return is correlated to its covariances with corresponding factors. However, since we mainly focus on the significance of λ_g, considering dependence between h_t and g_t, it is worthwhile to project C_g on C_h:

$$C_g = \iota_n \xi' + C_h \chi' + C_e, \tag{4}$$

where ξ and χ are coefficient matrices and C_e is the residual matrix.

Because there are a number of pricing factors in the real world (Fama–French factors, Hou–Xue–Zhang q-factors, etc.), we need to determine the components of h_t before we test g_t. It requires us to select factors with the most explanatory power to serve as the benchmark. Least Absolute Shrinkage Selection Operator (LASSO) regression is one of approaches to help us reduce the number of benchmark factors. LASSO is a modification of the RIDGE regression method where the constraint in the Lagrangian is the absolute values of the parameters and where the Lagrange multiplier governs the shrinkage Eric Ghysels [28]. Specifically, we run a cross-sectional LASSO regression of mean returns on the covariance between returns and all benchmark factors:

$$\min_{\gamma,\lambda}\left\{n^{-1}\left(\overline{r}-\iota_n\gamma-\widehat{C_h}\lambda\right)^2+\tau_0 n^{-1}|\lambda|_1\right\}, \tag{5}$$

where $\overline{r}$ is the time series mean of all test asset returns, $\widehat{C_h}$ is the covariance matrix of $\overline{r}$ and demeaned h_t, and τ_0 is the LASSO penalty parameter. We usually rely on cross validation to choose τ_0 and solve the Lasso coefficients λ with least angle algorithm [29] or coordinate descent algorithm [30]. LASSO regression will shrink some λ to zero and generate a sparse function specification. We denote the indices of remaining benchmark factors as set $\{\widehat{I}_1\}$.

However, in finite sample, traditional LASSO regression does not have the so-called "oracle property," that is, we cannot select the true model with the probability of one. It indicates the result given by previous LASSO regression suffers from severe omitted variable bias, which means our inference about loadings γ may be erroneous. Thus, performing a single LASSO step is not enough to make sure we can control all useful benchmark factors for testing g_t. We need to conduct another LASSO regression to search for factors which are omitted in the first step but still play a role in explaining test asset returns:

$$\min_{\xi_j,\chi_j}\left\{n^{-1}\left(\widehat{C_{g_i,\cdot}}-\iota_n\xi_j-\widehat{C_h}\chi_{j,\cdot}\right)^2+\tau_j n^{-1}\left|\chi_{j,\cdot}\right|_1\right\}, \tag{6}$$

where $\widehat{C_g}$ is the covariance matrix of $\overline{r}$ and demeaned g_t. Intuitively, for each new factor g_j, we identify the corresponding benchmark factors whose covariance with mean returns are closest to covariance between mean returns and factor g_j. Denote the union of all indices of selected benchmark factors as set $\widehat{I}_2=\cup_{j=1}^{d}\widehat{I_{2,j}}$.

So far, we have performed two rounds of LASSO regressions to select sufficient benchmark factors. Now, we can evaluate the marginal importance of new factors g_t with following cross-sectional regression. We are concerned about SDF loadings λ_g, which tell us whether the corresponding factors g_t are significant to price test asset returns:

$$\left(\widehat{\gamma_0},\widehat{\lambda_g},\widehat{\lambda_h}\right)=\underset{\gamma_0,\lambda_g,\lambda_h}{\arg\min}\left\{\left(\overline{r}-\iota_n\gamma_0-\widehat{C_g}\lambda_g-\widehat{C_h}\lambda_h\right)^2:\ \lambda_{h,j}=0,\forall j\notin\widehat{I}=\widehat{I}_1\cup\widehat{I}_2\right\}. \tag{7}$$

4.2. Factors and Test Portfolios. We construct our factor library based on previous discussion. Note that the aim of this section is to explore explanatory power of some new factors beyond a series of benchmark factors. Thus, the vector h_t includes all pricing factors appearing in Table 2, which come from mainstream factor models in latest literature. In addition, we also add another two factors. The first one is QMJ, a Quality-Minus-Junk factor proposed by Asness et al. [31]. The other is a liquidity factor AMI, which is constructed by Amihud's measure. Both factors are confirmed to have significant return spreads in A-share market, see details in Kang and Zhang [32] and Hu and Gu [33]. Our main empirical results are derived from these 18 benchmark factors in total. Someone may argue the volume of our factor library is too small to evaluate the contributions of new latent factors, especially compared to related literature in the US market. We respond to this doubt from two aspects. First, unlike the US market, there is no clear timeline of published date for newly discovered factors in China. Using the A-share market data, related studies mainly focus on testing the significance of factors discovered in developed countries. The purpose of our paper is to investigate whether we can add new factors into the baseline LSY4 model; hence, it is natural to denote current pricing factors as benchmark factors. Second, as a robustness check, we also expand the factor library to accommodate all the other anomalies in Table 1 except for anomalies we are going to test. However, we find our empirical result changes a little, indicating current pricing factors are powerful enough to subsume the main part of cross-sectional test asset returns.

For vector of new factors g_t, we firstly consider significant anomalies unexplained by the LSY4 model. Table 4 shows only 6 anomalies can yield significant LSY4 alphas. Among them, cumulative abnormal returns around earnings' announcement dates (**abr**), Dimson beta (**betad**), and one measure of seasonality (**ra25**) have significant return spreads. We suppose these three factors are proper candidates of latent new factors. Besides, we link to the work of Qiao [8], who replicates 231 anomalies in the A-share market. We test significant anomalies found in this paper, including quarterly assets-to-market (**amq**), changes in return on assets (**droa**), changes in return on equity (**droe**), dollar trading volume (**dtv**), idiosyncratic volatility per the FF3 model (**ivff**), long-term reversal (**lrev**), price momentum with prior 24 month returns (**m24**), maximum daily return (**mdr**), R&D expense to market equity (**rdm**), quarterly sales growth (**sgq**), quarterly sales-to-price ratio (**spq**), short-term reversal (**srev**), total volatility (**tv**), variation of share dollar trading volume (**vdtv**), and variation of share turnover (**vturn**). Note that some of these anomalies have been investigated by us in the previous section. Although they may generate insignificant return spreads according to Table 4, we still add them into the set of g_t since it will contribute to literature in this field. Certainly, we reconstruct these anomalies following the standard procedure of Fama and French [18] in order to compare with other factors. For each anomaly, we split stocks into 3 portfolios according to 30 and 70 percentiles of corresponding characteristic. Note that high rankings are associated with higher future average returns. Meanwhile, we also split stocks into two groups according to the median size. Factor returns are spreads between the average of two high-characteristic portfolios and the average of two low-characteristic portfolios.

We use 650 portfolios as test assets based on our factor library. Following the procedure of Fama and French [18],

stocks are independently sorted into 5×5 portfolios by the market value and one firm characteristic. In order to avoid imposing a specific factor structure on our test portfolios, we choose as many characteristics as possible, all of which are used to construct factors in h_t and g_t. The sorting characteristics include cumulative abnormal returns around earnings announcement dates, turnover, Amihud's liquidity measure, quarterly assets-to-market, Dimson beta, book-to-market ratio, changes in return on assets, changes in return on equity, dollar trading volume, expected investment growth, earnings-to-price ratio, investment growth, idiosyncratic volatility per the FF3 model, long-term reversal, price momentum with prior 24 month returns, maximum daily return, 11-month momentum, seasonality, R&D expense to market equity, return on equity, quarterly sales growth, quarterly sales-to-price ratio, short-term reversal, total volatility, variation of share dollar trading volume, and variation of share turnover. In robustness check, we also form another 156 3×2 portfolios.

4.3. Empirical Results. Now, we apply the double-selection LASSO approach to A-share market. We start by reporting summary statistics of factors in h_t and g_t, which are not introduced during previous discussion. Table 7 displays means, t-statistics, annualized Sharpe ratios, skewness, kurtosis, and Jarque–Bera statistics of totally 20 factors. To some extent, we find all these factors have significant return spreads except for long-term reversal (**lrev**) and price momentum with prior 24 month returns (**m24**). It is a little inconsistent with findings of Qiao [8]. Compared to their work, we exclude the smallest 30% stocks while constructing factors. On the contrary, Table 7 indicates our selection for tested factors is quite reasonable, since most of them yield extremely high t-statistics. Similar to discovered pricing factors we discussed before, these alternative factors also have leptokurtosis and fat tails.

We are interested in whether those factors with significant risk premia contribute to explaining cross-sectional returns. As emphasized by Feng et al. [11] and Cochrane [26], the key to answering this question is to compare factors' SDF loadings. A positive SDF loading indicates the high value of this factor captures good states of the world, and vice versa. Columns 1-2 of Table 8 report our SDF loading estimates as well as t-statistics using double LASSO approach. The two steps select totally 4 benchmark factors, including expected growth (R_{eg}), quality-minus-junk (QMJ), liquidity (AMI), and market excess return (MKT), where the first three factors are derived from the first step and the last one is derived from the second step. One may argue that why these four factors can serve as the benchmark instead of LSY4 factors. It needs to be emphasized that the members of selected benchmark factors highly depend on test assets. As for returns of our 650 test portfolios, the factors above show greater explanative ability than LSY4 factors. This inference is obtained from Table 8. In columns 1-2, we find 8 out of 18 tested factors have significant SDF loadings with the significance level of 0.05, most of which are categorized into trading frictions. In comparison, when we use LSY4 factors as controls, as shown in columns 3-4, the number of significant tested factors rises to 12. Since we do not select benchmark factors optimally, the estimate of SDF loadings suffers from omitted variable bias and is possibly imprecise. Furthermore, in columns 5-6, we add all 18 benchmark factors simultaneously into the regression to estimate SDF loadings. The result is similar to using LSY4 factors, with 11 significant factors. Selecting proper benchmark factors is important to evaluate the contributions of new factors. In fact, as pointed out by Feng et al. [11], controlling all available factors is an unbiased but inefficient approach, and the inefficiency becomes especially severe when the dimension of factor space is large. We will show this in Table 9.

We next show robustness of our results since double LASSO approach relies on two inputs: returns of test portfolios and the set of benchmark factors. In Table 10, we re-estimate SDF loadings by changing test portfolios and expanding the dimension of h_t. Firstly, we replace 650 test portfolios with fewer 156 portfolios. Then, we test extra benchmark factors which are anomalies presented in Table 4 but excluding the ones we need to test. Certainly, we follow the standard procedures to convert anomaly returns to factor returns, that is, we use spreads between two bivariate portfolios sorted by size and a specific characteristic instead of long-short decile spreads. In this way, we obtain another 81 benchmark factors and the total number increases to 99. Intuitively, we want to explore whether these tested factors can distinctively explain cross-sectional returns beyond the rest of characteristic-based anomalies. For convenience, Table 10 reports our baseline results in the left. Generally speaking, we find the results derived from 99 benchmark factors are similar to using only 18 benchmark factors. That indicates the original 18 factors can dominate the others when explaining variation of asset returns. On the contrary, it shows SDF loadings are less significant when we use only 156 portfolios. As we have mentioned before, the selection of test assets will influence the SDF loading estimates. However, we still find some tested factors kept significant over different cases, including cumulative abnormal returns around earnings' announcement dates (**abr**), Dimson beta (**betad**), changes in return on assets (**droa**), changes in return on equity (**droe**), and idiosyncratic volatility per FF3 model (**ivff**). We suppose these factors indeed contribute to explaining cross-sectional returns.

Finally, we investigate whether aforementioned factors can augment the performance of the LSY4 model. Besides the five newly found factors, we also consider quality-minus-junk (QMJ) and liquidity (AMI), both of which are selected from the first step of LASSO. In practice, we augment the LSY4 model to a 5-factor model by adding new factors into the original model one by one. Similar to the previous section, as for a new factor model, we mainly focus on its ability to explain anomalies and other pricing factors. Table 11 shows the results. The first two rows report the GRS statistics and average t-statistics in explaining 104 anomalies. The last two rows are derived from simplified spanning regressions that only report alphas of two pricing factors, R_{me} and PEAD. Note that these two factors still have significant alphas even if they are controlled by LSY4 factors. By comparison, we find

Table 7: Summary statistics of other factors. This table reports means, t-statistics, annualized Sharpe ratios, skewness, kurtosis, and Jarque–Bera statistics of factors which do not appear in previous models but are used in our double-selection LASSO approach. The sample period is from July 2000 to December 2019. AMI and QMJ denote the liquidity factor and Quality-Minus-Junk factor, respectively. Other factors are based on anomalies in Qiao [8] but reconstructed using bivariate 2×3 portfolios following Fama and French [18].

	(1)	(2)	(3)	(4)	(5)
Factor	AMI	QMJ	**abr**	**amq**	**betad**
Mean (%)	1.05	0.83	0.44	0.50	0.85
t-value	4.81	3.56	2.94	1.72	3.73
Sharpe ratio	1.09	0.81	0.67	0.39	0.84
Skewness	0.35	0.22	−0.07	0.01	0.88
Kurtusis	6.69	4.41	4.89	7.10	5.03
J-B stat.	137.30	21.30	34.86	164.12	70.80
	(6)	(7)	(8)	(9)	(10)
Factor	**droa**	**droe**	**dtv**	**ivff**	**lrev**
Mean (%)	0.85	0.88	0.75	1.04	0.16
t-value	7.42	7.15	3.82	5.30	0.72
Sharpe ratio	1.68	1.62	0.87	1.20	0.16
Skewness	0.19	0.16	0.06	0.12	0.34
Kurtusis	4.37	4.63	4.73	4.94	6.11
J-B stat.	19.81	26.93	29.45	37.15	98.84
	(11)	(12)	(13)	(14)	(15)
Factor	**m24**	**mdr**	**ra25**	**rdm**	**sgq**
Mean (%)	0.08	0.66	0.35	0.60	0.88
t-value	0.54	3.38	2.26	3.54	5.91
Sharpe ratio	0.12	0.77	0.51	0.80	1.34
Skewness	0.04	−0.13	0.71	0.19	0.27
Kurtusis	4.14	3.81	5.20	7.57	5.80
J-B stat.	12.78	7.02	67.01	205.41	79.24
	(16)	(17)	(18)	(19)	(20)
Factor	**spq**	**Srev**	**tv**	**vdtv**	**vturn**
Mean (%)	0.58	0.79	0.48	0.98	0.91
t-value	2.51	3.77	1.93	4.91	3.84
Sharpe ratio	0.57	0.85	0.44	1.11	0.87
Skewness	0.12	0.51	−0.45	0.51	−0.52
Kurtusis	7.65	4.89	4.12	6.23	6.71
J-B stat.	211.40	44.67	20.15	112.10	144.83

Table 8: Testing SDF loadings of g_t. This table reports SDF loading estimates of 18 tested factors with significant risk premia in the A-share market. We use 650 5×5 bivariate sorted portfolios as test assets. Benchmark factors include all factors appearing in Table 3 as well as the liquidity factor AMI and Quality-Minus-Junk (QMJ) factor. The sample period is from July 2000 to December 2019. The first column uses double-selection LASSO approach to determine controlling factors. The second column only uses factors of Liu–Stambaugh–Yuan 4-factor model [15] as controls. The third column uses all 18 benchmark factors as controls.

	(1) Double Lasso		(2) ChF4		(3) All	
	Coef	t-value	Coef	t-value	Coef	t-value
abr	8.77	3.29***	6.89	2.69***	−0.02	−0.01**
amq	−3.02	−0.89	−2.88	−0.96	12.74	2.22**
betad	8.97	3.86***	5.51	2.31**	7.03	2.84***
droa	52.64	3.81***	39.88	3.19***	41.02	2.71***
droe	−21.84	−1.77*	−5.85	−0.51	−9.19	−0.71
dtv	16.75	2.59***	33.54	5.29***	33.64	4.17***
ivff	27.06	7.35***	23.35	6.79***	15.74	3.86***
lrev	−2.82	−1.26	−4.5	−2.33**	0.29	0.1
m24	0.58	0.2	7.27	2.72***	6.09	1.68*
mdr	−13.95	−3.36***	−8.32	−2.09**	3.3	0.62
ra25	−4.53	−1.93*	−6.49	−2.83***	0.66	0.25
rdm	3.42	0.77	6.75	1.57	12.63	2.46**
sgq	−1.66	−0.49	−14.71	−4.31***	−13.99	−3.62***

Table 8: Continued.

	(1) Double Lasso		(2) ChF4		(3) All	
	Coef	t-value	Coef	t-value	Coef	t-value
spq	6.13	1.21	−2.43	−0.44	−6.55	−1.06
srev	3.8	1.97**	1.74	0.91	−0.2	−0.08
tv	−5.32	−1.13	−7.69	−1.64	−10.75	−1.91**
vdtv	−16.99	−2.32**	−28.3	−3.83***	−38.21	−4.24***
vturn	−1.19	−0.48	6.68	2.95***	10.99	3.68***

Table 9: Testing SDF loadings of g_t controlling all benchmark factors. This table reports testing results of SDF loading estimates controlling all 99 benchmark factors simultaneously. The sample period is from July 2000 to December 2019. In column 1, we use previous 650 5×5 bivariate sorted portfolios as test assets, while, in column 2, we use 156 2×3 bivariate sorted portfolios.

(1) 5×5 portfolios		(2) 2×3 portfolios	
Coef	t-value	Coef	t-value
16.53	2.372**	29.9	1.23
−13.71	−0.879	93.76	1.87*
17.46	3.294***	61.15	3.52***
52.44	1.774	161.05	1.43
3.08	0.127	−30.72	−0.342
−22.83	−1.431	−42.48	−0.836
−13.15	−0.854	−73.88	−1.297
−6.14	−0.825	−12.64	−0.441
14.89	1.925	−28.3	−0.957
11.13	1.162	−1.93	−0.048
2.78	0.413	34.41	1.358
13.48	1.211	9.68	0.276
2.18	0.224	89.84	2.21**
−8.73	−0.566	35.83	0.539
−6.46	−1.269	−8.55	−0.415
−8.9	−0.832	40.41	1.092
−6.43	−0.395	−12.07	−0.158
16.25	1.215	−51.08	−0.77

Table 10: Robustness for testing SDF loadings of g_t. This table reports robustness results of SDF loading estimates using double-selection LASSO approach. The sample period is from July 2000 to December 2019. In columns 1 and 2, we use previous 650 5×5 bivariate sorted portfolios as test assets, while, in columns 3 and 4, we use 156 2×3 bivariate sorted portfolios. Columns 1 and 3 select controls from factors appearing in Table 3 as well as the liquidity factor AMI and Quality-Minus-Junk (QMJ) factor, while columns 2 and 4 select controls from an expanding set including totally 99 benchmark factors.

	5×5 portfolios				2×3 portfolios			
	(1) 18 factors		(2) 99 factors		(3) 18 factors		(4) 99 factors	
	Coef	t-value	Coef	t-value	Coef	t-value	Coef	t-value
Abr	8.77	3.29***	6.47	2.28**	16.1	3.70***	15.44	3.35***
Amq	−3.02	−0.89	−13.89	−2.66**	−8.07	−1.5	−10.39	−1.07
Betad	8.97	3.86***	10.03	4.01***	7.38	1.89*	5.1	1.17
Droa	52.64	3.81***	60.94	4.58***	73.33	2.30**	78.06	2.43**
Droe	−21.84	−1.77*	−26.2	−2.18**	−48.56	−1.65*	−50.97	−1.72*
dtv	16.75	2.59***	7.38	1.17	−5.73	−0.37	−6.99	−0.44
ivff	27.06	7.35***	20.16	2.46**	43.09	5.85***	19.7	0.95
lrev	−2.82	−1.26	−2.3	−1.12	−1.87	−0.49	−2.49	−0.64
m24	0.58	0.2	3.88	1.35	10.83	2.07**	10.91	2.04**

Table 10: Continued.

	5 × 5 portfolios				2 × 3 portfolios			
	(1)		(2)		(3)		(4)	
	18 factors		99 factors		18 factors		99 factors	
	Coef	*t*-value	Coef	*t*-value	Coef	*t*-value	Coef	*t*-value
mdr	−13.95	−3.36***	−14.39	−3.39***	−12.15	−1.42	−7.88	−0.85
ra25	−4.53	−1.93**	−7.91	−3.05***	4	0.92	3.68	0.75
rdm	3.42	0.77	6.79	1.45	−4.41	−0.48	−1.15	−0.11
sgq	−1.66	−0.49	0.31	0.09	9.48	1.46	10.24	1.57
spq	6.13	1.21	9.11	1.66*	19.32	1.82*	13.77	1.19
srev	3.8	1.97**	2.97	1.52	−2.12	−0.54	−3.57	−0.86
tv	−5.32	−1.13	−0.73	−0.14	−15.54	−1.63	−22.45	−2.00**
vdtv	−16.99	−2.32**	−4.66	−0.63	10.26	0.6	7.32	0.38
vturn	−1.19	−0.48	−3.59	−1.28	−12.27	−2.38**	−9.95	−1.70**

Table 11: Comparision of new factors in explaining anomalies. This table reports results of explaining 104 anomalies using augmented Liu–Stambaugh–Yuan 4-factor models [15]. The first column corresponds to our baseline result. We add seven new factors into the Liu–Stambaugh–Yuan 4-factor model one by one, including Quality-Minus-Junk, liquidity, cumulative abnormal returns around earnings' announcement dates, Dimson beta, 4-quarter change in return on assets, and 4-quarter change in return on equity and idiosyncratic volatility per the FF3 model, in columns labeled "QMJ," "AMI," "abr," "betad," "droa," "droe," and "ivff." The first two rows display GRS statistics proposed by Shanken et al. [24] and absolute average *t*-statistics of alphas over 104 anomalies. The last two rows display alpha estimates and corresponding t-statistics (in parentheses) of regressing PEAD and R_{me} on augmented models. PEAD is the short horizon factor in the Daniel–Hirshleifer–Sun 3-factor model [22]. R_{me} is the size factor in the Hou–Xue–Zhang q5-factor model [2].

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	LSY4	QMJ	AMI	abr	LSY4+ betad	droa	droe	ivff
GRS	3.28	3.25	3.09	3.15	3.37	3.01	3.04	3.27
Average *t*	1	0.96	0.99	1.01	1.18	1.35	1.39	1.14
Alpha for PEAD	0.72%	0.21%	0.76%	0.41%	0.77%	0.41%	0.43%	0.77%
	−3.35	−3.15	−3.44	−2.33	−3.51	−1.95	−2.03	−3.63
Alpha for R_{me}	0.19%	0.20%	0.17%	0.09%	0.23%	0.10%	0.09%	0.24%
	−1.9	−2	−1.64	−0.99	−2.22	−1	−0.91	−2.56

abr, droa, and **droe** can improve the performance of the LSY4 model. All of the three factors can reduce the value of GRS statistic as well as alphas of R_{me} and PEAD. However, the improvement is limited. None of them can totally explain all anomalies and eliminate the significance of the factor PEAD. Thus, we suppose relying on current factor zoo is hard to enhance the pricing power of the LSY4 model. New characteristics beyond our data library are needed. We hope our paper can inspire future research in this field.

5. Conclusion

The "factor zoo" problem is a hot issue in empirical asset pricing. Given the short history and incomplete trade regulations of the Chinese A-share market, few studies focus on addressing this issue in China. Our study aims to fill this gap by testing the performance of pricing factors emerging in recent studies. Using A-share market data, we replicate a generation of pricing factors and construct a characteristic library. By explaining 104 anomalies and performing spanning regressions, we verify the Liu–Stambaugh–Yuan four-factor model dominates other counterparts in the A-share market, including the Fama–French six-factor factor model, the Hou–Xue–Zhang q5-factor model, the Stambaugh–Yuan four-factor model, and the Daniel–Hirshleifer–Sun three-factor model. Next, to explore whether there exist latent factors which have explanatory power beyond current factor models, we use a double-selection LASSO approach to estimate SDF loadings. Although we find some anomalies play a significant role in pricing cross-sectional returns, once we add them into the Liu–Stambaugh–Yuan four-factor model, the improvement is limited. We suppose future research should not be restricted in the current factor zoo when coming up with new factor models.

Data Availability

The trading data and financial data were collected from China Stock Market & Accounting Research Database (https://www.gtarsc.com/) and independently construct factors as well as anomalies in the A-share market. The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

The details to construct factors are provided in the Supplementary Appendix. (*Supplementary Materials*)

References

[1] J. H. Cochrane, "Presidential address: discount rates," *The Journal of Finance*, vol. 66, pp. 1047–1108, 2011.

[2] K. Hou, H. Mo, C. Xue, and L. Zhang, "Which factors?" *Review of Finance*, vol. 23, no. 1, pp. 1–35, 2018.

[3] F. E. Racicot and W. F. Rentz, "A panel data robust instrumental variable approach: a test of the new Fama-French five-factor model," *Applied Economics Letters*, vol. 24, pp. 410–416, 2017.

[4] F. E. Racicot, W. F. Rentz, and R. Théoret, "Testing the new fama and French factors with illiquidity: a panel data investigation," *Finance*, vol. 39, pp. 45–102, 2018.

[5] F. E. Racicot, W. F. Rentz, D. Tessier, and R. Théoret, "The conditional Fama-French model and endogenous illiquidity: a robust instrumental variables test," *PLoS One*, vol. 14, no. 9, Article ID e0221599, 2019.

[6] E. F. Fama and K. R. French, "A five-factor asset pricing model," *Journal of Financial Economics*, vol. 116, pp. 1–22, 2015.

[7] L. Pástor and R. F. Stambaugh, "Liquidity risk and expected stock returns," *Journal of Political Economy*, vol. 111, pp. 642–685, 2003.

[8] F. Qiao, "Replicating anomalies in China," *SSRN*, vol. 33, no. 5, pp. 2019–2133, 2019.

[9] K. Hou, C. Xue, and L. Zhang, "Replicating anomalies," *Review of Financial Studies*, 2018.

[10] E. F. Fama and K. R. French, "Choosing factors," *Journal of Financial Economics*, vol. 128, pp. 234–252, 2018.

[11] G. Feng, S. Giglio, and D. Xiu, *Taming the Factor Zoo: A Test of New Factors*, National Bureau of Economic Research, Cambridge, MA, USA, 2019.

[12] G. X. Hu, C. Chen, Y. Shao, and J. Wang, "Fama-French in China: size and value factors in Chinese stock returns," *International Review of Finance*, vol. 19, pp. 3–44, 2019.

[13] F. Jiang, X. Qi, and G. Tang, "Q-theory, mispricing, and profitability premium: evidence from China," *SSRN Electronic Journal*, vol. 87, pp. 135–149, 2018.

[14] C. Cheung, G. Hoguet, and S. Ng, "Value, size, momentum, dividend yield, and volatility in China's a-share market," *Journal of Portfolio Management*, vol. 41, pp. 57–70, 2014.

[15] J. Liu, R. F. Stambaugh, and Y. Yuan, "Size and value in China," *Journal of Financial Economics*, vol. 134, pp. 48–69, 2019.

[16] Y. Han, A. He, D. Rapach, and G. Zhou, "Expected stock returns and firm characteristics: E-LASSO, assessment, and implications," *SSRN*, 2019.

[17] J. Freyberger, A. Neuhierl, and M. Weber, "Dissecting characteristics nonparametrically," *The Review of Financial Studies*, vol. 33, 2019.

[18] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, pp. 3–56, 1993.

[19] N. Jegadeesh and S. Titman, "Return to buying winners and selling losers: implication for stock market efficiency," *The Journal of Finance*, vol. 48, pp. 65–91, 1993.

[20] K. Hou, C. Xue, and L. Zhang, "Digesting anomalies: an investment approach," *The Review of Financial Studies*, vol. 28, no. 3, pp. 650–705, 2015.

[21] R. F. Stambaugh and Y. Yuan, "Mispricing factors," *Review of Financial Studies*, vol. 30, pp. 1270–1315, 2017.

[22] K. D. Daniel, D. A. Hirshleifer, and L. Sun, "Short- and long-horizon behavioral factors," *Review of Financial Studies*, vol. 33, pp. 1673–1736, 2020.

[23] J. Shanken, "On the estimation of beta-pricing models," *Review of Financial Studies*, vol. 5, pp. 1–33, 1992.

[24] J. Shanken, M. R. Gibbons, and S. A. Ross, "A test of efficiency of a given portfolio," *Econometrica*, vol. 57, pp. 1121–1152, 1989.

[25] W. K. Newey and K. D. West, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, vol. 55, pp. 703–708, 1987.

[26] Cochrane, *Asset Pricing (Revised Edition)*, Princeton University Press, Princeton, NJ, USA, 2009.

[27] W. Ferson, *Empirical Asset Pricing: Models and Methods*, The MIT Press, Cambridge, MA, USA, 2019.

[28] E. Ghysels and M. Marcellino, *Applied Economic Forecasting Using Time Series Methods*, Oxford University Press, Oxford, UK, 2018.

[29] B. Efron, T. Hastie, I. M. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, pp. 407–499, 2004.

[30] J. H. Friedman, T. Hastie, H. Hofling, and R. Tibshirani, "Pathwise coordinate optimization," *Annals of Applied Statistics*, vol. 1, pp. 302–332, 2007.

[31] C. S. Asness, A. Frazzini, and L. H. Pedersen, *Quality Minus Junk*, Social Science Electronic Publishing, Waltham, MA, USA, 2013.

[32] W. Kang and H. Zhang, "Measuring liquidity in emerging markets," *Pacific-Basin Finance Journal*, vol. 27, pp. 49–71, 2014.

[33] Y. Hu and M. Gu, "Buffett's alpha: evidence from China stock market," *Management World*, vol. 8, pp. 41–54, 2018.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 6541902, 18 pages
https://doi.org/10.1155/2021/6541902

Research Article

The Cross-Correlations between Foreign Flows in Chinese A-Share Markets and Uncertainties in Home Markets

Tao Bing[1], **Fei Hu**[2] and **Hongkun Ma**[3,4]

[1]*School of Finance, Capital University of Economics and Business, Beijing 100070, China*
[2]*Chinese Academy of Macroeconomic Research, Beijing 100038, China*
[3]*Zhuhai Huafa Investment Holdings Group Co. Ltd., Hengqin 519000, China*
[4]*School of Business, Sun Yat-sen University, Guangzhou 510275, China*

Correspondence should be addressed to Hongkun Ma; mahongkun1212@163.com

Received 7 October 2021; Accepted 2 November 2021; Published 19 November 2021

Academic Editor: Dehua Shen

Copyright © 2021 Tao Bing 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 investigates the cross-correlations between the foreign flows in A-share market and the uncertainties of market, economy, and policy in home markets, namely, the VIX index and the US EPU index. By employing the cross-correlation statistics and MF-DCCA method, we find the existence of the cross-correlations between the foreign flows and the VIX index, the foreign flows, and the US EPU index from qualitative and quantitative perspectives, respectively. For the cross-correlation between the foreign flows and VIX, small fluctuations are persistent, while large fluctuations are antipersistent. In contrast, the cross-correlation between the foreign flows and US EPU is antipersistent and steady. These results are robust by dividing the foreign flows into two parts in Shenzhen and Shanghai stock exchanges, respectively, and by shortening the periods to after the implementation of Shenzhen-Hong Kong Stock Connect.

1. Introduction

Since China opened its door to the world in 1978, the financial liberalization has been increased (https://business.cornell.edu/hub/2020/05/11/chinas-financial-liberalization/). And the Keynote Speech by Xi Jinping, the president of the People's Republic of China, at the Opening of the Boao Forum for Asia Annual Conference 2018 initiates the new round of the opening-up of Chinese financial market. After that, the limited amounts of QFII and RQFII have been relaxed, the Shanghai-London Stock Connect Program has been implemented, and the foreign equity caps in the banking, securities, and insurance industries have also been raised. As a result, we can see that the foreign investment has been increasing in recent years (see the news from the Global Time: https://www.globaltimes.cn/page/202108/1231158.shtml.), and the holding of foreign investors in A-share market continuously increases (see Figure 1). Therefore, it is important to be clear about the factors affecting the foreign flows in A-share market.

In fact, this problem has been discussed since the early 1990s. Calvo et al. [1] and Fernandez-Arias [2] firstly introduce the framework of push and pull factors to analyze the factors that influence the international capital flows (the pull factors relate to the host economic forces that attract foreign capital, such as high host interest rates, stable political environment, and high growth potential). The push factors relate to the external or global or home forces that push international capital flows to the host markets, such as the low interest rates, high international political and market uncertainty, low potential growth, and the portfolio diversification. The later literature studies, such as Taylor and Sarno [3], Agénor [4], Chuhan et al. [5], Forbes and Warnock [6], Fratzscher [7], and Fuertes et al. [8], have further discussed this problem. See the review of Koepke Koepke [9] for more details. In recent years, the outbreaks of the global financial crisis, the catastrophic risk (such as the 9/11, China's 2008 earthquake, and Japan's 2011 earthquake), and the global public safety events (such as the Ebola and COVID-19) cause great uncertainties of market, economy, and policy. These

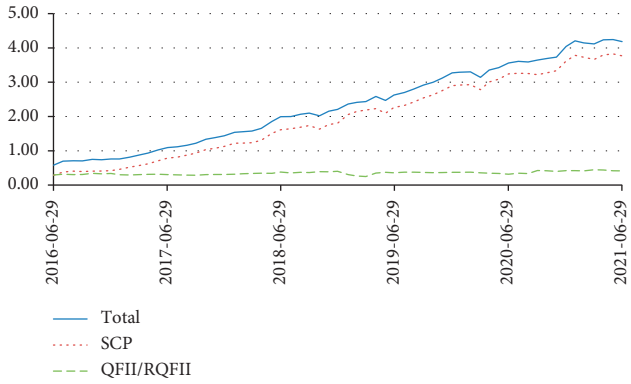


Figure 1: The holding share of foreign investors in A-share market from 29 June 2016 to 29 June 2021. The SCP refers to the foreign flows via the Stock Connect Program; QFII and RQFII refer to the Qualified Foreign Institutional Investors and RMB Qualified Foreign Institutional Investors, respectively.

uncertainties change the market sentiment and investors' expectations of economy and the fundamental value of assets and influence investors' trading behavior and their flows. And these effects are more pronounced for the trading behavior and flows of foreign investors in emerging markets. However, the results are not consistent. On the one hand, Borensztein and Gelos [10] report that foreign institutional investors are prone-panic due to their information disadvantage in emerging countries. Omay and Iren [11] find that foreign investors exhibit herding behavior, and the foreign flows tend to be unstable and prone to flee the Malaysian stock market during the 1997 Asian Crisis and the 2008 Global Financial Crisis. Hood et al. [12] find significant positive feedback and momentum trading patterns of foreign investors, namely, net foreign outflows, during Japan's 2011 earthquake. On the other hand, Gerlach and Yook [13] show that foreign investors deviate from the general positive feedback trading strategy and increase their holding of Korean stocks during the 13 North Korean military attacks from 1999 to 2010. And Bing and Ma [14] find significant negative feedback trading of foreign investors in A-share market during the COVID-19 crisis. Besides, Choe et al. [15] show that the strong positive feedback trading of foreign investors before Korea's financial crisis disappears during the crisis. Similar results are reported during the Asian crisis in Indonesia [16].

Therefore, it is necessary to investigate the relationships between flows and uncertainties. The existing literature mainly studies their linear relationship and rarely discusses the nonlinear relationship. However, many financial variables and their relations display properties of multifractality (see the review of Jiang et al. [17] for more details. Dai et al. [18] and Zhao and Dai [19] also discuss the multifractality of China EPU and US EPU). This paper will investigate the cross-correlations between the foreign flows in A-share markets and the uncertainties of market, economics, and policy in home market using the MF-DCCA approach proposed by Zhou [20] (the MF-DCCA approach has been widely used to study the nonlinear cross-correlation in financial markets, such as the cross-correlation between stock returns and news (Zhang et al. [21], Zhang et al. [22], and Zhang et al. [23]), online sentiment and market returns [24], EPU and exchange rate [25], exchange rate and market anxiety [26], China EPU and US EPU [19], carbon market price and EPU [27], cryptocurrency uncertainty indices and EPU [28], investors' attention and index futures [29], return-volume relationship of bitcoin market [30], online searches and bitcoin market [31], and online sentiment proxies [32]). The uncertainty of market is measured by the market panic index, namely, the VIX index (the VIX index has been widely used in the previous literature [26, 33]). And the uncertainty of economy and policy can be measured by the economic policy uncertainty (EPU) index developed by Baker et al. [34].

Our main findings are as follows: first, there are significant power-law relationship and a long-term cross-correlation between the foreign flows and VIX index, the foreign flows, and US EPU index. The Renyi exponent confirms the results. Second, the cross-correlation Hurst exponents show that cross-correlation between the foreign flows and US EPU is antipersistent and steady, while the cross-correlation between the foreign flows and VIX is antipersistent only when the fluctuations are large. When the fluctuations are small, the cross-correlation is persistent. In addition, the multifractal spectrum analysis shows that the strength of the cross-correlation between the foreign flows and VIX is larger than that between the foreign flows and US EPU. And the above results are robust by further analysis using the separate foreign flows, namely, the flows via Shanghai-Hong Kong Stock Connect Program and Shenzhen-Hong Kong Stock Connect Program, and the foreign flows after the implementation of Shenzhen-Hong Kong Stock Connect.

This paper contributes to the existing literature by verifying the cross-correlation between the foreign flows and VIX index, the foreign flows, and US EPU. To the best of our knowledge, this is the first paper to investigate the nonlinear relationships between the foreign flows in emerging markets and the uncertainty in home markets, which enriches linear analytical framework for the drivers of foreign flows to emerging markets. These results would be helpful for policy makers to forestall potential imported financial risks.

The remainder of the paper is organized as follows. In Section 2, we describe the methodology in detail. Section 3 describes the data. Section 4 shows the empirical results. For robustness, we shorten the sample period and replace the overall foreign flows with the separate flows in Section 5. Finally, Section 6 concludes the paper.

2. Methodology

In this section, the MF-DCCA approach used to explore the cross-correlation between the foreign flows and VIX index, the foreign flows, and the US EPU index will be described.

2.1. Cross-Correlation Analysis. To statistically examine the cross-correlation between the series, the cross-correlation statistic methodology proposed by Podobnik and Stanley [35] will be employed. The cross-correlation analysis $Q_{cc}(m)$

of the two series $\{x_i\}$ and $\{y_i\}$ is defined as shown in the following equation:

$$Q_{cc}(m) = N^2 \sum_{i=1}^{m} \frac{C_i^2}{N-i}, \tag{1}$$

where m is the degree of freedom, N is the length of the series $\{x_i\}$ and $\{y_i\}$, and C_i is the cross-correlation function and can be calculated as follows:

$$C_i = \frac{\sum_{j=i+1}^{N} x_j y_{j-i}}{\sqrt{\sum_{j=1}^{N} x_j^2 \sum_{j=1}^{N} y_j^2}}. \tag{2}$$

$Q_{cc}(m)$ is approximately $\chi^2(m)$ distributed with m degrees of freedom. The null hypothesis of the cross-correlation is that the cross-correlations are not significantly different from zero. Therefore, if $Q_{cc}(m)$ is greater than the critical value of $\chi^2(m)$, there is a cross-correlation between the two series. Otherwise, there is no cross-correlation.

2.2. MF-DCCA. Consider two series $\{x_i\}$ and $\{y_i\}$, where $i = (1, 2, \ldots, N)$ and N is the length. The MF-DCCA approach could be constructed by the following steps:

Step 1. Construct the profiles $\{X_i\}$ and $\{Y_i\}$ of the two series, respectively.

$$\begin{aligned} X_i &= \sum_{k=1}^{i} (x_i - \overline{x}), \\ Y_i &= \sum_{k=1}^{i} (y_i - \overline{y}), \end{aligned} \tag{3}$$

where $\overline{x}$ and $\overline{y}$ are the mean value of the two series, respectively.

Step 2. Divide the profiles into $N_s = \text{int}(N/s)$ nonoverlapping segments, where *int* denotes floor rounding function, and s is the length of the segments. As N is not always a multiple of s, a small part of the profile with length $(N - N_s * s)$ will be left. To make the most of the information contained in the data, the same procedure is repeated from the end of the profiles. Therefore, there are 2 Ns segments.

Step 3. Calculate the detrended covariance of every segment. First, we use the ordinary least squares (OLS) to obtain the local trend of every segment. Then, we calculate the detrended covariance. For the segment $v = 1, 2, \ldots, N_s$, the detrended covariance could be obtained by the following equation:

$$F^2(v, s) = \frac{1}{s} \sum_{j=1}^{s} \left(X_{(v-1)s+j} - \widetilde{X}_{(v-1)s+j}\right)\left(Y_{(v-1)s+j} - \widetilde{Y}_{(v-1)s+j}\right). \tag{4}$$

For the segment $v = N_s + 1, N_s + 2, \ldots 2N_s$, the detrended covariance is presented in the following equation:

$$F^2(v, s) = \frac{1}{s} \sum_{j=1}^{s} \left(X_{N-(v-N_s)s+j} - \widetilde{X}_{N-(v-N_s)s+j}\right)\left(Y_{N-(v-N_s)s+j} - \widetilde{Y}_{N-(v-N_s)s+j}\right), \tag{5}$$

where $\widetilde{X}_v$ and $\widetilde{Y}_v$ are the local trend of segment v.

Step 4. Calculate the qth-order fluctuation function by averaging all the detrended segment variances. For $q \neq 0$,

$$F_q(s) = \left[\frac{1}{2N_s} \sum_{v=1}^{2N_s} F^2(v, s)^{q/2}\right]^{1/q}. \tag{6}$$

For $q = 0$, the fluctuation function could be obtained by L'Hospital's rule:

$$F_0(s) = \exp\left[\frac{1}{4N_s} \sum_{v=1}^{2N_s} \ln F^2(v, s)\right]. \tag{7}$$

Step 5. Analyze the multifractal features. If the two series are long-term cross-correlated, there is a power-law relationship between $F_q(s)$ and the scale s:

$$F_q(s) \propto s^{H_{xy}(q)}, \tag{8}$$

where $H_{xy}(q)$ could be obtained by the log-log plots of $F_q(s)$ verse s though the OLS estimation. $H_{xy}(q)$ is the generalized Hurst exponent. If $H_{xy}(q)$ varies with q, the cross-correlation between the two series has multifractal features. If $H_{xy}(q)$ is greater than 0.5, the cross-correlation between the two series is persistent and positive. If $H_{xy}(q)$ is less than 0.5, the cross-correlation between them is antipersistent and negative. If $H_{xy}(q)$ is equal to 0.5, there is no significant consistent cross-correlation between them. In particular, when q is equal to 2, $H_{xy}(2)$ is the standard Hurst index.

Furthermore, we could get the relationship between the multifractal scaling Renyi exponent $\tau_{xy}(q)$ (provided by Shadkhoo and Jafari [36] and q by the following equation:

$$\tau_{xy}(q) = \text{qH}_{xy}(q) - 1. \tag{9}$$

If $\tau_{xy}(q)$ and q are linear, the cross-correlation between them is monofractal. Otherwise, their cross-correlation is multifractal. And we can get the multifractal spectrum $f_{xy}(\alpha)$ through the Legendre transformation:

$$\begin{aligned} \alpha &= H_{xy}(q) + \mathrm{qH}'_{xy}(q), \\ f_{xy}(\alpha) &= q\left(\alpha - H_{xy}(q)\right) + 1, \end{aligned} \tag{10}$$

where $H'_{xy}(q)$ is the derivative of $H_{xy}(q)$. And α is the singularity strength, which describes the strength of singularity spectrum. The value of the multifractal spectrum, $f_{xy}(\alpha)$, shows the fractal dimension of α.

According to Yuan et al. [37], the multifractal intensity could be measured by $\Delta H = H_{\max}(q) - H_{\min}(q)$, namely, the width of the multifractal spectrum. The higher the value of ΔH, the stronger multifractality and the greater corresponding market risk, and vice versa.

3. Data

In the A-share market, foreign investors trade mainly though the Qualified Foreign Institutional Investors (QFII) scheme and the Stock Connect Program (SCP). Compared to the QFII scheme, the SCP has advantages in transaction cost, transaction settlement, and the trading qualification authentication. Therefore, even though the QFII scheme was implemented as early as 2002, the SCP (the SCP includes Shanghai-Hong Kong Stock Connect and Shenzhen-Hong Kong Stock Connect, launched in 17 November 2014 and 5 December 2016, respectively.), launched in 17 November 2014, has been the preferred trading channel for foreign investors (see the holding share of foreign investors in Figure 1). Besides, there are great differences in the availability of data. The data of QFII could be only obtained by their quarterly position data, while Hong Kong Exchanges and Clearing Limited discloses the foreign investors' trading data every day. Based on the above two reasons, we select the flows though the SCP as the representative of foreign trading. The data of foreign flows could be collected from the WIND database. And we use the daily net inflows in the model (if the values is greater than zero, this indicates net inflows; otherwise, it indicates net outflows.).

We choose VIX index to measure market panic, which has been widely used in prior literature [38–41]. And the VIX index could also be collected from the WIND database. Since the foreign flow is daily net inflows, we adopt daily changes in VIX, which is similar to the prior literature [26, 33].

The EPU index (this index is constructed by counting the articles in 10 leading newspapers that contain the following three sets of terms: economic or economy, uncertain or uncertainty, and the government actions, such as White House, Federal Reserve, legislation, deficit, regulation, and congress. For more details, see Baker et al. [34] and the website http://www.policyuncertainty.com/index.html.), developed by Baker et al. [34], has been extensively used in the literature to measure the uncertainty of economic policy (the EPU has been extensively discussed in financial markets, such as its relationship with corporate investment [42], mergers and acquisitions [43], market returns [27], and exchange rate [25]. See the review of Al-Thaqeb and Algharabali [44] for details. Dai et al. [45] proposed a global economic policy uncertainty index to measure the economic policy uncertainty on a global scale. And this index has been used to predict crude oil futures volatility [46]. But this index is monthly frequency. We can not use it in this paper). Considering that the flows and VIX are daily data, we choose the daily EPU, namely, the US EPU index (the US EPU can be downloaded from this website: http://www.policyuncertainty.com/us_monthly.html). And we use daily changes in US EPU in this paper.

Since the SCP was launched in 17 November 2014, we choose the sample periods from 17 November 2014 to 30 May 2021. The vacant data due to China and US holidays are deleted. The descriptive statistics of the series are presented in Table 1. The Jarque–Bera tests of the three series all reject the normal distribution null hypothesis at 1% significance level.

4. Empirical Results

4.1. Cross-Correlation Test. We first employ the cross-correlation analysis to examine the cross-correlations between the two pairs of the series. Figure 2 shows the results of the cross-correlation statistics $Q_{cc}(m)$ with degrees of freedom m ranging from 1 to $N-1$. The critical values of the $\chi^2(m)$ at 5% significant level are shown with red lines to make a clear comparison. We can find that most values of $Q_{cc}(m)$ for both the two pairs exceed the critical value. These results reject the null hypothesis of no cross-correlations. Therefore, there is a long-term cross-correlation between the foreign flows in A-share market and VIX and the foreign flows in A-share market and US EPU.

4.2. Multifractal Detrended Cross-Correlation Analysis. In this section, we will examine the cross-correlation between the two pairs of the series using the MF-DCCA method. Consistent with the prior research [20, 21], we set the scale q from −10 to 10 with the step width of 1. Figure 3 shows the log-log plots of the fluctuation function $F_q(s)$ versus s between the foreign flows in A-share market and VIX and the foreign flows in A-share market and US EPU. The results show that $F_q(s)$ increases with the scale s. And the curves show the linearity within a certain interval. Therefore, there exist a significant power-law relationship and a long-term cross-correlation between the two pairs of series.

Figure 4 describes the relationship between the generalized Hurst exponent and q for the two pairs of series. The results show that the cross-correlation Hurst exponent $H_{xy}(q)$ decreases with the increase of q, indicating that there exists a multifractality in the cross-correlation between the foreign flows and VIX and the foreign flows and US EPU. In particular, for the series of foreign flows and VIX, when q is less than 4, $H_{xy}(q)$ is greater than 0.5, and its trend is relatively steady. This indicates that the cross-correlation is persistent when the fluctuation q is small. When $q \geq 4$, $H_{xy}(q)$ is less than 0.5, indicating that the cross-correlation

Table 1: Descriptive statistics of net foreign flows in A-share market, the change of VIX index, and US EPU index.

	VIX change	Net foreign flows	US EPU change
N	1511	1511	1511
Mean	0.002	9.208	0.076
Median	−0.080	6.284	0.140
Std. dev.	2.119	34.486	57.822
Maximum	24.860	217.233	267.360
Minimum	−17.640	−173.842	−282.840
Skewness	2.761	0.717	0.046
Kurtosis	35.903	10.308	6.610
Jarque–Bera	70079.860***	3491.684***	820.817***

***Statistical significance at the 0.001.

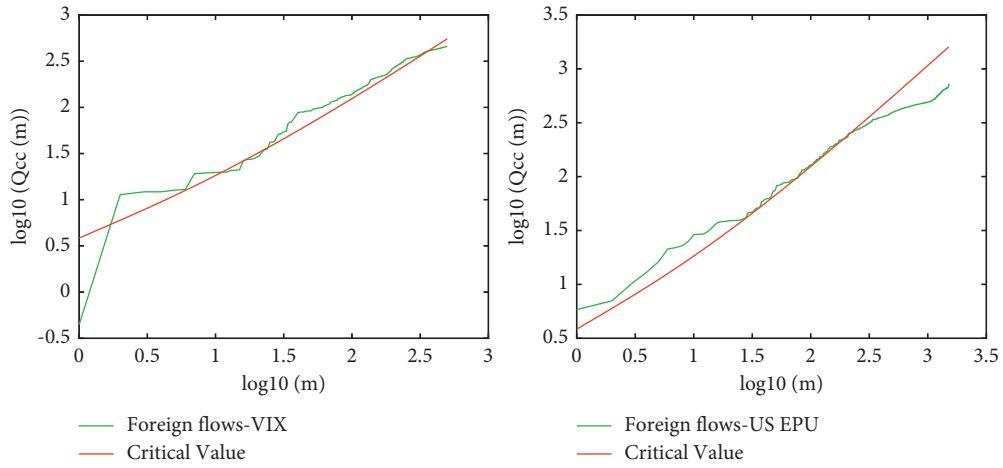


Figure 2: The cross-correlation statistics $Q_{cc}(m)$ between the foreign flows in A-share market and VIX and the foreign flows in A-share market and US EPU.

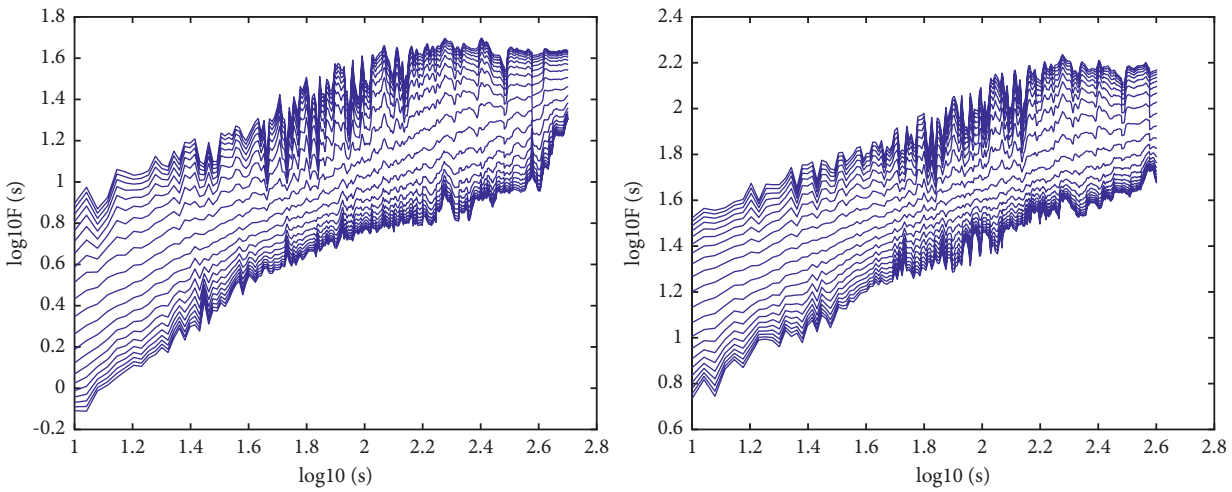


Figure 3: The log-log plots of $F_q(s)$ versus s between the foreign flows in A-share market and VIX and the foreign flows in A-share market and US EPU. The order of q increases from the bottom to the top.

is antipersistent. The change of the cross-correlation may be explained by the characteristics of the international financial flow. When the fluctuations are small, they are volatile at the host markets, namely, US market or global market for foreign markets. The emerging markets such as A-share market may be a haven for international capital. Therefore, foreign investors trade in A-share market. Namely, net inflows occur. However, the big fluctuations may indicate

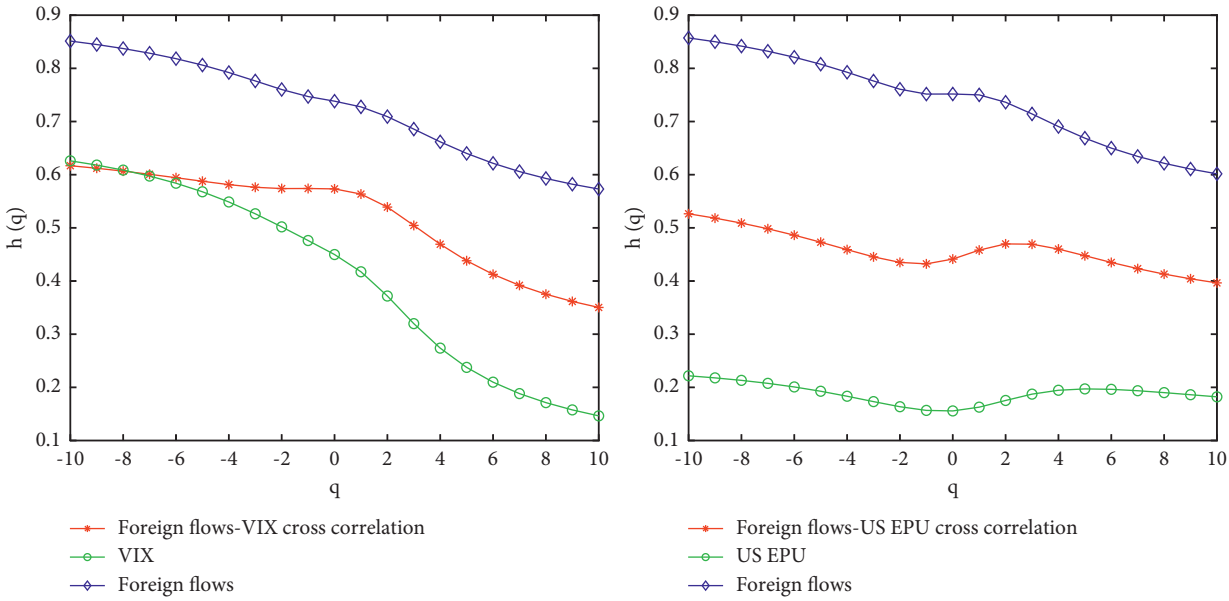


FIGURE 4: The generalized Hurst exponent of $H(q)$ versus q for the foreign flows in A-share market and VIX and the foreign flows in A-share market and US EPU.

global risk, or the risk may spread from US market to emerging markets. Therefore, there is a net outflow of international capital. For the series of foreign flows and US EPU, the Hurst exponent $H_{xy}(q)$ is less than 0.5 when q is greater than -8, indicating that their cross-correlation is antipersistent. Compared to that between the foreign flows and VIX, the cross-correlation between the foreign flows and US EPU is more steady.

Figure 5 presents the relationship between Renyi exponent $\tau(q)$ and q. The results show that the cross-correlations between $\tau(q)$ and q for the two pairs of series are nonlinear, confirming their multifractal cross-correlation properties.

Figure 6 shows the multifractal spectrum $f_{xy}(\alpha)$ versus α for two pairs of series. If the multifractal spectrum is a point, monofractality exists; otherwise, monofractality exists. We can see that all the multifractal spectrum varies significantly, confirming the cross-correlation between the foreign flows and VIX and the foreign flows and US EPU. Besides, the width of the multifractal spectrum, namely, the difference between the maximum of α and the minimum of α, measures the strength of the multifractal. As shown in Figure 6, the width of the spectrum of foreign flows and VIX is larger than that of foreign flows and US EPU, indicating that the strength of the cross-correlation between the foreign flows and VIX is larger.

5. Further Analysis

As mentioned above, the SCP could be divided into two stages: the implementation of Shanghai-Hong Kong Stock Connect Program and Shenzhen-Hong Kong Stock Connect Program. The cross-correlations between the overall foreign flows and VIX and the overall foreign flows and US EPU have been investigated in Section 4. In this section, we will further divide the foreign flows into two parts: the foreign flows via Shanghai-Hong Kong Stock Connect Program (HGT (here, we use HGT and SGT to denote the net foreign flows via Shanghai-Hong Kong Stock Connect Program and Shenzhen-Hong Kong Stock Connect Program, respectively)) and via Shenzhen-Hong Kong Stock Connect Program (SGT). And then, we will explore the cross-correlation between the separate foreign flows and VIX index and the separate foreign flows and US EPU index. Besides, the implementation of Shenzhen-Hong Kong Stock Connect increases the overall foreign flows in A-share market, which may affect the above cross-correlation. Therefore, we will investigate the cross-correlation with the overall foreign flows after the implementation of Shenzhen-Hong Kong Stock Connect. The above two analyses will further enrich our cognition about the cross-correlations between the foreign flows and VIX and foreign flows and US EPU. More importantly, these can also be a robust test for the conclusion in Section 4.

First, we show the cross-correlation analysis for the separate foreign flows via two stock exchange ((a) in Figures 7 to 10). Most values of $Q_{cc}(m)$ for all the pairs, except the HGT and US EPU, exceed the critical value, indicating a long-term cross-correlation among all the pairs. And the values of $Q_{cc}(m)$ between the HGT and US EPU are close to the critical value. The MF-DCCA analysis could give further evidence to show their cross-correlation relationship. All the log-log plots of $F_q(s)$ versus s ((b) in Figures 7–10) show that the fluctuation function $F_q(s)$ increases with s, indicating the significant power-law relationship and a long-term cross-correlation among the pairs of series.

And the Hurst exponent $H_{xy}(q)$ decreases with q ((c) in Figures 7–10), which indicates a multifractality in the cross-correlation among all the pairs. The decreasing

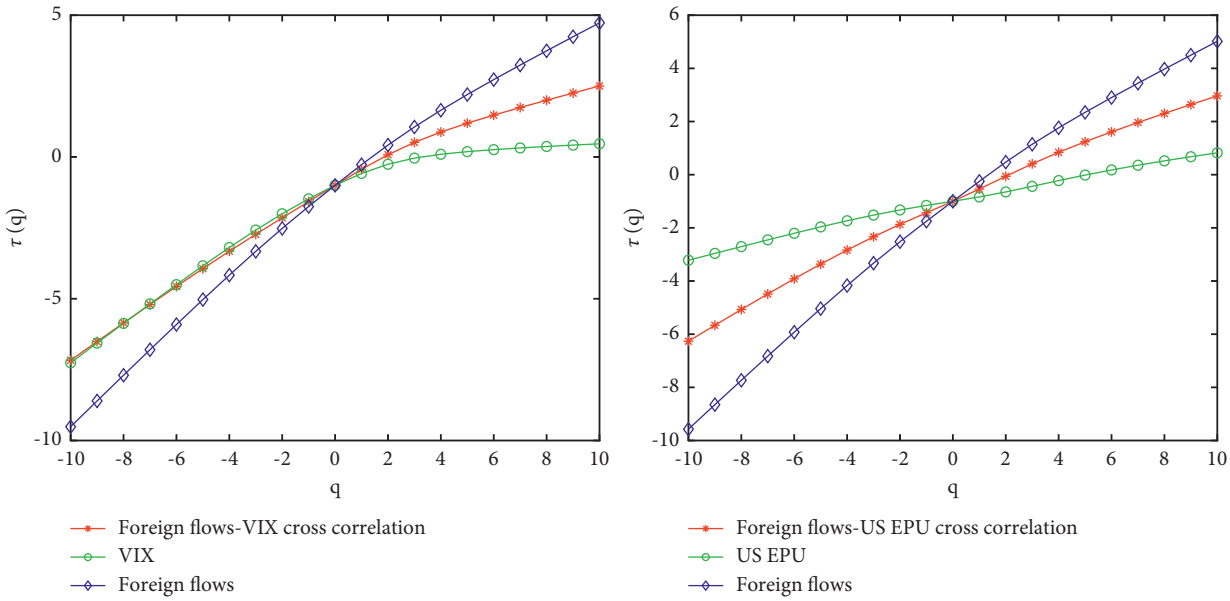


FIGURE 5: The Renyi exponent $\tau(q)$ versus q for the foreign flows in A-share market and VIX and the foreign flows in A-share market and US EPU.

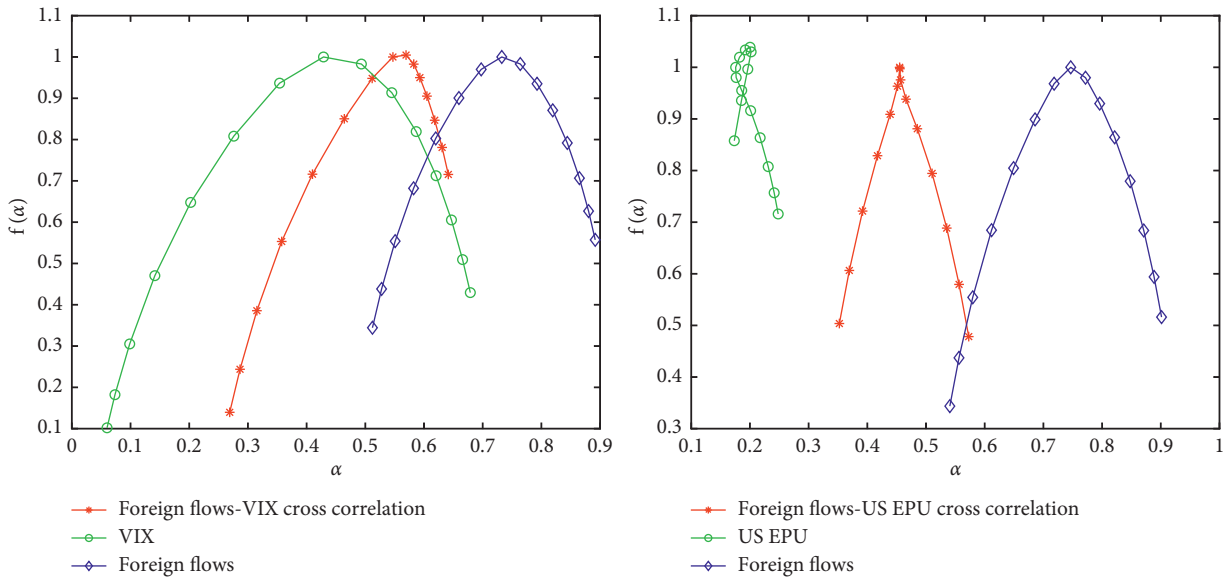


FIGURE 6: The multifractal spectrum $f(\alpha)$ versus α for the foreign flows in A-share market and VIX and the foreign flows in A-share market and US EPU.

amplitude of $H_{xy}(q)$ for HGT and VIX, and SGT and VIX is greater than that for HGT and US EPU, and SGT and US EPU. And most values of $H_{xy}(q)$ for HGT and US EPU, and SGT and US EPU are lower than 0.5. And the $H_{xy}(q)$ is more steady than that for HGT and VIX, and SGT and VIX. Besides, the values of $H_{xy}(q)$ for SGT and VIX, except $q = 10$, are greater than 0.5, indicating that the cross-correlation is persistent, while $H_{xy}(q)$ is greater than 0.5 when q is less than 2, and it is less than 0.5 when q is greater than 2. This is consistent with the results for the overall flows and VIX. All the curves of the Renyi exponent $\tau(q)$ versus q ((d) in Figures 7–10) for four pairs of series are nonlinear, which confirm the multifractal cross-correlation. Furthermore, the results of multifractal spectrum analysis ((e) in Figures 7–10) verify the cross-correlation. And the strength of the cross-correlation between the foreign flows and VIX is larger than that between the foreign flows and US EPU. In conclusion, the above results are consistent with those in Section 4.

Second, we analyze the cross-correlation for the period after the implementation of Shenzhen-Hong Kong Stock Connect. Overall, the results confirm the multifractal cross-correlation between the foreign flows and VIX and the foreign flows and US EPU, which are consistent with

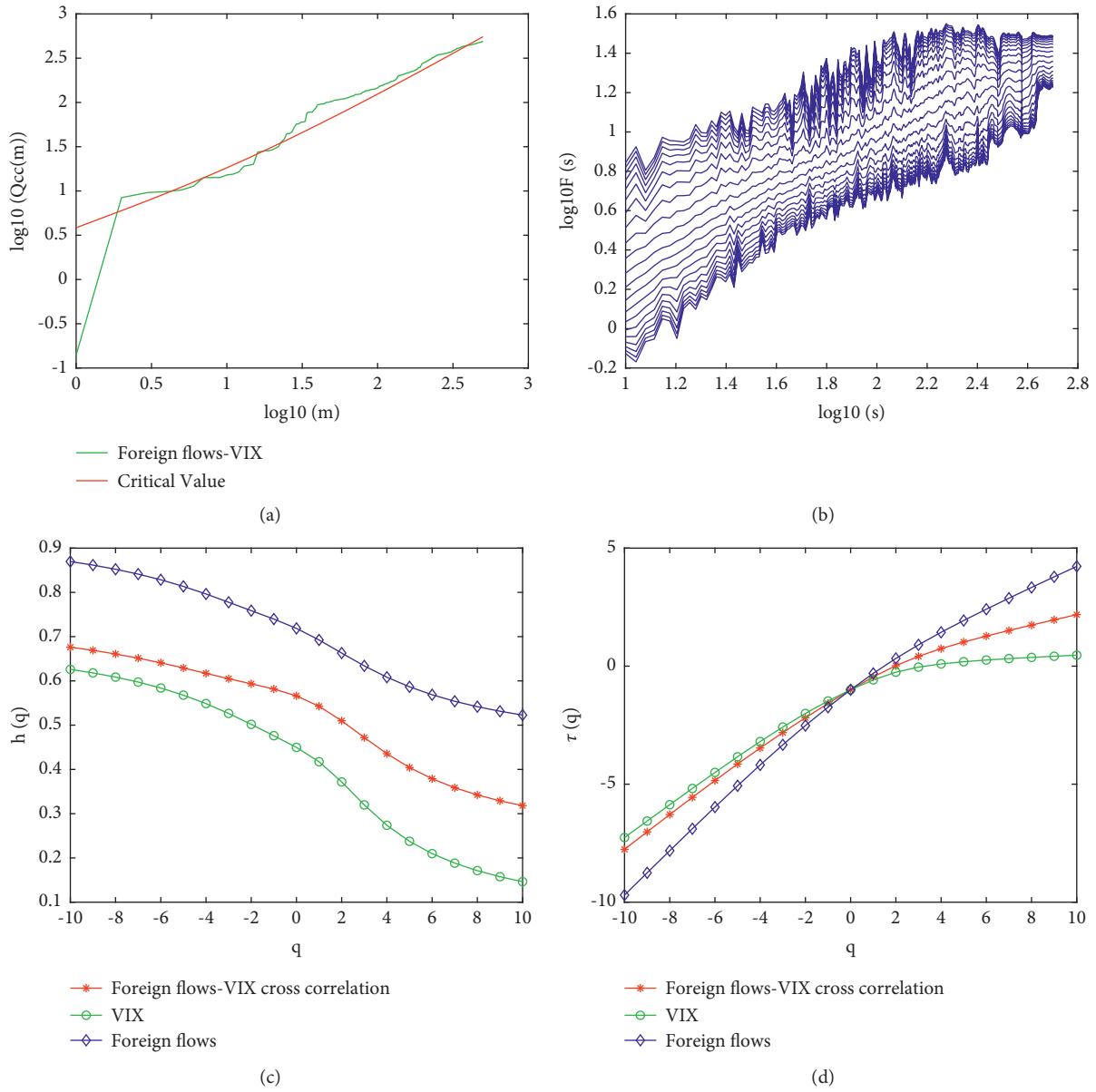
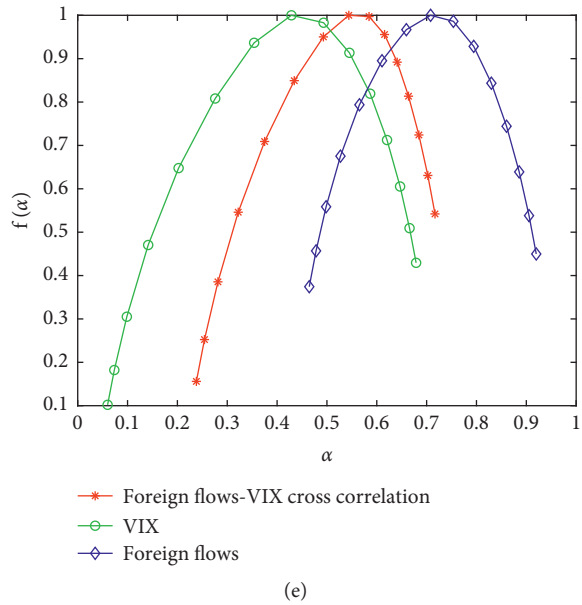
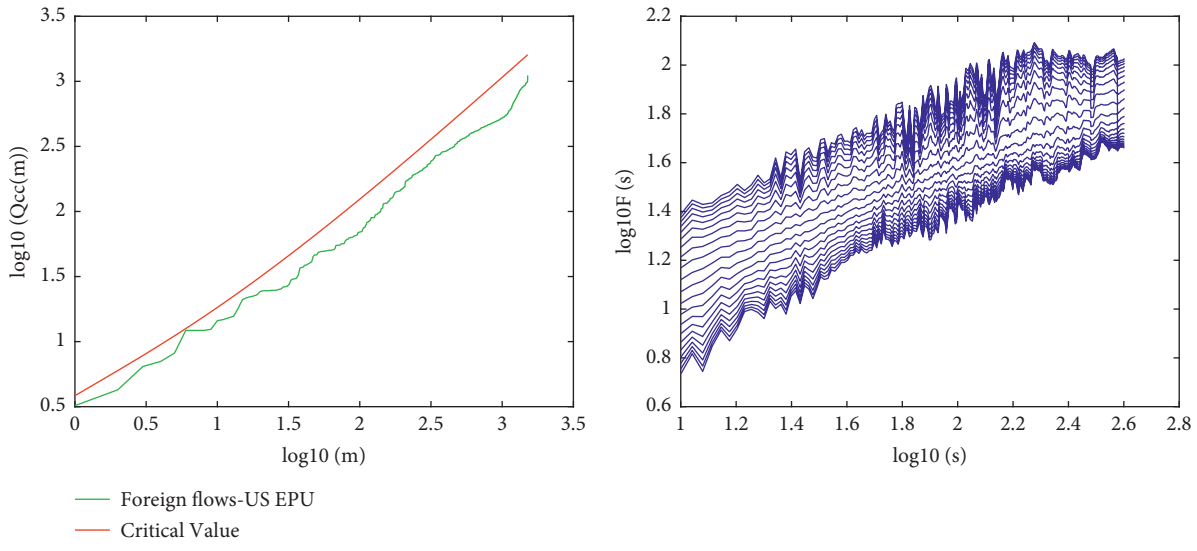


Figure 7: Continued.



(e)

Figure 7: Cross-correlation analysis and MF-DCCA analysis between HGT and VIX index. (a) Cross-correlation statistics $Q_{cc}(m)$. (b) Log-log plots of $F_q(s)$ versus s. (c) Generalized Hurst exponent of $H(q)$ versus q. (d) Renyi exponent of $\tau(q)$ versus q. (e) Multifractal spectrum $f(\alpha)$ versus α.



(a) (b)

Figure 8: Continued.

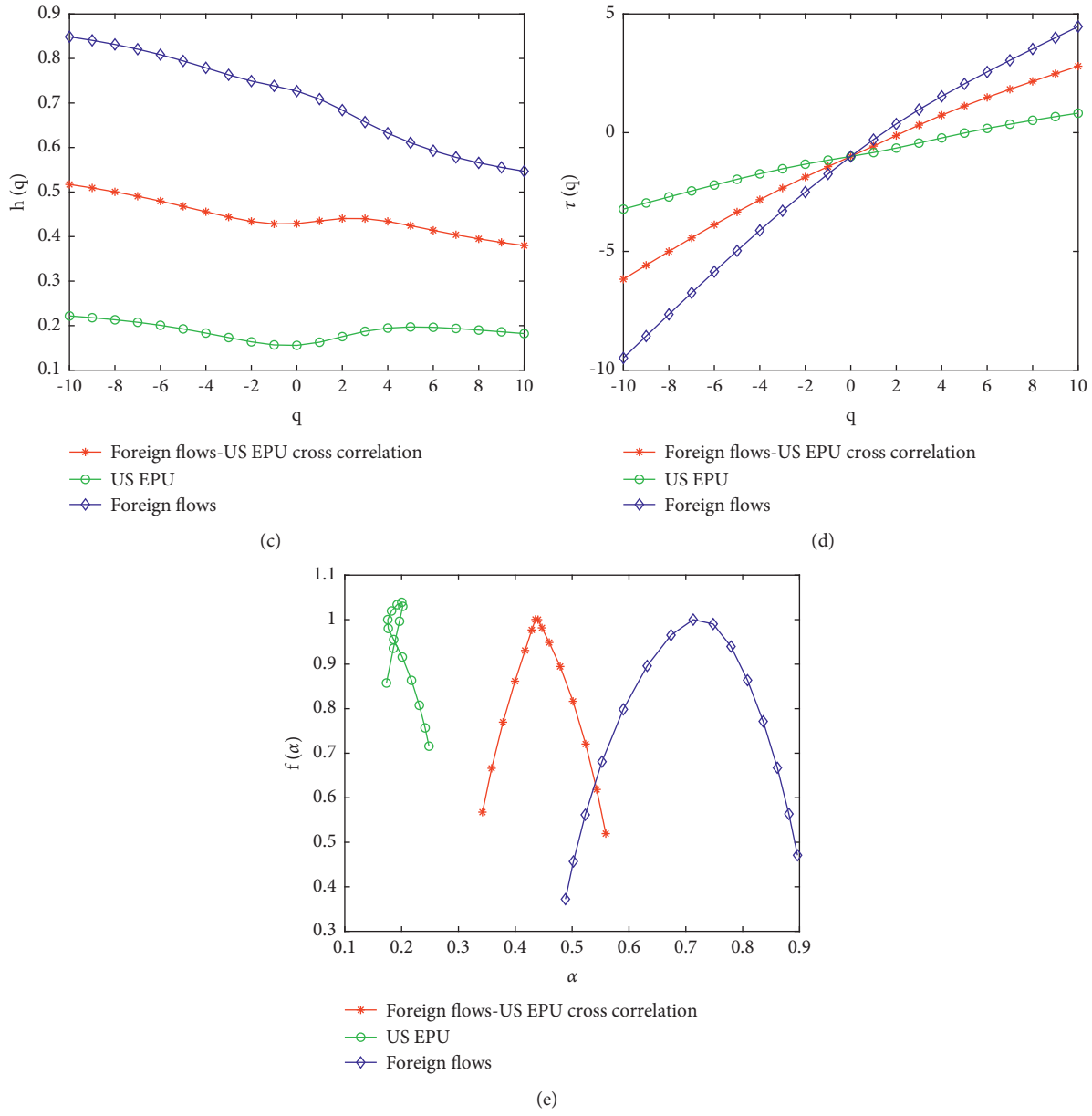


Figure 8: Cross-correlation analysis and MF-DCCA analysis between HGT and US EPU index. (a) Cross-correlation statistics $Q_{cc}(m)$. (b) Log-log plots of $F_q(s)$ versus s. (c) Generalized Hurst exponent of $H(q)$ versus q. (d) Renyi exponent of $\tau(q)$ versus q. (e) Multifractal spectrum $f(\alpha)$ versus α.

those for the whole period. After the implementation of Shenzhen-Hong Kong Stock Connect, most values of $H_{xy}(q)$ for foreign flows and VIX are greater than 0.5 (figure (c) in Figures 11 and 12), indicating the positive and persistent cross-correlation. This may be the result of the strong confidence in Chinese economy in recent years, especially the outstanding epidemic prevention and control after the break of COVID-19. And this result is consistent with the foreign investors' significant negative feedback trading during the COVID-19 crisis [14].

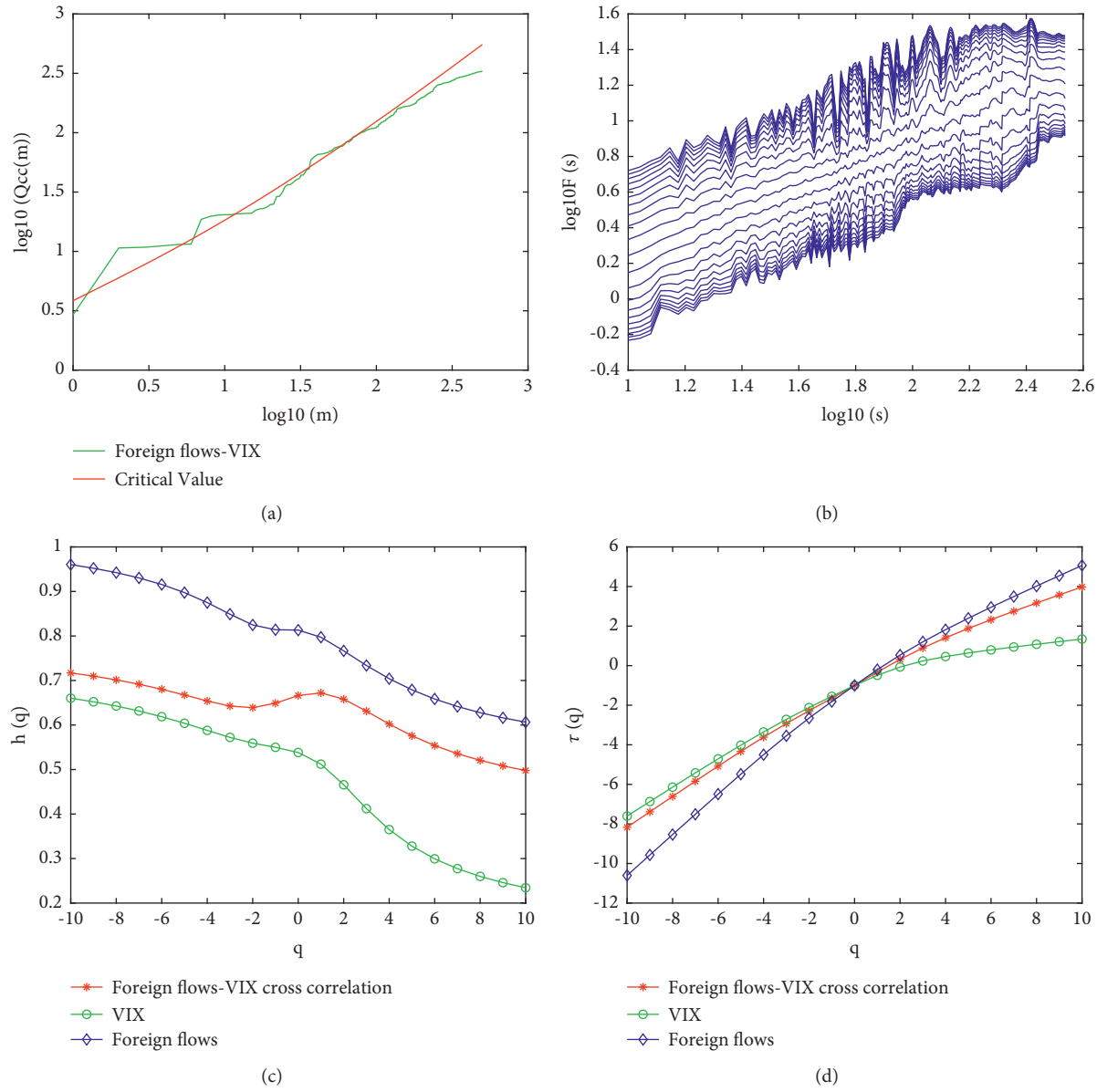
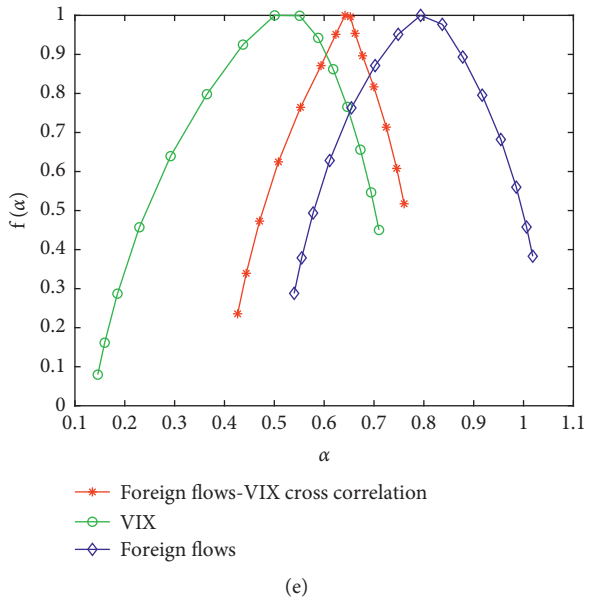
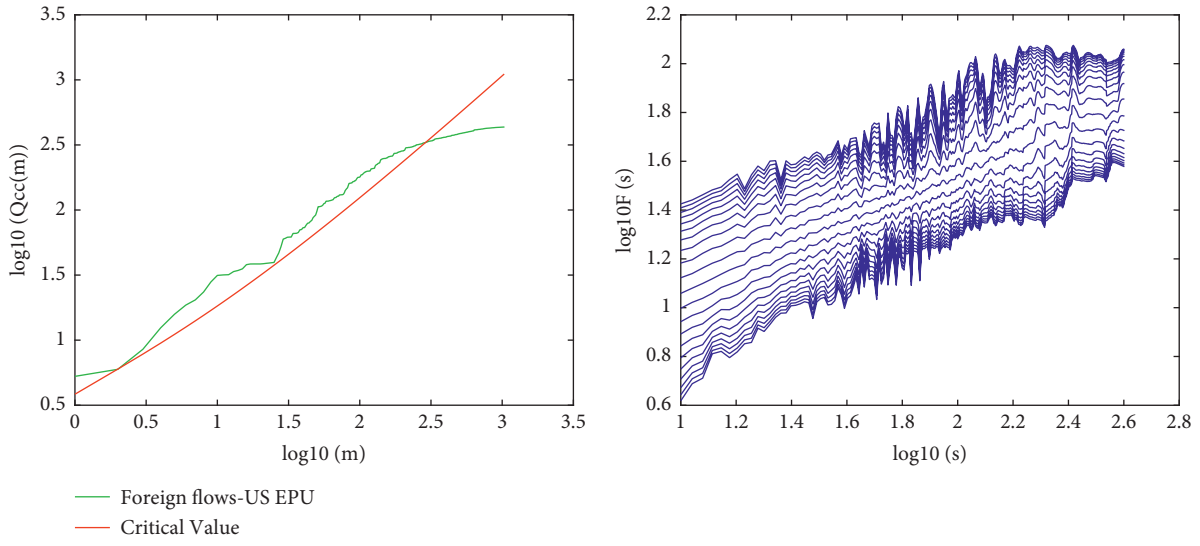


Figure 9: Continued.



(e)

Figure 9: Cross-correlation analysis and MF-DCCA analysis between SGT and VIX index. (a) Cross-correlation statistics $Q_{cc}(m)$. (b) Log-log plots of $F_q(s)$ versus s. (c) Generalized Hurst exponent of $H(q)$ versus q. (d) Renyi exponent of $\tau(q)$ versus q. (e) Multifractal spectrum $f(\alpha)$ versus α.



(a) (b)

Figure 10: Continued.

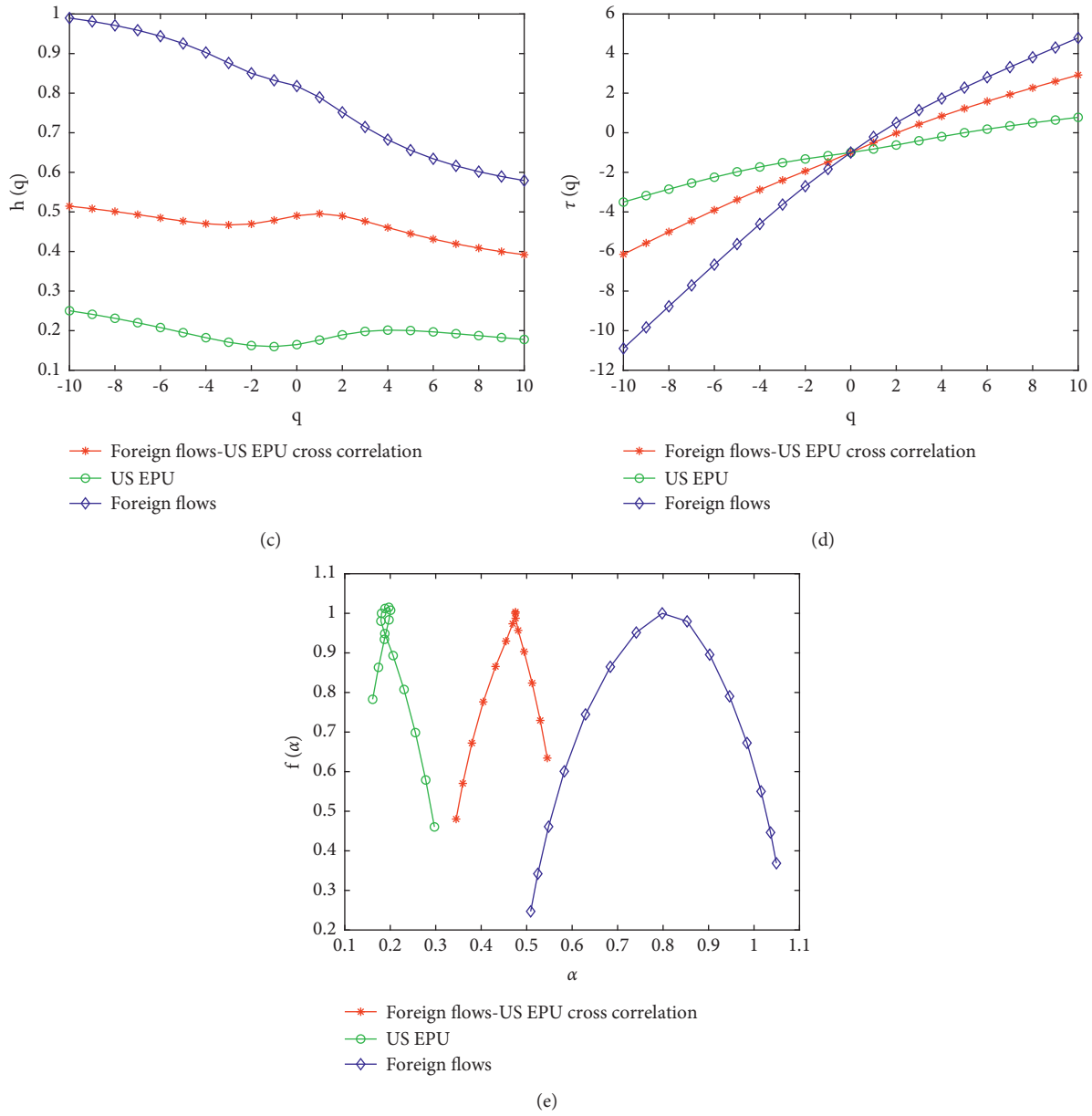


Figure 10: Cross-correlation analysis and MF-DCCA analysis between SGT and US EPU index. (a) Cross-correlation statistics $Q_{cc}(m)$. (b) Log-log plots of $F_q(s)$ versus s. (c) Generalized Hurst exponent of $H(q)$ versus q. (d) Renyi exponent of $\tau(q)$ versus q. (e) Multifractal spectrum $f(\alpha)$ versus α.

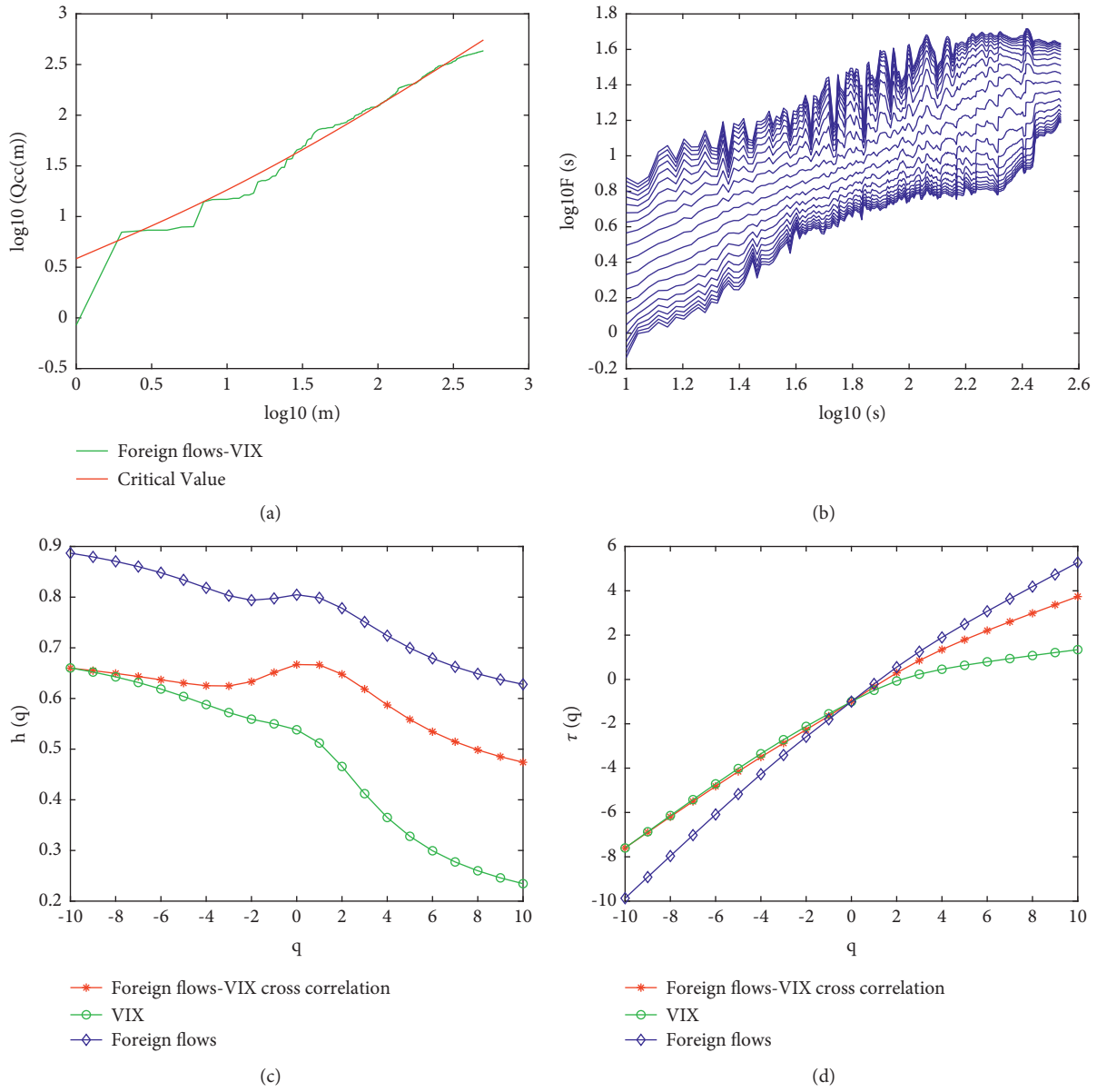
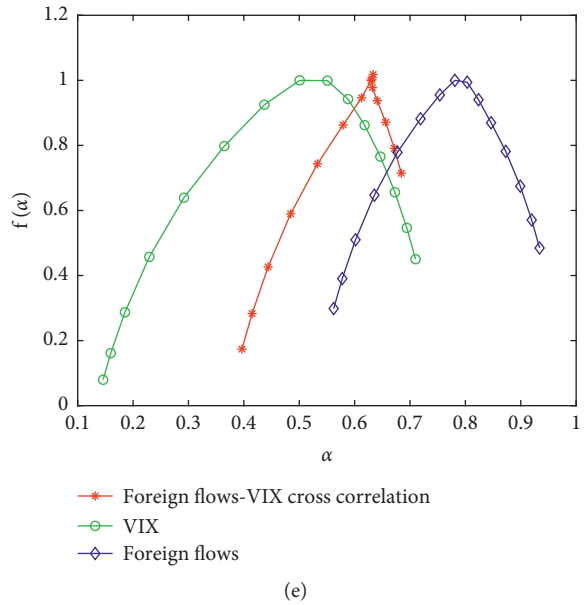
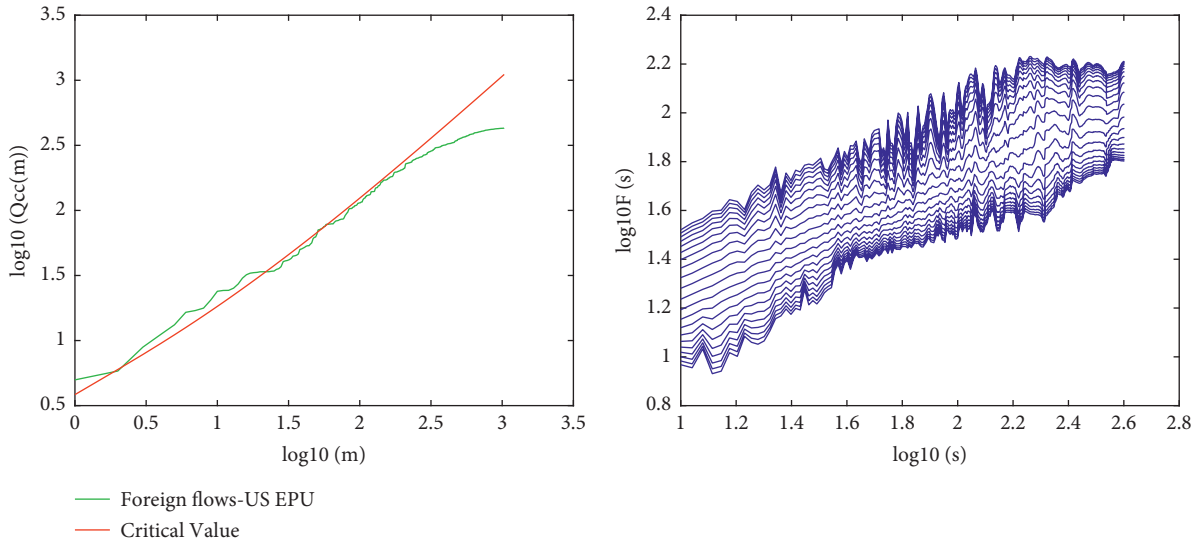


Figure 11: Continued.



(e)

Figure 11: Cross-correlation analysis and MF-DCCA analysis between the foreign flows after the implementation of Shenzhen-Hong Kong Stock Connect and VIX index. (a) Cross-correlation statistics $Q_{cc}(m)$. (b) Log-log plots of $F_q(s)$ versus s. (c) Generalized Hurst exponent of $H(q)$ versus q. (d) Renyi exponent of $\tau(q)$ versus q. (e) Multifractal spectrum $f(\alpha)$ versus α.



(a) (b)

Figure 12: Continued.

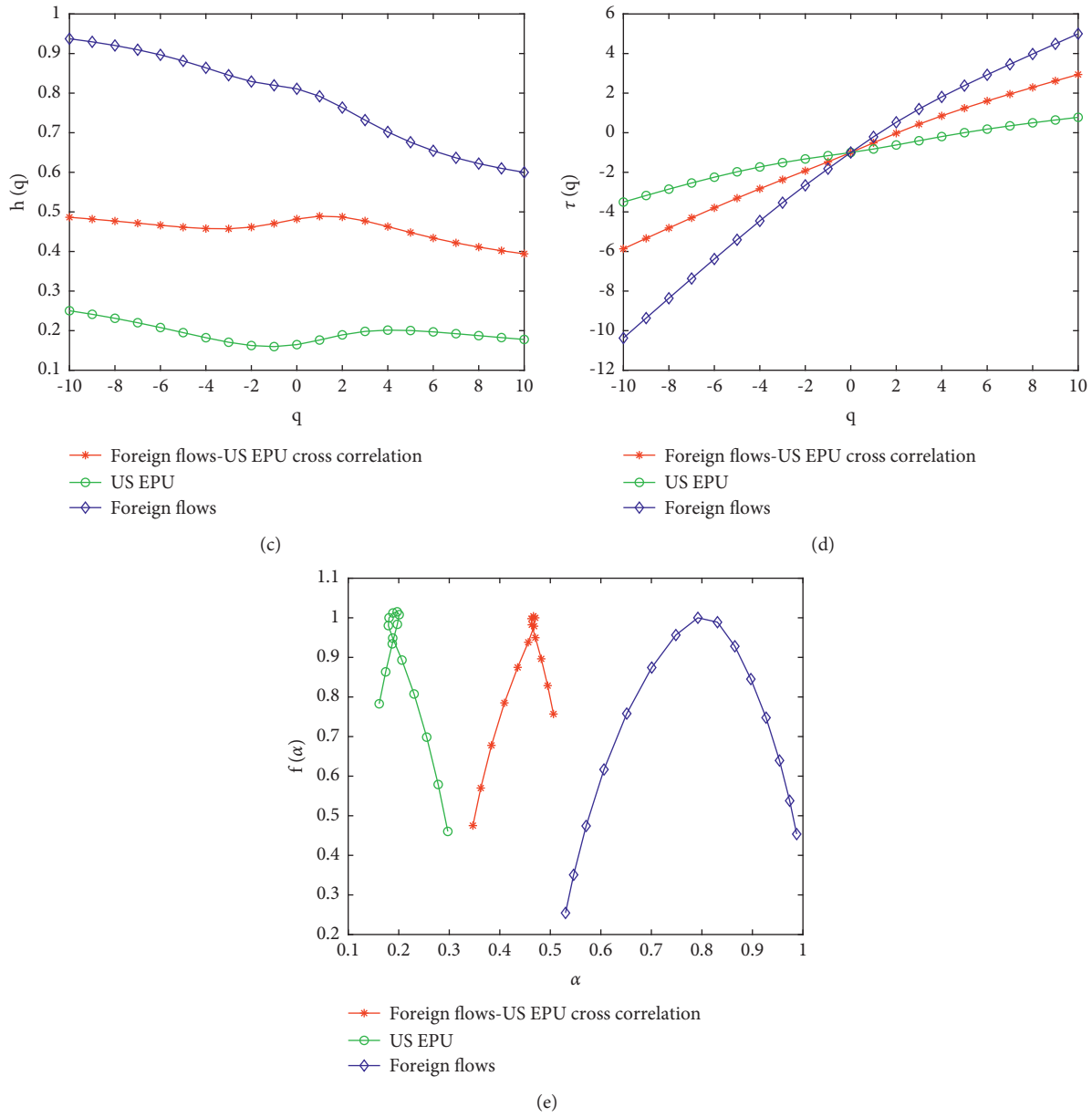


Figure 12: Cross-correlation analysis and MF-DCCA analysis between the foreign flows after the implementation of Shenzhen-Hong Kong Stock Connect and US EPU index. (a) Cross-correlation statistics $Q_{cc}(m)$. (b) Log-log plots of $F_q(s)$ versus s. (c) Generalized Hurst exponent of $H(q)$ versus q. (d) Renyi exponent of $\tau(q)$ versus q. (e) Multifractal spectrum $f(\alpha)$ versus α.

6. Conclusions

The relationship between the foreign flows in emerging markets and the influence factors in home market has been a main undertaking in international finance and microstructure literature. However, most literatures only explore the linear relationship. Few have focused on the nonlinear relationship. Using the MF-DCCA method, this paper investigates the cross-correlations between the foreign flows in A-share market and the push factors, namely, the uncertainties of market, economy, and policy in the home market, which are measured by the VIX index and the US EPU index, respectively. The cross-correlation statistics qualitatively show the existence of the cross-correlation between the foreign flows and the VIX index and the foreign flows and the US EPU index. Then, the MF-DCCA method confirms their long-term cross-correlation from the quantitative perspective. In addition, the cross-correlation between the foreign flows and VIX is persistent, when the fluctuations are small. However, the cross-correlation is antipersistent, when the fluctuations are large. In contrast, the cross-correlation between the foreign flows and US EPU is antipersistent and steady. Besides, the strength of the cross-correlation between the foreign flows and VIX is larger than that between the

foreign flows and US EPU. And the results are robust using the separate foreign flows, namely, the flows via Shanghai-Hong Kong Stock Connect Program and Shenzhen-Hong Kong Stock Connect Program, and flows with shortening period.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

The authors contributed equally to the writing of this paper. The authors read and approved the final manuscript.

Acknowledgments

This study was supported by the Fundamental Research Funds from Capital University of Economic and Business (Grant No. XRZ2020041), the National Natural Science Foundation of China (Grant No. 72103210), the Humanity and Social Science Youth Foundation of the Ministry of Education of China (Grant No. 18YJC790115), and the Soft Science Foundation of Science and Technology Department of Guangdong, China (Grant No. 2019A101002019).

References

[1] G. A. Calvo, L. Leiderman, and C. M. Reinhart, "Capital inflows and real exchange rate appreciation in Latin America: the role of external factors," *Staff Papers - International Monetary Fund*, vol. 40, no. 1, pp. 108–151, 1993.

[2] E. Fernandez-Arias, "The new wave of private capital inflows: push or pull?" *Journal of Development Economics*, vol. 48, no. 2, pp. 389–418, 1996.

[3] M. P. Taylor and L. Sarno, "Capital flows to developing countries: long- and short-term determinants," *The World Bank Economic Review*, vol. 11, no. 3, pp. 451–470, 1997.

[4] P.-R. Agénor, "The surge in capital flows: analysis of "pull" and "push" factors," *International Journal of Finance & Economics*, vol. 3, no. 1, pp. 39–57, 1998.

[5] P. Chuhan, S. Claessens, and N. Mamingi, "Equity and bond flows to Latin America and Asia: the role of global and country factors," *Journal of Development Economics*, vol. 55, no. 2, pp. 439–463, 1998.

[6] K. J. Forbes and F. E. Warnock, "Capital flow waves: surges, stops, flight, and retrenchment," *Journal of International Economics*, vol. 88, no. 2, pp. 235–251, 2012.

[7] M. Fratzscher, "Capital flows, push versus pull factors and the global financial crisis," *Journal of International Economics*, vol. 88, no. 2, pp. 341–356, 2012.

[8] A.-M. Fuertes, K. Phylaktis, and C. Yan, "Hot money in bank credit flows to emerging markets during the banking globalization era," *Journal of International Money and Finance*, vol. 60, pp. 29–52, 2016.

[9] R. Koepke, "What drives capital flows to emerging markets? A survey of the empirical literature," *Journal of Economic Surveys*, vol. 33, no. 2, pp. 516–540, 2019.

[10] E. Borensztein and R. G. Gelos, "A panic-prone pack? The behavior of emerging market mutual funds," *IMF Staff Papers*, vol. 50, no. 1, pp. 43–63, 2003.

[11] T. Omay and P. Iren, "Behavior of foreign investors in the Malaysian stock market in times of crisis: a nonlinear approach," *Journal of Asian Economics*, vol. 60, pp. 85–100, 2019.

[12] M. Hood, A. Kamesaka, J. Nofsinger, and T. Tamura, "Investor response to a natural disaster: evidence from Japan's 2011 earthquake," *Pacific-Basin Finance Journal*, vol. 25, pp. 240–252, 2013.

[13] J. R. Gerlach and Y. Yook, "Political conflict and foreign portfolio investment: evidence from North Korean attacks," *Pacific-Basin Finance Journal*, vol. 39, pp. 178–196, 2016.

[14] T. Bing and H. Ma, "COVID-19 pandemic effect on trading and returns: evidence from the Chinese stock market," *Economic Analysis and Policy*, vol. 71, pp. 384–396, 2021.

[15] H. Choe, B. C. Kho, and R. M. Stulz, "Do foreign investors destabilize stock markets? The Korean experience in 1997," *Journal of Financial Economics*, vol. 54, no. 2, pp. 227–264, 1999.

[16] M. Bowe and D. Domuta, "Investor herding during financial crisis: a clinical study of the Jakarta Stock Exchange," *Pacific-Basin Finance Journal*, vol. 12, no. 4, pp. 387–418, 2004.

[17] Z.-Q. Jiang, W.-J. Xie, W.-X. Zhou, and D. Sornette, "Multifractal analysis of financial markets: a review," *Reports on Progress in Physics*, vol. 82, no. 12, p. 125901, 2019.

[18] P.-F. Dai, X. Xiong, and W.-X. Zhou, "Visibility graph analysis of economy policy uncertainty indices," *Physica A: Statistical Mechanics and Its Applications*, vol. 531, p. 121748, 2019.

[19] R. Zhao and P.-F. Dai, "A multifractal cross-correlation analysis of economic policy uncertainty: evidence from China and US," *Fluctuation and Noise Letters*, vol. 20, no. 5, p. 2150041, 2021.

[20] W. X. Zhou, "Multifractal detrended cross-correlation analysis for two nonstationary signals," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 77, no. 6, p. 066211, 2008.

[21] W. Zhang, Y. Li, Z. Zhang, and D. Shen, "The dynamic cross-correlations between foreign news, local news and stock returns," *Physica A: Statistical Mechanics and Its Applications*, vol. 509, pp. 861–872, 2018.

[22] Z. Zhang, Y. Zhang, D. Shen, and W. Zhang, "The dynamic cross-correlations between mass media news, new media news, and stock returns," *Complexity*, vol. 2018, Article ID 7619494, 11 pages, 2018.

[23] Z. Zhang, Y. Zhang, and D. Shen, "The dynamic correlations between mass media news and new media news in stock market," *Frontiers in Physics*, vol. 8, p. 179, 2020.

[24] K. Xu, Y. Pang, and J. Han, "Dynamic cross-correlation between online sentiment and stock market performance: a global view," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 6674379, 11 pages, 2021.

[25] R. Zhao and Y. Cui, "Dynamic cross-correlations analysis on economic policy uncertainty and US dollar exchange rate: AMF-DCCA perspective," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 6668912, 9 pages, 2021.

[26] S. Li, X. Lu, and X. Liu, "Dynamic relationship between Chinese RMB exchange rate index and market anxiety: a new perspective based on MF-DCCA," *Physica A: Statistical Mechanics and its Applications*, vol. 541, p. 123405, 2020.

[27] S. Ye, P.-F. Dai, H. T. Nguyen, and N. Q. A. Huynh, "Is the cross-correlation of EU carbon market price with policy uncertainty really being? A multiscale multifractal perspective," *Journal of Environmental Management*, vol. 298, p. 113490, 2021.

[28] M. Foglia and P. F. Dai, ""Ubiquitous uncertainties": spillovers across economic policy uncertainty and cryptocurrency uncertainty indices," *Journal of Asian Business and Economic Studies*, 2021.

[29] X. Xiong, K. Xu, and D. Shen, "Dynamic cross-correlations between investors' attention and CSI300 index futures," *Fluctuation and Noise Letters*, vol. 18, no. 4, p. 1950022, 2019.

[30] W. Zhang, P. Wang, X. Li, and D. Shen, "Multifractal detrended cross-correlation analysis of the return-volume relationship of bitcoin market," *Complexity*, vol. 2018, Article ID 8691420, 20 pages, 2018.

[31] W. Zhang, P. Wang, X. Li, and D. Shen, "Quantifying the cross-correlations between online searches and Bitcoin market," *Physica A: Statistical Mechanics and its Applications*, vol. 509, pp. 657–672, 2018.

[32] Z. Zhang, Y. Zhang, D. Shen, and W. Zhang, "The cross-correlations between online sentiment proxies: evidence from Google Trends and Twitter," *Physica A: Statistical Mechanics and its Applications*, vol. 508, pp. 67–75, 2018.

[33] G. Sarwar, "Is VIX an investor fear gauge in BRIC equity markets?" *Journal of Multinational Financial Management*, vol. 22, no. 3, pp. 55–65, 2012.

[34] S. R. Baker, N. Bloom, and S. J. Davis, "Measuring economic policy uncertainty*," *Quarterly Journal of Economics*, vol. 131, no. 4, pp. 1593–1636, 2016.

[35] B. Podobnik and H. E. Stanley, "Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series," *Physical Review Letters*, vol. 100, no. 8, p. 084102, 2008.

[36] S. Shadkhoo and G. R. Jafari, "Multifractal detrended cross-correlation analysis of temporal and spatial seismic data," *The European Physical Journal B*, vol. 72, no. 4, pp. 679–683, 2009.

[37] Y. Yuan, X.-T. Zhuang, and X. Jin, "Measuring multifractality of stock price fluctuation using multifractal detrended fluctuation analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 388, no. 11, pp. 2189–2197, 2009.

[38] M. Arak and N. Mijid, "The VIX and VXN volatility measures: fear gauges or forecasts?" *Derivatives Use, Trading and Regulation*, vol. 12, no. 1, pp. 14–27, 2006.

[39] M. M. Copeland and T. E. Copeland, "Market timing: style and size rotation using the VIX," *Financial Analysts Journal*, vol. 55, no. 2, pp. 73–81, 1999.

[40] R. E. Whaley, "Understanding the VIX," *Journal of Portfolio Management*, vol. 35, no. 3, pp. 98–105, 2009.

[41] R. E. Whaley, "The investor fear gauge," *Journal of Portfolio Management*, vol. 26, no. 3, pp. 12–17, 2000.

[42] B. Julio and Y. Yook, "Political uncertainty and corporate investment cycles," *The Journal of Finance*, vol. 67, no. 1, pp. 45–83, 2012.

[43] Y. Sha, C. Kang, and Z. Wang, "Economic policy uncertainty and mergers and acquisitions: evidence from China," *Economic Modelling*, vol. 89, pp. 590–600, 2020.

[44] S. A. Al-Thaqeb and B. G. Algharabali, "Economic policy uncertainty: a literature review," *The Journal of Economic Asymmetries*, vol. 20, p. e00133, 2019.

[45] P.-F. Dai, X. Xiong, and W.-X. Zhou, "A global economic policy uncertainty index from principal component analysis," *Finance Research Letters*, vol. 40, p. 101686, 2021.

[46] P. F. Dai, X. Xiong, and W. X. Zhou, "The role of global economic policy uncertainty in predicting crude oil futures volatility: evidence from a two-factor GARCH-MIDAS model," 2020, https://arxiv.org/abs/2007.12838.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 7390887, 15 pages
https://doi.org/10.1155/2021/7390887

Research Article

Socially Responsible Investment Portfolio Construction with a Double-Screening Mechanism considering Machine Learning Prediction

Jun Zhang and Xuedong Chen

School of Management and Engineering, Capital University of Economics and Business, Beijing 100070, China

Correspondence should be addressed to Xuedong Chen; nmcxd@cueb.edu.cn

Received 3 September 2021; Accepted 26 October 2021; Published 17 November 2021

Academic Editor: Dehua Shen

Copyright © 2021 Jun Zhang and Xuedong Chen. 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.

Although socially responsible investment (SRI) has developed into an important investment style, only a small number of studies discuss SRI portfolio construction. In view of the overwhelming breakthrough of machine learning in prediction, this paper proposes SRI portfolio construction models by combining a double-screening mechanism considering machine learning prediction and an extended global minimum variance (GMV) model (or extended maximum Sharpe ratio (MSPR) model), which are, respectively, named double-screening socially responsible investment (DSSRI) portfolio models I and II. The proposed models consist of two stages, i.e., stock screening and asset allocation. First, this paper develops a novel double-screening mechanism incorporating environmental, social, and corporate governance (ESG) and return potential criteria to ensure that high-quality stocks with good ESG performance and high-return potential are input into the optimal portfolio. Specifically, to obtain accurate stock return predictions, an extreme learning machine model optimized by the genetic algorithm is employed to predict stock prices. Next, to trade off the financial and ESG objectives of SRI investors, an extended GMV model (or extended MSPR model) considering the ESG factor is introduced to determine the capital allocation proportion of the stocks. We take the A-share market of China as the sample to verify the effectiveness of the proposed models. The empirical results demonstrate that compared with alternative models, the proposed models can yield better annualized return and ESG score performance as well as competitive Sharpe ratio performance.

1. Introduction

Socially responsible investment (SRI) considers personal values and societal concerns in investment decisions [1, 2]. SRI investors consider investing as an extension of their lifestyle and thus are willing to incorporate their social beliefs and values into financial activities [3]. With the increasing salience of social and environmental issues and the gradual awakening of investors' consciousness of environmental, social, and corporate governance (ESG) factors, SRI has presented rapid growth [4] and has become an important investment style. As of 2020, there were 3038 signatories in the global financial market agreeing to abide by the Principles for Responsible Investment supported by the United Nations, encompassing management assets of 103.4 trillion dollars. Compared with the levels in 2019, this number of signatories represented an increase of 20%, and the management asset volume exhibited an increase of 666 trillion dollars (see Principles for Responsible Investment at http://www.unpri.org).

In fact, SRI has attracted wide attention from academia over the past decades. Researchers have explored SRI related topics from various perspectives, including the motivation of SRI investors [5, 6], socially responsible measurement [7, 8], and the relationship between ESG rating and finance returns of the firm [9, 10]. For more references, see [11]. However, only a small number of studies focuses on SRI portfolios. Unlike the conventional portfolio construction, SRI portfolio construction requires a stock screening process to ensure that the stocks in the portfolio are in line with SRI

investors' social values. A few researchers have made valuable explorations of stock screening. For example, Auer [12] applied different screening methods to test the performance of SRI. The empirical results showed that compared with positive screening, which can damage investment performance, negative screenings with low cut-off rates are better. Kempf and Osthoff [13] compared the three SRI screening mechanisms of negative, positive, and best-in-class screening through an investment strategy of buying stocks with high social responsibility ratings and selling stocks with low social responsibility ratings. The best-in-class screening was proven to generally obtain the highest abnormal returns. Barnett and Salomon [14] evaluated the effect of the intensity of social screening on the financial performance of SRI funds. They concluded that the financial returns of the SRI funds decrease with increasing screening intensity, while a rebound in financial returns occurs once the screening intensity reaches a maximum. It is worth mentioning that the stock screening conditions in previous studies were based solely on ESG criteria. However, as discussed by Diouf et al. [15], MacKenzie and Lewis [16], and Rosen et al. [17], SRI investors are concerned with both ESG and financial performance at the same time. There is no conclusive evidence that stocks screened based on ESG criteria can deliver high returns [2, 18–20]. Therefore, it is necessary to consider the return potential of candidate stocks in the screening process to ensure that the screened stocks meet both the ESG criteria and the return demand of SRI investors.

To screen the potential high-return stocks, reliable stock return prediction is critical. Machine learning methods with excellent learning capability have been widely used for stock return prediction. The representative machine learning models for stock prediction include support vector machine regression (SVR) [21, 22], artificial neural networks (ANNs) [23, 24], fuzzy logic system [25, 26], deep neural network (DNN) [27, 28], convolutional neural network [29, 30], long short-term memory model (LSTM) [31–33], and ensemble learning [34, 35]. Recently, extreme learning machine (ELM), which is a novel machine learning model, has received attention due to its powerful performance. For instance, Lie et al. [36] exploited the high accuracy and quick learning advantages of ELM to predict the stock price and then generate trading signals. Dash and Dash [37] employed ELM to train a computational efficient functional link artificial neural network and obtained better prediction effects for stock index movements than those with the back-propagation (BP) algorithm. Wei et al. [38] effectively improved the accuracy of exchange rate forecasting using a hybrid model of kernel-based ELM. Jiang et al. [39] developed two hybrid models for stock price prediction that employed ELM as the predictor. The empirical results showed that the proposed models had high prediction accuracy. Although ELM is capable of achieving good prediction accuracy, its main drawback is the mechanism of randomly generating input weights and biases of the hidden layer, which may lead to inferior solutions and thus undermine its stability and performance [40, 41]. To overcome this drawback, an effective strategy is to introduce an intelligent algorithm to optimize the input weights and biases of the hidden layer of ELM [40, 42, 43]. The genetic algorithm (GA), a classic and excellent intelligent algorithm, has been utilized to optimize machine learning models, i.e., ELM [44–46] and BP [47]. In particular, ELM optimized by the GA (GA-ELM) has been successfully applied in various prediction topics, such as CO_2 emission [44], wind power [45], and gas concentration [46]. Therefore, this paper introduces the GA-ELM to predict stock prices and then screens stocks with high-return potential based on the prediction results.

Once the stock screening is completed, SRI investors next need to determine the optimal investment weight of screened stocks. The mean-variance (MV) model, as the basis of modern portfolio theory, is still a common choice for SRI portfolio optimization. On this topic, only a handful of scholars have done relevant research. For example, Oikonomou et al. [48] tried to explore optimal ways to build SRI portfolios. They applied six models (three simplistic and three sophisticated) to construct SRI portfolios and compared the performance of each. The results indicated that the SRI portfolios based on the sophisticated optimization models represented by MV had superior risk-return trade-offs relative to the simplistic models (e.g., those with naive diversification). Branch et al. [49] analysed the performance of quantitative methodologies in ESG portfolio construction. They concluded that quantitative methods are effective at minimizing risk but could lead to unwanted exposure. Although the aforementioned studies performed meaningful and in-depth analyses, they either did not take the ESG factor into account in the portfolio optimization process or took ESG as a constraint condition instead of an optimization objective, thus failing to cater to investors' demand to maximize ESG. In [50], an extended MV was developed to construct an ESG portfolio, where minimizing the negative Sharpe ratio with the ESG score was taken as a novel variable in the objective function. The proposed model showed considerable superiority in ESG scores over the traditional MV model. However, this research lacked an effective screening process to guarantee that high-quality stocks with both high ESG and high-return potential were input into the SRI portfolio, leading to mediocre return performance. Similarly, several modified versions of MV models incorporating ESG objectives (see, for example, Utz et al. [51], Calvo et al. [52], Stephan et al. [53] and Bilbao et al. [54]) also fail to preselect high-quality stocks before SRI portfolio construction.

Based on the above analysis, this paper proposes two SRI portfolio construction models called double-screening socially responsible investment (DSSIR) portfolio models I and II (abbreviated as DSSIR-I and DSSIR-II), which combine a double-screening mechanism and portfolio optimization models. Specifically, the novel double-screening mechanism consists of ESG and return potential screening. First, in the ESG screening, stocks with inferior ESG scores are excluded from the investment universe to ensure that the screened stocks align with the social values of investors. Subsequently, in the return potential screening, to include stocks with high-return potential into the portfolio, the GA-ELM model is introduced to predict the price of the screened stock. Finally, the stocks with high ESG and high-return potential,

as high-quality stocks, are input into the portfolio. After this stock screening, the portfolio optimization model is employed to weigh the screened high-quality stocks. In this work, to trade off the financial and ESG objectives of SRI investors, an extended global minimum variance (GMV) model (or extended maximum Sharpe ratio (MSPR) model) incorporating the ESG factor, which we call GMV-ESG (MSPR-ESG), is introduced as a portfolio optimization model. The main contributions of this paper are as follows:

(1) Unlike the majority of studies on SRI portfolio construction, this paper incorporates stock prediction into the portfolio selection, thereby providing a new method for SRI portfolio construction.

(2) This paper develops two novel SRI portfolio models called DSSRI-I and DSSRI-II, in which a double-screening mechanism is designed to ensure high-quality stocks (both high ESG and return potential) being the component of the SRI portfolio.

(3) The proposed models are compared with alternative models, including SSRI-GMV (or SSRI-MSPR), DSSRI-I (ELM) (or DSSRI-II (ELM)), DSSRI-I (SVR) (or DSSRI-II (SVR)), DSSRI-I (DNN) (or DSSRI-II (DNN)), and DSSRI-I (LSTM) (or DSSRI-II (LSTM)). The comparisons demonstrate the feasibility and advantages of the proposed models.

The rest of this paper is as follows. In Section 2, we outline the related component models and introduce the proposed models. In Section 3, we describe datasets, parameter settings, and present and discuss the empirical results. In Section 4, we summarize the conclusions and prospects for future work.

2. Methodology

2.1. GA-ELM Model

2.1.1. Extreme Learning Machine. ELM, originally developed by Huang et al. [55], is an advanced algorithm for single-hidden layer feedforward neural network (SLFN), whose network topology consists of three layers, namely, the input layer, hidden layer, and output layer. The traditional algorithm for SLFN like BP algorithm determines the optimal weights through multiple iterations, while the ELM randomly generates input weights and biases of the hidden layer. The ELM can analytically calculate the output weights based on the minimum norm least squares solution. Therefore, compared with the BP algorithm, the ELM is capable of obtaining good generalized performance at an extremely fast speed. Given Q sample pairs (X_i, Y_i), where $i = 1, 2, \ldots, Q$, $X_i = [x_1, x_2, \ldots, x_n]^T$, and $Y_i = [y_1, y_2, \ldots, y_m]^T$, the following equation can be obtained for the ELM owning L neurons in the hidden layer [55]:

$$\sum_{i=1}^{L} \beta_i \varphi\left(w_i x_j + b_i\right) = y_j, \quad j = 1, \ldots, Q, \tag{1}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ denotes the input weight, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ represents the output weight, $\varphi(\cdot)$ is the activation function, and b_i is the bias of the ith neuron in the hidden layer. The matrix form of equation (1) is as follows [55]:

$$H\beta = Y, \tag{2}$$

where

$$H = \begin{bmatrix} \varphi(w_1 x_1 + b_1) & \cdots & \varphi(w_L x_1 + b_L) \\ \vdots & \cdots & \vdots \\ \varphi(w_1 x_N + b_1) & \cdots & \varphi(w_L x_N + b_L) \end{bmatrix}_{N \times L}. \tag{3}$$

H is the hidden layer output matrix. Accordingly, the output weight of the ELM can be obtained using the minimum norm least squares solution [55] as follows:

$$\widehat{\beta} = H^{\dagger} T, \tag{4}$$

where $H^{\dagger}$ represents the Moore–Penrose generalized inverse of H. The ELM can be implemented in the following four steps [55, 56]: (1) determine the number of neurons in the hidden layer; (2) randomly generate the input weights and biases of the hidden layer, w_i and b_i; (3) compute the hidden layer output matrix, H; and (4) estimate the output weight, $\widehat{\beta}$.

2.1.2. Genetic Algorithm. The genetic algorithm is a classic intelligent algorithm first proposed by Holland in 1975 [57]. GA is regarded as a very efficient tool for optimization problems since it is computationally simple and powerful, as well as there are no restrictive assumptions about the solution space [58]. The core idea of the GA, inspired by the theory of 'survival of the fittest', is to imitate the biological evolution process of a population to obtain a globally optimal solution [59]. In solving a practical problem, the GA randomly generates a population of individuals (called chromosomes) as candidate solutions of a practical problem to evolve in successive generations, aiming to eliminate poor solutions within the population; thus, the remaining good solutions are further developed into better ones [60]. As a result, after finite iterations of evolution (until a predefined termination criterion is satisfied), an optimal solution from the population can be obtained. In the process of evolution, the chromosomes of each generation determine those in the next generation through three genetic operations, i.e., selection, crossover, and mutation.

(1) Selection Operation. This operation chooses chromosomes from the current population as parents to produce offspring input within the new population. A common selection technique is the route wheel method in which the probability of selection is proportionate to the fitness of each individual [61]. The fitness is the measure of the quality of the chromosome (solution), which is calculated by presetting the fitness function defined according to the practical problem. Given that $f(C^i)$ is the fitness of the i-th chromosome, the total fitness of the population of G individuals can be expressed as follows [62]:

$$F = \sum_{i=1}^{G} f\left(C^i\right). \tag{5}$$

The accumulated probability of each chromosome is as follows [62]:

$$T_g = \frac{\sum_{i=1}^{g} f\left(C^i\right)}{F}. \tag{6}$$

A random number $R \in [0, 1]$ is generated, if $T_{g-1} < R < T_g$, individual g is selected as the parent.

(2) Crossover Operation. A crossover operation that exchanges some parts of the two parents to produce a new pair of offspring is conducted to seek promising solutions from the search space [63]. Given a crossover probability ε, if generating a random number R with a standard uniform distribution is less than ε, a locus is chosen between two neighbouring chromosomes, i.e., parents 1 and 2. The chromosomes after this locus are exchanged, thereby forming new offsprings 1 and 2 [64]. The crossover processing can be expressed as follows:

$$\begin{cases} \text{parent}^1 = 100110111 \\ \text{parent}^2 = 111000010 \end{cases} \Rightarrow \begin{cases} \text{offspring}^1 = 10011|0010, \\ \text{offspring}^2 = 11100|0111. \end{cases} \tag{7}$$

(3) Mutation Operation. The mutation operation ensures the diversity of individuals and prevents solutions from being too similar to avoid obtaining the local optimal solution [65]. Similar to the crossover operation, given a mutation probability ε, if a random number R with a standard uniform distribution is generated and is less than ε, a locus is chosen randomly on the parent, and the value of the locus is reversed, thereby producing a new offspring [64]. The mutation process can be expressed as follows:

$$\text{parent} = 100|\dot{1}|10101 \Longrightarrow \text{offspring} = 100|\dot{0}|10101. \tag{8}$$

2.1.3. Optimization of the ELM Model Using the GA. Although the ELM has a fast learning speed and good generalization performance, the stability of the prediction results is weak because it randomly selects the input weights and the biases of the hidden neurons in training [40]. To improve the prediction performance of the ELM, the best input weights and the biases of the hidden layer should be determined to replace those from random selection. The most direct way to do so is by constructing a hybrid framework to combine an optimization algorithm and the ELM, thereby optimizing the input weights and the biases of the hidden layer randomly generated by the ELM [42]. Following previous studies [44–46], considering the sound performance in global search, GA is utilized to optimize the input weights and the biases of the hidden layer to improve the performance of the ELM in stock prediction. A detailed procedure of the GA-ELM algorithm is given in Algorithm 1.

2.2. Extended Portfolio Optimization Models. Markowitz's mean-variance (MV) model, as a classic portfolio optimization method, has been popular in the practice of financial investment. Based on a set of different return or risk levels, MV can determine the efficient frontier, at which the portfolios can obtain the maximum return for the given risk level or achieve the minimum risk for the given return level. There are two commonly known portfolio models on the efficient frontier, i.e., the GMV model [66] and the MSPR model [50]. The former denotes the minimum risk portfolio of the efficient frontier, and the latter denotes portfolio that maximizes the Sharpe ratio [66]. This paper incorporates the ESG factor into the objective functions of the GMV and MSPR models to form the GMV-ESG and MSPR-ESG models, respectively. The GMV-ESG model can be defined as follows:

$$\begin{aligned} &\min \sigma_p, \\ &\max \varphi_p, \\ &\begin{cases} \sigma_p = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} w_i \sigma_{ij} w_j}, \\ \varphi_p = \sum_{i=1}^{n} w_i \varphi_i, \\ \sum_{i=1}^{n} w_i = 1, \\ 0 \le w_i \le 1, \quad \forall i = 1, 2, \ldots, n, \end{cases} \end{aligned} \tag{9}$$

where w_i is the weight of stock i, σ_{ij} represents the covariance between stocks i and j, and σ_p is the standard deviation of the portfolio. Following the advice of Auer [12], considering that investors usually have only a vague understanding of ESG, a synthetic ESG score is used to quantify the ESG performance. φ_p and φ_i denote the expected ESG score of portfolio and stock i, respectively. For the MSPR-ESG model, similarly to the Sharpe ratio, we trade-off ESG score maximization and risk minimization, thereby incorporating the ESG score under a unit of risk into the MSPR model. The MSPR-ESG model can be defined as follows:

Input: initialize the chromosome populations
Output: the optimal input weights and biases of the hidden layer of the ELM
(1) Begin
(2) Set the parameters of the ELM
(3) Define the fitness function
(4) Evaluate the fitness of all chromosomes in population
(5) C = the best chromosome
(6) $g = 1$
(7) **while** $g <$ maximum generation **do**
(8) **for** $i = 0 \longrightarrow P$ **do**
(9) Generate $rd1$ randomly in the interval $[0, 1]$
(10) **if** $rd1 <$ crossover probability **then**
(11) Update the chromosome according to equation (7)
(12) **end if**
(13) Generate $rd2$ randomly in the interval $[0, 1]$
(14) **if** $rd2 <$ mutation probability **then**
(15) Update the chromosome according to equation (8)
(16) **end if**
(17) Evaluate the chromosome fitness
(18) **end for**
(19) Update the chromosome populations according to equation (6)
(20) Rank the chromosome populations according to fitness
(21) C = the best current chromosome
(22) $g = g + 1$
(23) **end while**
(24) Return the optimal input weights and biases of the hidden layer of the ELM according to C
(25) End

Algorithm 1: GA-ELM algorithm.

$$\max \frac{u_p - r_f}{\sigma_p},$$

$$\max \frac{\varphi_p}{\sigma_p},$$

$$\begin{cases} u_p = \sum_{i=1}^{n} w_i u_i, \\ \varphi_p = \sum_{i=1}^{n} w_i \varphi_i, \\ \sigma_p = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} w_i \sigma_{ij} w_j}, \\ \sum_{i=1}^{n} w_i = 1, \\ 0 \le w_i \le 1, \quad \forall i = 1, 2, \ldots, n, \end{cases} \tag{10}$$

where u_i and u_p separately denote the expected return of stock i and the portfolio, respectively, and r_f denotes the risk-free return. In this study, the expected return u_i, covariance σ_{ij}, and ESG score φ_i are obtained by the historical data.

2.3. The Proposed Models. The input of high-quality stocks plays an important role in constructing an optimal portfolio [67]. Thus, it is necessary to exclude inferior stocks that cannot satisfy investor demands from the investment universe. For most SRI investors, high-quality stocks not only need to meet ESG expectations but also have high-return potential. In the proposed models, a novel double-screening mechanism considering ESG and return potential is implemented to guarantee that high-quality stocks are input into the optimal portfolio. To be specific, low-ESG stocks are first cleaned from the investment universe based on the ESG threshold given by the investor. Subsequently, the GA-ELM model is utilized to predict the price of the remaining stocks in the future period, thereby obtaining the corresponding return prediction. Finally, the predicted returns are sorted in a descending order, and stocks with top-ranked returns are regarded as the ultimate high-quality stocks obtained from double-screening.

Furthermore, as Oikonomou et al. [48] claimed, in addition to optimizing the investment universe through the screening process, SRI investors also carefully consider portfolio selection techniques. The conventional GMV and MSPR models cannot trade off the financial and ESG performance of SRI investors since they do not incorporate the ESG factor. Therefore, this paper includes the ESG score within the objective functions of the GMV and MSPR models to form extended portfolio optimization models, i.e., the GMV-ESG and MSPR-ESG models, to balance the ESG

and financial objectives of SRI investors. In summary, the proposed models are divided into the two stages of stock screening and asset allocation, involving four steps. The flowchart of the proposed models is shown in Figure 1, and a more detailed description of the proposed models is given as follows:

Step 1. Exclude low-ESG stocks. All sample stocks are ranked in descending order of ESG scores, i.e., $S = \{s_1, s_2, \ldots, s_i, s_j, \ldots, s_n | i < j, \mathrm{ESG}(s_i) > \mathrm{ESG}(s_j)\}$, where s_i denotes stock i. Then, the ESG score threshold ESG_T is set, and stocks with ESG scores lower than the threshold ESG_T are excluded from S to form a new stock pool $\overline{S} = \{s_1, s_2, \ldots, s_i, s_j, \ldots, s_h | i < j, \mathrm{ESG}(s_i) > \mathrm{ESG}(s_j),\ 1 < h < n, \mathrm{ESG}(s_h) > \mathrm{ESG}_T\}$.

Step 2. Predict stock prices. Because SRIs are usually medium-to long-term investments, we predict stock prices over a relatively long interval period instead of the next. First, stock price data are divided into a training set and a testing set, which are transformed into supervised learning forms. Let $P_i(t)$ denote the price of $s_i \in \overline{S}$ at time t $(t = 1, 2, \ldots, M)$, Δt denote the interval, and δ represent the time window size. Thus, the input matrix X_i and output matrix Y_i of s_i can be expressed as follows:

$$X_i = \begin{bmatrix} P_i(1) & P_i(2) & \ldots & P_i(\delta) \\ P_i(2) & P_i(3) & \ldots & P_i(\delta+1) \\ \ldots & \ldots & \ldots & \ldots \\ P_i(M-\Delta t-\delta+1) & P_i(M-\Delta t-\delta+2) & \ldots & P_i(M-\Delta t) \end{bmatrix},$$
$$Y_i = \begin{bmatrix} P_i(\delta+\Delta t) \\ P_i(\delta+\Delta t+1) \\ \ldots \\ P_i(M) \end{bmatrix}. \tag{11}$$

We stack $X_i\,(i = 1, 2, \ldots, h)$ in one large feature metric V with dimension $w \times \delta\,(w = h \times (M - \Delta t - \delta + 1))$. Similarly, we can obtain one large target matrix O with dimension $w \times 1$. Subsequently, the feature metric V and target matrix O are fed to the GA-ELM model for training. Then, the optimal GA-ELM model obtained is applied to predict the prices of all stocks for the next Δt periods, i.e., $\widehat{P}_i(t + \Delta t)$, $(i = 1, \ldots, h)$.

Step 3. Screen high-quality stocks. Converting predicted stock prices to predicted returns, $\widehat{R}_{s_i}(t + \Delta t) = (\widehat{P}_i(t + \Delta t) - P_i(t))/P_i(t)$, $(i = 1, \ldots, h)$. Then, the obtained predicted returns $\widehat{R}_{s_i}(t + \Delta t)$ are sorted in descending order, and the top k stocks are considered the ultimate high-quality stocks $\widehat{S} = \{s_1, s_2, \ldots, s_i, s_j, \ldots, s_k | i < j, \widehat{R}_{s_i}(t + \Delta t) > \widehat{R}_{s_j}(t + \Delta t),\ 1 < k < h\}$.

Step 4. Compute the proportion of capital allocation. The GMV-ESG and MSPR-ESG models are employed to determine the proportion of capital allocated to the high-quality stocks $\widehat{S}$, thereby obtaining the optimal SRI portfolios. For convenience in solving, a multi-objective model is often transformed into a single-objective model. Equations (9) and (10) can be expressed as follows:

$$\max \sum_{i=1}^{n} w_i \varphi_i - \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i \sigma_{ij} w_j},$$
$$\begin{cases} \sum_{i=1}^{n} w_i = 1, \\ 0 \le w_i \le 1, \quad \forall i = 1, 2, \ldots, n, \end{cases}$$
$$\max \frac{u_p - r_f}{\sigma_p} + \frac{\varphi_p}{\sigma_p},$$
$$\begin{cases} u_p = \sum_{i=1}^{n} w_i u_i, \\ \varphi_p = \sum_{i=1}^{n} w_i \varphi_i, \\ \sigma_p = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i \sigma_{ij} w_j}, \\ \sum_{i=1}^{n} w_i = 1, \\ 0 \le w_i \le 1, \quad \forall i = 1, 2, \ldots, n. \end{cases} \tag{12}$$

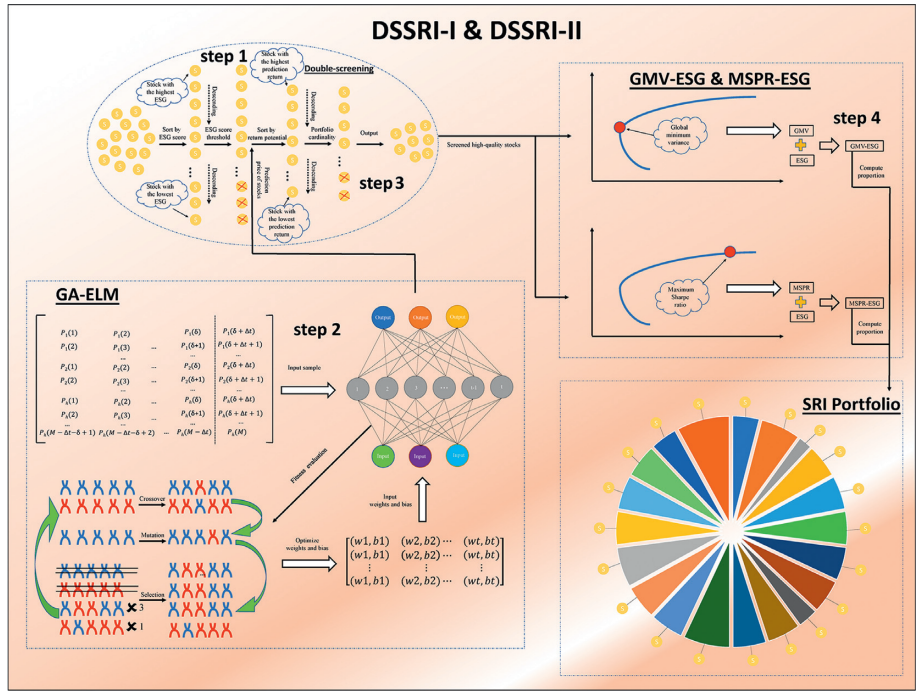


Figure 1: Flowchart of the proposed DSSRI-I and DSSRI-II.

3. Empirical Results and Analysis

3.1. Data Description. In this paper, we use the ESG score from Thomson Reuters ASSET 4, one of the most prominent professional ESG rating agencies. We obtain the monthly closing prices of stocks from January 1, 2010 to December 31, 2019 in Choice Database. At present, the Thomson Reuters ASSET 4 database reports the ESG scores of only a few Chinese listed companies. After excluding some companies whose ESG scores are much lower or whose stock price data are notably incomplete, we select 50 companies as the study samples, as shown in Table 1. The stock closing price time series is split into training and testing sets based on the yearly rolling window approach, thereby forming three study periods. Specifically, in the first study period, we apply the stock prices from 2010 to 2016 for the training set and from 2017 for the testing set. Then, we move to the next year in turn until the third study period, using the stock prices from 2012 to 2018 for the training set and the stock prices from 2019 for the testing set.

3.2. Parameter Settings. In our experiments, the interval Δt is set to 3 and the time window size δ is set to 6; that is, the stock price from the first 6 months is used to predict the price 3 months ahead to achieve the quarterly prediction. The parameters of all relative models are presented in Table 2. The asset size of the portfolio is a critical factor to be considered by investors. Because inappropriate portfolio cardinality weakens the performance of the portfolio, it can either increase the fee or worsen the risk. Some studies provide us with valuable references on this subject; for example, Zulkifli et al. [68] argued that the inclusion of 15 stocks can diversify away the diversifiable risk. Tang [69] found that portfolios containing 20 stocks can eliminate 95% of the diversifiable risk. Considering the above conclusions and trading off the fee and risk, we set the portfolio cardinality $C = 15$, 16, 17, 18, 19, and 20. To guarantee the reliability of the prediction results, the ultimate prediction results of all models are obtained by computing the average value of 20 independent runs. The empirical study is implemented using Python 3.7 and TensorFlow 2.0 environments on a computer with 1.80 GHz CPU and 16.0 GB RAM.

3.3. Stock Double-Screening. In this section, double-screening, including ESG screening and return potential screening, is performed. First, considering the limited stock universe in this study, we censor the 20% of stocks with the lowest ESG scores to complete the ESG screening. Next, for the remaining stocks, we use the GA-ELM model to predict the stock prices, further screening out stocks based on their return potential. To demonstrate the performance of the GA-ELM in terms of stock screening, we compare it with four advanced machine learning models: ELM, SVR, DNN, and LSTM. For prediction evaluation, several statistical criteria, such as the mean square error, root mean square error, and mean absolute error, are commonly used to measure the error of the prediction model. However, considering the purpose of the quantitative investment, these statistical metrics may not truly meet the financial goals of the investor. As mentioned in [70], practitioners focus more on potential trading profitability than on the prediction accuracy of models. A straightforward fact is that the more high-return stocks a portfolio contains, the greater is its probability of yielding high returns. Therefore, in this paper,

Table 1: The stock exchange codes of the selected companies.

000002.SZ	000039.SZ	000063.SZ	000157.SZ	000338.SZ
000776.SZ	000898.SZ	002142.SZ	002594.SZ	600011.SH
600016.SH	600026.SH	600028.SH	600029.SH	600030.SH
600036.SH	600188.SH	600196.SH	600362.SH	600377.SH
600600.SH	600808.SH	600875.SH	601088.SH	601111.SH
601186.SH	601238.SH	601288.SH	601318.SH	601328.SH
601336.SH	601390.SH	601398.SH	601600.SH	601607.SH
601618.SH	601628.SH	601727.SH	601766.SH	601800.SH
601808.SH	601857.SH	601866.SH	601898.SH	601899.SH
601919.SH	601939.SH	601988.SH	601991.SH	601998.SH

we screen high-return stocks to evaluate the prediction models. In other words, a good prediction model for quantitative investment can screen more high-return stocks than an inferior prediction model can. Obviously, there are two extreme cases: the best model can screen all high-return stocks, and the worst model cannot screen any.

The top N stocks with the highest return among candidate stocks are regarded as high-return stocks. To verify the accuracy and generalizable performance of the GA-ELM model, we can give N different values. Considering that the value of N directly determines the asset size of the SRI portfolio constructed, we set $N = 15$, 16, 17, 18, 19, and 20. Table 3 shows the minimum, maximum, mean, and average accuracy (average accuracy = mean/N) of the screened number of high-return stocks from 2017Q1 to 2019Q4 using different models. We find that the GA-ELM outperforms the other models in nearly all cases in terms of minimum, maximum, mean, and average accuracy. Specifically, when the number of high-return stocks $N = 15$, 16, 17, 18, 19, and 20, the values of minimum, maximum, mean, and average accuracy obtained by the GA-ELM are almost uniformly the highest. For example, when $N = 15$, the GA-ELM can screen 6.25 high-return stocks on average, while the ELM screens 5.25. The average accuracy of the GA-ELM is 6.67% higher than that of the ELM. The comparison results imply that using GA to optimize ELM can effectively improve the performance of ELM. Furthermore, compared with conventional machine learning models, i.e., SVR, and deep learning methods, i.e., DNN and LSTM, the average number of high-return stocks obtained by GA-ELM separately increase by 0.58, 1.17, and 0.83, and the corresponding average accuracy also improves by 3.89%, 7.78%, and 5.56%. Accordingly, the superiority of the GA-ELM is further confirmed. When the number of high-return stocks N equals other values, i.e., 16, 17, 18, 19, or 20, we obtain similar conclusions, which indicate that the GA-ELM model has good generalizability. Therefore, the GA-ELM is a promising tool for use in the practice of stock screening.

3.4. SRI Portfolio Construction. In this section, the screened high-quality stocks are input into the GMV-ESG and MSPR-ESG models to construct SRI portfolios. The proposed models are compared with several alternative models at different portfolio cardinalities along the dimensions of return (annualized return), risk-adjusted return (Sharpe ratio), and ESG (ESG score).

3.4.1. Outline of Alternative Models. To test the performance of the proposed models, two kinds of alternative models are introduced for comparison:

(1) Alternative Model I: SSRI-GMV (or SSRI-MSPR). These models represent conventional SRI strategies that do not consider the return potential of stocks in the screening stage and do not introduce the ESG factor into the portfolio optimization model. In these models, low-ESG stocks are first excluded using ESG screening. Next, to keep the portfolio cardinality consistent with that of the proposed models, the same number of stocks as in the proposed models are selected from the screened stocks. Then, the GMV model (or MSPR model) is utilized to construct the SRI portfolio.

(2) Alternative Model II: DSSRI-I (Based on Other Machine Learning Models) (or DSSRI-II (Based on Other Machine Learning Models)). The purpose of introducing these models is to prove that the screening results obtained by different machine learning prediction models can affect the performance of the SRI portfolio. Specifically, we first exclude some stocks with low-ESG scores from the investment universe. Subsequently, several advanced machine learning models, such as ELM, SVR, DNN, and LSTM, are employed separately to predict stock prices, thereby further screening stocks with high-return potential. Finally, as with the proposed models, the GMV-ESG model (or MSPR-ESG model) is utilized to construct the SRI portfolio.

3.4.2. Alternative Model I Comparison

(1) Comparison of the Proposed DSSRI-I and the SSRI-GMV. Table 4 reports the annualized return, Sharpe ratio, and ESG score of the proposed DSSRI-I and the SSRI-GMV at different portfolio cardinalities ($C = 15$, 16, 17, 18, 19, and 20). The performance comparison results of the two models are shown in Figure 2. It is obvious that the proposed DSSRI-I provides the highest values on the three evaluation metrics. If we take portfolio cardinality $C = 15$ as an example, the annualized return, Sharpe ratio, and ESG score of the SSRI-GMV are 0.0444, 0.3198, and 55.3456, respectively. Correspondingly, these metrics for the proposed DSSRI-I are 0.0982, 0.4930, and 61.9809, representing increases of 0.0538, 0.1733, and 6.6353. For other portfolio cardinalities ($C = 16$, 17, 18, 19, and 20), the proposed DSSRI-I still generate good performance on all evaluation metrics. Specifically, compared with the SSRI-GMV, the annualized return of the proposed DSSRI-I increases in the range of 0.0320–0.0428, the Sharpe ratio increases in the range of 0.0800–0.1510, and the ESG score increases in the range of 5.3184–6.3273. Furthermore, at the portfolio cardinality $C = 17$, the proposed DSSRI-I provides the highest annualized return and Sharpe ratio of 0.1113 and 0.5517, respectively. The highest ESG score is generated by the proposed model at 61.9809 at portfolio cardinality $C = 15$. In summary, the proposed DSSRI-I can yield better ESG and financial performance than the SSRI-GMV.

Table 2: Parameter settings of different models.

Prediction models	Parameter	Value
SVR	Kernel function	rbf
	γ	$[2^{-15}, 2^{-13}, \ldots, 2^{13}, 2^{15}]$
	C	$[2^{-15}, 2^{-3}, \ldots, 2^{13}, 2^{15}]$
DNN	Layer structure (number of neurons)	Dense (30) – Dense (3) – Dense (1)
	Activation function	Relu
	Dropout	0.2
	Learning rate	0.01
	The number of epochs	200
LSTM	Layer structure (number of neurons)	LSTM (60) – Dense (16) – Dense (1)
	Dropout	0.1
	Learning rate	0.01
	Number of epochs	200
ELM	Activation function	Sigmoid
	Neuron number of hidden layer	100
GA	Population size	25/20
	Maximum generation	100/500
	Crossover probability	0.8/0.5
	Mutation probability	0.01/0.01

Table 3: Comparison of screening results obtained by different models.

Model	The number of high-return stock = 15				The number of high-return stock = 16			
	Minimum	Maximum	Mean	Average accuracy (%)	Minimum	Maximum	Mean	Average accuracy (%)
SVR	3.00	**9.00**	5.67	37.78	3.00	9.00	6.25	39.06
DNN	3.00	7.00	5.08	33.89	3.00	8.00	5.83	36.46
LSTM	3.00	**9.00**	5.42	36.11	3.00	**10.00**	5.83	36.46
ELM	3.00	7.00	5.25	35.00	3.00	8.00	6.08	38.02
GA-ELM	**4.00**	8.00	**6.25**	**41.67**	**4.00**	9.00	**6.83**	**42.71**
Model	The number of high-return stock = 17				The number of high-return stock = 18			
	Minimum	Maximum	Mean	Average accuracy (%)	Minimum	Maximum	Mean	Average accuracy (%)
SVR	4.00	10.00	7.17	42.16	5.00	10.00	7.50	41.47
DNN	3.00	10.00	6.75	39.71	4.00	12.00	7.92	43.98
LSTM	3.00	10.00	6.58	38.73	4.00	11.00	7.67	42.59
ELM	4.00	9.00	7.17	42.16	5.00	10.00	8.08	44.91
GA-ELM	**5.00**	**11.00**	**7.92**	**46.57**	**5.00**	**13.00**	**8.83**	**49.07**
Model	The number of high-return stock = 19				The number of high-return stock = 20			
	Minimum	Maximum	Mean	Average accuracy (%)	Minimum	Maximum	Mean	Average accuracy (%)
SVR	6.00	11.00	8.33	43.86	6.00	12.00	9.25	46.25
DNN	5.00	13.00	8.75	46.05	7.00	14.00	9.50	47.50
LSTM	6.00	12.00	8.67	45.61	7.00	13.00	9.50	47.50
ELM	7.00	11.00	9.00	47.37	8.00	12.00	9.58	47.92
GA-ELM	**7.00**	**14.00**	**9.75**	**51.32**	**9.00**	**15.00**	**11.00**	**55.00**

The best values are demonstrated with bold font.

(2) Comparison of the Proposed DSSRI-II and the SSRI-MSPR. The annualized return, Sharpe ratio, and ESG score of the two models are shown in Table 5. Figure 3 compares the performance of the two models. Obviously, DSSRI-II has higher evaluation metrics than the SSRI-MSPR at different portfolio cardinalities ($C = 15$, 16, 17, 18, 19, and 20). For example, the annualized returns of the proposed DSSRI-II are 0.0882, 0.0780, 0.1057, 0.0856, 0.0905, and 0.0887 at portfolio cardinalities $C = 15$ to 20, representing increases of 63.02%, 67.76%, 94.95%, 82.01%, 62.13%, and 65.61% over those of the SSRI-MSPR. For the Sharpe ratio, the DSSRI-II performs similarly. Specifically, the best Sharpe ratio (0.5339) is obtained by the DSSRI-II with portfolio cardinality $C = 17$. In contrast, the best Sharpe ratio of the SSRI-MSPR is 0.3750. When the portfolio cardinality $C = 15$, the DSSRI-II generates the highest ESG score of 62.0703, which is 11.50% higher than the best score of the SSRI-MSPR, at 55.6684. In short, the proposed DSSRI-II outperforms the SSRI-MSPR in terms of annualized return, Sharpe ratio, and ESG score.

3.4.3. Alternative Model II Comparison

(1) Comparison of the Proposed DSSRI-I and the DSSRI-I (Based on Other Machine Learning Models). In this section, we contrast the proposed DSSRI-I with the DSSRI-I (ELM), DSSRI-I (SVR), DSSRI-I (DNN), and DSSRI-I (LSTM) to

Table 4: Comparison of the DSSRI-I and the SSRI-GMV.

Model	Portfolio cardinality $C = 15$			Portfolio cardinality $C = 16$			Portfolio cardinality $C = 17$		
	AR	SR	ESG	AR	SR	ESG	AR	SR	ESG
DSSRI-I	**0.0982**	**0.4930**	**61.9809**	**0.0904**	**0.4701**	**61.8525**	**0.1113**	**0.5517**	**61.9619**
SSRI-GMV	0.0444	0.3198	55.3456	0.0583	0.3901	55.5345	0.0686	0.4281	55.6346
Model	Portfolio cardinality $C = 18$			Portfolio cardinality $C = 19$			Portfolio cardinality $C = 20$		
	AR	SR	ESG	AR	SR	ESG	AR	SR	ESG
DSSRI-I	**0.0947**	**0.4952**	**61.5369**	**0.0962**	**0.5049**	**61.2844**	**0.0994**	**0.5222**	**60.9385**
SSRI-GMV	0.0614	0.3931	55.5110	0.0564	0.3710	55.3652	0.0567	0.3712	55.6201

AR = annualized return; SR=Sharpe ratio; ESG = ESG score. The best values are demonstrated with bold font.

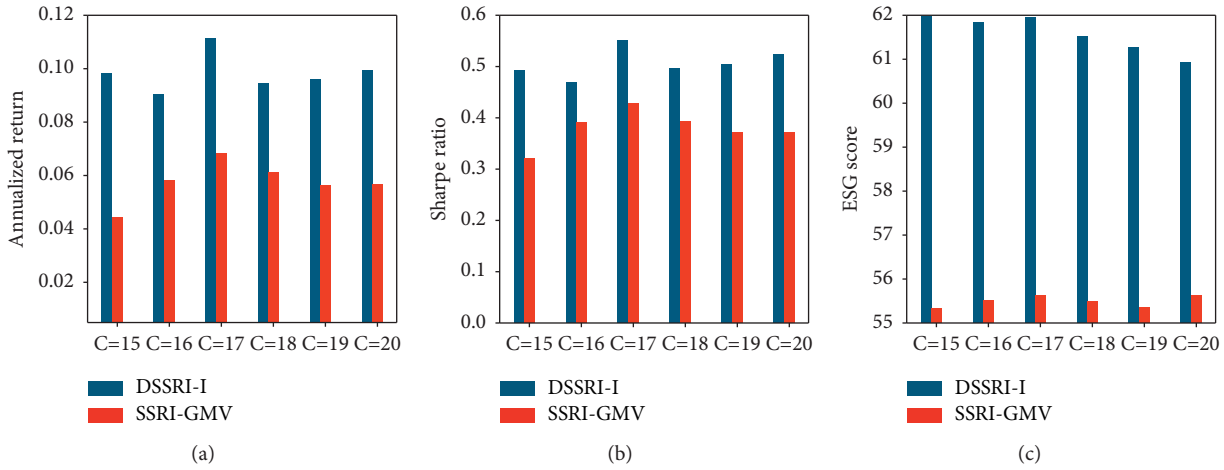


Figure 2: Performance of the DSSRI-I and the SSRI-GMV at different portfolio cardinalities.

Table 5: Comparison of the DSSRI-II and the SSRI-MSPR.

Model	Portfolio cardinality $C = 15$			Portfolio cardinality $C = 16$			Portfolio cardinality $C = 17$		
	AR	SR	ESG	AR	SR	ESG	AR	SR	ESG
DSSRI-II	**0.0882**	**0.4575**	**62.0703**	**0.0780**	**0.4188**	**61.9743**	**0.1057**	**0.5339**	**61.8762**
SSRI-MSPR	0.0541	0.3748	55.5466	0.0465	0.3311	55.4773	0.0542	0.3725	55.3819
Model	Portfolio cardinality $C = 18$			Portfolio cardinality $C = 19$			Portfolio cardinality $C = 20$		
	AR	SR	ESG	AR	SR	ESG	AR	SR	ESG
DSSRI-II	**0.0856**	**0.4620**	**61.6706**	**0.0905**	**0.4862**	**61.2793**	**0.0887**	**0.4883**	**60.8618**
SSRI-MSPR	0.0470	0.3293	55.4834	0.0558	0.3750	55.6684	0.0536	0.3694	55.5192

AR = annualized return; SR=Sharpe ratio; ESG = ESG score. The best values are demonstrated with bold font.

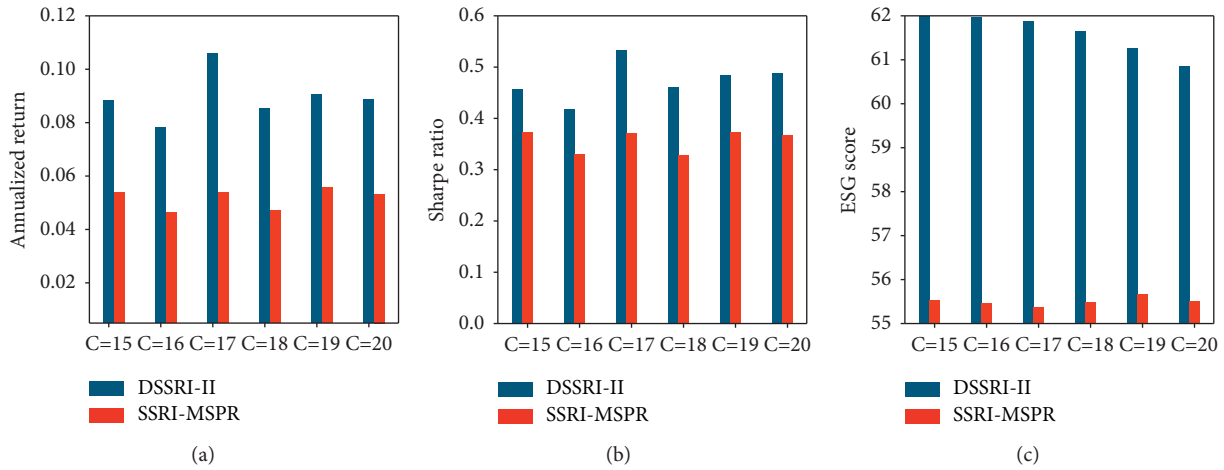


Figure 3: Performance of the DSSRI-II and the SSRI-MSPR at different portfolio cardinalities.

Table 6: Comparison of the DSSRI-I model and the DSSRI-I (based on other machine learning models).

Model	Portfolio cardinality $C = 15$			Portfolio cardinality $C = 16$			Portfolio cardinality $C = 17$		
	AR	SR	ESG	AR	SR	ESG	AR	SR	ESG
DSSRI-I	**0.0982**	0.4930	**61.9809**	0.0904	0.4701	**61.8525**	**0.1113**	**0.5517**	**61.9619**
DSSRI-I (ELM)	0.0274	0.2269	59.8493	0.0446	0.2969	59.4059	0.0525	0.3320	59.2332
DSSRI-I (SVR)	0.0070	0.1304	58.6909	0.0098	0.1426	58.7900	0.0162	0.1741	58.6922
DSSRI-I (DNN)	0.0901	**0.5623**	57.8888	**0.0941**	**0.6063**	57.5421	0.0705	0.4730	57.6859
DSSRI-I (LSTM)	0.0455	0.3156	59.0560	0.0208	0.1949	58.9020	0.0260	0.2202	58.4004
Model	Portfolio cardinality $C = 18$			Portfolio cardinality $C = 19$			Portfolio cardinality $C = 20$		
	AR	SR	ESG	AR	SR	ESG	AR	SR	ESG
DSSRI-I	**0.0947**	**0.4952**	**61.5369**	**0.0962**	**0.5049**	**61.2844**	**0.0994**	**0.5222**	**60.9385**
DSSRI-I (ELM)	0.0498	0.3241	59.2538	0.0480	0.3193	59.0777	0.0537	0.3501	59.4691
DSSRI-I (SVR)	-0.0026	0.0765	58.6725	0.0052	0.1166	58.4987	0.0158	0.1711	58.5493
DSSRI-I (DNN)	0.0688	0.4666	57.8780	0.0602	0.4293	57.6697	0.0567	0.4077	57.5114
DSSRI-I (LSTM)	0.0388	0.2843	58.5475	0.0247	0.2145	58.4433	0.0567	0.3688	58.3013

AR = annualized return; SR=Sharpe ratio; ESG = ESG score. The best values are demonstrated with bold font.

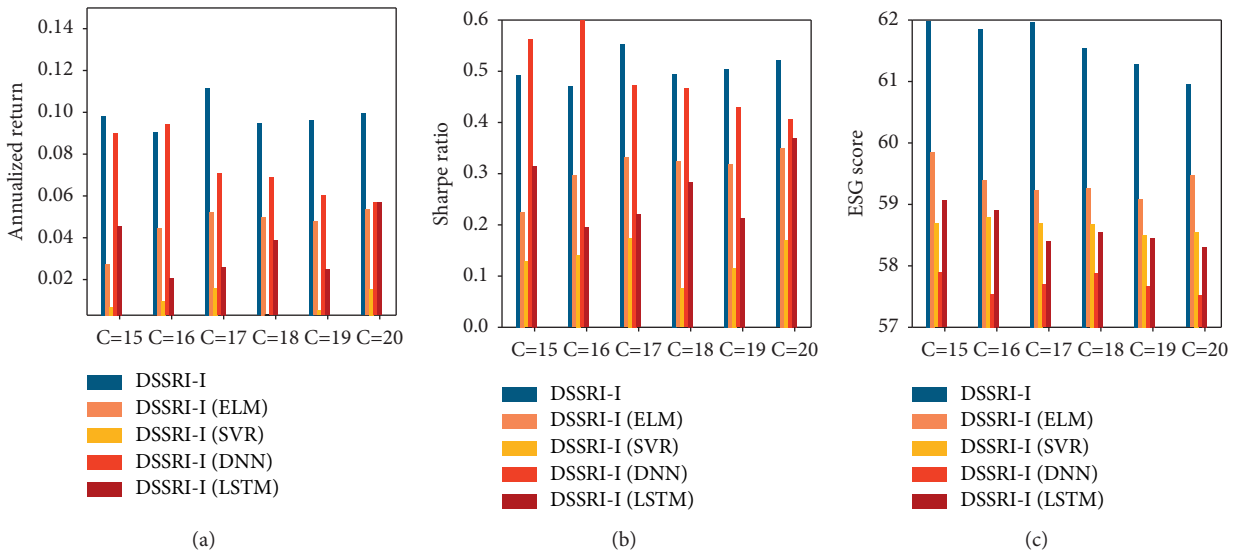


Figure 4: Performance of the DSSRI-I model and the DSSRI-I (based on other machine learning models) at different portfolio cardinalities.

verify the superiority of return potential screening based on the GA-ELM. The annualized return, Sharpe ratio, and ESG score of all models at different portfolio cardinalities $C = 15$, 16, 17, 18, 19, and 20 are presented in Table 6. The comparison of the evaluation metrics for the different models is displayed in Figure 4. First, we compare the proposed DSSRI-I with the DSSRI-I (ELM). As shown in Table 6 and Figure 4, when the portfolio cardinality $C = 15$, the annualized return of the proposed DSSRI-I is 0.0708 higher than that of the DSSRI-I (ELM). The Sharpe ratio of the proposed DSSRI-I is 0.4930, which is 117.32% higher than that of the DSSRI-I (ELM). The highest ESG score (61.9809) is still obtained by the proposed DSSRI-I. For other portfolio cardinalities ($C = 16$, 17, 18, 19, and 20), similar conclusions are obtained. In general, the proposed model is superior to the DSSRI-I (ELM) in terms of annualized return, Sharpe ratio, and ESG score and thus proves again that GA optimization can improve the performance of the ELM. Subsequently, the proposed DSSRI-I is compared with SRI portfolio models based on double-screening considering other advanced machine learning models (SVR, DNN, and LSTM), i.e., DSSRI-I (SVR), DSSRI-I (DNN), and DSSRI-I (LSTM). From Table 6 and Figure 4, we can find that the highest annualized returns at almost all the given portfolio cardinality levels are obtained by the DSSRI-I. When the portfolio cardinalities are $C = 17$, 18, 19 and 20, the proposed DSSRI-I provides the highest Sharpe ratio at 0.5517, 0.4952, 0.5049, and 0.5222, respectively. The proposed DSSRI-I generates the second-highest Sharpe ratio with portfolio cardinalities $C = 15$ and 16, following the DSSRI-I (DNN). The DSSRI-I generates the highest ESG scores from portfolio cardinalities $C = 15$ to 20 with values of 61.9809, 61.8525, 61.9619, 61.5369, 61.2844, and 60.9385, respectively. In general, compared with other SRI portfolio models based on advanced machine learning models, the proposed DSSRI-I presents obvious advantages in annualized return and ESG performance and has a competitive Sharpe ratio performance. Therefore, the proposed DSSRI-I based on the GA-ELM is capable of improving the ESG and financial performance of SRI portfolios.

Table 7: Comparison of the DSSRI-II model and the DSSRI-II (based on other machine learning models).

Model	Portfolio cardinality $C = 15$			Portfolio cardinality $C = 16$			Portfolio cardinality $C = 17$		
	AR	SR	ESG	AR	SR	ESG	AR	SR	ESG
DSSRI-II	**0.0882**	0.4575	**62.0703**	0.0780	0.4188	**61.9743**	**0.1057**	**0.5339**	**61.8762**
DSSRI-II (ELM)	0.0360	0.2659	59.8217	0.0458	0.3028	59.3619	0.0407	0.2845	59.1623
DSSRI-II (SVR)	0.0110	0.1499	58.5464	0.0130	0.1580	58.9834	0.0212	0.1981	58.7566
DSSRI-II (DNN)	0.0873	**0.5465**	57.8571	**0.1024**	**0.6361**	57.6394	0.0697	0.4616	57.7544
DSSRI-II (LSTM)	0.0537	0.3505	59.1856	0.0298	0.2392	58.7573	0.0268	0.2240	58.6118
Model	Portfolio cardinality $C = 18$			Portfolio cardinality $C = 19$			Portfolio cardinality $C = 20$		
	AR	SR	ESG	AR	SR	ESG	AR	SR	ESG
DSSRI-II	**0.0856**	0.4620	**61.6706**	**0.0905**	0.4862	**61.2793**	**0.0887**	**0.4883**	**60.8618**
DSSRI-II (ELM)	0.0442	0.3002	59.2208	0.0444	0.3031	59.0595	0.0596	0.3728	59.4935
DSSRI-II (SVR)	0.0019	0.1015	58.5811	0.0098	0.1399	58.5797	0.0162	0.1738	58.5194
DSSRI-II (DNN)	0.0677	**0.4653**	57.8591	0.0698	**0.4922**	57.8490	0.0565	0.4030	57.4872
DSSRI-II (LSTM)	0.0342	0.2612	58.6037	0.0306	0.2438	58.3080	0.0521	0.3441	58.4704

AR = annualized return; SR=Sharpe ratio; ESG = ESG score. The best values are demonstrated with bold font.

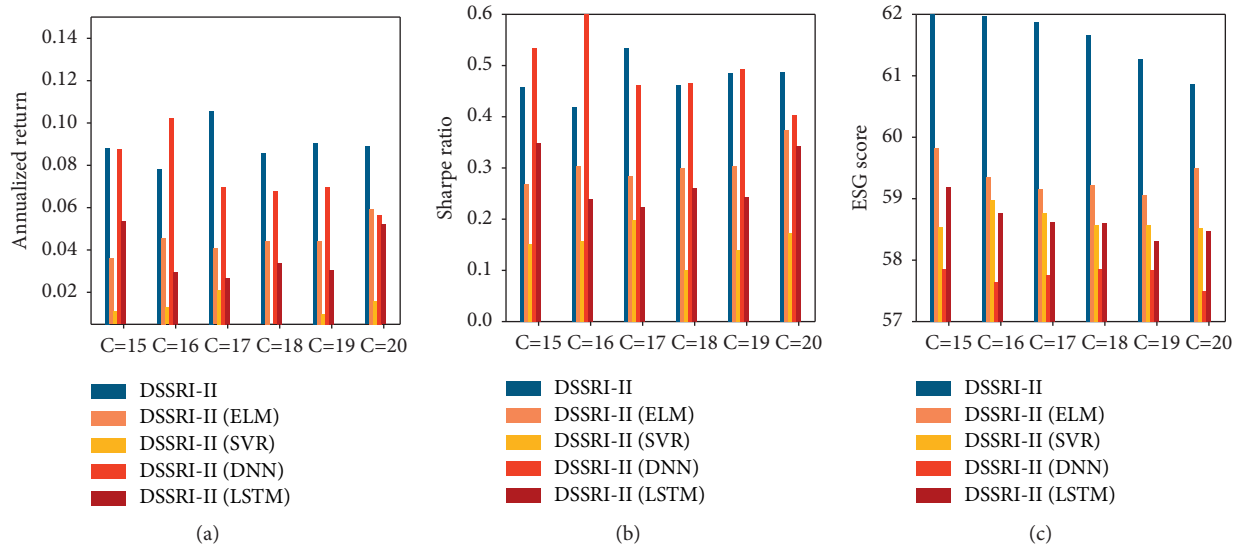


Figure 5: Performance of the DSSRI-II model and the DSSRI-II model (based on other machine learning models) at different portfolio cardinalities.

(2) Comparison of the Proposed DSSRI-II and the DSSRI-II (Based on Other Machine Learning Models). As shown in Table 7 and Figure 5, the proposed DSSRI-II performs well in terms of annualized return, Sharpe ratio, and ESG score. Specifically, for the annualized return, the proposed DSSRI-II obtains the highest values at almost all the given portfolio cardinalities. However, an exception is when the portfolio cardinality $C = 16$, in which case the DSSRI-II (DNN) yields the highest annualized return of 0.1024. The proposed DSSRI-II yields the second-highest annualized return with 0.0780. The proposed DSSRI-II generates the second-highest Sharpe ratio at portfolio cardinalities $C = 15$, 16, 18, and 19 and the highest Sharpe ratio at portfolio cardinalities $C = 17$ and 20. For the ESG score, with the portfolio cardinality $C = 15$ taken as an example, the ESG score of the proposed model is 62.0703, representing an improvement of 3.76%, 6.02%, 7.28%, and 4.87% over the DSSRI-II (ELM), DSSRI-II (SVR), DSSRI-II (DNN), and DSSRI-II (LSTM), respectively. At the other portfolio cardinality levels, the proposed DSSRI-II is consistently ahead of the comparison models by at least 2.30% on the ESG score. In summary, similarly to the DSSRI-I, the DSSRI-II can provide good annualized return and ESG performance as well as competitive Sharpe ratio performance in comparison with the performance of the DSSRI-II (based on other machine learning models).

4. Conclusion and Future Work

In this work, we propose two SRI portfolio construction models called DSSRI-I and DSSRI-II based on machine learning, intelligent algorithm, and portfolio theory. The models consist of two stages: stock screening and asset allocation. In the stock screening stage, a double-screening mechanism is developed to generate the input of the SRI portfolio. Specifically, low-ESG stocks are first excluded from the investment universe to guarantee that the

assets are in line with the social values of SRI investors. Then, an ELM model with GA optimization is introduced to predict the prices of the remaining stocks and thus further screen high-return potential stocks. In the asset allocation stage, the extended GMV model (or extended MSPR model) considering the ESG factor is introduced to determine the investment proportion of the screened stocks from the first stage. To test the performance, we compared the proposed models with alternative models in terms of annualized return, Sharpe ratios, and ESG scores. The conclusions show the following: (1) compared with the conventional SSRI-GMV (or SSRI-MSPR), the proposed DSSRI-I (or DSSRI-II) performs better on both ESG and financial measures. (2) The proposed DSSRI-I (or DSSRI-II) can obtain a higher annualized return and ESG score than the alternative models based on advanced machine learning models (i.e., DSSRI-I (ELM) (or DSSRI-II (ELM)), DSSRI-I (SVR) (or DSSRI-II (SVR)), DSSRI-I (DNN) (or DSSRI-II (DNN)), and DSSRI-I (LSTM) (or DSSRI-II (LSTM)). (3) The proposed models can generate a competitive Sharpe ratio performance in comparison with that of the above models. Therefore, considering the superior performance of the proposed models, they can be regarded as effective tools to construct SRI portfolios.

In the future, our work may focus on the following three aspects. First, other stock markets in different political, economic, and social environments should be studied to test the generalizability of the proposed models. Second, the performance of the proposed models can be further analysed by introducing various intelligent algorithms to optimize the ELM. Finally, the proposed models can be extended to a general framework for portfolio construction based on the personal preference of investors.

Abbreviations

ANNs: Artificial neural networks
BP: Back propagation
DNN: Deep neural network
DSSRI: Double-screening socially responsible investment
ELM: Extreme learning machine
ESG: Environmental, social, and corporate governance
GA: Genetic algorithm
GMV: Global minimum variance
LSTM: Long short-term memory
MSRP: Maximum Sharpe ratio
MV: Mean-variance
SLFN: Single-hidden layer feedforward neural network
SRI: Socially responsible investment
SVR: Support vector regression.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Graduate Science and Technology Innovation Foundation of the Capital University of Economics and Business, China.

References

[1] Social Investment Forum Foundation, *Report on Socially Responsible Investing Trends in the united states*, Social Investment Forum Foundation, Washington, DC, 1999.
[2] M. Statman, "Socially responsible indexes," *Journal of Portfolio Management*, vol. 32, no. 3, pp. 100–109, 2006.
[3] K. Pattit, "Understanding socially responsible investing: the effect of decision frames and trade-off options," *Journal of Business Ethics*, vol. 87, no. 4, pp. 41–55, 2009.
[4] L. Renneboog, J. Horst, J. Ter Horst, and C. Zhang, "Socially responsible investments: institutional aspects, performance, and investor behavior," *Journal of Banking & Finance*, vol. 32, no. 9, pp. 1723–1742, 2008.
[5] D. J. Beal, M. Goyen, and P. Philips, "Why do we invest ethically?" *Journal of Investing*, vol. 14, no. 3, pp. 66–78, 2005.
[6] W. R. Pasewark and M. E. Riley, "It's a matter of principle: the role of personal values in investment decisions," *Journal of Business Ethics*, vol. 93, no. 2, pp. 237–253, 2010.
[7] T. A. Hart and M. Sharfman, "Assessing the concurrent validity of the revised kinder, lydenberg, and domini corporate social performance indicators," *Business & Society*, vol. 54, no. 5, pp. 575–598, 2015.
[8] L. Widyawati, "Measurement concerns and agreement of environmental social governance ratings," *Accounting and Finance*, vol. 61, no. S1, pp. 1589–1623, 2021.
[9] N. Sachin and R. Rajesh, "An empirical study of supply chain sustainability with financial performances of indian firms," *Environment Development and Sustainability*, vol. 7, 2021.
[10] F. Garcia, J. Gonzalez-Bueno, F. Guijarro, and J. Oliver, "Forecasting the environmental, social, and governance rating of firms by using corporate financial performance variables: a rough set approach," *Sustainability*, vol. 12, no. 8, 2020.
[11] J. Feng, J. W. Goodell, and D. Shen, "Esg rating and stock price crash risk: evidence from China," *Finance Research Letters*, Article ID 102476, 2021, In press.
[12] B. Auer, "Do socially responsible investment policies add or destroy european stock portfolio value?" *Journal of Business Ethics*, vol. 135, pp. 1–17, 2014.
[13] A. Kempf and P. Osthoff, "The effect of socially responsible investing on portfolio performance," *European Financial Management*, vol. 13, no. 5, pp. 908–922, 2007.
[14] M. L. Barnett and R. M. Salomon, "Beyond dichotomy: the curvilinear relationship between social responsibility and financial performance," *Strategic Management Journal*, vol. 27, no. 11, pp. 1101–1122, 2006.
[15] D. Dominique, T. Hebb, and E. L. H. Touré, "Exploring factors that influence social retail investors decisions: evidence from desjardins fund," *Journal of Business Ethics*, vol. 134, no. 1, pp. 45–67, 2014.
[16] C. MacKenzie and A. Lewis, "Morals and markets: the case of ethical investing," *Business Ethics Quarterly*, vol. 9, no. 3, pp. 439–452, 1999.
[17] B. N. Rosen, D. M. Sandler, and D. Shani, "Social issues and socially responsible investment behavior: a preliminary empirical investigation," *Journal of Consumer Affairs*, vol. 25, no. 2, pp. 221–234, 1991.

[18] J. Nofsinger and A. Varma, "Socially responsible funds and market crises," *Journal of Banking & Finance*, vol. 48, pp. 180–193, 2014.

[19] C. Calvo, C. Ivorra, and V. Liern, "Finding socially responsible portfolios close to conventional ones," *International Review of Financial Analysis*, vol. 40, pp. 52–63, 2015.

[20] M. von Wallis and C. Klein, "Ethical requirement and financial interest: a literature review on socially responsible investing," *Business Research*, vol. 8, no. 1, pp. 61–98, 2014.

[21] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers & Operations Research*, vol. 32, no. 10, pp. 2513–2522, 2005.

[22] R. I. Rasel, N. Sultana, and P. Meesad, "An efficient modelling approach for forecasting financial time series data using support vector regression and windowing operators," *International Journal of Computational Intelligence Studies*, vol. 4, no. 2, p. 134, 2015.

[23] E. Guresen, G. Kayakutlu, and T. U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10389–10397, 2011.

[24] M. Qiu, L. Cheng, and Y. Song, "Application of the Artifical Neural Network in predicting the direction of stock market index," in *Proceedings of the 2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems*, Fukuoka, Japan, 2016.

[25] M. R. Hassan, "A combination of hidden markov model and fuzzy model for stock market forecasting," *Neurocomputing*, vol. 72, pp. 3439–3446, 2009.

[26] S. Chakravarty and P. K. Dash, "A pso based integrated functional link net and interval type-2 fuzzy logic system for predicting stock market indices," *Applied Soft Computing*, vol. 12, no. 2, pp. 931–941, 2012.

[27] O. Lachiheb and M. S. Gouider, "A hierarchical deep neural network design for stock returns prediction," *Procedia Computer Science*, vol. 126, pp. 264–272, 2018.

[28] R. Singh and S. Srivastava, "Stock prediction using deep learning," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 18569–18584, 2017.

[29] W. Zhang, K. Tao, J. Li, Y. Zhu, L. Jing, and H. Cheng, "Modeling and prediction of stock price with convolutional neural network based on blockchain interactive information," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 6686181, 10 pages, 2020.

[30] W. Chen, M. Jiang, W.-G. Zhang, and Z. Chen, "A novel graph convolutional feature based convolutional neural network for stock trend prediction," *Information Sciences*, vol. 556, pp. 67–94, 2021.

[31] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2017.

[32] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: a case study of china stock market," in *Proceedings of the IEEE International Conference on Big Data*, pp. 2823-2824, Santa Clara, CA, USA, 2015.

[33] Y. Zhang, G. Chu, and D. Shen, "The role of investor attention in predicting stock prices: the long short-term memory networks perspective," *Finance Research Letters*, vol. 38, Article ID 101484, 2021.

[34] F. Zhou, Q. Zhang, D. Sornette, and L. Jiang, "Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices," *Applied Soft Computing*, vol. 84, 2019.

[35] W. Chen, H. Zhang, M. Kumar Mehlawat, and L. Jia, "Meancvariance portfolio optimization using machine learning-based stock price prediction," *Applied Soft Computing*, vol. 100, 2021.

[36] X. Li, H. Xie, R. Wang et al., "Empirical analysis: stock market prediction via extreme learning machine," *Neural Computing & Applications*, vol. 27, no. 1, pp. 67–78, 2014.

[37] R. Dash and P. Dash, "Stock price index movement classification using a ceflann with extreme learning machine," *Power, Communication & Information Technology Conference IEEE*, vol. 10, pp. 22–28, 2015.

[38] Y. Wei, S. Sun, J. Ma, S. Wang, and K. K. Lai, "A decomposition clustering ensemble learning approach for forecasting foreign exchange rates," *Journal of Management Science and Engineering*, vol. 4, no. 1, pp. 45–54, 2019.

[39] M. Jiang, L. Jia, Z. Chen, and W. Chen, "The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm," *Annals of Operations Research*, 2020, In press.

[40] B. Zhu, Y. Feng, D. Gong, S. Jiang, L. Zhao, and N. Cui, "Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data," *Computers and Electronics in Agriculture*, vol. 173, Article ID 105430, 2020.

[41] H. Wei, J. Chen, J. Zhu, X. Yang, and H. Chu, "A novel algorithm of nested-elm for predicting blasting vibration," *Engineering with Computers*, vol. 7, 2020.

[42] P. Mohapatra, S. Chakravarty, and P. K. Dash, "An improved cuckoo search based extreme learning machine for medical data classification," *Swarm and Evolutionary Computation*, vol. 24, pp. 25–49, 2015.

[43] L. L. Li, J. Sun, M. L. Tseng, and Z. G. Li, "Extreme learning machine optimized by whale optimization algorithm using insulated gate bipolar transistor module aging degree evaluation," *Expert Systems with Applications*, vol. 127, 2019.

[44] Y. Li and H. Hu, "Influential factor Analysis and projection of industrial CO2 emissions in China based on extreme learning machine improved by genetic algorithm," *Polish Journal of Environmental Studies*, vol. 29, no. 3, pp. 2259–2271, 2020.

[45] H. Wang, Y. Wang, and Z. Ji, "Simulation of wind power prediction based on improved elm," *Journal of System Simulation*, vol. 30, pp. 4437–4447, 2018.

[46] Y. Chen, Z.. Wang, Z.. Wang, and L. Xiao, "Research on concentration retrieval of gas ftir spectra by interval extreme learning machine and genetic algorithm," *Spectroscopy and Spectral Analysis*, vol. 34, pp. 1244–1248, 2014.

[47] N. Zhu, C. Zhu, and E. Ali, "A combined machine learning algorithms and dea method for measuring and predicting the efficiency of Chinese manufacturing listed companies," *Journal of Management Science and Engineering*, 2020, In press.

[48] I. Oikonomou, E. Platanakis, and C. Sutcliffe, "Socially responsible investment portfolios: does the optimization process matter?" *The British Accounting Review*, vol. 50, no. 4, pp. 379–401, 2018.

[49] M. Branch, L. R. Goldberg, and P. Hand, "A guide to esg portfolio construction," *Journal of Portfolio Management*, vol. 45, no. 4, pp. 61–66, 2019.

[50] N. N. Y. Vo, X. He, S. Liu, and G. Xu, "Deep learning for decision making and the optimization of socially responsible investments and portfolio," *Decision Support Systems*, vol. 124, Article ID 113097, 2019.

[51] S. Utz, M. Wimmer, and R. E. Steuer, "Tri-criterion modeling for constructing more-sustainable mutual funds," *European Journal of Operational Research*, vol. 246, no. 1, pp. 331–338, 2015.

[52] C. Calvo, C. Ivorra, and V. Liern, "Fuzzy portfolio selection with non-financial goals: exploring the efficient frontier," *Annals of Operations Research*, vol. 245, pp. 31–46, 2016.

[53] S. Gasser, M. Rammerstorfer, and K. Weinmayer, "Markowitz revisited: Social portfolio engineering," *European Journal of Operational Research*, 2017.

[54] A. Bilbao-Terol, M. Arenas-Parra, V. Cañal-Fernández, and C. Bilbao-Terol, "Selection of socially responsible portfolios using hedonic prices," *Journal of Business Ethics*, vol. 115, no. 3, pp. 515–529, 2013.

[55] G. B. Huang, Y. Qin, and K. Chee, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[56] P. Du, J. Wang, W. Yang, and T. Niu, "Container throughput forecasting using a novel hybrid learning method with error correction strategy," *Knowledge-Based Systems*, vol. 182, p. 10, 2019.

[57] J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, USA, 1975.

[58] J.-S. Chen, J.-L. Hou, S.-M. Wu, and Y.-W. Chang-Chien, "Constructing investment strategy portfolios by combination genetic algorithms," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3824–3828, 2009.

[59] F. Amini and G. Hu, "A two-layer feature selection method using genetic algorithm and elastic net," *Expert Systems with Applications*, vol. 166, Article ID 114072, 2021.

[60] E. Tekeli, N. özbay, and S. Kaçiranlar, "Implementation and validation of new optimization methods by genetic algorithm for two-parameter ridge estimator," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 9, p. 5, 2021.

[61] S. S. Moghaddasi and N. Faraji, "A hybrid algorithm based on particle filter and genetic algorithm for target tracking," *Expert Systems with Applications*, vol. 147, Article ID 113188, 2020.

[62] B. Chen, J. Zhong, and Y. Chen, "A hybrid approach for portfolio selection with higher-order moments: empirical evidence from shanghai stock exchange," *Expert Systems with Applications*, vol. 145, Article ID 113104, 2020.

[63] E. C. Brown and R. T. Sumichrast, "Evaluating performance advantages of grouping genetic algorithms," *Engineering Applications of Artificial Intelligence*, vol. 18, no. 1, pp. 1–12, 2005.

[64] Y. Zheng, M. Huang, Yi Lu, and W. Li, "Fractional stochastic resonance multi-parameter adaptive optimization algorithm based on genetic algorithm," *Neural Computing & Applications*, vol. 32, no. 22, p. 11, 2020.

[65] H. Park, D. Son, B. Koo, and B. Jeong, "Waiting strategy for the vehicle routing problem with simultaneous pickup and delivery using genetic algorithm," *Expert Systems with Applications*, vol. 165, Article ID 113959, 2021.

[66] Y. Chen and X. Wang, "A hybrid stock trading system using genetic network programming and mean conditional value-at-risk," *European Journal of Operational Research*, vol. 240, no. 3, pp. 861–871, 2015.

[67] W. Wang, W. Li, N. Zhang, and K. Liu, "Portfolio formation with preselection using deep learning from long-term financial data," *Expert Systems with Application*, vol. 143, pp. 113042.1–113042.17, 2020.

[68] Z. Mohamed, D. Mohamad, and O. Samat, "Maximizing portfolio diversification benefit via extended mean-variance model," in *Proceedings of the IEEE Symposium on Industrial Electronics and Applications*, p. 10, Penang, Malaysia, 2010.

[69] G. Y. N. Tang, "How efficient is naive portfolio diversification? an educational note," *Omega*, vol. 32, no. 2, pp. 155–160, 2004.

[70] S. Sun, S. Wang, and Y. Wei, "A new ensemble deep learning approach for exchange rates forecasting and trading," *Advanced Engineering Informatics*, vol. 46, Article ID 101160, 2020.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 9240374, 15 pages
https://doi.org/10.1155/2021/9240374

Research Article

Incentive and Coordination in the Two-Sided Market: Evidence from the P2P Lending Market

Yingxiu Zhao and Baojuan Shi

North China University of Science and Technology School of Economics, Tangshang 063210, China

Correspondence should be addressed to Baojuan Shi; shibaojuan@ncst.edu.cn

Received 13 September 2021; Revised 22 October 2021; Accepted 22 October 2021; Published 13 November 2021

Academic Editor: Xiao Li

Copyright © 2021 Yingxiu Zhao and Baojuan Shi. 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.

Two-sided markets serve as information intermediaries by connecting participants on both sides. In this study, we focus on the coordination of participants in the P2P lending market using a coupon strategy as an incentive to attract investment. Using a two-sided market model, we find that when a platform adopts the coupon strategy, (i) the platform utility and participants' utility are both greater and (ii) the number of participants is greater. In addition, as most research on two-sided markets and coupon strategy focuses on theoretical models, our study provides empirical support using data from Renrendai.com over 2018 to 2019.

1. Information

A two-sided market is roughly defined as a market in which one or several platforms enable interactions between end-users and attempt to get two (or more) sides "on the board" by appropriately pricing each side [1]. Due to cross-market benefits, market participation by the two sides depends on their expectations of the extent to which the other side of the market will participate. As market intermediaries, two-sided platforms face a challenging coordination problem that consists of attracting both participants. To examine this coordination problem, in this study, we introduce the investment in the coupon strategy by two-sided platforms.

There are several examples of two-sided platform markets adopting the coupon strategy, including e-commerce platforms such as Alibaba, Meituan.com, and Xiecheng.com, as well as online lending (e.g., Renrendai.com) in China and Amazon, eBay, and Lending Club outside China. However, studies of the coordination of the coupon strategy in these markets are scarce. Hagiu and Spulber introduce investment in first-party content as a strategic variable for two-sided platforms, show the interplay of platforms' pricing strategies, and define the coordination is the platform that matches the supplier and demander through a first-party content strategy so that the platform utility can be maximized when the utility of both sides is not less than 0 [2]. Dinerstein et al. explore the role of platform design in online markets, emphasizing the trade-off between reducing search friction by matching buyers to their most desired products and intensifying price competition among sellers [3]. Jung et al. found that a government-provided subsidy is optimal for platforms, consumers, and service providers [4].

Many studies of the coordination of the coupon strategy in a single market exist. The researcher provides a mathematical model that addresses the coordination of the schedules between a manufacturer and its suppliers in a decentralised supply chain [5, 6]. Ali explores the coordination and incentive issues that may arise from the different types of collaborations between telecommunication firms and banks [7]. Reimers and Xie develop a model of consumer demand to empirically study whether (i) firms can use coupons as a means of price discrimination to attract new consumers without cannibalizing revenue from existing ones and (ii) these new consumers return to the firm after price promotion [8]. Edelman et al. analyse the conditions under which a business can benefit from issuing e-coupons [9], and Dholakia provides survey-based evidence that some businesses benefit from using e-coupons by finding new consumers [10]. Zhou et al. [11] studied the value of a

cashback website to an e-shop in both centralised and decentralised settings. However, as two-sided markets are more complex than single markets [12, 13], and since it is necessary to coordinate the market participants on both sides, it is worth studying more in detail whether the coupon strategy in two-sided markets can coordinate both sets of participants.

In particular, we focus on the P2P lending market, which serves as an intermediary between borrowers and investors. And the participants have limited attention [14, 15]. In the P2P lending market, borrower and investor participation represent a coordination game for which there are typically many Nash equilibria [2]. The use of the coupon strategy could thus serve as a strategic instrument to induce borrower and investor participation. We focus on the extent to which using the coupon strategy affects the number of participants and their utility. To this end, we propose an analytical model based on the two-sided market model [2, 16]. Through the model analysis, we make three main findings. First, we find that the number of borrowers and investors is greater when the platform adopts the coupon strategy. Second, the utility of borrowers and investors is greater than for platforms that do not adopt the coupon strategy. Third, the profit difference between a platform adopting the coupon strategy and the one not doing so is inverted U-shaped with the coupon investment. Thus, the benefit to the platform increases under an appropriate coupon strategy.

We further build on the standard notion of unfavourable and favourable expectations [2, 17–19] to explore the ease with which a platform adopts the coupon strategy. The model analysis shows that if using a coupon strategy raises the number of borrowers and investors, the platform invests more into its coupon strategy when it is expected to be favourable. If platform utility decreases between favourable and unfavourable expectations, investments in the coupon strategy increase when the expectation are unfavourable.

Our empirical analysis employs detailed data from Renrendai.com, a leading P2P lending platform in China. Renrendai.com is an online market for unsecured personal loans. We analyse the annual reports of the platform from 2012 to 2019 and find that the platform has adopted the coupon strategy since the beginning of 2018. Therefore, this study mainly uses data from January 2018 to December 2019. Using regression analysis, we confirm that the results are consistent with the earlier results of the model. We find that the number of borrowers and investors is higher and that they have greater utility in the months with the coupon strategy. These results suggest that platform utility is greater with the coupon strategy. In addition, the empirical results show that the platform (Renrendai.com) adopts a coupon strategy under unfavourable expectations.

Theoretically and empirically, the main contribution is that we find two-sided market platforms promote the coordination of participants using the coupon strategy. Firstly, the results highlight that an appropriate coupon strategy increases the number of participants and their utility. Thus, it is verified that the impact of coupon on the externalities is not only a simple question of "quantity" but also a question of "quality", which extended the conclusion of Trabucchi et al. [20]. Although the platform is an information intermediary, it can optimize the trading results of participants, using strategies such as coupons, and promote the healthy development of the entire trading system. This result also complements the research of management perspective on two-sided platforms [12]. Secondly, the difference of the platform's benefits between adopting the coupon strategy and not doing so is inverted U-shaped with the coupon strategy investment. This implies that the effectiveness of using coupons to increase profits depends on the number of participants and the coordination ability of the platform, which supplements the conclusion of Bo et al. [6]. In addition, as most research on two-sided markets or coupon strategy focuses on theoretical models, our study bridges this gap in the literature by providing empirical support, shedding light on how managers can determine the number of coupons to be used under a coordination strategy.

The remainder of this paper proceeds as follows. Section 2 reviews the relevant literature. The two-sided market model is built, and the hypotheses are proposed in Section 3. Section 4 describes the data and regression model. The empirical results are presented in Section 5. Section 6 provides robustness checks. Finally, we discuss our findings in Section 7.

2. Literature Review

Two primary streams of the literature relate most closely to our work: studies of two-sided markets and studies of the coupon strategy. First, several branches of the literature on two-sided markets yield insights; Trabucchi et al. [20] present an overview of the dynamics governing value mechanisms within a platform through the theoretical lens of the business model by using two prominent cases of multisided platforms. The other research also includes network externalities, network effects, and multiproduct pricing. Rochet and Tirole, for example, emphasize the importance of the relative price elasticity of demand on the two sides of the platform for determining pricing structures [1, 16]. In addition, most analyses apply to specific industries: payment systems [21, 22], Internet-related intermediation [23–25], software and video games [18, 26, 27], and media markets [28]. Jullien [29] and Rochet and Tirole [1] propose useful roadmaps, while some researchers discuss the online lending market as a two-sided market, including Wei and Lin [30], Gao et al. [31], and Liu et al. [32]. However, studies of how to promote coordination in two-sided markets are scarce [2–4] as noted earlier. We study the influence of coupon strategy on the number of participants and their utility, and it is verified that the impact of coupons on the externalities is not only a simple question of "quantity" but also a question of "quality" [20].

Second, most research on the coupon strategy focuses on a single market. Lu and Moorthy [33] compared the effectiveness of coupons and rebates as price discrimination tools. The models of Levedahl [34], Narasimhan [35], and Dube et al. [36] pertain to coupons or rebates. Arcelus et al. develop a single-period decision model for a retailer facing uncertain and price-dependent demand by introducing a

rebate to address excess demand [37]. Song and Fai Tso [38] research the influence of marketing strategy on the consumers' decision. Zhou et al. [11] developed a model to study the value of a cashback website to an e-shop in both centralised and decentralised settings. Moreover, they also demonstrate and compare the two parties' optimal strategies and their corresponding profits in both settings. In addition, Zhang et al. [39] studied how online retailers employing discounted prices affected customer behaviour in the short and long term using a randomised field experiment with data from the Alibaba Group. Due to the complexity of two-sided markets, the impact of coupon strategy should be further studied, which complements the research of management perspective on two-sided platforms [12].

3. Two-Sided Market Model

To model a two-sided market, we focus on the P2P lending market. P2P lending connects a continuum of investors i uniformly distributed over $[0, N]$ and a continuum of borrowers j uniformly distributed over $[0, M]$. Pricing for borrowers is decided by the platform, that is, the expenditure of the borrower φ, including pay for investor w, and other management rates for the platform. In addition, the platform provides investors with x coupon strategy. The platform incurs the costs of providing a coupon $c(x)$, which are increasing and convex in x.

3.1. Benchmark Model. In this section, we provide the case in which platforms do not adopt the coupon strategy as the benchmark model. Investor i's net benefit from joining the platform is

$$U_i(W_i, P_i) = u(W_i) - P_i, \tag{1}$$

where $u(W_i)$ is the utility of investor i, which comes from borrowers. $u(\cdot)$ is increasing, and W_i is the total return of investor i by investing in borrowers. The platform charges service fee P_i to investor i. The details are shown in Appendix.

Borrower j's net benefit from joining the platform is

$$V_j(n_j, \varphi_j) = U_j - \varphi_j, \tag{2}$$

where $U_j = \sum_{i=1}^{n_j} u_{ji}$ is the utility of borrower j, u_{ji} is the utility of borrower j from investor i, and n_j is the number of investors who invest in borrower j. φ_j is the expenditure of the borrower j, including the pay for investors and other management rates for the platform.

We assume that borrower participation $m = m(P, \varphi)$ and investor participation $n = n(P, W)$.

The platform's profits are

$$\Pi(P, \varphi) = Pn(P, W) + (\varphi - W)m(\varphi, P) - C, \tag{3}$$

where C is the platform cost. The platform charges an average service fee P to investors; W is the average benefit of investors from borrowers, and φ is the average expenditure of borrowers.

In particular, the number of investors in the platform is $n = n(P, W) = \max\{(u(W) - P/A)D, 0\}$, which has a positive correlation with investor utility, A is the average investment of investors, and D is a constant. The number of borrowers in the platform is $m = m(\varphi, P) \in [0, M]$.

3.2. Model under the Coupon Strategy. When the platform adopts the coupon strategy, investor i's net benefit from joining the platform is

$$U_i(W_i, x, P_i) = u(W_i, x) - P_i, \tag{4}$$

where $u(W_i, x)$ is the utility that investor i obtains from the platform and borrowers, including the coupon strategy of the platform and the return by investing borrowers. W_i is the return of investor i by investing in borrowers. We suppose that the investors' utility function $u(\cdot,\cdot)$ is increasing, twice continuously differentiable, and concave or linear in each of its arguments. As is natural, investor utility is increasing in coupons and the benefit to borrowers. Thus, an investor joins the platform if and only if the expected net benefit is non-negative.

Borrower j's net benefit from joining the platform is

$$V_j(n_j, x, \varphi_j) = U_j - \varphi_j - \phi, \tag{5}$$

where $U_j = \sum_{i=1}^{n_j} \pi_{ji}(x) u_{ji}$ is the utility of borrower j. $\pi_{ji}(x)$ is the degree of investor i's willingness to invest in borrower j, $\pi_{ji}(0) = 1$, and $\pi_{ji}(x) \in [0, H)$. n_j is the same as equation (2). The function $\pi(x)$ can be increasing or decreasing depending on the coupon strategy. A borrower chooses to participate if the expected benefits are non-negative.

The game with the simultaneous entry of the borrower and investor has two stages. First, the profit-maximizing platform chooses the amount of coupon x and the average fees P and φ for investors and borrowers, respectively. Second, borrowers and investors choose whether to join the platform based on the prices chosen by it and their expectations about participation by borrowers on the other side of the market. The equilibrium of the full game consists of the first-stage choices of platform, (P^*, φ^*, x^*), and the resulting Nash equilibrium of the borrower and investor from their participation decisions in the second stage. There are multiple self-fulfilled participation equilibria for any vector (P, φ, x) of platform choices.

The platforms' profits are

$$\Pi(p, \varphi, x) = \mathrm{Pn}(x, P, W) + (\varphi - W)m(x, P, \varphi) - c(x) - C. \tag{6}$$

The platform chooses coupon x as well as the average service fee of investor P and the average expenditure of borrower φ to maximize profits, given the effects of its actions on the outcome of the participation subgame played by borrowers and investors. Therefore, maximizing the utility yields

$$\max_{P,\varphi,x}\{\mathrm{Pn}(x,P,W)+\varphi m(x,P,\varphi)-c(x)-C\},$$
$$s.t.\ U_j-\varphi_j-\phi\geq 0, \tag{7}$$
$$u(W_i,x)-u(P_i)\geq 0.$$

According to $U_i(W_i,x,P_i)\geq 0$ and $V_j(n_j,x,\varphi_j)\geq 0$, the platform utility is maximized as follows:

$$\max_{P,\varphi,x}\{\mathrm{Pn}(x,P,W)+(\varphi-W)m(x,P,\varphi)-c(x)-C\},$$
$$s.t.\ \pi(x)\overline{\mathrm{un}}-\varphi\geq 0, \tag{8}$$
$$u(W,x)-P\geq 0,$$

where $\pi(x)$ is the average degree of investors' willingness to invest in borrowers, $\overline{u}$ is the average utility of borrowers, and $u(W,x)$ is the average utility of investors. $\overline{n}=u(W,x)-P/\mathrm{Am}D$ is the average number of investors investing. The participation of borrowers is $m=m(P,\varphi,x)$ when $V_j(n_j,x,\varphi_j)<0$, $m=0$, and the platform reaches the service limit when $V_j(n_j,x,\varphi_j)>0$, $m=M$. And the participation of investors is $n=n(P,W,x)$, where $n=\max\{u(W,x)-P/AD,0\}$.

3.3. Comparison of the Two Situations. Comparing equations (1) and (4), we find that the utility of investors is greater under the coupon strategy. Additional investigation leads to the first proposition. Table 1 summaries the notation.

Proposition 1. *Under the coupon strategy, the utility of investors is greater $U_i(W_i,x,P_i)>U_i(W_i,P_i)$, and there are more investors.*

Proposition 1 is proven as shown in Appendix. The platform increases its costs if it adopts the coupon strategy.

To study the influence of the coupon strategy on platform utility. Let equation (6) subtract equation (3), we have equation (9) as follows:

$$\begin{aligned}\Pi(P,\varphi,x)-\Pi(P,W)=&P[n(x,P,W)-n(P,W)]\\&+(\varphi-W)[m(x,P,\varphi)\\&-m(P,\varphi)]-c(x).\end{aligned} \tag{9}$$

We suppose $\Pi=\Pi(P,\varphi,x)-\Pi(P,W)$, equation (9) obtains the partial derivative of x as follows:

$$\Pi'=\mathrm{Pn}_x(x,P,W)+(\varphi-W)m_x(x,P,\varphi)-c'(x). \tag{10}$$

When $c'(x)<Pn_x(x,P,W)+(\varphi-W)m_x(x,P,\varphi)$, the difference between the platform's benefit from investing in the coupon strategy and not investing increases with x; when $c'(x)>Pn_x(x,P,W)+(\varphi-W)m_x(x,P,\varphi)$, the difference decreases with x; and when $c'(x)=Pn_x(x,P,W)+(\varphi-W)m_x(x,P,\varphi)$, the difference is the maximized. Furthermore, we also find that the coupon investment x is related to the number of participants. Proposition 2 summarizes these findings.

Proposition 2. *The profit difference between a platform adopting the coupon strategy and not doing so is inverted U-shaped with coupon investment x. Therefore, the platform adopts an appropriate coupon strategy, and this brings about positive benefits.*

3.4. Favourable and Unfavourable Expectations. Whether a platform has favourable or unfavourable expectations depends on the level of participation by both sides. A platform has favourable expectations (F platform) when borrowers and investors always coordinate at the equilibrium at the highest possible level of participation on both sides. On the contrary, a platform has unfavourable expectations (UF platform) when borrowers and investors always coordinate at the equilibrium at the lowest possible level of participation on both sides.

We first consider the F platform. It maximizes profit, which is

$$\max_{P,\varphi,x}\left\{P\left[\frac{u(\mathscr{W},x)-P}{A}D\right]+(\varphi-\mathscr{W})M-c(x)-C\right\},$$
$$s.t.\ \pi(x)\overline{u}\left[\frac{u(\mathscr{W},x)-P}{\mathrm{AM}}D\right]-\varphi\geq 0,$$
$$u(\mathscr{W},x)-P\geq 0 \tag{11}$$

where $\mathscr{W}$ is the maximum benefits of investors by investing in borrowers when the platform has the largest number of borrowers. Then, the platform's maximization problem becomes

$$\begin{aligned}\max_{P,x}\pi^F(P,x,\varphi)=\max_{P,x,\varphi}\Bigg\{&[P+\overline{u}\pi(x)]\left[\frac{u(\mathscr{W},x)-P}{A}D\right]\\&-c(x)-C-\mathscr{W}M\}.\end{aligned} \tag{12}$$

The first-order conditions determining the F platform's optimal choices of investor price and coupon strategy are

$$P_f=\frac{u\left(x_f,\mathscr{W}\right)-\overline{u}\pi\left(x_f\right)}{2},$$
$$c'\left(x_f\right)=\frac{\left[u\left(x_f,\mathscr{W}\right)+\overline{u}\pi\left(x_f\right)\right]\left[u_x\left(x_f,\mathscr{W}\right)+\pi_x(x)\overline{u}\right]}{2A}D, \tag{13}$$

where $u_x(W,x)=(\partial u/\partial x)(W,x)$; $\pi_x(x)=(\mathrm{d}\pi/\mathrm{d}x)(x)$.

Next, we consider the UF platform, which has few borrowers and investors. Assume that the number of borrowers Y in the platform is not much larger than 0. The optimization problem is

$$\max_{P,w,x}\left\{P\left[\frac{u(\mathrm{W}',x)-P}{A}D\right]+(\varphi-W)Y-c(x)-C\right\},$$
$$s.t.\ \pi(x)\overline{\mathrm{un}}-\varphi\geq 0,$$
$$u(W,x)-P\geq 0, \tag{14}$$

where $\mathrm{W}'\leq\mathscr{W}$. The resulting platform profits are

Table 1: Parameter definitions.

Parameter	Define
i	Investor i
j	Borrower j
r_{ij}	The real interest rate of investor i from borrower j
r_{aij}	The nominal interest rate for an investor to invest in borrower j priced by the platform
A_{ij}	The biding amount of investor i investing in borrower j
m_i	The number of borrowers invested by investor i
e_{ij}	The platform's service fee rate
W_i	The return of investor i by investing in m_i borrowers
P_i	The service fee charged by the platform to investor i
x	An amount of the coupon strategy
U_j	The utility of borrower j
n_j	The number of investors who invest in borrower j
φ_j	The expenditure of borrower j
$\pi_{ji}(x)$	The investment intention of investor i who invests in borrower j
A	The average investment of investors
P	The average service of investors
φ	The average expenditure of borrowers
C	The platform cost
$\overline{u}$	The average utility of borrowers
m, n	The number of borrowers and investors, respectively
W	The average benefit of investors from borrowers
$c(x)$	The platform's costs of providing coupons
$\mathscr{W}$	The return of investors by investing in borrowers when the platform has the largest number of borrowers

$$\max_{P,x}\left\{P\left[\left(\frac{u(\mathrm{W}',x)-P}{A}\right)D\right]+Y\pi(x)\,\overline{u}max\left[\left(\frac{u(W,x)-P}{\mathrm{AY}}\right)D,0\right]-c(x)-\mathrm{YW}-C\right\}. \tag{15}$$

The optimal investor price and coupon strategy for the UF platform are (x_{UF}, P_{UF}). Calculating the first-order conditions in equation (15), we obtain (x_{UF}, P_{UF}) as follows:

$$P_{UF}=\frac{u(x,\mathrm{W}')-\overline{u}\pi(x)}{2},$$
$$c'\left(x_{uf}\right)=D\frac{\left[u_x(x,\mathrm{W}')-\pi_x(x)\overline{u}\right]\left[u\left(x,\mathscr{W}^{\cdot}\right)-\overline{u}\pi(x)\right]}{2A}+D\frac{\overline{u}\pi(x)u_x(x,W)+\overline{u}\pi_x(x)u(x,W)}{A}. \tag{16}$$

Based on equations (12)–(16), we propose two lemmas that characterise the platform's optimal choice of coupon strategy.

Lemma 1. *With the entry of borrowers and investors, the F platform's optimization problem is equivalent to solving* $\max_x \Pi^F(x)$*, where*

$$\Pi^F(x)=\max_x\left\{\frac{\left[u(x,\mathscr{W})+\overline{u}\pi(x)\right]^2}{4A^2}D-c(x)-C-\mathscr{W}M\right\}. \tag{17}$$

Lemma 2. *With the entry of borrowers and investors, the UF platform's optimization problem is equivalent to solving* $\max_x \Pi^{UF}(x)$*, where*

$$\Pi^{UF}(x)\max_x\frac{\left[u(x,\mathscr{W})-\overline{u}\pi(x)\right]^2}{4A}D+\overline{u}\pi(x)\frac{u(W,x)}{A}-c(x)-C-\mathrm{WY}. \tag{18}$$

Let $\Delta(x)=\Pi^F(x)-\Pi^{UF}(x)>(<)0$ *represent the gain (loss) in profitability due to unfavourable expectations. The UF platform selects a higher (lower) investment in x when* $\Delta(x)$ *is decreasing (increasing). Furthermore, we derive our main result (proven in Appendix), which compares platforms' investment under the coupon strategy.*

Proposition 3. *If coupon strategy x raises the number of borrowers and investors, the F platform invests more in the coupon strategy. In addition, if* $\Delta(x)=\Pi^F(x)-\Pi^{UF}(x)>0$*, the UF platform selects a higher (lower) investment in x when* $\Delta(x)$ *is decreasing (increasing).*

3.5. Hypotheses. Vana et al. analyse panel data from a large cashback company and show that an increase in the probability that consumers make an additional online purchase raises the size of that purchase [40]. Moreover, the authors find that consumers are more likely to spend the money returned to them at generalist retailers, such as department stores than at other retailers. Sahni et al. found that discounts increase the demand for promoted products and revenue [41]. Overall, we thus propose the following hypotheses.

H1. The platform adopting the coupon strategy increases the number of investors.

H2. To attract participants, the platform adopts the coupon strategy regardless of whether it faces unfavourable or favourable expectations.

H3. The utility of borrowers, investors, and the platform is larger when adopting the coupon strategy.

In a two-sided market, buyers experience positive network effects depending on the number of suppliers and suppliers receive similar positive network effects depending on the number of buyers [25, 27]. Therefore, we hypothesise the following:

H4. The platform adopting the coupon strategy increases the number of borrowers.

4. Data and Methodology

4.1. Data. China has regulated the P2P lending market in recent years. From the end of 2017, regulatory authorities issued many regulatory documents, and the regulations became stronger in 2019. Table 2 shows the main regulatory documents in 2018 and 2019. With the increase in supervision, many P2P lending platforms have closed, leading to unfavourable expectations in the P2P lending industry. Table 2 shows that the regulatory authorities adopted more policies in 2019 than in 2018 as regulatory efforts became stronger.

The data are from Renrendai.com, one of the largest P2P lending platforms in China. Renrendai.com was founded in October 2010 and has the largest number of registered members for P2P lending in the country. By the end of 2019, it had about five million confirmed loans with a lending amount of 50 billion RMB and 200 million registered investors. In 2018, the platform has launched a coupon strategy as an incentive to attract investors and implemented more coupons in 2019 than in 2018, and the results are shown in Table 3. Therefore, the platform adopts the coupon strategy when facing unfavourable expectations, which implies that Hypothesis 3 is unsuitable for P2P lending markets.

To compare the changes before and after the coupon strategy, we use data from January 2017 to December 2019 in Renrendai.com. Table 4 shows the growth in annual transaction amount and the number of transactions between 2016 and 2019 (see also Figure 1). We find that the transaction amount increased from 2016 to 2017, then decreased in 2018, and the growth rate became negative in 2019. However, the growth in the annual number of transactions rapidly increased in 2019, with growth in the transaction number from 2016 to 2018.

Table 5 summarizes the descriptive statistics of investors (Panel A) and borrowers (Panel B) in 2017, 2018, and 2019. Panel A of Table 5 shows that the average number of investors does not decrease over time but does come more dispersed. Panel B shows that the number of borrowers increases and the amount of borrowing decreases, investors prefer to invest in short-term objects, and the yield decreases little. Hence, given the unfavourable expectations of P2P lending, whether the situation of Renrendai.com is related to the incentive strategy adopted by the platform deserves exploration through an empirical analysis.

4.2. Variables and Methodology. The data set from Renrendai.com includes loan information of 717 days. The main independent variable is the coupon strategy. The dependent variables include the number of borrowers and investors on day t, average investors' return on day t, and average borrower utility and platform utility on day t. Rysman uses advertising demand to express the effectiveness of advertisers. Thus, the utility of borrowers is a proxy for the number of investors in this study [22]. Qiu et al. study platform utility using trading volume or management fees [42]. Thus, we choose the average daily trading volume as a proxy of platform utility. In addition, as there are network externalities in two-sided markets [1, 21, 43], we use the number of borrowers and investors on day $t-1$ and other listing information as the control variables. Table 6 defines the variables used in our analysis, and Table 7 provides the descriptive statistics for those variables between 2018 and 2019.

To study the influence of the coupon strategy on participants, the following empirical model is adopted:

$$\text{Borrower}_t = \beta_0 + \beta_1 \text{Coupon} + \beta_2 \text{interest}_{t-1} + \beta_3 \text{Borrower}_{t-1} + \beta_4 \text{lender}_{t-1} + \xi_{it}, \tag{19}$$

$$\text{Lender}_t = \beta_0 + \beta_1 \text{Coupon} + \beta_2 \text{Borrower}_t + \beta_3 \text{Lender}_{t-1} + \beta_4 X + \xi_{it}, \tag{20}$$

$$L\text{return}_t = \beta_0 + \beta_1 \text{Coupon} + \beta_2 \text{Borrower}_t + \beta_3 X + \xi_{it}, \tag{21}$$

$$\text{Trading}_t = \beta_0 + \beta_1 \text{Coupon} + \beta_2 \text{Borrower}_t + \beta_3 \text{Lender}_t + \beta_3 X + \xi_{it}, \tag{22}$$

where Coupon is coded 1 if the platform adopts a coupon strategy and 0 otherwise. interest_{t-1} is the average interest rate at time $t-1$. Borrower_{t-1} is the number of borrowers at time $t-1$, and lender_{t-1} is the number of investors at time $t-1$. X includes information on listings such as loan amount, loan terms, and interest rate. Borrower_t is the number of borrowers at time t, Lender_t is the number of investors and average utility of borrowers at time t, and $Tra\,di\,ng_t$ is the trading volume and the average platform utility at time t. $L\text{return}_t$ is the average utility of investors, which is calculated as follows:

Table 2: The policy of online lending in 2018 and 2019.

Time	Policy
2018.1.	"Notice of special rectification and acceptance of P2P lending risk"
2018.4.	"Notice of strengthening rectification and acceptance of asset management business through Internet"
2018.8.	"Notice of compliance inspection of P2P lending institutions" issued by the Internet finance regulation office
2019.1.	"No.1 Document" issued by the Internet finance regulation office
2019.4.	"Pilot work plan for the qualified record of online lending information intermediary institutions"
2019.7.	"Forum of the special rectification of network lending risk"
2019.9.	"Notice of strengthening the construction of a credit reporting system in the P2P lending field"
2019.11.	"Guiding opinions on the pilot transformation of online lending information intermediaries into small loan companies"

Table 3: The coupon strategies time provided by the platform.

2018	2019
1.10.–1.17.	1.1.–1.15.
5.8.–5.21.	4.1.–4.30.
5.28.–6.18.	5.8.–5.14.
7.2.–7.17.	5.15.–5.30.
8.8.–8.28.	6.6.–6.12.
9.13.–9.30.	6.13.–6.26.
10.1.–10.8.	7.4.–7.10.
11.8.–11.21.	8.1.–8.21.
	9.1.–9.30.
	10.1.–10.25.
	11.1.–11.29.
	12.1.–12.25.

Note. There are 120 days and 226 days in which the platform adopts coupon strategy in 2018 and 2019 years, respectively.

Table 4: The growth of the transaction amount and number of transactions.

Year	Transaction amount (%)	Number of transactions (%)
2019	−78.78	179.8
2018	38.47	49.34
2017	94.86	105.16
2016	84.70	16.17

Note. Transaction amount is the year-on-year increase in transaction amount; and the number of transactions is the year-on-year increase in year-on-year increase in the number of transactions.

$$L\text{return}_t = \text{bidamount}_t * \text{rate}_t, \tag{23}$$

where bidamount_t is the average investment at time t and $rate_t$ is the interest rate at time t.

5. Empirical Results

We begin our empirical analysis using OLS regression. Tables 8–10 summarize the estimation results for borrowers, investors, and the platform, respectively. Table 8 shows the influence of coupon strategy on the number of investors and investors' return at time t. The relationship between the coupon strategy and the number of investors is significantly positive at the 1% level, as shown in column (1) of Table 8. There are similar results for investors' returns in column (5) of Table 8. When controlling for the information on listings, such as interest rate, we find that the coupon strategy has a positive impact on the number of investors and investors' returns (see columns (2) and (6) of Table 8). As shown in columns (3) and (7) of Table 8, the results are significantly positive at the 5% and 1% levels, respectively, when participant information is controlled. We obtain similar results in columns (4) and (8), which are controlled for using listing information and participant information at the same time. Furthermore, the influence of the control variables on the number of investors and the investors' returns is reasonable. For example, although a higher interest rate attracts more investors, the possibility of borrower default is higher. Therefore, there is a positive correlation with the number of investors and a negative correlation with the investors' return [44]. In columns (3), (4), (7), and (8), the number of investors at time $t-1$ has a positive impact on the number of investors at time t, suggesting herding behaviour. However, having more investors leads to less personal income. Therefore, Hypotheses 1 and 3 are supported.

Table 9 reports the results of the relationship between the coupon strategy and platform utility. As shown in column (1), the coupon strategy has a positive effect on the trading volume, which is significantly at the 1% level. As shown in columns (2)–(8), the results are also significantly positive at the 1% level when information on listings and the number of participants is controlled for. Meanwhile, when we control for information on listings and participants at the same time, the relationship between the coupon strategy and trading volume remains significantly positive, which is shown in column (9) of Table 9. Furthermore, we find that the loan term has a negative influence on the trading volume, whereas the number of borrowers and investors has a positive influence. Overall, the influence of the coupon strategy remains unchanged regardless of the influence of the other control variables. Platform utility is higher when the platform adopts the coupon strategy, supporting Hypothesis 3.

Table 10 shows the baseline analysis of the influence of the coupon strategy on the number of borrowers at time t. We find that using coupons promotes the number of borrowers in the P2P lending platform in column (1) of Table 10, which is significant at the 1% level. When the interest rate, the number of borrowers, and the number of investors at time $t-1$ are individually controlled for, the relationship between the coupon strategy and the number of borrowers at time t is significant and positive at the 1% level, while those control variables are significant in columns (2), (3), and (4), respectively. Meanwhile, the result is significant at the 5% level when controlled, the variables are controlled for at the same time, and the influence of the control variables is

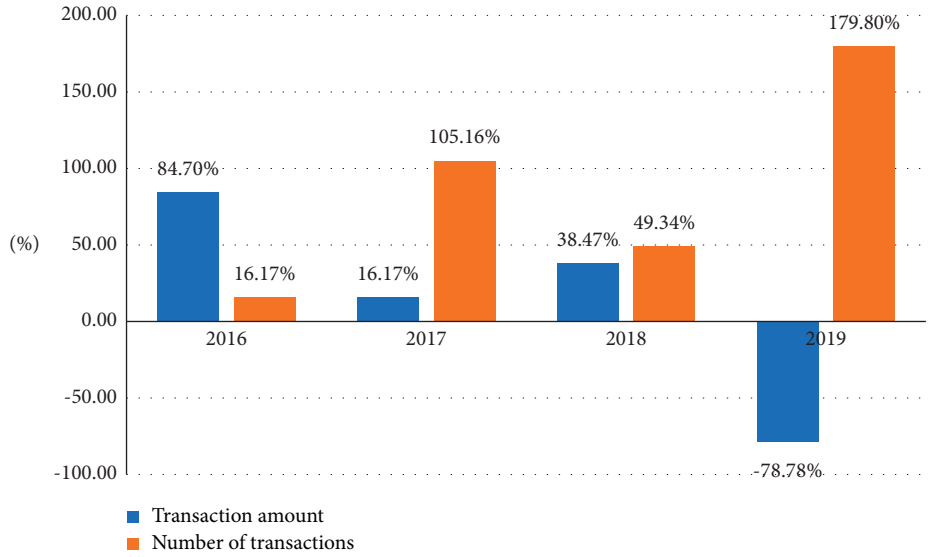


Figure 1: The growth of transaction amount and number of transactions.

Table 5: Descriptive statistics of participation number.

Year	Obs. (Times)	Max	Min	Mean	Median
Panel A. Descriptive statistic of the annual number of investors					
2017	475840 (28860637)	18337750	50	45823	10300
2018	589791 (27328645)	23735150	50	51191	11200
2019	439806 (51176179)	16047650	50	52553	12050
Panel B. Descriptive statistics of borrowers					
$Interest_{17}$	272107 (274269)	11.76	8	9.82	10.2
$Month_{17}$	272107 (274269)	48	3	32.13	36
$Amount_{17}$	272107 (274269)	376000	1000	79501	73700
$Interest_{18}$	394618 (409593)	10.2	9	9.75	9.6
$Month_{18}$	394618 (409593)	48	12	31.01	36
$Amount_{18}$	394618 (409593)	197500	5000	73712	63600
$Interest_{19}$	1680165 (3139456)	10.8	9.6	10.21	10.2
$Month_{19}$	1680165 (3139456)	36	3	9.66	9
$Amount_{19}$	1680165 (3139456)	200000	300	7362	2100

Panel A is descriptive statistics of the number of investors every year, the Obs. is the number of investors, and the Times is the number of bids times. Panel B is descriptive statistics of borrowers including listings' information, the Obs. is the number of borrowers, and the Times is the loan times. $Interest_{17,18,19}$ is the interest rate in 2017, 2018, and 2019 year, respectively. $Month_{17,18,19}$ is the loan term in 2017, 2018, and 2019 year, respectively. $Amount_{17,18,19}$ is the borrowing amount in 2017, 2018, and 2019 year, respectively.

Table 6: Definition of variable.

Variable	Definition
Borrower	The daily number of borrowers
Lender	The daily number of investors
Lreturn	Average daily investors' return
Trading	Average daily platform utility (average daily trading volume)
Coupon	The platform adopts the coupon strategy; 1 is the coupon strategy adopted in the month, otherwise 0.
Interest	Average daily interest rate
Amount	Average daily loan amount
Month	Average daily loan term
Bidamount	Average daily investment amount

Note. Borrower, Lender, and Lreturn are the dependent variables; Coupon is the main independent variable; and the others are the control variables.

constant in column (5). The results suggest a herding effect among borrowers and that investors have positive cross-network externalities with borrowers. This evidence supports Hypothesis 4.

In addition, we use the difference-in-difference methodology to test the causal effect of adopting the coupon strategy on the number of participants. Our treatment group comprises the months for which the platform adopted its

Table 7: Summary statistics.

Variable	Obs.	Mean	Std.	Min	Max	Med
Borrower	717	4906.496	7595.286	1	68114	1393
Lender	717	50012.85	57445.14	2	235533	10608
Lreturn	717	93.643	82.754	14.881	1170	60.172
Coupon	717	0.529	0.499	0	1	1
Interest	717	9.905	7.994	9	10.344	9.868
Month	717	25.306	9.39	7.994	36	29.835
Amount	717	68226.14	38517.03	1748.521	139886.8	80472.38
Bidamount	717	955.412	871.48	145.888	13000	619.397

Note. We identify participants for 717 days between January 2018 and December 2019 on Renrendai.com.

Table 8: The influence of coupon strategy on the number of investors and their return.

	Lender				Lreturn			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coupon	1.213***	0.253 * *	0.187 * *	1.877***	0.326***	0.017*	0.991***	0.681***
	(8.96)	(2.39)	(2.52)	(6.64)	(6.29)	(1.83)	(4.99)	(3.33)
Interest		35.75***		6.514***		−10.61***		−1.525
		(10.96)		(3.23)		(−7.54)		(−1.20)
Month		−3.918***		−0.064		0.466***		0.323 * *
		(−10.11)		(−0.25)		(2.79)		(2.04)
Amount		1.400***		0.197 * *		0.003		0.260***
		(8.58)		(1.98)		(0.04)		(4.18)
Lender_{t-1}			0.642***	0.716***			−0.213***	−0.193***
			(32.98)	(29.14)			(−14.92)	(−12.84)
Borrower			0.485***	0.271***			−0.023***	−0.028
			(16.20)	(7.15)			(−1.15)	(−1.41)
B∗Coupon			0.103***	0.063			0.087***	0.132***
			(2.83)	(1.59)			(2.67)	(3.99)
C	6.840***	−7.532***	−0.101	−16.741***	4.437***	27.10***	6.491***	7.998***
	(4.51)	(−9.69)	(−0.48)	(−3.56)	(7.63)	(8.08)	(2.83)	(2.69)
N	717	717	716	716	717	717	716	716
R^2	0.10	0.533	0.844	0.848	0.10	0.376	0.525	0.574

Note. The dependent variables are the number of investors and their return (remarked as Lender and Lreturn) at time t. The independent variable is the coupon (remarked as Coupon) at time t. Lender_{t-1} is the control variable, which is the number of investors at time $t-1$. Month, Amount, and Borrower are the control variable at time t, which are the listings' loan term, amount, and the number of borrowers, respectively. B∗Coupon is the cross-term of the number of borrowers and coupons.

coupon strategy (defined as 1), and our control group includes noncoupon strategy months (defined as 0). Renrendai.com adopted its coupon strategy on 1 January 2018. Therefore, the control sample runs from 1 January 2017 to 31 December 2017, and the treatment sample runs from 1 January 2018 to 31 December 2018.

We estimate the effect of the coupon strategy in a regression framework using the following model:

$$Y = \beta_0 + \beta_1 \text{time}_{it} \times \text{Treated}_{it} + \beta_2 \text{time}_{it} + \beta_3 \text{Treated}_{it} + \beta_4 \text{Control}_{it} + \xi_{it}, \tag{24}$$

where Y represents the number of borrowers and investors. $time_{it}$ is a dummy variable defined as 1 for 2018 and 0 otherwise. Treated_{it} is a dummy variable defined as 1 if the platform adopts the coupon strategy at time t and 0 otherwise. Control_{it} is a vector of the control variables including those in regressions (19) and (20), and ξ_{it} is the error term. The coefficient of the double interaction term captures the causal effect of the coupon strategy.

Without considering the control variables, Table 11 lists the differences between the treatment and control groups, which is the number of participants before and after the platform adopts the coupon strategy. When the platform adopts the coupon strategy, the average number of investors and borrowers increases significantly in the treatment group in column (6). The value of the difference in the average number of borrowers and investors is 0.678 and 0.108, respectively, significant at the 1% level. Furthermore, in the difference-in-difference model, we consider the other control variables controlled for in the basic regression. The results are shown in Table 12. The coefficient β_1 is positively significant at the l% level, which implies that the platform adopts an appropriate coupon strategy to raise the number of participants. Thus, Tables 11 and 12 provide support for Hypotheses 1 and 4.

6. Robustness Tests

In this section, we provide robustness checks for our main results of the influence of the coupon strategy on the number of borrowers and investors as well as participants' utility.

Table 9: The influence of coupon strategy on the trading number.

	Trading								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Coupon	1.192***	0.415***	0.263***	0.288***	0.279***	0.605***	0.392***	0.373***	0.183 * *
	(10.05)	(4.48)	(2.87)	(3.00)	(3.14)	(5.74)	(4.16)	(3.96)	(2.21)
Interest		54.37***			23.94***				13.24***
		(4.16)			(6.30)				(3.51)
Month			−2.110 * *		−2.81***				−3.166***
			(−6.19)		(−6.65)				(−7.43)
Amount				−0.866***	0.639***				0.887***
				(−3.93)	(3.70)				(5.40)
Lender						0.477***		0.087*	0.125***
						(7.23)		(2.37)	(3.33)
Borrower							0.684***	0.609***	0.434***
							(4.01)	(4.29)	(10.57)
C	6.687***	−7.70***	3.63***	6.21***	−5.84***	2.403***	2.023***	1.757***	−4.77***
	(9.58)	(−3.29)	(5.01)	(4.02)	(−5.05)	(9.30)	(9.92)	(7.56)	(−2.74)
N	715	715	715	715	715	715	715	715	715
R^2	0.123	0.525	0.553	0.513	0.584	0.380	0.515	0.518	0.645

Note. The dependent variables are the average daily utility of platforms (remarked as Trading, the average daily Trading volume) at time t. The independent variable is the coupon (remarked as Coupon) at time t. Lender and borrower are the control variable, which are the number of investors and borrowers at time $t-1$, respectively. Month, Amount, and Borrower are the control variable at time t, which are the listings' loan term, amount, and the number of borrowers, respectively. t statistics are given in parentheses. $^{***}p$ <0.01, $^{**}p$ <0.05, and $^{*}p$ <0.1.

Table 10: The influence of coupon strategy on the number of borrowers.

	Borrower				
	(1)	(2)	(3)	(4)	(5)
Coupon	0.972***	0.543***	0.524***	0.238***	0.180 * *
	(8.08)	(5.29)	(5.28)	(3.15)	(2.41)
Interest_{t-1}		39.60***			12.36***
		(17.83)			(5.89)
Lender_{t-1}			0.511***		0.071 * *
			(19.66)		(2.27)
Borrower_{t-1}				0.795***	0.757***
				(34.90)	(22.18)
C	6.840***	−83.66***	2.219***	1.394***	−25.94***
	(4.51)	(−16.49)	(9.15)	(8.54)	(−5.58)
N	717	716	716	716	716
R^2	0.18	0.34	0.428	0.674	0.759

Note. The dependent variable is the number of borrowers (remarked as Borrower) at time t. The independent variable is the coupon (remarked as Coupon) at time t. Interest_{t-1}, Lender_{t-1}, and Borrower_{t-1} are the control variables which are the interest rate, the number of investors, and borrowers at time $t-1$, respectively. t statistics are given in parentheses. $^{***}p$ <0.01, $^{**}p$ <0.05, and $^{*}p$ <0.1.

Because the data do not satisfy the normal distribution, we use a nonparametric Wilcoxon test to study the number of participants and trading volume before and after the implementation of the coupon strategy. Using data from January 2018 to December 2019, we divide the number of participants and trading volume into two groups: the platform does not adopt a coupon strategy at time t (group 1) and the platform adopts a coupon strategy at time t (group 2). Panel A of Table 13 shows that the number of borrowers in group 2 is significantly more than that in group 1. In 2018, 2019, or both years, the p-value is significant at the 5% and 1% levels. Panel B shows that the number of investors in group 2 is larger than that in group 1 at a significance level of 1%, which implies that the coupon strategy can increase the number of investors. Panel C reports that the trading volume of group 2 is significantly higher than that of group 1, which means that the platform increases its utility by adopting an appropriate coupon strategy.

Recall that our main results in Tables 8–10 use the full data sample (January 2018 to December 2019) to verify the hypothesis. We now examine whether these results are robust to including the subsamples of January 2018 to December 2018 and January 2019 to December 2019. We reestimate the regressions (19)–(21) and find that the results remain significantly positive at the 1% and 10% levels, as shown in Table 14. Although some of the results are not significant in columns (6), (9), (10), and (12), they are significant at the 10% level. In other words, these results are significant overall. Therefore, Hypotheses 1, 3, and 4 are again supported.

Table 11: The difference of participants between experimental group and control group under coupon strategy.

	Control		Treat		Difference		DID
	(1)	(2)	(3)	(4)	(5) = (2)−(1)	(6) = (4)−(3)	(7) = (6)−(5)
Lender	9.813	9.921	9.606	10.393	0.108	0.786***	0.678***
					(0.480)	(4.04)	(2.50)
Borrower	5.853	6.369	6.020	6.782	0.516***	0.762 * *	0.108***
					(2.76)	(2.43)	(2.61)

Note. The Control group is that the day of the participants not affected by the coupon strategy, and the Treat group is that the day of the participants affected by the coupon strategy. t statistics are given in parentheses. ***p <0.01, **p <0.05, and *p <0.1.

Table 12: The regression result of DID.

Variables	Borrower	Lender
$time \times Treate\ d$	0.669***	0.731***
	(3.85)	(2.63)
$time$	0.697*	−0.243
	(1.83)	(−1.06)
$Treate\ d$	0.540***	0.086
	(4.52)	(0.380)
$Control$	Y	Y
C	−9.311***	7.131***
N	(−8.10)	(3.02)
R^2	0.461	0.350

Note. The table provides estimation results from a regression in which the dependent variable is the number of borrowers and investors. $time \times Coupon$ is the cross-term of the time and coupon. $time$ is a dummy variable equal to one if the time is 2018 year, otherwise 0. $Coupon$ is a dummy variable equal to one if platform adopts the coupon. $Control$ is a vector of control variables that include the control variables in regressions (I) and (II). t statistics are given in parentheses. ***p <0.01, **p <0.05, and *p <0.1.

Table 13: Wilcoxon test of the number of participants.

	G1			G2			Z	P
	Obs.	Mean	Std.	Obs.	Mean	Std.		
Panel A. The number of borrowers								
B2019	129	3453.07	4680.706	236	11286.75	10007.894	−1.888	0.034
B2018	210	1104.633	623.34	143	1238.83	676.88	−8.752	0.000
B	338	2003.35	3140.31	379	7495.58	9285.62	−9.419	0.000
Panel B. The number of investors								
L2019	126	74709.89	44990.68	236	104564.35	49969.51	−3.871	0.000
L2018	210	3850.89	2558.79	143	5150.01	3412.31	−3.871	0.000
L	338	30904.06	44280.81	379	67054.45	62328.84	−9.112	0.000
Panel C. The number of trading volume								
T2019	140	3609.1	410.58	224	11548.77	672.24	−9.060	0.000
T2018	216	877.59	32.99	135	978.23	46.45	−1.967	0.029
T	356	1951.78	177.122	359	7573.77	499.18	−9.861	0.000

Note. Table 13 shows the Wilcoxon test of the number of participants. Group 1 is that the platform does not adopt the coupon strategy in time t, and group 2 is that the platform adopts the coupon strategy in time t. B2019, L2019, and T2019 are the number of borrowers and investors in the 2019 year, and the trading volume in the 2019 year, respectively. B2018, L2018, and T2018 are the similar implication in the 2018 year. B, L, and T are the overall situation of two years.

Table 14: The influence of coupon strategy on the participant by using the subset of the data sample.

	Lender				Lreturn				Borrower			
	2018		2019		2018		2019		2018		2019	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Coupons	0.374***	0.985***	0.660***	1.262***	0.326***	0.017*	0.025 * *	1.901***	0.184*	0.108*	1.581***	0.052*
	(3.31)	(2.79)	(4.99)	(5.59)	(6.29)	(1.83)	(2.36)	(3.92)	(1.76)	(1.89)	(8.88)	(1.95)
Interest		−7.488***		−17.11***		−10.61***		3.451				
		(−2.81)		(−6.01)		(−7.54)		(1.22)				
Month		2.089***		1.456***		0.466***		−2.340***				
		(3.97)		(6.73)		(2.79)		(−4.18)				
Amount		−0.920***		−0.208***		0.003		0.892***				
		(−5.12)		(−3.03)		(0.04)		(4.54)				
Lender_{t-1}		0.283***		0.167***				−0.215***		−0.053		0.0004*
		(9.92)		(7.48)				(−6.91)		(−0.6)		(1.67)
L*coupon								−0.499***				
								(−6.11)				
Borrower		0.707***		0.946***				0.017				
		(22.45)		(31.79)				(0.50)				
B*coupon		0.131 * *		0.166***				0.341***				
		(2.54)		(5.81)				(4.48)				
Lender_{t-1}										−3.811		9.969 * *
										(−1.06)		(2.00)
Borrower_{t-1}										0.593***		0.964***
										(7.49)		(12.34)
C	7.913***	2.13***	10.67***	9.24***	4.437***	27.10***	4.746***	−3.621	6.646***	11.79	7.043***	−19.42*
	(10.25)	(3.72)	(10.02)	(5.86)	(7.63)	(8.08)	(8.29)	(−0.60)	(2.96)	(1.45)	(9.19)	(−1.68)
N	353	352	365	365	717	717	353	352	353	352	365	365
R^2	0.073	0.784	0.062	0.889	0.10	0.376	0.050	0.327	0.063	0.302	0.176	0.794

Table 14 shows the regression analysis in 2018 and 2019, respectively. The number of investors, borrowers, and trading volume are the dependent variables. The independent variable is the coupon, and the control variables are similar to those in Tables 8-10. L*Coupon is the cross-term of the number of lenders and coupons. t statistics are given in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

7. Conclusions

In this paper, we examine how adopting the coupon strategy affects investors' investment decisions and willingness of borrowers in two-sided platforms. Most research on two-sided markets is based on theoretical models. Hence, we use the two-sided market model, which focuses on the incentive strategy of the platform, to produce empirically verifiable hypotheses, which we test using data from Renrendai.com.

Firstly, the benchmark model is established and then the adoption of the coupon strategy is added into the benchmark model. Finally, we compare the utility of investors, borrowers, and the platform as well as the number of participants in the two models. Under the coupon strategy, participants' utility and the number of participants were larger. Although the platform enjoys more participation and more benefits under the coupon strategy, it is further affected by the additional costs of running the coupon strategy. Overall, the platform utility is higher when the platform adopts an appropriate coupon strategy. Additionally, we find that the platform adopts a coupon strategy to attract investors regardless of whether it is expected to be favourable or not. If the difference between the utility under unfavourable and favourable expectations decreases gradually, the platform invests more in the coupon strategy when it is expected to be unfavourable.

Furthermore, the empirical results confirm our theoretical predictions, showing that the number of investors and borrowers is significant when the platform adopts a coupon strategy. The utility of participants is greater in the months in which the coupon strategy is adopted. Moreover, we find that Renrendai.com is more likely to adopt a coupon strategy when it is expected to be unfavourable.

While our analysis focuses on the P2P lending market, the conclusions can also be generalized to other markets. The two-side platform invests a coupon strategy that has yielded imperative guidance for the behaviour of managerial and participants. For example, for the managers of the two-sided market, who can analyse and summarize the number of participants every day and then put forward the most appropriate coupon strategy. Finally, this paper has some limitations. For instance, we did not consider borrowers, defaulting (credit risk), or other complex situations. Therefore, the influence of other management strategies and the corresponding problems are worthy of study in future research.

Appendix

A. Proofs

The return of investor i is

$$R_i = \sum_{j=1}^{m_i} r_{ij} A_{ij} - \sum_{j=1}^{m_i} \max\left\{\left(r_{ij} - r_{aij}\right), 0\right\} A_{ij}, \tag{A.1}$$

where remark $e_{ij} = r_{ij} - r_{aij}$, m_i is the number of borrowers which invested by investor i, r_{ij} is the real interest rate of investor i from borrower j. A_{ij} is the biding amount of investor i investing borrower j. r_{aij} is the nominal interest rate for an investor invested in the borrower j, which priced by the platform. The platform charges service fee rate e_{ij} to investor i when invest borrower j. Then equation (A.1) is rewired as follows:

$$R_i = \sum_{j=1}^{m_i} r_{ij} A_{ij} - \sum_{j=1}^{m_i} e_{ij} A_{ij}. \tag{A.2}$$

We suppose that $W_i = \sum_{j=1}^{m_i} r_{ij} A_{ij} = \sum_{j=1}^{m_i} w_{ij}$ indicates that the investor i obtains the total return by investing in m_i borrowers, and $P_i = \sum_{j=1}^{m_i} e_{ij} A_{ij} = \sum_{j=1}^{m_i} p_{ij}$ is the service fee charged by the platform to investor i.

Proof of Proposition 1. Comparing equation (1) to (4), we obtain

$$U_i(W_i, x, P_i) - U^b{}_i(W_i, P_i) = u(W_i, x) - u_b(W_i), \tag{A.3}$$

$u(\cdot)$ is increasing, and the investors' utility function $u(\because)$ is increasing, twice continuously differentiable, and concave or linear in each of its arguments. Thus, $u(W_i, x) - u_b(W_i)$, and $n \geq n_b$. □

Proof of Proposition 4. Assume that x_f, x_{uf} are uniquely defined. We compare the first-order condition determining x_f (equation (13)) with the one determining x_{uf} (equation (16)). Let us plug x_f instead of x_{uf} in (16) taking (13) into account, then

$$\begin{aligned}
&\frac{\left[u_x\left(x_f, \mathscr{W}\right) - \overline{u}\pi_x\left(x_f\right)\right]\left[u\left(x_f, \mathscr{W}\right) - \overline{u}\pi\left(x_f\right)\right]}{2A} D + D \\
&\quad \frac{\overline{u}\pi(x)u_x(x, W) + \overline{u}\pi_x(x)u(x, W)}{A} - c'\left(x_f\right) \\
&= \frac{\overline{u}D}{A}\left\{\pi\left(x_f\right)\left[u_x(x, W) - u_x\left(x_f, \mathscr{W}\right)\right] + \pi_x\left(x_f\right)\right. \\
&\quad \left.\left[u(x, W) - u\left(x_f, \mathscr{W}\right)\right]\right\}.
\end{aligned} \tag{A.4}$$

The investors' utility function $u(\because)$ is increasing, twice continuously differentiable, and concave or linear in each of its arguments$u_x(x_f, W) - u_x(x_f, \mathscr{W}) < 0, u(x_f, W) - u(x_f, \mathscr{W}) < 0$. Therefore, the expression on the right side of the equation is negative, so that $x_f > x_{uf}$. □

Data Availability

The data were taken from the Renrendai.com platform.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J.-C. Rochet and J. Tirole, "Two-sided markets: a progress report," *The RAND Journal of Economics*, vol. 37, no. 3, pp. 645–667, 2006.

[2] A. Hagiu and D. Spulber, "First-Party content and coordination in two-sided markets," *Management Science*, vol. 59, no. 4, pp. 933–949, 2013.
[3] M. Dinerstein, L. Einav, J. Levin, and N. Sundaresan, "Consumer price search and platform design in internet commerce," *The American Economic Review*, vol. 108, no. 7, pp. 1820–1859, 2018.
[4] Q. Jung, B. Kim, M. Park, and W. Straub, "Innovation and policy support for two-sided market platforms: can government policy makers and executives optimize both societal value and profits?" *Information Systems Research*, vol. 30, no. 3, pp. 1–14, 2019.
[5] C. Shi and H. Lee, "Incentive alignment and coordination of project supply Chains," *Management Science*, vol. 63, no. 4, pp. 1011–1025, 2016.
[6] H. Bo, Y. Chen, H. Li, and L. Qi, "Time-sensitive Supply Chain disruption recovery and resource sharing incentive strategy," *Journal of Management Science and Engineering*, vol. 6, no. 2, pp. 165–176, 2021.
[7] T. Ali, *Incentive and Coordination Issues in the Mobile Banking Industry of Pakistan*, MIT Sloan School of Management, Cambridge, MA, USA, 2014.
[8] I. Reimers and C. Xie, "Do coupons expand or cannibalize revenue? Evidence from an e-Market," *Management Science*, vol. 65, no. 1, pp. 1–15, 2019.
[9] B. Edelman, S. Jaffe, and S. D. Kominers, "To Groupon or not to Groupon: the profitability of deep discounts," *Marketing Letters*, vol. 27, no. 1, pp. 39–53, 2014.
[10] M. Dholakia, *How Effective Are Groupon Promotions for Businesses?*, Rice University, Houston, TX, USA, 2010.
[11] Y.-W. Zhou, B. Cao, Q. Tang, and W. Zhou, "Pricing and rebate strategies for an e-shop with a cashback website," *European Journal of Operational Research*, vol. 262, no. 1, pp. 108–122, 2017.
[12] D. Trabucchi and T. Buganza, "Landlords with no lands: a systematic literature review on hybrid multi-sided platforms and platform thinking," *European Journal of Innovation Management*, 2021.
[13] D. Trabucchi, S. Sanasi, A. Ghezzi, and T. Buganza, "Idle asset hunters-the secret of multi-sided platforms," *Research-Technology Management*, vol. 64, no. 1, pp. 33–42, 2021.
[14] Y. Li, J. W. Goodell, and D. Shen, "Comparing search-engine and social-media attentions in finance research: evidence from cryptocurrencies," *International Review of Economics & Finance*, vol. 75, no. 4, pp. 723–746, 2021.
[15] Y. Hu, X. Li, W. Goodell, and D. Shen, "Investor Attention Shocks and Stock Co-movement: Substitution or Reinforcement?" *International Review of Financial Analysis*, vol. 73, 2021.
[16] J.-C. Rochet and J. Tirole, "Platform competition in two-sided markets," *Journal of the European Economic Association*, vol. 1, no. 4, pp. 990–1029, 2003.
[17] B. Caillaud and B. Jullien, "Chicken & egg: competition among intermediation service providers," *The RAND Journal of Economics*, vol. 34, no. 2, pp. 309–328, 2003.
[18] A. Hagiu, "Pricing and commitment by two-sided platforms," *The RAND Journal of Economics*, vol. 37, no. 3, pp. 720–737, 2006.
[19] H. Halaburda and Y. Yehezkel, "Platform Competition under Asymmetric Information," *American Economic Journal: Microeconomics*, vol. 5, 2011.
[20] D. Trabucchi, T. Buganza, and R. Verganti, "Quantity or quality? Value creation in two-sided platforms," *Technology Analysis & Strategic Management*, vol. 33, no. 2, pp. 162–175, 2021.
[21] J. Wright and U. Kaiser, "Price structure in two—sided markets: evidence from the magazine industry," *International Journal of Industrial Organization*, vol. 24, no. 1, pp. 1–28, 2006.
[22] M. Rysman, "Competition between networks: a study of the market for yellow pages," *The Review of Economic Studies*, vol. 71, no. 2, pp. 483–512, 2004.
[23] M. Armstrong, *Competition in Two-Sided Markets*, Industrial Organization, Vienna, Austria, 2006.
[24] M. Rysman, "An empirical analysis of payment card usage," *The Journal of Industrial Economics*, vol. 55, no. 1, pp. 1–36, 2007.
[25] S. Li, Y. Liu, and S. Bandyopadhyay, "Network effects in online two-sided market platforms: a research note," *Decision Support Systems*, vol. 49, no. 2, pp. 245–249, 2010.
[26] G. G. Parker and M. W. Van Alstyne, "Two-sided network effects: a theory of information product design," *Management Science*, vol. 51, no. 10, pp. 1494–1504, 2005.
[27] J. M. Gallaugher and Y.-M. Wang, "Understanding network effects in software markets: evidence from web server pricing," *MIS Quarterly*, vol. 26, no. 4, pp. 303–327, 2002.
[28] M. Reisinger, L. Ressner, and R. Schmidtke, "Two-sided markets with pecuniary and participation externalities," *The Journal of Industrial Economics*, vol. 27, no. 1, pp. 1–26, 2009.
[29] B. Jullien, "Two-sided markets and electronic intermediaries," *CESifo Economic Studies*, vol. 51, no. 2, pp. 233–260, 2005.
[30] Z. Wei and M. Lin, "Market mechanisms in online Peer-to-Peer lending," *Management Science*, vol. 63, no. 12, pp. 1–22, 2017.
[31] G.-X. Gao, Z.-P. Fan, X. Fang, and Y. F. Lim, "Optimal Stackelberg strategies for financing a supply chain through online peer-to-peer lending," *European Journal of Operational Research*, vol. 267, no. 2, pp. 585–597, 2018.
[32] H. Liu, H. Qiao, S. Wang, and Y. Li, "Platform competition in Peer-to-Peer lending considering risk control ability," *European Journal of Operational Research*, vol. 274, no. 1, pp. 280–290, 2019.
[33] Q. Lu and S. Moorthy, "Coupons versus rebates," *Marketing Science*, vol. 26, no. 1, pp. 67–82, 2007.
[34] J. W. Levedahl, "Marketing, price discrimination, and welfare: Comment," *Southern Economic Journal*, vol. 50, no. 3, pp. 886–891, 1984.
[35] C. Narasimhan, "Competitive promotional strategies," *Journal of Business*, vol. 61, no. 4, pp. 427–449, 1988.
[36] J.-P. Dubé, Z. Fang, N. Fong, and X. Luo, "Competitive price targeting with smartphone coupons," *Marketing Science*, vol. 36, no. 6, pp. 944–975, 2017.
[37] F. J. Arcelus, R. Gor, and G. Srinivasan, "Price, rebate and order quantity decisions in a newsvendor framework with rebate-dependent recapture of lost sales," *International Journal of Production Economics*, vol. 140, no. 1, pp. 473–482, 2012.
[38] L. Song and G. K. Fai Tso, "Consumers can learn and can forget - modeling the dynamic decision procedure when watching TV," *Journal of Management Science and Engineering*, vol. 5, no. 2, pp. 87–104, 2020.
[39] D. J. Zhang, H. Dai, L. Dong et al., "The long-term and spillover effects of price promotions on retailing platforms: evidence from a large randomized experiment on Alibaba," *Management Science*, vol. 66, no. 6, pp. 2589–2609, 2020.
[40] P. Vana, A. Lambrecht, and M. Bertini, "Cashback is cash forward: delaying a discount to entice future spending,"

Journal of Marketing Research, vol. 55, no. 6, pp. 852–868, 2018.

[41] N. Sahni, D. Zou, and P. Chintagunta, "Do targeted discount offers serve as advertising? Evidence from 70 field experiments," *Management Science*, vol. 63, no. 8, pp. 1–18, 2016.

[42] J. Qiu, F. Nie, M. Tong, and G. Hu, "The characteristics of two side market of third-party electronic trading platform—an empirical analysis based on online personal lending market," *Journal of Management Science in China*, vol. 19, no. 1, pp. 47–59, 2016.

[43] E. Argentesi and L. Filistrucchi, "Estimating market power in a two-sided market: the case of newspapers," *Journal of Applied Econometrics*, vol. 22, no. 7, pp. 1247–1266, 2007.

[44] F. Fisher and B. Guimaraes, "U.S. real interest rates and default risk in emerging economies," *Journal of Money, Credit, and Banking*, vol. 45, no. 5, pp. 967–975, 2013.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 2010705, 20 pages
https://doi.org/10.1155/2021/2010705

Research Article

Dependence and Risk Spillover among Hedging Assets: Evidence from Bitcoin, Gold, and USD

Jiang Yu,[1] **Yue Shang**,[2] **and Xiafei Li**[1]

[1]*School of Economics and Management, Southwest Jiaotong University, Chengdu, China*
[2]*School of Marxism, Yunnan University of Finance and Economics, Kunming, China*

Correspondence should be addressed to Xiafei Li; 546602717@qq.com

Received 2 August 2021; Accepted 30 August 2021; Published 13 September 2021

Academic Editor: Dehua Shen

Copyright © 2021 Jiang Yu 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.

Understanding the dependence and risk spillover among hedging assets is crucial for portfolio allocation and regulatory decision making. Using various copula and conditional Value-at-Risk (*CoVaR*) measures, this paper quantifies the dependence and risk spillover effects between three traditional and emerging hedging assets: Bitcoin, gold, and USD. Furthermore, we investigate these effects at various short- and long-term horizons using a variational model decomposition (VMD) method. The empirical results show that there is strong negative dependence between gold and USD, but Bitcoin and gold are weakly and positively connected. Secondly, risk spillovers exist only between Bitcoin and gold and between gold and USD. The risk spillover effect between Bitcoin and gold are not stable, that is, if Bitcoin or gold faces the downward or upward risk, both the downward and upward risk of another asset have the chance to increase. The negative risk spillover between gold and USD is stable, especially in long-term horizons. Finally, the risk spillover between Bitcoin and gold as well as between gold and USD are asymmetric at downward and upward market environment.

1. Introduction

In recent years, the global financial and energy markets have frequently experienced huge price volatility due to the increasing uncertainties in real economy and international geopolitical conflicts, etc. To deal with the increasing volatility/risk in financial and energy markets, investors usually choose hedging assets like gold and US dollar (USD) to offset market risk and lock in profits [1–10]. Since Nakamoto [11] introduced the concept of Bitcoin, more and more investors have paid attentions to digital cryptocurrency, making digital cryptocurrency transactions play an increasingly prominent role in financial press and investment scene. Of all digital cryptocurrencies, Bitcoin typically has the largest trading volume. In December 2017, the Chicago Board Options Exchange (CBOE) and the Chicago Mercantile Exchange (CME) Group even launched the Bitcoin-based futures contracts. The financialization of Bitcoin and its characteristics of high returns and high volatility mean that any evidence of risk spillovers between Bitcoin and other financial assets would have major impact on risk management, investment portfolio management, and policy making [12–14]. Specially, Bitcoin shares some similarities with gold and USD. The Commodity Futures Trading Commission (CFTC) officially declared that though Bitcoin does not have legal tender status, it is a digital representation of value that can be used as a medium of exchange, a unit of account, and/or a store of value. Besides, several famous financial media like CNN and Bloomberg have labeled Bitcoin as the New Gold, and CFTC also declared virtual money a commodity like gold. These mean that if users of Bitcoin are rational, Bitcoin has some intrinsic value like gold. Besides, as a peer-to-peer electronic cash system, Bitcoin allow one party of transactions pays another directly online without going through a financial institution, thus having payment function and liquidity like USD. The research of Dyhrberg [15] has also proved that gold maximizes the attribute of value storage by sacrificing liquidity while USD maximizes liquidity by sacrificing storage value, but Bitcoin combines the characteristics of gold and USD and can be used for payments, portfolio management, and risk hedging like gold and USD.

Other literature has also attempted to explore Bitcoin's ability to hedge market risks and found that it can be used for payment, portfolio management, and risk hedging [15–23]. However, some studies about the hedging capabilities of Bitcoin have found different results. For example, Baur et al. [13] found no significant relationship between Bitcoin and traditional asset classes by a regression analysis, suggesting that Bitcoin is more suitable for using as a speculative but not as a hedging asset. Klein et al. [19] also showed that Bitcoin is not the New Gold and has no hedging ability for stock market. This means that whether Bitcoin can replace the role of gold and USD as hedging assets in financial market remains controversial. So, it still makes sense to explore the linkage between Bitcoin and gold/USD.

As two commonly used hedging assets, the relationship between gold and USD has been widely explored [1, 5, 24–34]. These research studies demonstrate negative correlations between gold and USD by a variety of linear and nonlinear methods like VAR framework, ARDL model, spillover index, value-at-risk measures, copula functions, multivariate GARCH-class model, etc., suggesting that gold can be used as a hedge against a falling USD. The relationship between Bitcoin and gold as well as between Bitcoin and USD has also been explored in recent years [13, 19, 23, 35–41]. However, research studies on the relationship between Bitcoin and gold as well as between Bitcoin and USD yield some controversial results. For example, Baur et al. [13] found that Bitcoin return is uncorrelated to gold and USD returns by a linear regression analysis, whereas Bouri et al. [36] demonstrated that Bitcoin return correlated closely with gold and USD returns by a smooth transition VAR-GARCH-in-mean model. Besides, Bouri et al. [36] also showed that the sign of the spillovers between Bitcoin and gold/USD and the significance of the spillovers are different during bear and bull markets. However, Wang et al. [23] proved that there is no return spillover between Bitcoin and gold/currency, and volatility spillover exist only between Bitcoin and gold by a VAR-GARCH-BEKK model. Interesting, using different extensions to the NARDL model, Bouri et al. [37] found significant negative influence of gold and USD prices on Bitcoin price, whereas Jareño et al. [38] found significant positive connectedness between Bitcoin and gold. The findings of Bouri et al. [37] and Jareño et al. [38] both showed that the short-term and long-term influence of gold to Bitcoin is asymmetric. These two opposite results may due to the fact that the correlation between Bitcoin and gold/USD are dynamics, while these results are obtained in static models. The studies of Zwick and Syed [41] and Klein et al. [19] have illustrated this point. Zwick and Syed [41] showed that there is a structural break for the relationship between Bitcoin and gold prices. This structural break occurred on 5th October 2017. Gold negative correlated with Bitcoin before this time, whereas it correlated positively with Bitcoin after that. Klein et al. [19] obtained the dynamic correlation of Bitcoin and gold by a dynamic BEKK-GARCH model, observing that the correlation between Bitcoin and gold fluctuate around 0. Overall, the spillover between Bitcoin and gold/USD remains unclear, and exploring the complex relationship between them remains a key topic.

The research studies above suggest that the relationship between Bitcoin, gold, and USD may be dynamic and nonlinear. Besides, it may be asymmetry between bull and bear markets, and between short-term and long-term. This means that it would make sense to capture these characteristics simultaneously in a research process. For these considerations, we capture the nonlinear dependent structure between Bitcoin, gold, and USD by various static and dynamic bivariate copula functions and explore the risk spillover effect between them in upward and downward market environment by conditional value-at-risk ($CoVaR$) and delta$CoVaR$ ($\Delta CoVaR$) measurements described by Adrian and Brunnermeier [42]. Then, VMD technology is further utilized to explore the dependence structure and risk spillover effect in short-term and long-term horizons. The$CoVaR$ and $\Delta CoVaR$ measurements of Adrian and Brunnermeier [42] are widely used for quantifying the positive risk spillover effect from one financial market to another, that is, the possibility of market 1 facing the downward (upward) risk when a related market 2 facing a downward (upward) risk [43–49]. However, gold and USD are negatively correlated, and Klein et al. [19], Bouri et al. [37], and Zwick and Syed [41] also found that gold may be negatively correlated with Bitcoin. So, we extend the $CoVaR$ measurements used in Ji et al. [50] and Ji et al. [51], further exploring the possibility of target market 1 facing the downward (upward) risk when the related market 2 facing an upward (downward) risk, i.e., negative risk spillover effect from market 2 to market 1. As far as we know, only a few studies have explored the nonlinear interdependence and risk spillover effects between Bitcoin, gold, and USD by combining these approaches.

This paper contributes to the existing literature as we find some interesting results using the copula functions and the extended $CoVaR$ measurements together with the VMD method. Our empirical results first find strong negative dependence between gold and USD, followed by the weak positive dependence between Bitcoin and gold but nearly no dependence between Bitcoin and USD. Besides, there is no risk spillover effect between Bitcoin and USD but unstable risk spillover effect between Bitcoin and gold. In other words, both the downward and the upward risk of Bitcoin (gold) have the chance to increase when gold (Bitcoin) facing the downward/upward risk. It confirms the view of Baur et al. [13] and Klein et al. [19] that Bitcoin may be more speculative and can hardly be used as an alternative currency and may have no stable hedging ability like gold, either. Moreover, there is stable negative risk spillover effect between gold and USD, especially in long-term horizons. Interestingly, in short-term horizons, when USD (gold) is facing the downward or upward risk, both the downward and upward risk of gold (USD) have the possibility to increase, indicting the stable hedging ability of gold in long-term horizons but unstable hedging power in short-term horizons. The risk spillover between gold and USD and from Bitcoin to gold is stronger in the upward market environment, while the risk spillover from gold to Bitcoin is stronger in the downward market environment. These empirical

results have some important implications for investors and policy makers.

The rest of this paper is organized as follows: Section 2 describes the empirical methods utilized in this paper, Section 3 describes the data, Sections 4 and 5 discuss the empirical results, and Section 6 concludes the work.

2. Methodology

2.1. Model for Marginal Distribution. Before estimating the copula function, we employ the AR-GARCH-skew-t model to construct the marginal distribution of Bitcoin, gold, and USD index. In this model, the conditional mean is calculated by an AR (m) model:

$$r_t = \phi_0 + \sum_{i=1}^{m} \phi_i r_{t-i} + \varepsilon_t, \tag{1}$$

in which ϕ_0 and ϕ_i are the parameters to be estimated, m is the lag order of returns, and $\varepsilon_t = \sigma_t z_t$, $z_t \sim$ i.i.d. $D(0,1)$ is the error item. The conditional variance, σ_t^2, is given by a GARCH model:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{2}$$

where ω, α_i, β_j, and λ_i are the parameters to be estimated and $\sum_i^p \alpha_i + \sum_j^q \beta_j$ should be smaller than 1, p and q are the lag order of ARCH term and GARCH term, respectively. Brooks [52] proves that using a GARCH-class model with one lag order is sufficiently enough to describe the volatility clustering in asset returns. So, we set m, p, and q, to be one, constructing an AR (1)-GARCH (1, 1) model. To capture the fat tail and asymmetry distribution feature of Bitcoin, gold, and USD returns, we assume the i.i.d. random variable, z_t, follows the skewed-t distribution [53]. The density of skew-t distribution is given as follows:

$$f(z_t; v, \eta) = \begin{cases} bc\left(1 + \frac{1}{v-2}\left(\frac{bz_t + a}{1-\eta}\right)^2\right)^{-((v+1)/2)}, & z_t < -\frac{a}{b}, \\ bc\left(1 + \frac{1}{v-2}\left(\frac{bz_t + a}{1+\eta}\right)^2\right)^{-((v+1)/2)}, & z_t \geq -\frac{a}{b}, \end{cases} \tag{3}$$

where $a = 4\eta c(v-2)(v-1)^{-1}$, $b^2 = 1 + 3\eta - a^2$, and $c = \Gamma(2^{-1}(v+1))/\sqrt{\pi(v-2)}$; v is the degree of freedom $(2 < v \leq \infty)$; and η is the asymmetric parameter $(-1 < \eta < 1)$. This skewed-t distribution will converge to a standard Gaussian distribution if $\eta = 0$ and $v \longrightarrow \infty$ but will converge to the symmetric Student-t distribution if $\eta = 0$ and v is finite.

2.2. Copula Function. We explore the dependence structure between Bitcoin, gold, and USD using some bivariate copula functions. According to Sklar [54], a bivariate copula can couple two marginal distributions to represent a joint distribution function as

$$F_{XY}(x, y) = C\left(F_X(x), F_Y(y)\right), \tag{4}$$

where $F_{XY}(x, y)$ is the joint distribution function of two random variables X and Y; $F_X(x)$ and $F_Y(y)$ are the marginal distribution functions of X and Y, respectively; and $C(u, v)$ for $u = F_X(x)$ and $v = F_Y(y)$ is uniquely determined on $\mathrm{Ran}F_X \times \mathrm{Ran}F_Y$ when margins are continuous.

Copula functions allow modeling for the dependence structures and marginal distributions, providing flexibility in characterizing dependence. Some copula functions can also capture the tail dependence, that is, the probability that two variables experience extreme upward or downward movement. The upper and lower tail dependence can be obtained by the following:

$$\lambda_U = \lim_{u \longrightarrow 1} \Pr\left[X \geq F_X^{-1}(u) \mid Y \geq F_Y^{-1}(u)\right] = \lim_{u \longrightarrow 1} \frac{1 - 2u + C(u,u)}{1-u},$$
$$\lambda_L = \lim_{u \longrightarrow 0} \Pr\left[X \leq F_X^{-1}(u) \mid Y \leq F_Y^{-1}(u)\right] = \lim_{u \longrightarrow 0} \frac{C(u,u)}{u}, \tag{5}$$

where $\lambda_U, \lambda_L \in [0,1]$. If $\lambda_L > 0$ ($\lambda_U > 0$), there is a lower (upper) tail dependence.

To obtain more accurate dependence characteristics, seven copula functions with different tail dependence characteristics are considered in this study. These copula functions are specified as follows:

(1) *Gaussian Copula*. Gaussian copula function is given by

$$C_N(u, v; \rho) = \Phi\left(\Phi^{-1}(u), \Phi^{-1}(v)\right), \tag{6}$$

where $\rho \in [-1, 1]$ is the correlation parameter; Φ is the standard Gaussian cumulative distribution function; $\Phi^{-1}(u)$ and $\Phi^{-1}(v)$ are standard Gaussian quantile functions. The tail dependence of this copula function is zero, i.e., $\lambda_U = \lambda_L = 0$.

(2) *Student-t Copula*. Student-t copula function is defined as

$$C_{\mathrm{ST}}(u, v; \rho, v) = T\left(t_v^{-1}(u), t_v^{-1}(v)\right), \tag{7}$$

in which v is the parameter of degree-of-freedom, $t_v^{-1}(u)$ and $t_v^{-1}(v)$ are the quantile functions of Student-t, T is the bivariate Student-t cumulative distribution function. This copula function has symmetric nonzero tail dependence, and the tail dependence can be obtained by

$$\lambda_U = \lambda_L = 2t_{v+1}\left(\frac{-\sqrt{v+1}\sqrt{1-\rho}}{\sqrt{1+\rho}}\right). \tag{8}$$

(3) *Clayton Copula*. Clayton copula function is defined as

$$C_{\mathrm{CL}}(u, v; \delta) = \max\left\{\left(u^{-\delta} + v^{-\delta} - 1\right)^{-1/\delta}, 0\right\}. \tag{9}$$

It allows for lower tail dependence and upper tail independence, i.e., $\lambda_L = 2^{-1/\delta}$ $(\delta \geq -1)$ and $\lambda_U = 0$.

(4) *Rotated Clayton Copula*. Rotated Clayton copula is defined as

$$C_{\mathrm{RCL}}(u,v;\delta) = u + v - 1 + C_{\mathrm{CL}}(1-u, 1-v;\delta), \quad (10)$$

and it is a copula function with lower tail independence and upper tail dependence, that is $\lambda_L = 0$, $\lambda_U = 2^{-1/\delta}$ $(\delta \geq -1)$.

(5) *Gumbel Copula*. Gumbel copula is defined as

$$C_G(u,v;\delta) = \exp\left(-\left((-\log u)^{\delta} + (-\log v)^{\delta}\right)^{1/\delta}\right). \quad (11)$$

The tail dependence of this copula function is asymmetric with lower tail independence and upper tail dependence, thus, $\lambda_L = 0$, $\lambda_U = 2 - 2^{1/\delta}$ $(\delta \geq 1)$.

(6) *Rotated Gumbel Copula*. Rotated Gumbel copula is defined as

$$C_{\mathrm{RG}}(u,v;\delta) = u + v - 1 + C_G(1-u, 1-v;\delta). \quad (12)$$

In contrast to the Gumbel copula, this copula function has lower tail dependence but no upper tail dependence, i.e., $\lambda_L = 2 - 2^{1/\delta}$ $(\delta \geq 1)$, $\lambda_U = 0$.

(7) *SJC (symmetrized Joe–Clayton) Copula*. SJC copula is given by

$$C_{\mathrm{SJC}}(u,v;\lambda_U,\lambda_L) = 0.5\left(C_{\mathrm{JC}}(u,v;\lambda_U,\lambda_L) + C_{\mathrm{JC}}(1-u,1-v;\lambda_U,\lambda_L) + u + v - 1\right), \quad (13)$$

where $C_{\mathrm{JC}}(u,v;\lambda_U,\lambda_L)$ is the Joe–Clayton,

$$C_{\mathrm{JC}}(u,v;\lambda_U,\lambda_L) = 1 - \left(1 - \left\{\left[1-(1-u)^{\kappa}\right]^{-\gamma} + \left[1-(1-v)^{\kappa}\right]^{-\gamma} - 1\right\}^{-1/\gamma}\right)^{1/\kappa}, \quad (14)$$

in which $\lambda_L \in (0,1)$, $\lambda_U \in (0,1)$, $\kappa = 1/\log_2(2-\lambda_U)$ and $\gamma = -1/\log_2(\lambda_L)$. The SJC copula has the special symmetric tail dependence, that is, $\lambda_L = \lambda_U$.

Considering the possible time-varying dependence between two variables, we further construct the time-varying copulas by allowing the dependence parameters of the seven copula functions to be dynamic. For the Gaussian copula function with time-varying parameters (TVP Gaussian), the dynamic of dependence parameter, ρ_t, follow an ARMA (1, q) process [55]:

$$\rho_t = \Lambda\left(\psi_0 + \psi_1\rho_{t-1} + \psi_2\frac{1}{q}\sum_{j=1}^{q}\Phi^{-1}\left(u_{t-j}\right)\cdot\Phi^{-1}\left(v_{t-j}\right)\right), \quad (15)$$

where $\Lambda(x) = (1-e^{-x})(1+e^{-x})^{-1}$ is the modified logistic transformation, which can keep ρ_t in (–1, 1) and ψ_0, ψ_1, and ψ_2 are the parameters to be estimated.

For the Student-t copula, two dynamic forms of dependency parameter are considered. One is to make the correlation parameter ρ to change over time, but the degree of freedom υ remain constant. In this time-varying Student-t copula, the dynamic of ρ is given in equation (15) by substituting $\Phi^{-1}(x)$ with $t_{\upsilon}^{-1}(x)$. However, research have shown that the degree of freedom parameter is unlikely to remain constant over time [56–59]. Because both ρ and υ are important parameters for measuring the copula cumulative distribution function and the tail dependence, we further consider another time-varying form for Student-t copula function. This copula function allows both ρ and υ to be time varying, and υ is evolved in [60]:

$$\upsilon_t = \Psi\left(\theta_0 + \theta_1\upsilon_{t-1} + \theta_2\frac{1}{q}\sum_{j=1}^{q}t_{t-j}^{-1}\left(u_{t-j}\right)\cdot t_{t-j}^{-1}\left(v_{t-j}\right)\right), \quad (16)$$

where $\Psi(x) = 2(1+50e^{x})(1+e^{x})^{-1}$ is the logistic transformation for keeping υ_t in (0, 100).

For the SJC copula, the evolutions of the tail dependence parameters follow

$$\lambda_t^U = \Delta\left(\omega_U + \beta_U\rho_{t-1} + \alpha_U\frac{1}{q}\sum_{i=1}^{q}\left|u_{t-i} - v_{t-i}\right|\right),$$
$$\lambda_t^L = \Delta\left(\omega_L + \beta_L\rho_{t-1} + \alpha_L\frac{1}{q}\sum_{i=1}^{q}\left|u_{t-i} - v_{t-i}\right|\right), \quad (17)$$

where $\Delta(x)$ denotes the logistic transformation, $\Delta(x) = (1+e^{-x})^{-1}$ that keeps the parameters of the time-varying SJC copula in (0, 1).

For other four copula functions, the parameters evolve as

$$\delta_t = \omega + \beta\delta_{t-1} + \alpha\frac{1}{q}\sum_{i=1}^{q}\left|u_{t-i} - v_{t-i}\right|. \quad (18)$$

2.3. Measurements for VaR and CoVaR. In this study, we quantify the downside and upside risk of Bitcoin, gold and USD by the downside and upside Value-at-Risk (*VaR*). The downside risk, $\mathrm{VaR}_t^{D,\alpha}$, for a confidence level $1-\alpha$ at time t is given by $\Pr(r_t \leq \mathrm{VaR}_t^{D,\alpha}) = \alpha$, which is the α quantile of the distribution of return series. Similarly, the upside risk, $\mathrm{VaR}_t^{U,\alpha}$, is the $(1-\alpha)$ quantile of the distribution of return series, that is $\Pr(r_t \geq \mathrm{VaR}_t^{U,\alpha}) = \alpha$. These two risks can be measured from marginal distribution models as $\mathrm{VaR}_t^{D,\alpha} = \mu_t + \sigma_t t_{\upsilon,\eta}^{-1}(\alpha)$ and $\mathrm{VaR}_t^{U,\alpha} = \mu_t + \sigma_t t_{\upsilon,\eta}^{-1}(1-\alpha)$, respectively, where, μ_t is the conditional mean from equation (1) σ_t is the

conditional standard deviation from equation (2), $t_{v,\eta}^{-1}(\alpha)$ is the α quantile of the skewed Student t distribution.

To quantify the risk spillover effects between Bitcoin, gold, and USD, the conditional value-at-risk (*CoVaR*) proposed by Adrian and Brunnermeier [42] is utilized. *CoVaR* considers the correlation between two assets and quantifies the risk of asset 1 when a correlated asset 2 suffers extreme risks. Given two potentially correlated asset return series, $r_{1,t}$ and $r_{2,t}$, the downside and upside *CoVaR* of asset 1 are generally expressed as

$$\Pr\left(r_{1,t} \le \mathrm{CoVaR}_{1|2\ D,t}^{D,\alpha} \mid r_{2,t} \le \mathrm{VaR}_{2,t}^{D,\beta}\right) = \alpha, \tag{19}$$

$$\Pr\left(r_{1,t} \ge \mathrm{CoVaR}_{1|2U,t}^{U,\alpha} \mid r_{2,t} \ge \mathrm{VaR}_{2,t}^{U,\beta}\right) = \alpha, \tag{20}$$

respectively, where α and β are the significant level of asset 1 and asset 2, respectively, and $\mathrm{CoVaR}_{1|2D,t}^{D,\alpha}$ ($\mathrm{CoVaR}_{1|2U,t}^{U,\alpha}$) is the downward (upward) risk of asset 1 when the correlated asset 2 is extreme downward (upward) movement. In combination with the copula function, equations (19) and (20) can be rewritten as

$$C\left(F_{1,t}\left(\mathrm{CoVaR}_{1|2\ D,t}^{D,\alpha}\right), F_{2,t}\left(\mathrm{VaR}_{2,t}^{D,\beta}\right)\right) - \alpha\beta = 0, \tag{21}$$

$$F_{1,t}\left(\mathrm{CoVaR}_{1|2U,t}^{U,\alpha}\right) - C\left(F_{1,t}\left(\mathrm{CoVaR}_{1|2U,t}^{U,\alpha}\right), F_{2,t}\left(\mathrm{VaR}_{2,t}^{U,\beta}\right)\right) - \beta + \alpha\beta = 0, \tag{22}$$

respectively, where $F_{1,t}$ and $F_{2,t}$ are the marginal distribution of asset 1 and asset 2, respectively, $C(F_{1,t}, F_{2,t})$ is the joint distribution function of the best-fitted copula function. (We measure the *CoVaR* based on the best-fitted copula function mainly because that the more accurate copula cumulative distribution is helpful for obtaining the more accurate risk spillover characteristics.)

Equations (19)–(22) are useful for capturing the positive risk spillovers between two markets. However, because gold and USD are negatively correlated, the *CoVaR* measurements above may no longer applicable. In order to capture the possible negative risk spillovers between two markets, we refer to Ji et al. [50] and Ji et al. [51] and consider another two kinds of CoVaRs, $\mathrm{CoVaR}_{1|2U,t}^{D,\alpha}$ and $\mathrm{CoVaR}_{1|2D,t}^{U,\alpha}$. $\mathrm{CoVaR}_{1|2U,t}^{D,\alpha}$ is the downside risk of asset 1 conditional on the fact that the correlated asset 2 exhibits the extreme upward movement, and $\mathrm{CoVaR}_{1|2D,t}^{U,\alpha}$ is the upside risk of asset 1 conditional on the extreme downward movement of asset 2. These two CoVaRs can be expressed as

$$\begin{aligned} \Pr\left(r_{1,t} \le \mathrm{CoVaR}_{1|2U,t}^{D,\alpha} \mid r_{2,t} \ge \mathrm{VaR}_{2,t}^{U,\beta}\right) &= \alpha, \\ \Pr\left(r_{1,t} \ge \mathrm{CoVaR}_{1|2D,t}^{U,\alpha} \mid r_{2,t} \le \mathrm{VaR}_{2,t}^{D,\beta}\right) &= \alpha, \end{aligned} \tag{23}$$

respectively. Similarity, $\mathrm{CoVaR}_{1|2U,t}^{D,\alpha}$ and $\mathrm{CoVaR}_{1|2\ D,t}^{U,\alpha}$ can be rewritten as the following equations:

$$F_{1,t}\left(\mathrm{CoVaR}_{1|2U,t}^{D,\alpha}\right) - C\left(F_{1,t}\left(\mathrm{CoVaR}_{1|2U,t}^{D,\alpha}\right), F_{2,t}\left(\mathrm{VaR}_{2,t}^{U,\beta}\right)\right) - \alpha\beta = 0, \tag{24}$$

$$C\left(F_{1,t}\left(\mathrm{CoVaR}_{1|2\ D,t}^{U,\alpha}\right), F_{2,t}\left(\mathrm{VaR}_{2,t}^{D,\beta}\right)\right) - \beta + \alpha\beta = 0, \tag{25}$$

Following the two-step procedure of Reboredo and Ugolini [48], we can obtain the *CoVaR*s by measuring equations (21) and (22) and (24) and (25). If *CoVaR*s stronger than corresponding *VaR*s, the risk spillover exists. The significance of risk spillover can then be test by comparing the difference between the *CoVaR*s and *VaR* of asset 1, employing theeeee Kolmogorov–Smirnov (KS) bootstrapping test of Abadie [61].

Finally, to quantify the systemic risk contribution (risk spillover) of asset 2 to asset 1, four delta *CoVaR*s ($\Delta CoVaR$) are employed. $\Delta CoVaR$s can describe the difference between the risk of asset 1 conditional on the extreme movement of the correlated asset 2 and the risk of asset 1 conditional on the normal state of asset 2. Let $\Delta\mathrm{CoVaR}_{1|2\ D,t}^{D,\alpha}$ ($\Delta\mathrm{CoVaR}_{1|2U,t}^{U,\alpha}$) represents the growth rate of the downside (upside) risk of asset 1 when asset 2 is from normal state to extreme downward (upward) movement. Similarly, let $\Delta\mathrm{CoVaR}_{1|2U,t}^{D,\alpha}$ ($\Delta\mathrm{CoVaR}_{1|2\ D,t}^{U,\alpha}$) be the growth rate of the downside (upside) risk of asset 1 when asset 2 is from normal state to extreme upward (downward) movement. Positive $\Delta CoVaR$ indicates that there may be positive or negative risk spillover from asset 2 to asset 1. These four $\Delta CoVaR$s can be defined as the following four equations:

$$\begin{aligned} \Delta\mathrm{CoVaR}_{1|2\ D,t}^{D,\alpha} &= \frac{\left(\mathrm{CoVaR}_{1|2\ D,t}^{D,\alpha} - \mathrm{CoVaR}_{1|2M,t}^{D,\alpha}\right)}{\mathrm{CoVaR}_{1|2M,t}^{D,\alpha}}, \\ \Delta\mathrm{CoVaR}_{1|2U,t}^{U,\alpha} &= \frac{\left(\mathrm{CoVaR}_{1|2U,t}^{U,\alpha} - \mathrm{CoVaR}_{1|2M,t}^{U,\alpha}\right)}{\mathrm{CoVaR}_{1|2M,t}^{U,\alpha}}, \\ \Delta\mathrm{CoVaR}_{1|2U,t}^{D,\alpha} &= \frac{\left(\mathrm{CoVaR}_{1|2U,t}^{D,\alpha} - \mathrm{CoVaR}_{1|2M,t}^{D,\alpha}\right)}{CoVaR_{1|2M,t}^{D,\alpha}}, \\ \Delta\mathrm{CoVaR}_{1|2D,t}^{U,\alpha} &= \frac{\left(\mathrm{CoVaR}_{1|2\ D,t}^{U,\alpha} - \mathrm{CoVaR}_{1|2M,t}^{U,\alpha}\right)}{\mathrm{CoVaR}_{1|2M,t}^{U,\alpha}}, \end{aligned} \tag{26}$$

where $\mathrm{CoVaR}_{1|2M}^{D,\alpha}$ and $\mathrm{CoVaR}_{1|2M}^{U,\alpha}$ satisfies $\Pr(r_{1,t} \le \mathrm{CoVaR}_{1|2M,t}^{D,\alpha} | F_{2,t}(r_{2,t}) = 0.5) = \alpha$ and $\Pr(r_{1,t} \ge \mathrm{CoVaR}_{1|2M,t}^{U,\alpha} \mid F_{2,t}(r_{2,t}) = 0.5) = \alpha$, respectively.

3. Data and Summary Statistics

The data we utilized in this article include the daily prices for Bitcoin, gold, and USD index from 1 May, 2013 to 31 July, 2019. Bitcoin price is downloaded from https://coinmarketcap.com/. The gold prices and USD index are

collected from Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/). The price and return dynamics of daily Bitcoin, gold and USD indices are drawn in Figures 1 and 2, respectively.

Table 1 presents the summary statistics of Bitcoin, gold, and USD returns. We find in Table 1 that the mean and median of gold return series are negative, whereas the mean and median of the return of other two assets are positive, and Bitcoin returns have the largest mean and median. Besides, the standard deviation of Bitcoin return is also bigger than other two series. Positive skewness statistics indicate that the distribution of our three asset return series is right skewed. The kurtosis statistics for all series are greater than 3, indicating that their distribution has the characteristic of peak. We can also find that all the return series are not normally distributed because their Jarque–Bera statistics are significant at 1% levels. ADF statistics proved the stationarity of all series. The statistics of Ljung–Box test for returns and squared returns indicate the presence of autocorrelation in gold and Bitcoin returns. Specially, although the squared return of USD is autocorrelative, the return of it is weak autocorrelative. The LM tests for ARCH effect indicate the significant conditional heteroscedasticity effect for all return series.

4. Tests for Risk Spillover Effect

4.1. Estimation Results for Marginal Models. The estimation results for the parameters of three marginal models (AR (1)-GARCH (1, 1)-skew-t) are presented in Table 2. Table 2 shows firstly that the AR coefficient (ϕ_1) of our marginal distribution models are not significant for three asset return series. Besides, α_1 and β_1 are significant for all marginal distribution models, implying obvious ARCH effect and volatility clustering in three return series. Furthermore, the tail parameters of three models are significantly positive, meaning that the residual series of these models are fat tailed. The asymmetry parameter for USD model is significantly positive, indicating that the standardized residual of this model is right skewed. However, the asymmetry parameters for gold and Bitcoin returns are not significant. The p values of the KS test are larger than 0.1, indicating that the skew-t distribution is a good description for three marginal distributions, and then, the dependence structure between Bitcoin, gold, and USD can be measured by the copula functions in the next step.

4.2. Estimation Results for Copula Functions. In this subsection, we estimate the dependence structure of each assets pair using various static and time-varying copula functions. The best-fitted copula function is selected by the AIC criterion, and the AIC for each copula function is presented in Table 3. Table 3 shows that time-varying copula functions fit better than static ones. The dependence structure between Bitcoin and gold and between gold and USD can be better described by the Student-t copula function with two dynamic parameters, ρ_t and υ_t (i.e., "TVP-Student 2" in Table 3), indicating the symmetric tail dependence of gold and other two assets. However, TVP-Gaussian copula function is the best-fitted one for describing the dependence structure between Bitcoin and USD, indicating an average dependence of these two assets.

Figure 3 shows the evolutions of the dependence of three asset pairs and presents the mean of each dynamic dependence parameter. The dependence structures draw in Figure 3 comes from the best-fitted copula functions. The dynamic dependence parameter between Bitcoin and gold drawn in the top panel of Figure 3 shows that the dependence between these two assets fluctuates around a small positive mean at 0.032. The middle panel of Figure 3 shows that the dependence between Bitcoin and USD has a mean that is very close to 0. The dependence between gold and USD, however, fluctuates around a strong negative mean, at -0.295, as shown in the bottom panel of Figure 3. In summary, gold and USD show the strongest dependency, followed by gold and Bitcoin, whereas there is weakest dependency between Bitcoin and USD.

4.3. Tests for Risk Spillover Effect. To explore the risk spillover effect between gold, Bitcoin, and USD, we measure the *VaR*s and *CoVaR*s for each asset. The evolution of *VaR*s and *CoVaR*s are drawn in Figure 4. Panel A of Figure 4 show the *VaR*s and *CoVaR*s of Bitcoin-gold pair, and the left one draws the *CoVaR*s of Bitcoin conditional on gold, the right one shows the *CoVaR*s of gold conditional on Bitcoin. Panel B draws the *VaR*s and *CoVaR*s of Bitcoin-USD pair, and the left one and the right one shows the *CoVaR*s of Bitcoin conditional on USD and USD conditional on Bitcoin, respectively. Panel C presents the *VaR*s and *CoVaR*s of gold-USD pair, and the right one draws the *CoVaR*s of gold conditional on USD, the left draws the *CoVaR*s of USD conditional on gold. We can find in Figure 4 that the risk spillover between Gold and USD are more pronounced; however, the negative risk spillovers between other asset pairs are still need to be verified mathematically.

Columns 2–4 and columns 7–9 of Table 4 list the mean and standard deviations of *VaR*s and *CoVaR*s of each asset pairs. We can find in Table 4 that both the mean of $\mathrm{CoVaR}^{D,\alpha}_{1|2\ D}$ and $\mathrm{CoVaR}^{D,\alpha}_{1|2U}$ of Bitcoin conditional on gold are stronger than the downside *VaR* of Bitcoin, and the mean of $\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ and $\mathrm{CoVaR}^{U,\alpha}_{1|2\ D}$ are also stronger than the upside *VaR*. Similarly, $\mathrm{CoVaR}^{D,\alpha}_{1|2\ D}$ and $\mathrm{CoVaR}^{D,\alpha}_{1|2U}$ of gold conditional on Bitcoin are stronger than the downside *VaR* of gold, and the $\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ and $\mathrm{CoVaR}^{U,\alpha}_{1|2\ D}$ are also stronger than the upside *VaR*. These suggest that there may be unstable risk spillovers between Bitcoin and gold. As for the Bitcoin-USD pairs, we can only find a weak increase of $\mathrm{CoVaR}^{D,\alpha}_{1|2\ D}$ and $\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ of Bitcoin with respect to *VaR*s of Bitcoin, meaning a weak positive risk spillover effect between Bitcoin and USD. For the gold-USD pairs, we found that $\mathrm{CoVaR}^{D,\alpha}_{1|2\ D}$ and $\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ for each of these two asset are lower than the corresponding *VaR*s, while $\mathrm{CoVaR}^{D,\alpha}_{1|2U}$ and $\mathrm{CoVaR}^{U,\alpha}_{1|2\ D}$ of them are stronger than the corresponding *VaR*s, indicating a negative risk spillover effect between gold and USD.

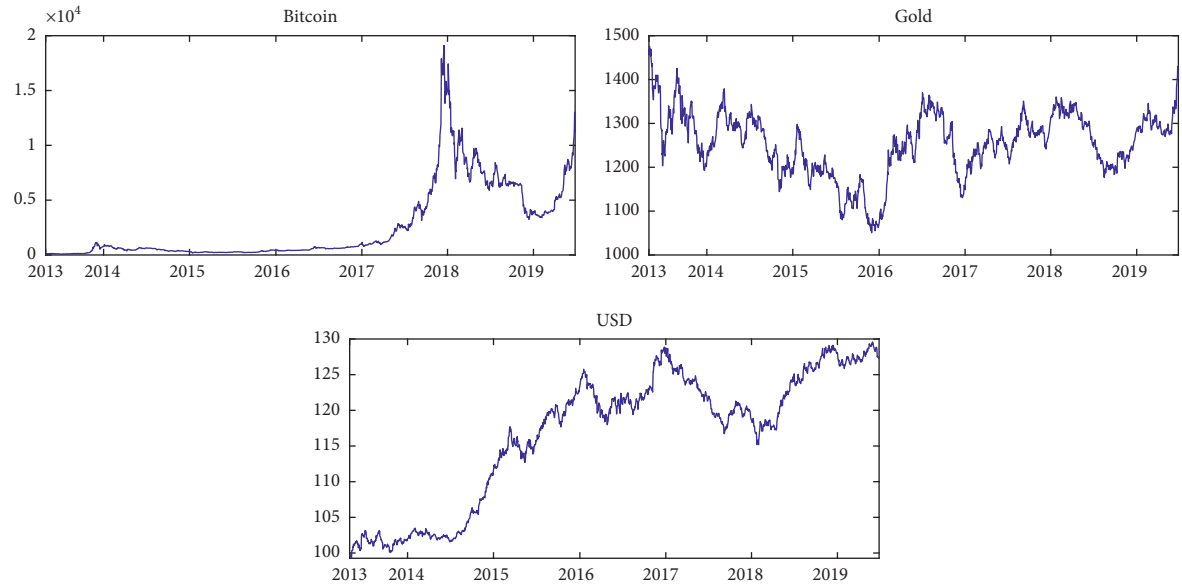


Figure 1: Time variations of daily prices of (a) Bitcoin, (b) gold, and (c) USD indices.

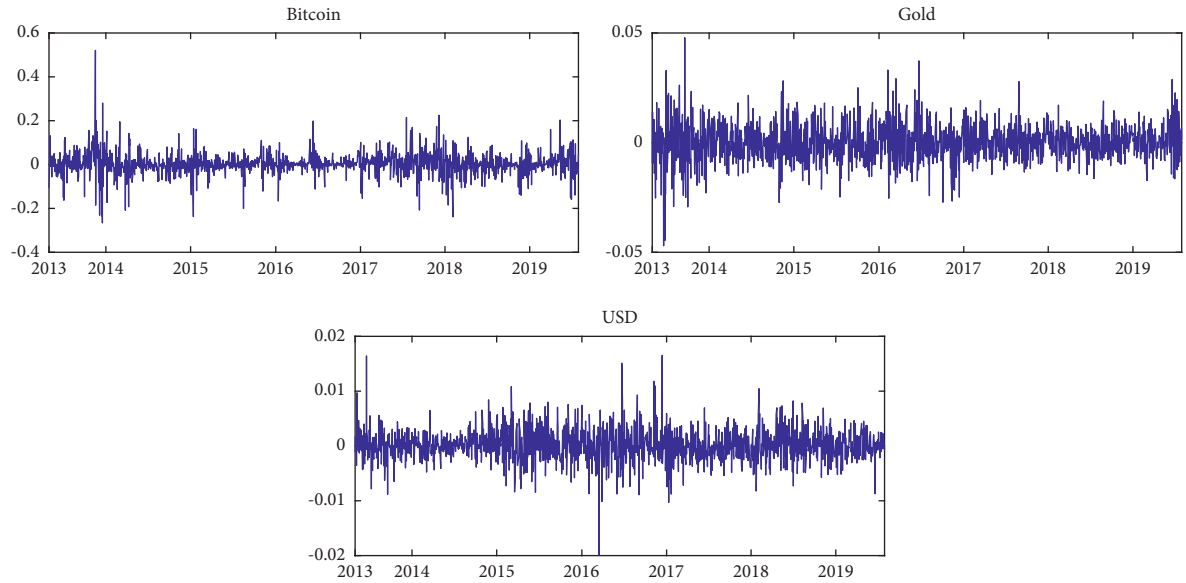


Figure 2: Daily returns of (a) Bitcoin, (b) gold, and (c) USD.

To further confirm the existence of risk spillover effect between gold, Bitcoin, and USD, Kolmogorov–Smirnov (KS) test of Abadie [61] is utilized. The KS statistic can be defined as

$$\mathrm{KS}_{mn} = \left(\frac{mn}{m+n}\right)^{1/2} \sup_{x} \left|F_m(x) - F_n(x)\right|, \tag{27}$$

where m and n are the size of two samples, respectively. $F_m(x)$ and $F_n(x)$ are the corresponding cumulative distribution functions of these two samples. The KS test in this subsection test the null hypothesis that the $CoVaR$ of asset 1 conditional on asset 2 is no stronger than the VaRs of asset 1. If the null hypothesis is rejected, the risk spillover from asset 2 to asset 1 is significantly exist. The results of KS test are listed in Table 5. The p values that are lower than 0.1 are marked with bold. As shown in Table 5, the null hypothesis of no positive and negative system risk spillover between Bitcoin and gold market are rejected at 10% significant level, indicating the unstable risk spillover effect between Bitcoin and gold market. Then, the hypothesis of no positive and negative risk spillover between Bitcoin and USD cannot be rejected, suggesting that there are no significant risk spillover effects between these two assets. Finally, the results of KS test cannot reject the null hypothesis of no positive system risk spillover between gold and USD market, whereas the null hypothesis of no negative system risk spillover between these two assets are

Table 1: Summary statistics and unconditional correlations.

	Bitcoin	Gold	USD
Obs	1579	1579	1579
Mean (%)	0.282	−0.002	0.016
Max	0.521	0.048	0.017
Min	−0.266	−0.047	−0.020
Median (%)	0.212	−0.013	0.000
Std. dev	0.051	0.008	0.003
Skewness	0.408	0.033	0.159
Kurtosis	13.7053	5.453	6.717
J-B	7583.263***	396.093***	915.751***
ADF	−41.240***	−38.952***	−39.084***
Q (5)	18.502***	8.342	2.801
Q (10)	24.306***	26.263***	9.752
Q^2 (5)	103.103***	92.277***	15.979***
ARCH (5)	87.66***	76.376***	13.650***

Note: this table presents the summary statistics of gold, Bitcoin and USD returns and lists the unconditional correlations between gold and Bitcoin and USD returns. Obs., Mean, Max., Min, Median, and Std. dev. describe the observations, mean value, maximum, minimum, median, and standard deviation of each return series, respectively. J-B is the Jarque–Bera statistics [62] for normality test. ADF summarizes the statistic of Augmented Dickey–Fuller [63] unit root test. Q (5) and Q (10) denote the Ljung–Box statistics [64] for autocorrelation test of the return series for up to 5^{th} and 10^{th} lags, respectively. Q^2 (5) is the Ljung–Box statistics for autocorrelation test of the squared return series. ARCH (5) refers to the statistics of Engle's [65] LM test for autoregressive conditional heteroscedasticity (ARCH) effect. The symbols ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 2: Parameters estimates of marginal models.

	Bitcoin	Gold	USD
ϕ_0	0.002 (9.029e−04)	−1.933e−05 (1.929e−04)	1.540e−04 (6.504e−05)
ϕ_1	−0.022 (0.040)	0.036 (0.025)	0.003 (0.024)
ω	3.124e−05 (1.950e−05)	6.013e−07 (3.148e−07)	3.049e−08 (2.173e−08)
α_1	0.144*** (0.020)	0.036*** (0.011)	0.031*** (0.007)
β_1	0.856*** (0.029)	0.955*** (0.013)	0.966*** (0.007)
Asymmetry	−0.027 (0.040)	0.042 (0.057)	0.064* (0.032)
Tail	3.054*** (0.134)	9.349*** (2.026)	5.794*** (0.855)
KS test	[0.107]	[0.955]	[0.333]

Note: this table shows the parameters estimates and standard errors (in parentheses) for the marginal models with skewed-t distribution. KS test reports the p value (in square brackets) of the Kolmogorov–Smirnov test, representing the adequacy of the skewed-t distribution. The null hypothesis of the Kolmogorov–Smirnov test cannot be rejected when p values is greater than 0.1. The symbols ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

rejected at 10% significant level, confirming the stable negative risk spillover between USD and gold market.

Then, we calculate the Δ*CoVaRs* for each asset pairs to quantify the risk contribution (risk spillover) of each asset to another. The dynamics of Δ*CoVaR*s are drawn in Figure 5, and the mean and standard deviation of Δ*CoVaR*s are shown in column 5–6 and column 10–11 of Table 4. Figure 5 and Table 4 show that $\Delta\text{CoVaR}^{D,\alpha}_{1|2\ D}$, $\Delta\text{CoVaR}^{U,\alpha}_{1|2U}$, $\Delta\text{CoVaR}^{D,\alpha}_{1|2U}$, and $\Delta\text{CoVaR}^{U,\alpha}_{1|2\ D}$ from gold to Bitcoin and from Bitcoin to gold fluctuate around positive means, indicating the possible unstable risk spillover between Bitcoin and gold. $\Delta\text{CoVaR}^{D,\alpha}_{1|2\ D}$ and $\Delta\text{CoVaR}^{U,\alpha}_{1|2U}$ from gold to Bitcoin as well as from Bitcoin to gold have stronger mean than $\Delta\text{CoVaR}^{D,\alpha}_{1|2U}$ and $\Delta\text{CoVaR}^{U,\alpha}_{1|2\ D}$, meaning the dominant of positive risk spillover effect. The Δ*CoVaR*s of Bitcoin-USD pair, however, fluctuates around the values, which are very close to zero. For the gold-USD pairs, $\Delta\text{CoVaR}^{D,\alpha}_{1|2U}$ and $\Delta\text{CoVaR}^{U,\alpha}_{1|2\ D}$ from USD to gold as well as from gold to USD fluctuate around the higher positive means, whereas $\Delta\text{CoVaR}^{D,\alpha}_{1|2\ D}$ and $\Delta\text{CoVaR}^{U,\alpha}_{1|2U}$ fluctuate around smaller negative means, indicating a major negative risk spillover between gold and USD. These findings further confirm the results of KS test.

4.4. Tests for Asymmetric Downside and Upside Risk Spillover Effect. Table 4 also shows that downside Δ*CoVaR*s from gold to Bitcoin are larger than upside Δ*CoVaR*s, whereas

Table 3: Results of copula function.

	Bitcoin and gold	Bitcoin and USD	Gold and USD
Gaussian	−1.627	0.001	−138.552
Clayton	−0.536	0.003	0.068
Rotated Clayton	−4.044	−0.010	0.061
Gumbel	−3.287	0.010	0.078
Rotated Gumbel	−1.434	0.007	0.126
Student	−4.288	0.766	−165.082
SJC	−3.408	3.557	37.320
TVP-Gaussian	−2.013	**−2.416**	−145.130
TVP-Clayton	−1.248	−1.396	0.071
TVP-rotated Clayton	−4.108	−0.713	0.063
TVP-Gumbel	−3.466	−0.175	−0.093
TVP-rotated Gumbel	−1.946	−0.868	0.128
TVP-student	−4.623	−1.680	−170.553
TVP-student 2	**−5.211**	−1.858	**−174.458**
TVP-SJC	−0.118	10.593	69.498

Note: this table reports the AIC values of different copula functions. The AICs of the best-fitted copula functions (i.e., the minimum AICs) are bold and underlined.

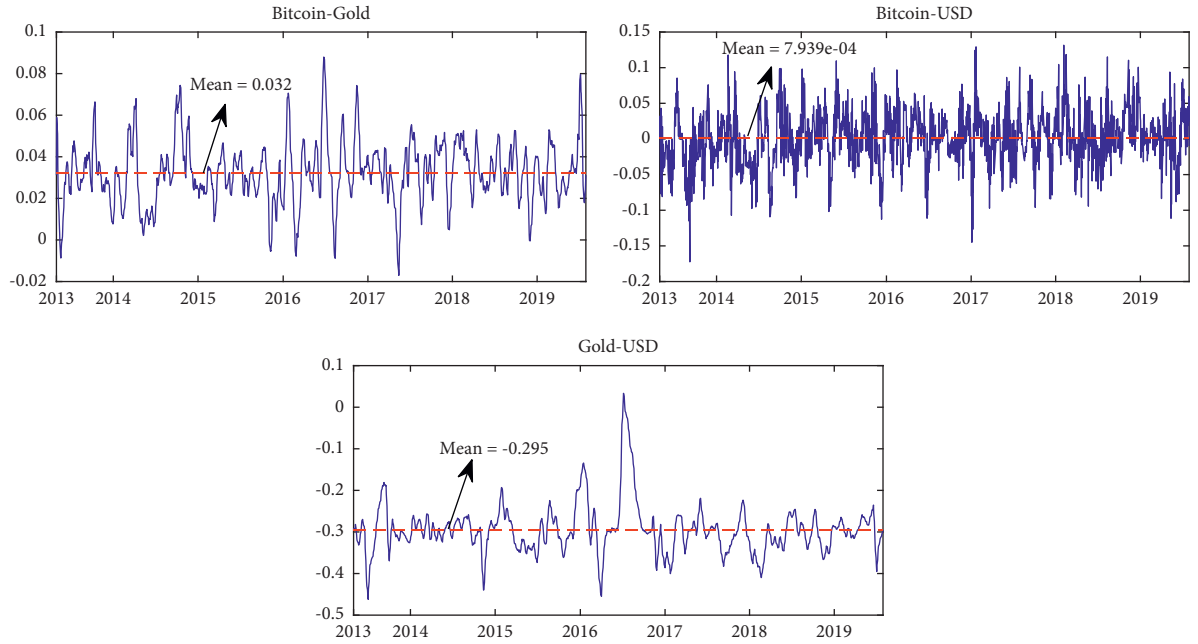


Figure 3: Time evolutions of the dependence of Bitcoin-gold, Bitcoin-USD, and gold-USD pairs. Note: the time-varying dependence structure is based on the best-fitted bivariate copula.

upside Δ*CoVaR*s from Bitcoin to gold are larger than downside Δ*CoVaR*s. Besides, downside Δ*CoVaR*s between gold and USD are smaller than upside Δ*CoVaR*s. These summary statistics indicates the possible existence of asymmetric downside and upside risk spillover between Bitcoin and gold markets and between gold and USD markets. Table 6 presents the results of KS test, which test the null hypothesis of equal upside and downside risk spillovers. Because the positive risk spillover between gold and Bitcoin dominates, and there is no positive risk spillover between gold and USD, we only explore the upside and downside asymmetry of the positive risk spillover between gold and Bitcoin, and the negative risk spillover between gold and USD. As shown in column 2, the null hypothesis of symmetric positive upside and downside risk spillover between gold and Bitcoin can be rejected at 1% significant level. The results in column 3 also reject the null hypothesis of equal negative upside and downside risk spillover between gold and USD. The stronger downside risk spillover from gold to Bitcoin is due to the facts that the standardized residual of Bitcoin is left skewed. Similarly, because the standardized residuals of gold and USD are right skewed, the upside risk spillovers from Bitcoin to gold and between gold and USD are stronger.

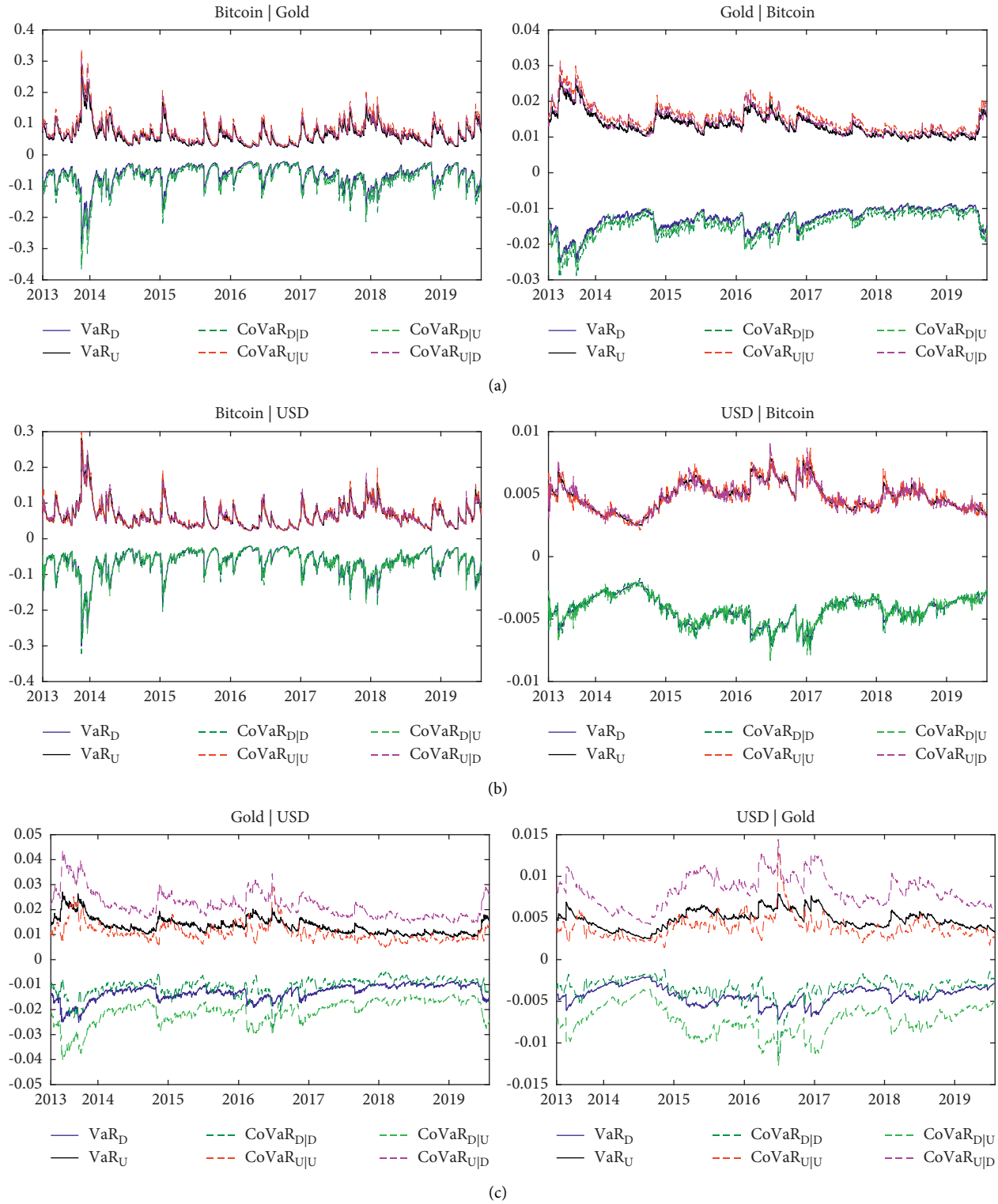


Figure 4: Downside and upside *VaR*s and *CoVaR*s. Note: Panel A shows the downside and upside *VaR*s and *CoVaR*s of Bitcoin-gold pair. Panel B shows the downside and upside *VaR*s and *CoVaR*s of Bitcoin-USD pair. Panel C shows the downside and upside *VaR*s and *CoVaR*s of gold-USD pair. *D* and *U* denote the downside and upside risk, respectively. $\text{CoVaR}_{D|D}$ and $\text{CoVaR}_{D|U}$ describe the downside risk of one market when another market is extreme downward and upward movement, respectively. $\text{CoVaR}_{U|U}$ and $\text{CoVaR}_{U|D}$ means the upside risk of one market when another market facing the extreme upward and downward risk, respectively. (a) Panel A *VaR*s and *CoVaR*s of Bitcoin-gold pair. (b) Panel B *VaR*s and *CoVaR*s of Bitcoin-USD pair. (c) Panel C *VaR*s and *CoVaR*s of gold-USD pair.

Table 4: Summary statistics for *VaRs*, *CoVaRs*, and Δ*CoVaRs*.

Symbol	Downside					Upside				
1|2	$\mathrm{VaR}_1^{D,\alpha}$	$\mathrm{CoVaR}_{1\|2D}^{D,\alpha}$	$\mathrm{CoVaR}_{1\|2U}^{D,\alpha}$	$\Delta\mathrm{CoVaR}_{1\|2D}^{D,\alpha}$	$\Delta\mathrm{CoVaR}_{1\|2U}^{D,\alpha}$	$\mathrm{VaR}_1^{U,\alpha}$	$\mathrm{CoVaR}_{1\|2U}^{U,\alpha}$	$\mathrm{CoVaR}_{1\|2D}^{U,\alpha}$	$\Delta\mathrm{CoVaR}_{1\|2U}^{U,\alpha}$	$\Delta\mathrm{CoVaR}_{1\|2D}^{U,\alpha}$
Bitcoin|gold	−0.064	−0.079	−0.070	**0.198**	**0.111**	0.065	0.079	0.070	**0.181**	**0.101**
	(0.034)	(0.042)	(0.038)	**(0.043)**	**(0.080)**	(0.033)	(0.041)	(0.037)	**(0.040)**	**(0.073)**
Gold|	−0.013	−0.015	−0.014	**0.127**	**0.073**	0.013	0.016	0.014	**0.134**	**0.077**
Bitcoin	(0.003)	(0.004)	(0.003)	**(0.027)**	**(0.051)**	(0.003)	(0.004)	(0.004)	**(0.028)**	**(0.054)**
Bitcoin|USD	−0.064	−0.065	−0.064	0.001	−0.001	0.065	0.066	0065	0.001	−0.000
	(0.034)	(0.036)	(0.034)	(0.050)	(0.050)	(0.033)	(0.034)	(0.033)	(0.046)	(0.046)
USD|Bitcoin	−0.004	−0.004	−0.004	0.001	−0.001	0.005	0.005	0.005	0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.038)	(0.038)	(0.001)	(0.001)	(0.001)	(0.038)	(0.038)
Gold|USD	−0.013	−0.011	−0.021	−0.021	**0.415**	0.013	0.011	0.022	−0.022	**0.436**
	(0.003)	(0.003)	(0.005)	(0.149)	**(0.035)**	(0.003)	(0.004)	(0.005)	(0.157)	**(0.036)**
USD|gold	−0.004	−0.003	−0.007	−0.020	**0.503**	0.005	0.004	0.008	−0.020	**0.508**
	(0.001)	(0.001)	(0.002)	(0.163)	**(0.044)**	(0.001)	(0.001)	(0.002)	(0.164)	**(0.043)**

Note: this table shows the means and the standard deviations (in parenthesis) of *VaR*s, *CoVaR*s and Δ*CoVaR*s of Bitcoin, gold, and USD. $CoVaR_{1|2D}^{D,\alpha}$ and $CoVaR_{1|2U}^{D,\alpha}$ describe the downside risk of market 1 when market 2 is extreme downward and upward movement, respectively. $CoVaR_{1|2U}^{U,\alpha}$ and $CoVaR_{1|2D}^{U,\alpha}$ are the upside risk of market 1 when market 2 facing the extreme upward and downward risk, respectively. $\Delta CoVaR_{1|2D}^{D,\alpha}$ and $\Delta CoVaR_{1|2U}^{D,\alpha}$ are the risk contribution of the downward and upward movement of market 2 to the downward movement of market 1, respectively. $\Delta CoVaR_{1|2U}^{U,\alpha}$ and $\Delta CoVaR_{1|2\,D}^{U,\alpha}$ are the risk contribution of the upward and downward movement of market 2 to the upward movement of market 1, respectively.

5. Short-Term and Long-Term Dependence and Risk Spillover

In this section, we check the robustness in the main findings of Section 4 by exploring the short- and long-term dependence and risk spillover between the three assets through a variational mode decomposition (VMD) approach. Because our analysis above found no significant risk spillover effect between Bitcoin and USD, we just examine the short- and long-term dependence structure and risk spillover effect between Bitcoin and gold and between gold and USD.

5.1. Results for Short-Term and Long-Term Dependence. Before exploring the dependence and risk spillover of Bitcoin-gold and Bitcoin-USD pairs in short-term and long-term horizons, we utilized the variational mode decomposition (VMD) approach to distinguish the short-term and long-term dynamics of Bitcoin, gold, and USD return series. (The advantage of obtaining short-term and long-term sequences by VMD method is that this method can decompose the short-term and long-term variation trends implicit in the original sequence without loss of sample length.) The VMD method can be mathematically described as a solution of the constrained variational problem [66]:

$$\begin{aligned}\min_{\{u_k\},\{\omega_k\}} &= \left\{\sum_k \left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\},\\ \text{s.t.}\sum_k u_k &= f,\end{aligned} \tag{28}$$

where k is the total number of modes; u_k is the k th mode; ω_k is the center frequency of the k th mode; ∂_t, $\delta(t)$, and $*$ are the partial derivatives, the Dirac distribution and the convolution, respectively; f is the original series. Then, the constraint in equation (28) is addressed by an augmented Lagrangian function:

$$\begin{aligned}\mathscr{L}(\{u_k\},\{\omega_k\},\lambda) = \alpha\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]\right\|_2^2\\ +\left\|f-\sum u_k\right\|_2^2+\langle\lambda, f-\sum u_k\rangle,\end{aligned} \tag{29}$$

where λ is the Lagrange multiplier, α is the balancing parameter of the data-fidelity constraint, and $\|\bullet\|_p$ represents the usual vector ℓ_p norm where $p=2$. The solutions for u and ω can be found in Fourier domain and updated by

$$u_n^{n+1} = \frac{\left(f-\sum_{i\neq k}u_i+(\lambda/2)\right)}{1+2\alpha\left(\omega-\omega_k\right)^2}, \tag{30}$$

$$\omega_n^{n+1} = \frac{\int_0^\infty \omega\left|u_k(\omega)\right|^2 \mathrm{d}_\omega}{\int_0^\infty \left|u_k(\omega)\right|^2 \mathrm{d}_\omega}, \tag{31}$$

respectively, where n is the number of iterations, which is 500 for default. λ in equations (29) and (30) is updated by

$$\widehat{\lambda}^{n+1}\leftarrow\widehat{\lambda}^n+\tau\left(\widehat{f}-\sum_k\widehat{u}_k^{n+1}\right), \tag{32}$$

until convergence: $\sum_k\|\widehat{u}_k^{n+1}-\widehat{u}_k^n\|_2^2/\|\widehat{u}_k^n\|_2^2$.

Following Lahmiri [67], Shahzad et al. [68], and Mensi et al. [46], the standardized residuals, z_t, obtained from the marginal models in Table 2 are decomposed into 10 subseries (i.e., $k=10$), which are compressed around ten different center frequencies. (The reason for this article to decompose the standardized residuals but not the raw return series is that the standardized residual series is an independent and identically distributed series with weak typical features like autocorrelation and heteroscedasticity. By decomposing the standardized residuals, the short-term and long-term

Table 5: KS tests of difference between *CoVaRs* and *VaRs*.

Symbol	Downside		Upside	
1|2	H_0: $\mathrm{CoVaR}^{D}_{1\|2D} \geq \mathrm{VaR}^{D,\alpha}_{1}$ H_1: $\mathrm{CoVaR}^{D}_{1\|2D} < \mathrm{VaR}^{D,\alpha}_{1}$	H_0: $\mathrm{CoVaR}^{D,\alpha}_{1\|2U} \geq \mathrm{VaR}^{D,\alpha}_{1}$ H_0: $\mathrm{CoVaR}^{D,\alpha}_{1\|2U} < \mathrm{VaR}^{D,\alpha}_{1}$	H_0: $\mathrm{CoVaR}^{U,\alpha}_{1\|2U} \leq \mathrm{VaR}^{U,\alpha}_{1}$ H_0: $\mathrm{CoVaR}^{U,\alpha}_{1\|2U} > \mathrm{VaR}^{U,\alpha}_{1}$	H_0: $\mathrm{CoVaR}^{U,\alpha}_{1\|2D} \leq \mathrm{VaR}^{U,\alpha}_{1}$ H_0: $\mathrm{CoVaR}^{U,\alpha}_{1\|2D} > \mathrm{VaR}^{U,\alpha}_{1}$
Bitcoin|gold	**0.185** **[0.000]**	**0.074** **[0.000]**	**0.182** **[0.000]**	**0.070** **[0.000]**
Gold|Bitcoin	**0.247** **[0.000]**	**0.113** **[0.000]**	**0.250** **[0.000]**	**0.121** **[0.000]**
Bitcoin|USD	0.015 [0.700]	0.022 [0.437]	0.018 [0.592]	0.022 [0.451]
USD|Bitcoin	0.024 [0.385]	0.028 [0.291]	0.026 [0.355]	0.027 [0.310]
Gold|USD	0.006 [0.953]	**0.735** **[0.000]**	0.006 [0.952]	**0.756** **[0.000]**
USD|gold	0.016 [0.641]	**0.740** **[0.000]**	0.017 [0.646]	**0.774** **[0.000]**

Note: this table lists the statistic and the p value (in squared brackets) of the Kolmogorov–Smirnov (KS) tests, which verified the null hypothesis of no systemic risk spillover effect between Bitcoin, gold, and USD.

dependencies between two time series captured by copula functions would be less affected by their typical features.) Then, the ten modes are marked as VMD1 to VMD10 in the order of center frequency from small to large, in which the mode with the lowest center frequency (VMD1) is chosen to be the long-term series and the mode with the highest center frequency (VMD10) is used as the short-term series. The long-term dynamics (VMD1) and short-term dynamics (VMD10) for Bitcoin, gold, and USD are drawn in Figure 6.

Then, the short- and long-term dynamics of standardized residual series are employed to estimate the copula functions. The AIC of the copula functions based on short-term and long-term dynamics are listed in Table 7. Table 7 shows that the time-varying Gaussian copula function is the best-fitted one for describing the short-term and long-term dependence structure between gold and Bitcoin. The long-term dependence structure between gold and USD can be well described by the time-varying Gaussian copula function, whereas the short-term dependence between them can be well described by the Student-t copula with two time-varying parameters. Then, the dynamics of the best-fitted short-term and long-term dependencies are drawn in Panel A and Panel B of Figure 7, respectively. For the short-term dependence parameters drawn in Panel A of Figure 7, the dependence between gold and other two assets both fluctuate around a small average value, and the mean of the short-term dependence parameters between Bitcoin and gold and between gold and USD are 0.029 and -0.052, respectively. For the long-term dependence parameters shown in Panel B of Figure 7, the dependence between Bitcoin and gold and between gold and USD fluctuate around 0.046 and -0.397, respectively.

Overall, the short-term and long-term dependence between gold and USD are still stronger than those between Bitcoin and gold. Besides, the means of long-term dependence are stronger than those of short-term dependence, it reminds us that long-term investors should pay more attention to the dependency between gold and other two assets than short-term investors.

5.2. Tests for Short-Term and Long-Term Spillover Effect. To find the short-term and long-term risk spillover effect between Bitcoin and gold and between gold and USD, we measure *CoVaR*s based on short-term and long-term dependence structure and test the null hypothesis of no risk spillover between Bitcoin and gold and between gold and USD. The means and standard deviations are listed in columns 2–3 and columns 6–7 of Table 8, and the results of KS test, which test the hypothesis of no short-term and long-term risk spillover effects, are listed in Table 9. Table 9 firstly reject the null hypothesis of no positive and negative short-term and long-term system risk spillover effect between Bitcoin and gold market, further indicating the instability of risk spillover between Bitcoin and gold. Besides, both the null hypothesis of no positive and negative short-term risk spillover between gold and USD are reject at 10% significant level, suggesting that the short-term risk spillover effect between gold and USD is instable. The null hypothesis of no positive long-term risk spillovers between gold and USD cannot be rejected, whereas the null hypothesis of no negative long-term risk spillovers is rejected at 1% significant level, indicating the stable long-term negative risk spillover between gold and USD.

Then, we measure $\Delta CoVaR$s based on short-term and long-term dependence structure to quantify the short-term and long-term risk spillover effect between Bitcoin and gold and between gold and USD. The dynamics of short-term and long-term $\Delta CoVaR$s are drawn in Figure 8. (We only draw $\Delta\mathrm{CoVaR}^{D,\alpha}_{1|2D}$ and $\Delta\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ for Bitcoin-gold pairs and $\Delta\mathrm{CoVaR}^{D,\alpha}_{1|2U}$ and $\Delta\mathrm{CoVaR}^{U,\alpha}_{1|2D}$ for gold-USD pairs.) The mean of them are list in columns 4–5 and columns 8–9 of Table 8. We can find that both short-term and long-term $\Delta\mathrm{CoVaR}^{D,\alpha}_{1|2D}$ and $\Delta\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ from gold to Bitcoin have positive mean, whereas the mean of $\Delta\mathrm{CoVaR}^{D,\alpha}_{1|2U}$ and $\Delta\mathrm{CoVaR}^{U,\alpha}_{1|2D}$ of them are negative, further suggesting the dominant of positive risk spillover from gold to Bitcoin. For the $\Delta CoVaR$s from Bitcoin to gold, only the mean of short-

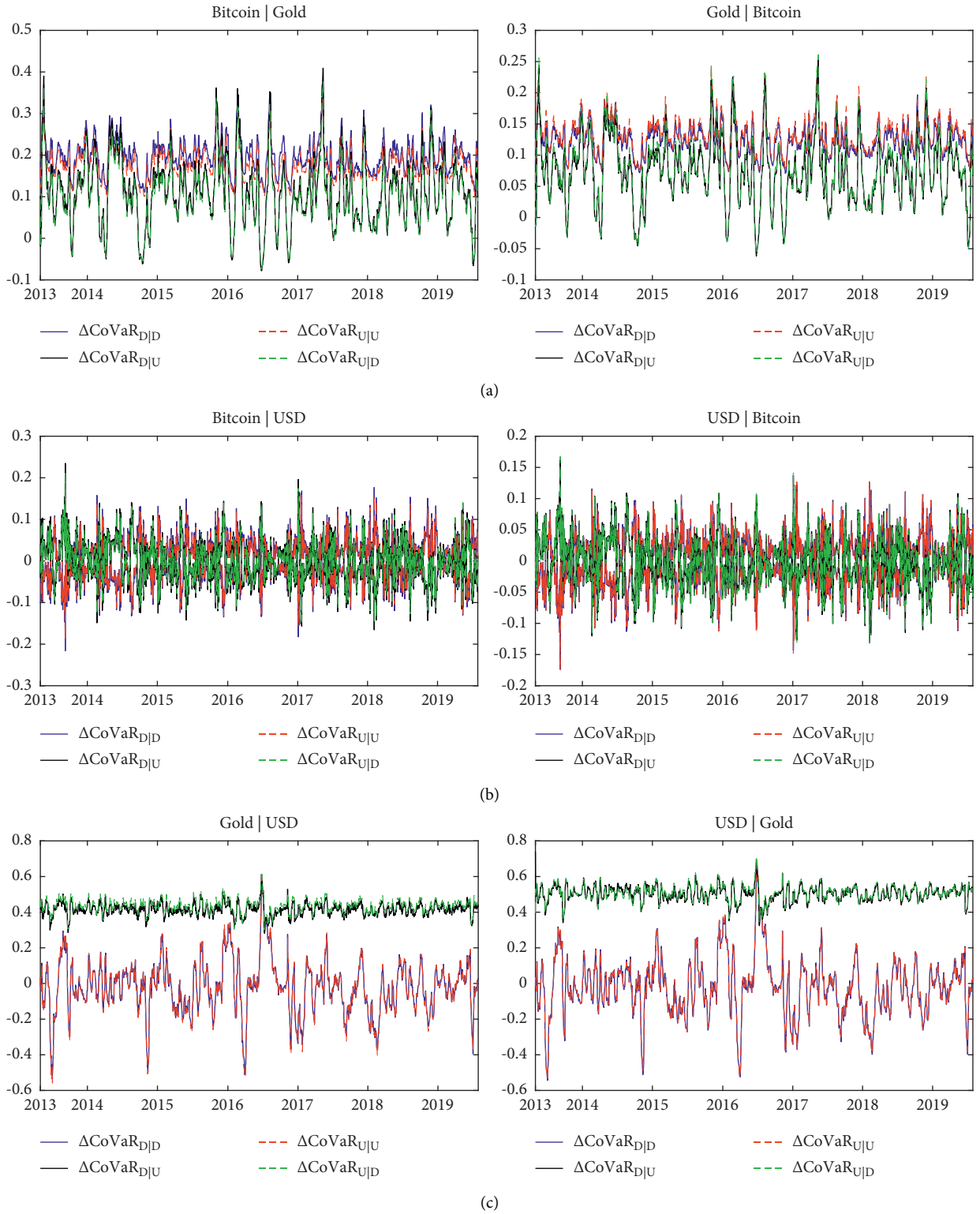


Figure 5: Downside and upside *ΔCoVaRs*. Note: Panel A shows the downside and upside Δ*CoVaRs* of Bitcoin-gold pair. Panel B shows the downside and upside Δ*CoVaRs* of Bitcoin-USD pair. Panel C shows the downside and upside Δ*CoVaRs* of gold-USD pair. $\Delta\text{CoVaR}^{D,\alpha}_{1|2D}$ and $\Delta\text{CoVaR}^{D,\alpha}_{1|2U}$ are the risk contribution of the downward and upward movement of market 2 to the downward movement of market 1, respectively. $\Delta\text{CoVaR}^{U,\alpha}_{1|2U}$ and $\Delta\text{CoVaR}^{U,\alpha}_{1|2D}$ are the risk contribution of the upward and downward movement of market 2 to the upward movement of market 1, respectively. (a) Panel A *ΔCoVaRs* of Bitcoin-gold pair. (b) Panel B Δ*CoVaR*s of Bitcoin-USD pair. (c) Panel C Δ*CoVaR*s of gold-USD pair.

Table 6: The KS tests for equalities of the downside and upside risk spillover effect.

Symbol 1|2	H_0: $\Delta\text{CoVaR}^{D,\alpha}_{1\|2\,D} = \Delta\text{CoVaR}^{U,\alpha}_{1\|2\,U}$ H_1: $\Delta\text{CoVaR}^{D,\alpha}_{1\|2\,D} \neq \Delta\text{CoVaR}^{U,\alpha}_{1\|2\,U}$	H_0: $\Delta\text{CoVaR}^{D,\alpha}_{1\|2\,U} = \Delta\text{CoVaR}^{U,\alpha}_{1\|2\,D}$ H_1: $\Delta\text{CoVaR}^{U,\alpha}_{1\|2\,U} \neq \Delta\text{CoVaR}^{U,\alpha}_{1\|2\,D}$
Bitcoin | gold	**0.198** **[0.000]**	
Gold | Bitcoin	**0.115** **[0.000]**	
Gold | USD		**0.262** **[0.000]**
USD | gold		**0.047** **[0.059]**

Note: this table summary the statistic and the p value (in squared brackets) of Kolmogorov–Smirnov (KS) test, testing the null hypothesis of equalities of the downside and upside $\Delta CoVaR$s.

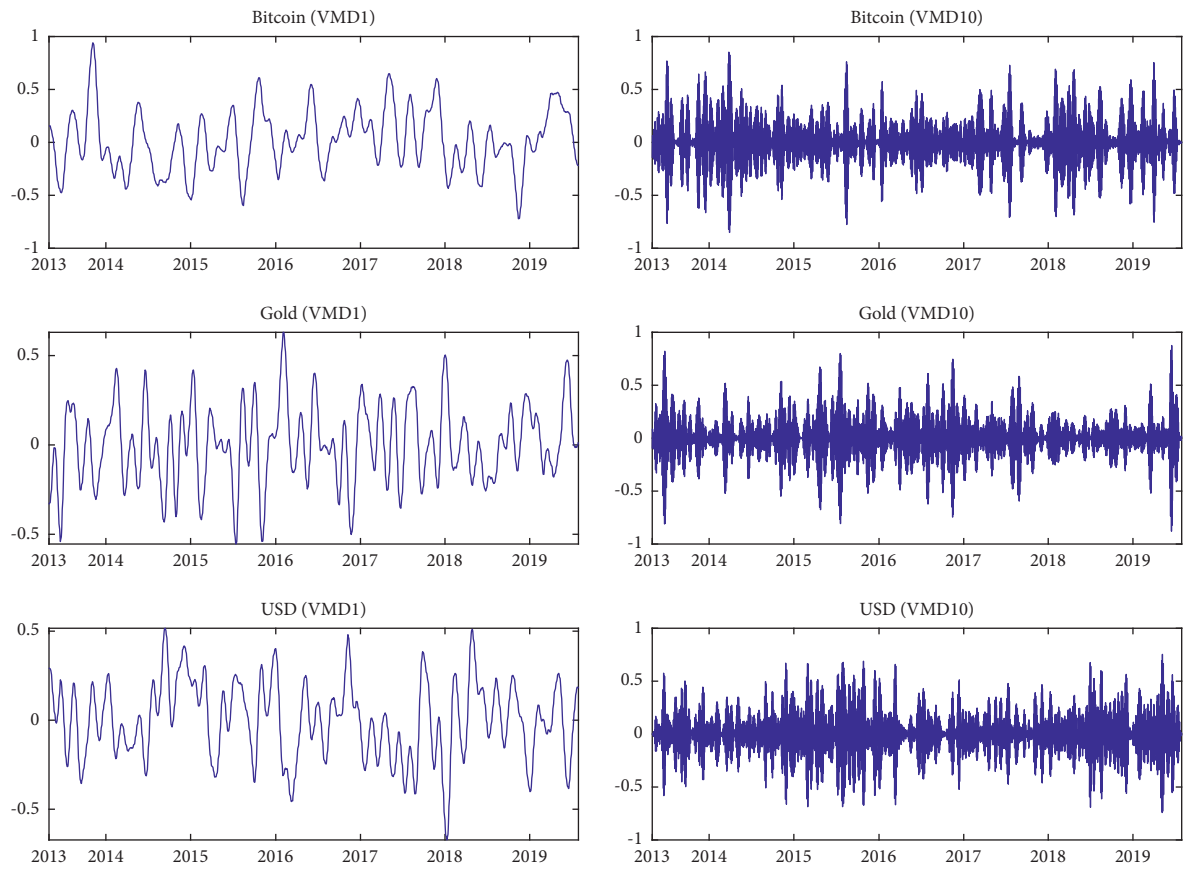


Figure 6: Results of the variational mode decomposition for the Bitcoin, gold, and USD series.

term $\Delta\text{CoVaR}^{D,\alpha}_{1|2\ D}$ and $\Delta\text{CoVaR}^{U,\alpha}_{1|2\,U}$ are positive, meaning the dominant of positive short-term risk spillover from Bitcoin to gold. Moreover, both short-term and long-term $\Delta\text{CoVaR}^{D,\alpha}_{1|2\,U}$ and $\Delta\text{CoVaR}^{U,\alpha}_{1|2D}$ between gold and USD have positive means, whereas both short-term and long-term $\Delta\text{CoVaR}^{D,\alpha}_{1|2D}$ and $\Delta\text{CoVaR}^{U,\alpha}_{1|2\,U}$ between gold and USD have negative means, further indicating the major negative risk spillover effect between gold and USD. What's more, the mean of the long-term $\Delta CoVaR$s between gold and USD are stronger than the short-term $\Delta CoVaR$s.

Table 7: Results of copula functions based on short-term and long-term dynamics.

	Bitcoin and gold		Gold and USD	
	Short term	Long term	Short term	Long term
Gaussian	−3.101	−4.370	−23.055	−368.472
Clayton	−0.638	0.008	0.0284	0.119
Rotated Clayton	−0.630	0.018	0.0286	0.116
Gumbel	−0.070	0.031	0.047	0.189
Rotated Gumbel	−0.063	0.002	0.047	0.192
Student	−0.436	−6.190	−29.096	−368.852
SJC	0.936	10.003	15.020	51.806
TVP-Gaussian	**−172.095**	**−445.316**	−137.640	**−542.407**
TVP-Clayton	−103.705	−285.717	−43.128	−44.847
TVP-rotated Clayton	−103.492	−270.566	−42.965	−41.757
TVP-Gumbel	−116.064	−146.486	−48.482	−46.911
TVP-rotated Gumbel	−116.223	−324.886	−48.343	−50.494
TVP-Student	−167.218	−440.555	−137.062	−537.223
TVP-Student 2	−158.283	−420.278	**−138.356**	−514.019
TVP-SJC	−167.651	−180.610	34.598	107.045

Note: this table reports the AIC values of different copula functions. The AICs of the best-fitted copula functions (i.e., the minimum AICs) are bold and underlined.

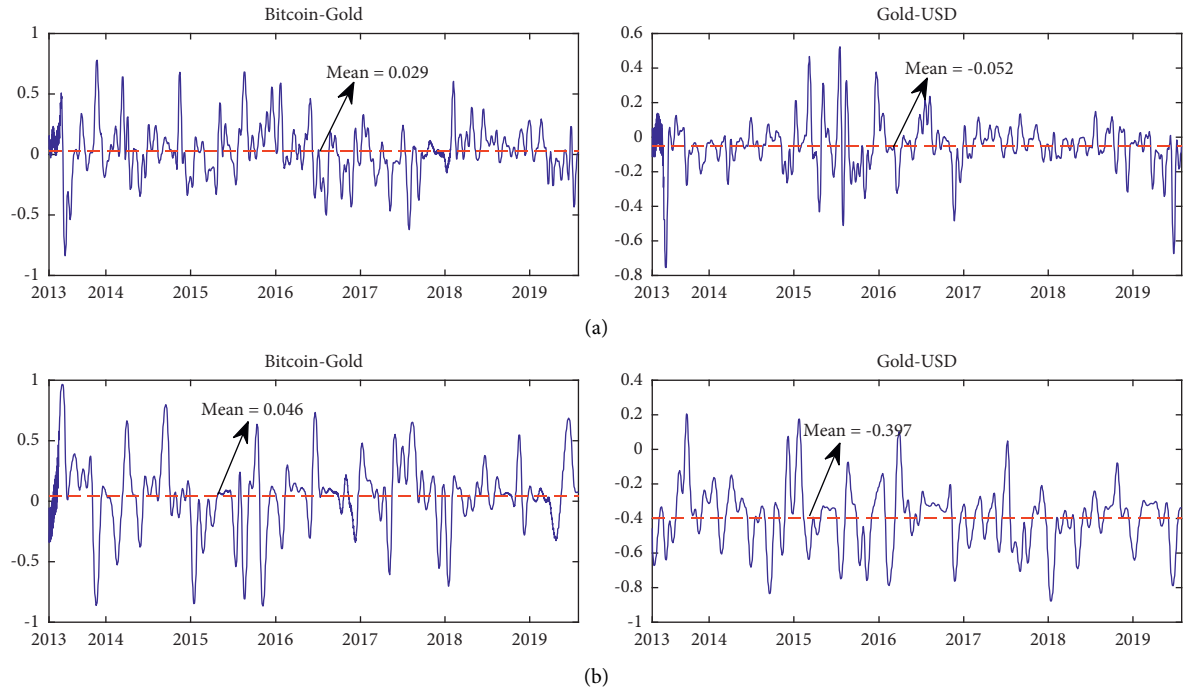


Figure 7: Time evolutions of short-term and long-term dependence between Bitcoin, USD, and gold returns. Note: the time-varying dependence structure is based on the best-fitted bivariate copula. (a) Panel A short-term dependence between Bitcoin and gold and between gold and USD. (b) Panel B long-term dependence between Bitcoin and gold and between gold and USD.

Table 8: Summary statistics of the short-term and long-term *CoVaRs* and Δ*CoVaRs*.

Symbol	Downside				Upside			
1|2	$\mathrm{CoVaR}^{D,\alpha}_{1\|2D}(S)$	$\mathrm{CoVaR}^{D,\alpha}_{1\|2D}(L)$	$\Delta\mathrm{CoVaR}^{D,\alpha}_{1\|2D}(S)$	$\Delta\mathrm{CoVaR}^{D,\alpha}_{1\|2D}(L)$	$\mathrm{CoVaR}^{U,\alpha}_{1\|2U}(S)$	$\mathrm{CoVaR}^{U,\alpha}_{1\|2U}(L)$	$\Delta\mathrm{CoVaR}^{U,\alpha}_{1\|2U}(S)$	$\Delta\mathrm{CoVaR}^{U,\alpha}_{1\|2U}(L)$
Bitcoin|gold	−0.073	−0.073	**0.045**	**0.046**	0.073	0.073	**0.044**	**0.049**
	(0.069)	(0.061)	**(0.326)**	**(0.559)**	(0.066)	(0.058)	**(0.300)**	**(0.492)**
Gold|Bitcoin	−0.013	−0.014	**0.006**	−0.019	0.014	0.014	**0.006**	−0.022
	(0.005)	(0.007)	**(0.234)**	(0.422)	(0.005)	(0.008)	**(0.241)**	(0.450)
Gold|USD	−0.012	−0005	−0.023	−0.519	0.013	0.005	−0.025	−0.545
	(0.004)	(0.005)	(0.172)	(0.413)	(0.004)	(0.005)	(0.183)	(0.428)
USD|gold	−0.004	−0.002	−0.023	−0.558	0.005	0.002	−0.021	−0.523
	(0.002)	(0.001)	(0.190)	(0.460)	(0.002)	(0.001)	(0.186)	(0.379)
	$\mathrm{CoVaR}^{D,\alpha}_{1\|2U}(S)$	$\mathrm{CoVaR}^{D,\alpha}_{1\|2U}(L)$	$\Delta\mathrm{CoVaR}^{D,\alpha}_{1\|2U}(S)$	$\Delta\mathrm{CoVaR}^{D,\alpha}_{1\|2U}(L)$	$\mathrm{CoVaR}^{U,\alpha}_{1\|2D}(S)$	$\mathrm{CoVaR}^{U,\alpha}_{1\|2D}(L)$	$\Delta\mathrm{CoVaR}^{U,\alpha}_{1\|2D}(S)$	$\Delta\mathrm{CoVaR}^{U,\alpha}_{1\|2D}(L)$
Bitcoin|gold	−0.063	−0.068	−0.034	−0.147	0.064	0.068	−0.028	−0.084
	(0.049)	(0.085)	(0.314)	(5.706)	(0.047)	(0.080)	(0.288)	(0.688)
Gold|Bitcoin	−0.013	−0.012	−0.048	−0.143	0.013	0.012	−0.051	−0.151
	(0.006)	(0.006)	(0.235)	(0.911)	(0.006)	(0.007)	(0.247)	(2.093)
Gold|USD	−0.015	−0.020	**0.071**	**0.313**	0.015	0.021	**0.075**	**0.328**
	(0.005)	(0.005)	**(0.135)**	**(0.133)**	(0.005)	(0.006)	**(0.144)**	**(0.139)**
USD|gold	−0.005	−0.007	**0.085**	**0.381**	0.005	0.008	**0.086**	**0.385**
	(0.002)	(0.002)	**(0.158)**	**(0.169)**	(0.002)	(0.002)	**(0.159)**	**(0.171)**

Note: this table shows the means and the standard deviations (in parenthesis) of short-term and long-term *CoVaR*s and Δ*CoVaR*s of Bitcoin-gold and gold-USD pairs. $\mathrm{CoVaR}^{D,\alpha}_{1|2D}$ and $\mathrm{CoVaR}^{D,\alpha}_{1|2U}$ describe the downside risk of market 1 when market 2 is extreme downward and upward movement, respectively. $\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ and $\mathrm{CoVaR}^{U,\alpha}_{1|2D}$ are the upside risk of market 1 when market 2 facing the extreme upward and downward risk, respectively. $\Delta\mathrm{CoVaR}^{D,\alpha}_{1|2D}$ and $\Delta\mathrm{CoVaR}^{D,\alpha}_{1|2U}$ are the risk contribution of the downward and upward movement of market 2 to the downward movement of market 1, respectively. $\Delta\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ and $\Delta\mathrm{CoVaR}^{U,\alpha}_{1|2D}$ are the risk contribution of the upward and downward movement of market 2 to the upward movement of market 1, respectively. *S* and *L* represent short-term and long-term, respectively.

Table 9: Results of the KS tests for the difference of short-term and long-term *CoVaR*s and *VaR*s.

Symbol	Downside				Upside			
	H_0: $\mathrm{CoVaR}^{D,\alpha}_{1\|2D} \geq \mathrm{VaR}^{D,\alpha}_{1}$		H_0: $\mathrm{CoVaR}^{D,\alpha}_{1\|2U} \geq \mathrm{VaR}^{D,\alpha}_{1}$		H_0: $\mathrm{CoVaR}^{U,\alpha}_{1\|2U} \leq \mathrm{VaR}^{U,\alpha}_{1}$		H_0: $\mathrm{CoVaR}^{U,\alpha}_{1\|2D} \leq \mathrm{VaR}^{U,\alpha}_{1}$	
	H_0: $\mathrm{CoVaR}^{D,\alpha}_{1\|2D} < \mathrm{VaR}^{D,\alpha}_{1}$		H_0: $\mathrm{CoVaR}^{D,\alpha}_{1\|2U} < \mathrm{VaR}^{D,\alpha}_{1}$		H_0: $\mathrm{CoVaR}^{U,\alpha}_{1\|2U} > \mathrm{VaR}^{U,\alpha}_{1}$		H_0: $\mathrm{CoVaR}^{U,\alpha}_{1\|2D} > \mathrm{VaR}^{U,\alpha}_{1}$	
1|2	Short	Long	Short	Long	Short	Long	Short	Long
Bitcoin|gold	**0.082** [**0.000**]	**0.107** [**0.000**]	**0.046** [**0.030**]	**0.068** [**0.000**]	**0.076** [**0.000**]	**0.105** [**0.000**]	**0.046** [**0.041**]	**0.064** [**0.004**]
Gold|Bitcoin	**0.110** [**0.000**]	**0.162** [**0.000**]	**0.078** [**0.000**]	**0.062** [**0.002**]	**0.119** [**0.000**]	**0.179** [**0.000**]	**0.083** [**0.000**]	**0.067** [**0.000**]
Gold|USD	**0.043** [**0.056**]	0.006 [0.953]	**0.153** [**0.000**]	**0.635** [**0.000**]	**0.040** [**0.074**]	0.006 [0.952]	**0.166** [**0.000**]	**0.651** [**0.000**]
USD|gold	**0.057** [**0.009**]	0.004 [0.960]	**0.160** [**0.000**]	**0.600** [**0.000**]	**0.062** [**0.002**]	0.004 [0.969]	**0.167** [**0.000**]	**0.632** [**0.000**]

Note: this table lists the statistic and the p value (in squared brackets) of the Kolmogorov–Smirnov (KS) tests, which verified the null hypothesis of no systemic risk spillover effect between Bitcoin and gold and between gold and USD.

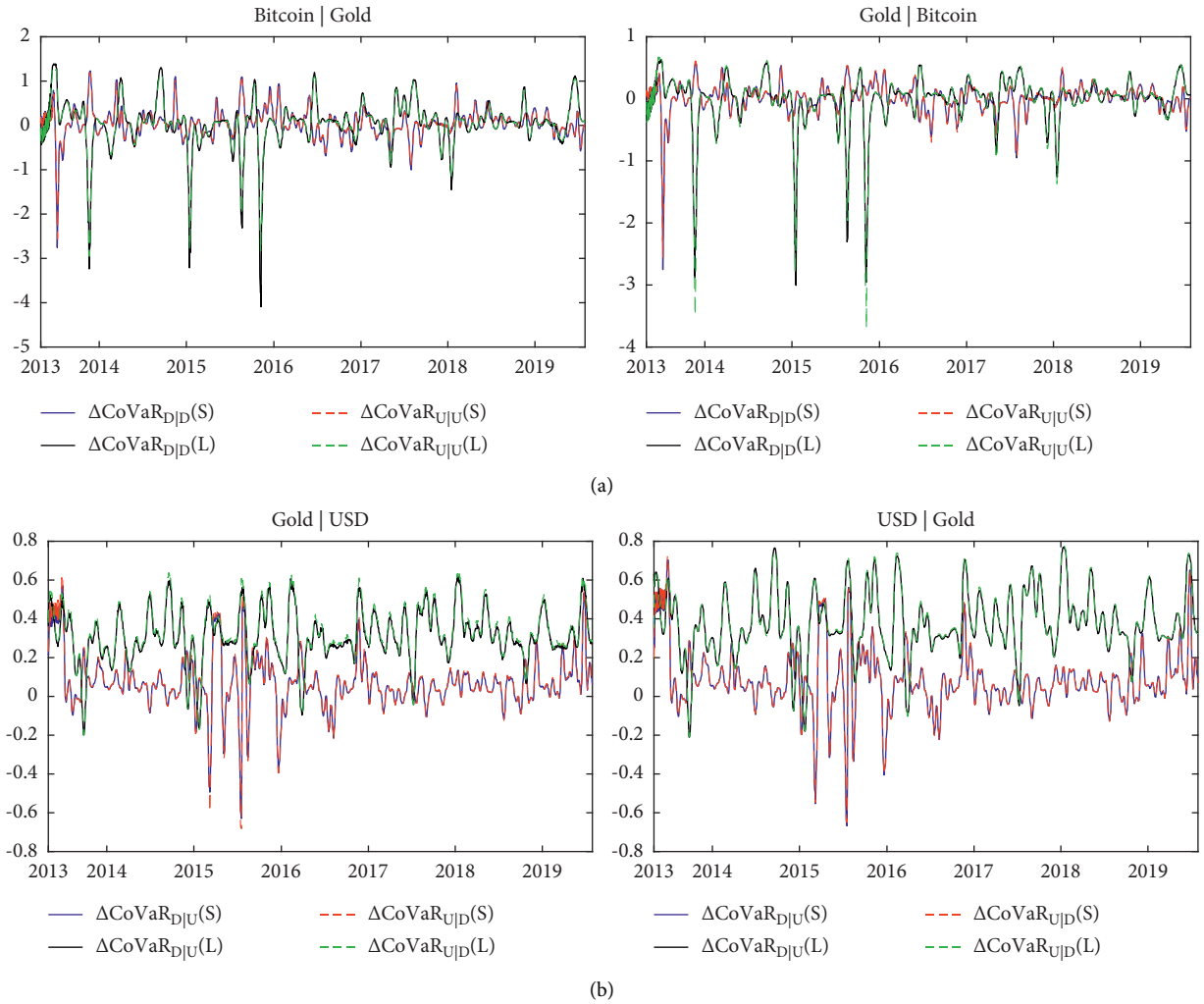


Figure 8: Short-term and long-term Δ*CoVaRs*. Notes: Panel A shows the short-term and long-term Δ*CoVaRs* of Bitcoin-gold pair. Panel B shows the short-term and long-term Δ*CoVaRs* of gold-USD pair. $\Delta\mathrm{CoVaR}^{D,\alpha}_{1|2D}$ and $\Delta\mathrm{CoVaR}^{D,\alpha}_{1|2U}$ are the risk contribution of the downward and upward movement of market 2 to the downward movement of market 1, respectively. $\Delta\mathrm{CoVaR}^{U,\alpha}_{1|2U}$ and $\Delta\mathrm{CoVaR}^{U,\alpha}_{1|2D}$ are the risk contribution of the upward and downward movement of market 2 to the upward movement of market 1, respectively. S and L denote the short-term and long-term Δ*CoVaRs*, respectively. (a) Panel A Short-term and long-term *ΔCoVaRs* of Bitcoin-gold pair. (b) Panel B Short-term and long-term Δ*CoVaR*s of gold-USD pair.

6. Conclusions

This article focuses on exploring the dependence structure and risk spillover effect between Bitcoin, gold, and USD. The dependence between Bitcoin, gold, and USD are captured using various static and dynamic copula functions, and the risk spillover effects are quantified with some measures and extensions about conditional Value-at-Risk ($CoVaR$). Besides, we also utilize the variational mode decomposition (VMD) technology to explore the short-term and long-term dependence and risk spillover effect. The main findings are as follow.

Firstly, there is strongly negative dependence between gold and USD, whereas Bitcoin is weakly positively correlated with gold and nearly not correlated with USD. The positive dependence between Bitcoin and gold supports the findings of Jareño et al. [38], suggesting that Bitcoin has potential hedging ability like gold. Interesting, the risk spillover effect between Bitcoin and gold are not stable, that is, the downside and the upside risk of Bitcoin (gold) have the chance to increase when gold (Bitcoin) facing the downward or upward risk, but the positive risk spillovers are dominant. Thirdly, there is no risk spillover effect between Bitcoin and USD, but strong negative risk spillover between gold and USD. The results of the dependence and risk spillover between Bitcoin and USD confirm the findings about return and volatility spillover between Bitcoin and USD in Wang et al. [23]. The unstable risk spillover between Bitcoin and gold and the nonexistent risk spillover between Bitcoin and USD further demonstrate that Bitcoin may far less of a hedge than gold and USD, and it is more suitable for using as a speculative asset because of its high return and high volatility. Fourthly, the risk spillover between Bitcoin and gold and between gold and USD are asymmetric at downward and upward market environment. The downside risk spillover from gold to Bitcoin are stronger than the upside risk spillover, whereas the upside risk spillover from Bitcoin to gold and between gold and USD are stronger than the downside risk spillover. For the short-term and long-term risk spillover effect, the risk spillover effect between Bitcoin and gold are still unstable. Specially, the risk spillover effect between gold and USD are not stable in short-term horizons, whereas there is a stable negative risk spillover between them in long-term horizons. Combined with the fact that gold can be used as a hedge against a falling USD, this finding implies that gold can be used to hedge the long-term risk of USD but is not stable enough to hedge the short-term risk of the USD.

The findings of this article have several important implications for investors. Firstly, in the current economic and digital cryptocurrency trading environment, gold and USD continue to be the preferred safe-haven assets for investors. Secondly, due to the greater speculative nature of digital cryptocurrencies, investors have the opportunity to earn greater returns by investing in cryptocurrencies, but they also face greater risks. Therefore, investors should be cautious about Bitcoin investments. Besides, this article finds a potential hedging ability for Bitcoin but its hedging ability still needs to be strengthened. To strength the hedging ability of Bitcoin, policymakers need to speed up its legalization process and improve the systems related to the supply, management, and trading of Bitcoin. In addition, the stronger downside risk spillover from gold to Bitcoin means that Bitcoin investors with long positions face bigger risk than those with short positions. The stronger upside risk spillover from Bitcoin to gold means that gold investors with short positions face bigger risk than those with long positions. However, due to the instability of risk spillover between bitcoin and gold, bitcoin (gold) investors cannot use the abnormal changes in gold (bitcoin) price as a basis to judge whether they should close their positions in time. Moreover, the stronger upside risk spillover between gold and USD means that gold and USD investors with short positions both face bigger risk than those with long positions. Because there are primarily negative risk spillovers between gold and USD, gold investors with short positions should close their positions timely when steep fall of USD begins. Similarly, USD investors with short positions should close their positions timely when observing the start of steep fall in gold price, too. Finally, we conclude that gold can be used to hedge the long-term risk of USD but is not stable enough to hedge the short-term risk of USD. In fact, USD that is often referred to in the literatures is usually the USD index or the dollar exchange rate, which is the ratio of USD to other currencies. If the corresponding other currencies are not depreciating/appreciating at the consistent rate as USD, the USD and gold will rise and fall together in the short term. That is why we find gold an unstable hedge against USD in the short term. However, USD index and gold remain negatively correlated in the long term. As a result, we find that gold can hedge the long-term risk of USD. So, investors should make a broad judgment on the relative depreciation or appreciation rate of USD and local currencies before deciding whether to use USD as a hedge.

Data Availability

All data used in this article are publicly available. Bitcoin price was downloaded from https://coinmarketcap.com/. The gold prices and USD index are collected from Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/).

Conflicts of Interest

The authors declare no conflicts of interest in this article.

Authors' Contributions

J. Y. prepared the initial manuscript and the data collection. Y. S. offered the pivotal idea of this research and the literature review. X.-F. L. made the empirical model estimations and supervised the writing process of this paper. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (71671145 and 71971191), Humanities and Social Science Fund of Ministry of Education of China (17YJA790015, 17XJA790002, 18YJC790132, and

18XJA790002), Science and Technology Innovation Team of Yunnan Provincial Universities (2019014), and Yunnan Fundamental Research Projects (202001AS070018).

References

[1] M. Arfaoui and A. Ben Rejeb, "Oil, gold, US dollar and stock market interdependencies: a global analytical insight," *European Journal of Management and Business Economics*, vol. 26, no. 3, pp. 278–293, 2017.

[2] T. Conlon, B. M. Lucey, and G. S. Uddin, "Is gold a hedge against inflation? A wavelet time-scale perspective," *Review of Quantitative Finance and Accounting*, vol. 51, no. 2, pp. 317–345, 2017.

[3] A. Kaul and S. Sapp, "Y2K fears and safe haven trading of the U.S. dollar," *Journal of International Money and Finance*, vol. 25, no. 5, pp. 760–779, 2006.

[4] C.-S. Liu, M.-S. Chang, X. Wu, and C. M. Chui, "Hedges or safe havens-revisit the role of gold and USD against stock: a multivariate extended skew-tcopula approach," *Quantitative Finance*, vol. 16, no. 11, pp. 1763–1789, 2016.

[5] X. Ma, R. Yang, D. Zou, and R. Liu, "Measuring extreme risk of sustainable financial system using GJR-GARCH model trading data-based," *International Journal of Information Management*, vol. 50, pp. 526–537, 2020.

[6] A. A. Salisu and I. Adediran, "Gold as a hedge against oil shocks: evidence from new datasets for oil shocks," *Resources Policy*, vol. 66, Article ID 101606, 2020.

[7] A. A. Salisu, I. D. Raheem, and U. B. Ndako, "The inflation hedging properties of gold, stocks and real estate: a comparative analysis," *Resources Policy*, vol. 66, 2020.

[8] M. H. Shakil, I. H. M. Mustapha, M. Tasnia, and B. Saiti, "Is gold a hedge or a safe haven? An application of ARDL approach," *Journal of Economics, Finance and Administrative Science*, vol. 23, no. 44, pp. 60–76, 2018.

[9] S.-Y. Wang and S.-K. Lin, "The pricing and hedging of structured notes with systematic jump risk: an analysis of the USD knock-out reversed swap," *International Review of Economics & Finance*, vol. 19, no. 1, pp. 106–118, 2010.

[10] X. Wen and H. Cheng, "Which is the safe haven for emerging stock markets, gold or the US dollar?" *Emerging Markets Review*, vol. 35, pp. 69–90, 2018.

[11] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," 2008, https://Bitcoin.org/Bitcoin.pdf.

[12] C. Baek and M. Elbeck, "Bitcoins as an investment or speculative vehicle? A first look," *Applied Economics Letters*, vol. 22, no. 1, pp. 30–34, 2014.

[13] D. G. Baur, K. Hong, and A. D. Lee, "Bitcoin: medium of exchange or speculative assets?" *Journal of International Financial Markets, Institutions and Money*, vol. 54, pp. 177–189, 2018.

[14] X. Gong and B. Lin, "Effects of structural changes on the prediction of downside volatility in futures markets," *Journal of Futures Markets*, vol. 41, no. 7, pp. 1124–1153, 2021.

[15] A. H. Dyhrberg, "Bitcoin, gold and the dollar-a GARCH volatility analysis," *Finance Research Letters*, vol. 16, pp. 85–92, 2016a.

[16] E. Bouri, S. J. H. Shahzad, D. Roubaud, L. Kristoufek, and B. Lucey, "Bitcoin, gold, and commodities as safe havens for stocks: new insight through wavelet analysis," *The Quarterly Review of Economics and Finance*, vol. 77, 2020.

[17] D. Das, C. L. Le Roux, R. K. Jana, and A. Dutta, "Does Bitcoin hedge crude oil implied volatility and structural shocks? A comparison with gold, commodity and the US dollar," *Finance Research Letters*, vol. 36, Article ID 101335, 2020.

[18] A. H. Dyhrberg, "Hedging capabilities of Bitcoin. Is it the virtual gold?" *Finance Research Letters*, vol. 16, pp. 139–144, 2016.

[19] T. Klein, H. Pham Thu, and T. Walther, "Bitcoin is not the new gold-a comparison of volatility, correlation, and portfolio performance," *International Review of Financial Analysis*, vol. 59, pp. 105–116, 2018.

[20] A. Kliber, P. Marszałek, I. Musiałkowska, and K. Świerczyńska, "Bitcoin: safe haven, hedge or diversifier? Perception of bitcoin in the context of a country's economic situation-a stochastic volatility approach," *Physica A: Statistical Mechanics and Its Applications*, vol. 524, pp. 246–257, 2019.

[21] R. Selmi, W. Mensi, S. Hammoudeh, and J. Bouoiyour, "Is Bitcoin a hedge, a safe haven or a diversifier for oil price movements? A comparison with gold," *Energy Economics*, vol. 74, pp. 787–801, 2018.

[22] S. J. H. Shahzad, E. Bouri, D. Roubaud, L. Kristoufek, and B. Lucey, "Is Bitcoin a better safe-haven investment than gold and commodities?" *International Review of Financial Analysis*, vol. 63, pp. 322–330, 2019.

[23] G. Wang, Y. Tang, C. Xie, and S. Chen, "Is Bitcoin a safe haven or a hedging asset? Evidence from China," *Journal of Management Science and Engineering*, vol. 4, no. 3, pp. 173–188, 2019.

[24] R. Bedoui, S. Braiek, K. Guesmi, and J. Chevallier, "Retracted: on the conditional dependence structure between oil, gold and USD exchange rates: nested copula based GJR-GARCH model," *Energy Economics*, vol. 80, pp. 876–889, 2019.

[25] F. Capie, T. C. Mills, and G. Wood, "Gold as a hedge against the dollar," *Journal of International Financial Markets, Institutions and Money*, vol. 15, no. 4, pp. 343–352, 2005.

[26] H.-F. Chang, L.-C. Huang, and M.-C. Chin, "Interactive relationships between crude oil prices, gold prices, and the NT-US dollar exchange rate-A Taiwan study," *Energy Policy*, vol. 63, pp. 441–448, 2013.

[27] Q. He, Y. Guo, and J. Yu, "Nonlinear dynamics of gold and the dollar," *The North American Journal of Economics and Finance*, vol. 52, Article ID 101160, 2020.

[28] F.-L. Lin, Y.-F. Chen, and S.-Y. Yang, "Does the value of US dollar matter with the price of oil and gold? A dynamic analysis from time-frequency space," *International Review of Economics & Finance*, vol. 43, pp. 59–71, 2016.

[29] B. Mo, H. Nie, and Y. Jiang, "Dynamic linkages among the gold market, US dollar and crude oil market," *Physica A: Statistical Mechanics and Its Applications*, vol. 491, pp. 984–994, 2018.

[30] P. Y. Ping, M. H. B. Ahmad, and N. B. Ismail, "Volatility spillover effect study in U.S. dollar and gold market based on bivariate-BEKK model," *Advances in Industrial and Applied Mathematics*, vol. 1750, Article ID 060006, 2016.

[31] J. C. Reboredo, "Is gold a safe haven or a hedge for the US dollar? Implications for risk management," *Journal of Banking & Finance*, vol. 37, no. 8, pp. 2665–2676, 2013.

[32] S. K. Samanta and A. H. M. Zadeh, "Co-movements of oil, gold, the U.S. dollar, and stocks," *Modern Economy*, vol. 3, no. 1, pp. 111–117, 2012.

[33] Y. S. Wang and Y. L. Chueh, "Dynamic transmission effects between the interest rate, the US dollar, and gold and crude oil prices," *Economic Modelling*, vol. 30, pp. 792–798, 2013.

[34] Y. Wei, C. Liang, Y. Li, X. Zhang, and G. Wei, "Can CBOE gold and silver implied volatility help to forecast gold futures

volatility in China? Evidence based on HAR and ridge regression models," *Finance Research Letters*, vol. 35, Article ID 101287, 2020.

[35] L. Bai, Y. Wei, G. Wei, X. Li, and S. Zhang, "Infectious disease pandemic and permanent volatility of international stock markets: a long-term perspective," *Finance Research Letters*, vol. 40, Article ID 101709, 2021.

[36] E. Bouri, M. Das, R. Gupta, and D. Roubaud, "Spillovers between Bitcoin and other assets during bear and bull markets," *Applied Economics*, vol. 50, no. 55, pp. 5935–5949, 2018.

[37] E. Bouri, R. Gupta, A. Lahiani, and M. Shahbaz, "Testing for asymmetric nonlinear short- and long-run relationships between Bitcoin, aggregate commodity and gold prices," *Resources Policy*, vol. 57, pp. 224–235, 2018.

[38] F. Jareño, M. d. l. O. González, M. Tolentino, and K. Sierra, "Bitcoin and gold price returns: a quantile regression and NARDL analysis," *Resources Policy*, vol. 67, Article ID 101666, 2020.

[39] S. H. Kang, R. P. McIver, and J. A. Hernandez, "Co-movements between Bitcoin and gold: a wavelet coherence analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 536, Article ID 120888, 2019.

[40] Y. Wei, S. Qin, X. Li, S. Zhu, and G. Wei, "Oil price fluctuation, stock market and macroeconomic fundamentals: evidence from China before and after the financial crisis," *Finance Research Letters*, vol. 30, pp. 23–29, 2019.

[41] H. S. Zwick and S. A. S. Syed, "Bitcoin and gold prices: a fledging long-term relationship," *Theoretical Economics Letters*, vol. 9, no. 7, pp. 2516–2525, 2019.

[42] T. Adrian and M. K. Brunnermeier, "CoVaR," *The American Economic Review*, vol. 106, no. 7, pp. 1705–1741, 2016.

[43] G. Boako and P. Alagidede, "Systemic risks spillovers and interdependence among stock markets: international evidence with covar-copulas," *South African Journal of Economics*, vol. 86, no. 1, pp. 82–112, 2018.

[44] G. Girardi and A. Tolga Ergün, "Systemic risk measurement: multivariate GARCH estimation of CoVaR," *Journal of Banking & Finance*, vol. 37, no. 8, pp. 3169–3180, 2013.

[45] Q. Ji, E. Bouri, D. Roubaud, and S. J. H. Shahzad, "Risk spillover between energy and agricultural commodity markets: a dependence-switching CoVaR-copula model," *Energy Economics*, vol. 75, pp. 14–27, 2018.

[46] W. Mensi, S. Hammoudeh, S. J. H. Shahzad, and M. Shahbaz, "Modeling systemic risk and dependence structure between oil and stock markets using a variational mode decomposition-based copula method," *Journal of Banking & Finance*, vol. 75, pp. 258–279, 2017.

[47] W. Mensi, A. Tiwari, E. Bouri, D. Roubaud, and K. H. Al-Yahyaee, "The dependence structure across oil, wheat, and corn: a wavelet-based copula approach using implied volatility indexes," *Energy Economics*, vol. 66, pp. 122–139, 2017.

[48] J. C. Reboredo and A. Ugolini, "Systemic risk in European sovereign debt markets: a CoVaR-copula approach," *Journal of International Money and Finance*, vol. 51, pp. 214–244, 2015.

[49] X. Sun, C. Liu, J. Wang, and J. Li, "Assessing the extreme risk spillovers of international commodities on maritime markets: a GARCH-Copula-CoVaR approach," *International Review of Financial Analysis*, vol. 68, Article ID 101453, 2020.

[50] Q. Ji, B.-Y. Liu, H. Nehler, and G. S. Uddin, "Uncertainties and extreme risk spillover in the energy markets: a time-varying copula-based CoVaR approach," *Energy Economics*, vol. 76, pp. 115–126, 2018.

[51] Q. Ji, B.-Y. Liu, and Y. Fan, "Risk dependence of CoVaR and structural change between oil prices and exchange rates: a time-varying copula model," *Energy Economics*, vol. 77, pp. 80–92, 2019.

[52] C. Brooks, *Introductory Econometrics for Finance*, Cambridge University Press, Cambridge, UK, 2002.

[53] B. E. Hansen, "Autoregressive conditional density estimation," *International Economic Review*, vol. 35, no. 3, pp. 705–730, 1994.

[54] M. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut Statistique de l'Universite de Paris*, vol. 8, pp. 229–231, 1959.

[55] A. J. Patton, "Modelling asymmetric exchange rate dependence*," *International Economic Review*, vol. 47, no. 2, pp. 527–556, 2006.

[56] T. Berger and G. S. Uddin, "On the dynamic dependence between equity markets, commodity futures and economic uncertainty indexes," *Energy Economics*, vol. 56, pp. 374–383, 2016.

[57] R. Jammazi and J. C. Reboredo, "Dependence and risk management in oil and stock markets. A wavelet-copula analysis," *Energy*, vol. 107, pp. 866–888, 2016.

[58] J. C. Reboredo and A. Ugolini, "Quantile dependence of oil price movements and stock returns," *Energy Economics*, vol. 54, pp. 33–49, 2016.

[59] B. Tong, C. Wu, and C. Zhou, "Modeling the co-movements between crude oil and refined petroleum markets," *Energy Economics*, vol. 40, pp. 882–897, 2013.

[60] O. C. D. Silva Filho, F. A. Ziegelmann, and M. J. Dueker, "Modeling dependence dynamics through copulas with regime switching," *Insurance: Mathematics and Economics*, vol. 50, no. 3, pp. 346–356, 2012.

[61] A. Abadie, "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 284–292, 2002.

[62] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review/Revue Internationale de Statistique*, vol. 55, no. 2, pp. 163–172, 1987.

[63] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 427–431, 1979.

[64] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.

[65] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1008, 1982.

[66] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[67] S. Lahmiri, "Long memory in international financial markets trends and short movements during 2008 financial crisis based on variational mode decomposition and detrended fluctuation analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 437, pp. 130–138, 2015.

[68] S. J. H. Shahzad, R. R. Kumar, S. Ali, and S. Ameer, "Interdependence between Greece and other European stock markets: a comparison of wavelet and VMD copula, and the portfolio implications," *Physica A: Statistical Mechanics and Its Applications*, vol. 457, pp. 8–33, 2016.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 2800671, 12 pages
https://doi.org/10.1155/2021/2800671

Research Article

The Impacts of the Infectious Disease Epidemic on the Permanent Volatility of Precious Metal and Crude Oil Futures Markets: A Long-Term Perspective

Yue Shang,[1] **Xiaodan Chen**,[2] **Yifeng Zhang**,[2] **and Yu Wei**[2]

[1]*School of Marxism, Yunnan University of Finance and Economics, Kunming, China*
[2]*School of Finance, Yunnan University of Finance and Economics, Kunming, China*

Correspondence should be addressed to Yu Wei; weiyusy@126.com

Received 31 July 2021; Revised 17 August 2021; Accepted 28 August 2021; Published 3 September 2021

Academic Editor: Dehua Shen

Copyright © 2021 Yue Shang 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 aim of this paper is to identify the quantitative impacts of the infectious disease pandemic on the permanent volatility of precious metal and crude oil futures from a long-term perspective by using a recently constructed Infectious Disease Equity Market Volatility Tracker (ID-EMV) to capture the epidemic severity and with a novel mixed data sampling GARCH (GARCH-MIDAS) method. Different from the extant literature only focusing on the short-term influences of the COVID-19 epidemic on commodity futures market, this paper shows that the infectious disease pandemic does have significant and positive impacts on the permanent (long-term) volatilities of precious metal and crude oil futures markets lasting for at least up to 12 months. In addition, these specific impacts on crude oil futures are greater than those on precious metal futures. Finally, we find that the infectious disease epidemic has larger impacts on gold (WTI oil) futures than those on silver (Brent oil) futures. All these findings are robust after controlling the negative influences of lagged long-run realized volatility in commodity futures markets.

1. Introduction

The outbreak of COVID-19 has caused huge losses in the global real economy and financial markets, especially in the commodity futures markets due to its direct impacts on both the demand and supply chain of commodities [1–17]. Crude oil markets are hit the hardest for the collapse in travelling and massive plant shutdowns arising from mitigation measures [7, 11, 14, 16, 18–21]. On 20 April 2020, the NYMEX WTI crude oil futures settled at −37.63 USD, not only the biggest drop in history but also the first time for crude oil futures price falling below zero value. Accompanied with the plunge in oil prices, however, precious metal prices keep on an upward trend in general during the COVID-19 pandemic. For example, the COMEX gold futures price increases from near 1500 USD per ounce in January 2020 to about 1900 USD per ounce at the end of 2020 and reaches the highest price over 2000 USD in August 2020. Thus, precious metals are largely used as hedge assets against market risk in crude oil markets [18, 20, 22–25]. Although the hedging role of precious metals in crude oil assets have been widely discussed in the literature, no research focuses on the quantitative impacts of the infectious disease epidemic on the volatility, especially the permanent volatility, of precious metal and crude oil futures market from a long-term perspective. To the best knowledge of the authors, this paper is the first research focusing on the quantitative impacts of the infectious disease pandemic on the permanent volatility of precious metal and crude oil futures markets by using a recently invented Infectious Disease Equity Market Volatility Tracker (ID-EMV) to capture the epidemic severity and with a long-term data sample through 2010 to 2020.

Most related research in investigating the influence of the infectious disease pandemic on precious metal and crude oil markets arose in the recent COVID-19 epidemic [7, 18, 20, 21]. That means most of these studies focus on the short-term effects of COVID-19 on those commodity

markets only by comparing the qualitative changes in market statuses (i.e., returns, volatilities, or correlations) before and after the outbreak of the COVID-19 epidemic. For example, Adekoya et al. [18] explored the hedging abilities of gold against oil and stock markets risks during the COVID-19 epidemic period by using the dynamic and regime dependent models. They found that gold can offset against oil and stock market risks across the times of global emergency, and short-term investors hoping portfolio diversification can find good hedging capability in gold. Dutta et al. [19] examined the safe-haven property of gold for crude oil market during the COVID-19 outbreak, and they showed that gold outperforms Bitcoin as a safe-haven asset. Mensi et al. [20] examined the impacts of COVID-19 on the asymmetric multifractality of gold and oil. They revealed that, in the pre-COVID-19 time period, the gold and oil markets are talentless in downward and upward trends, respectively. But, after the COVID-19 epidemic outbreak, gold and oil are even more talentless in upward and downward trends, respectively. Gold and oil have become more inefficient during the pandemic outbreak compared to the pre-COVID-19 period. Salisu et al. [21] evaluated the safe haven potential of gold vs. oil price risk by the asymmetric VARMA-GARCH model by using daily-frequency data. They found that gold presents major safe-haven characteristics to oil price risks. Corbet et al. [5] showed that the volatility connections among the Chinese stock markets, gold, oil and Bitcoin evolve significantly during this COVID-19 period. Such time-varying dependences at the periods of stress show further fact to deliberately support the effectiveness of the development of Bitcoin. In addition, before the COVID-19 pandemic, there also has been plenty of research in exploring the return and volatility dependencies among precious metal and crude oil markets by using various econometric models [26–49]. For instance, Al-Yahyaee et al. [27] examined the multiscale associated movement and connectedness among precious and nonferrous metals by employing Diebold and Yilmaz's [50] spillover index and some wavelet approaches. They found strong comovements among nonferrous and precious metals across medium-term and low frequencies, and gold can offer the best diversification effect on nonferrous metals.

Although a great deal of work has been carried out in the field of volatility modelling for precious metal and crude oil markets [51–53], it is worth noting that no research has been conducted to test the *direct impacts* of the infectious disease epidemic on the long-term or permanent volatility of both precious metal and crude oil futures market by now. Furthermore, we think it is extremely important for policy makers, commodity producers, and investors to know the quantitative influences of infectious disease epidemic on long-run volatility in these commodity markets. In addition, identifying the differences in these influences between precious metal and crude oil assets can also help us to get deeper understanding in the heterogeneity among risky and hedging instruments when facing the same risk resources and then contribute to make better regulation measures and portfolio decisions [29, 37, 38, 40]. Therefore, this paper contributes to the extant literature in the three ways: firstly, instead of comparing the qualitative changes in returns, volatilities, or correlations in precious metal and crude oil futures markets during the recent COVID-19 pandemic, this paper focuses on the quantitative impacts of the infectious disease epidemic on the volatilities of precious metal and crude oil futures markets. There is no doubt that quantitative information of such impacts can help a lot for policy makers and investors in their decision-making processes. Secondly, most of recent related research is only interested in the short-term effects of the COVID-19 pandemic on commodity futures markets, e.g., a few months before and after the outbreak of COVID-19. This paper, however, explores the long-term influence of the infectious disease epidemic by using a data sample through January 2000 to November 2020, which can offer us more general pictures about the shocks of public health emergencies on volatilities of precious metal and crude oil futures markets. Finally, this paper provides marginal information about the impacts of infectious disease epidemic on the long-run (i.e., monthly) volatility of precious metal and crude oil markets by utilizing a new mixed data sampling GARCH model, i.e., GARCH-MIDAS, proposed by Engle and Rangel [54] and Engle et al. [55], while most extant research only care about the short-run (i.e., daily) volatilities in commodity markets during the COVID-19 turmoil by using ordinary GARCH-type or HAR-type models, while as pointed by Corsi [56], investors in financial markets are heterogeneous with different investment horizons, i.e., short, medium, and long run, for their different portfolio allocation purposes and structures.

The rest of this paper is organized as follows: Section 2 introduces the GARCH-MIDAS model, Section 3 provides descriptive statistics of the data, Section 4 discusses the empirical results, Section 5 presents some robustness checks of the major findings, and finally Section 6 concludes this paper.

2. Methodology

In traditional volatility modelling, GARCH-type models are commonly employed. However, these traditional models can only capture the impacts of exogenous variables on the volatility of endogenous variables with both of them sampling at the same time frequency, e.g., daily, weekly or monthly. In this paper, we aim at examining the impacts of the low-frequency infectious diseases epidemic (ID-EMV) on long-term volatility of precious metal and crude oil futures markets, which is out of the scope of those traditional GARCH-type models. Therefore, a novel mixed data sampling method (MIDAS), the GARCH-MIDAS proposed by Engle and Rangel [54] and Engle et al. [55], is employed here to solve the problem of inconsistent sampling frequencies of exogenous and endogenous variables in volatility modelling. In the GARCH-MIDAS model, the conditional volatility of an asset is constructed by the product of a short-term volatility component and a long-term volatility one. To be specific, the short-term (e.g., daily) volatility element is caused by a simple GARCH (1, 1) process while the permanent (long-term) element is determined by a MIDAS regression on a low-frequency (e.g., monthly) exogenous variable. The concrete modelling process of the GARCH-MIDAS model is as follows:

In the GARCH-MIDAS model, the logarithmic return on day i in month t, $r_{i,t}$, is calculated as

$$r_{i,t} - \omega = \sqrt{g_{i,t}\tau_t} z_{i,t}, \quad \forall i = 1, \ldots, N_t, \\ z_{i,t} | I_{i-1,t} \sim N(0,1), \tag{1}$$

where $\omega = E_{i-1,t}(r_{i,t})$ is the conditional mean of $r_{i,t}$ given the information set, $I_{i-1,t}$, which is attainable at $(i-1)$ day in month t. N_t is the total number of trading days in month t, and $z_{i,t}$ is an *i.i.d.* innovation term. Moreover, $g_{i,t}$ in equation (1) is the short-term volatility component following a GARCH (1, 1) process as follows:

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{\left(r_{i-1,t} - \omega\right)^2}{\tau_t} + \beta g_{i-1,t}, \tag{2}$$

where $\alpha > 0$, $\beta > 0$ and $1 - \alpha - \beta > 0$. τ_t in equation (1) is the permanent (long-term) volatility component. To keep it taking a positive value, we usually model the logarithmic τ_t by this MIDAS regression:

$$\log(\tau_t) = m + \theta \sum_{k=1}^{K} \varphi_k(w) X_{t-k}, \tag{3}$$

where K is the time lags for the permanent volatility and X_{t-k} is a low-frequency exogenous impactor that may influence this permanent volatility; see, for example, monthly ID-EMV utilized in this paper. $\varphi_k(w)$ is a weighting function to determine the decaying influence of X_{t-k} and is defined by a one-parameter Beta polynomial:

$$\varphi_k(w) = \frac{((1-k)/K)^{w-1}}{\sum_{j=1}^{K} ((1-j)/K)^{w-1}}, \tag{4}$$

where $w \geq 1$. Furthermore, as pointed by Engle et al. [55], the permanent volatility is also easily impacted by its own past realized volatilities RV_{t-k}, which is defined as $RV_{t-k} = \sum_{i=1}^{N_t} r_{i,t-k}^2$ and thus we rewrite equation (3) to contain two low-frequency impactors as follows:

$$\log(\tau_t) = m + \theta_1 \sum_{k=1}^{K} \varphi_k(w_1) RV_{t-k} + \theta_2 \sum_{k=1}^{K} \varphi_k(w_2) X_{t-k}, \tag{5}$$

where RV_{t-k} can be treated as a control variable, which enables us to identify the quantitative influence of the low-frequency exogenous variable X_{t-k} on permanent volatility.

To estimate the GARCH-MIDAS model, two methods, fixed window and rolling window, are commonly employed. The difference between these two approaches is that the rolling-window method sets the long-run volatility to vary every high-frequency time; however, the fixed-window approach sets the permanent volatility in a calendar time window, i.e., one week/month/quarter/year. In the empirical analysis section, following the suggestions, we employ a rolling-window approach to identify the monthly factors in the empirical estimations. It means that the long-run or permanent volatility in this study can vary in each time period.

3. Data

In this empirical study, we use two categories of data sampling at daily and monthly frequencies, respectively. The first one is *daily* data of four commodity futures, i.e., precious metal and crude oil. To be specific, gold and silver futures traded in COMEX are utilized to represent the precious metal markets. In the same way, ICE Brent and NYMEX WTI crude oil futures are selected to indicate the performances of crude oil markets. All the daily futures prices are recorded for contracts at the earliest specified delivery date (for instance, for WTI crude oil, each contract expires on the third business day before the 25th calendar day of the month preceding the delivery month. If the 25th calendar day of the month is a nonbusiness day, trading ceases on the third business day before the business day preceding the 25th calendar day. After a contract expires, the nearest expiration contract for the remainder of that calendar month is the second following month).

The second one is the *monthly* data of the Infectious Disease Equity Market Volatility Tracker (ID-EMV) developed by Baker et al. [57]. This ID-EMV tracker is constructed in the following steps: first, Baker et al. [57] specify related text terms in four sets as follows (term variants also included): (1) E: [economic, economy, and financial]; (2) M: [stock market, equity, equities, standard, and poor]; (3) V: [volatility, volatile, uncertain, uncertainty, risk, and risky] and (4) ID: [epidemic, pandemic, virus, flu, disease, coronavirus, MERS, SARS, EBOLA, H5N1, and H1N1]. Second, they obtain daily counts of newspaper articles that contain at least one term in each of E, M, V, and ID across approximately 3,000 US Newspapers. Third, the raw ID-EMV counts are scaled by the count of all articles in the same day. In a final step, they multiplicatively rescale the resulting series mirroring this approach to scale a Categorical EMV series in the overall EMV tracker. That is, they match the level of the VIX between 1990 and 2016 using the overall EMV index and then scale this ID-EMV index to reflect the ratio of the ID-EMV articles to total EMV articles. Baker et al. [57] released the ID-EMV tracker at daily frequency in the website of Economic Policy Uncertainty (http://www.policyuncertainty.com), while we convert it into monthly frequency for the reason that there are many zero values of it at daily frequency. More specifically, there are 8240 zero records out of total 13118 (i.e., about 63%) daily ID-EMV trackers through January 1, 1985 to November 30, 2020. Thus, it is difficult to track the quantitative fluctuations of the infectious diseases epidemic with this daily tracker containing too many zero values. Moreover, the major purpose of this research is to quantify the impact of the infectious disease epidemic on the long-term volatility of precious metal and crude oil markets. So, we transform daily ID-EMV$_{i,t}$ at the ith day in month t into monthly ID-EMV$_t$ by summing all daily ID-EMV$_{i,t}$ within a month t. The data sample used in this study goes through January 5, 2000 to November 30, 2020, which is by far the longest research sample concerning impacts of ID-EMV on commodity futures market volatility as we know. The

precious metal and crude oil futures price is available at the Wind database, and the ID-EMV data are collected from the website http://www.policyuncertainty.com/infectious_EMV.html.

Figure 1 shows the time evolutions of returns for gold, silver, Brent, and WTI oil futures from January 2000 to November 2020. We can see that there is clear volatility clustering in these return series occasionally, e.g., during the time periods of the 2008 subprime mortgage crisis and the recent COVID-19 pandemic. Additionally, crude oil markets seem to fluctuate more violently than precious metal markets, especially during the recent COVID-19 pandemic time.

Table 1 presents the descriptive statistics of monthly ID-EMV and four futures returns. Firstly, in general, for the reason of being recorded at different time frequencies, the monthly change rate of ID-EMV has larger mean and standard deviation than those of daily futures returns. In addition, ID-EMV has smaller skewness and kurtosis than futures returns and, thus, presents normal distribution. Secondly, in terms of commodity futures, Brent and WTI oil have larger standard deviations than gold and silver returns, indicating greater risk is observed in oil markets than in precious metal markets. This is also reasonable due to the facts that gold and silver are commonly used as hedging assets against risk assets, such as crude oil. Moreover, all the four commodity futures returns have negative skewness and kurtosis larger than 3, providing confirmative evidence of nonnormality distributions of them. Additionally, we can also see that gold and silver have no significant auto-correlations in the time series but Brent and WTI crude oil have, implying long-memory features in oil markets. Finally and most importantly, all the ADF and P-P tests reject the null hypothesis of unit root in these time series, indicating that all the return series are stationary and can be modelled without further transforms.

4. Empirical Analysis

In this section, we use the GARCH-MIDAS + ID-EMV method defined by equations (1)–(4) to quantify the impacts of infectious disease epidemic on permanent volatility of precious metal and crude oil markets. That means only one exogenous variable, i.e., monthly ID-EMV tracker, is incorporated in equation (3) and the parameter θ indicating the quantitative impacts is what we are most interested in.

We present in Figure 2 the estimated total and long-run (permanent) volatility of four futures assets with 6-month lags in ID-EMV. Firstly, it shows that, in most time periods, the total volatility fluctuates around permanent component, except for silver futures. Secondly, regarding precious futures markets, in the most volatile time periods, i.e., the 2008 global financial crisis and the recent COVID-19 pandemic, the long-run volatility increases accordingly in gold futures, while it seems to have no obvious ascents in silver futures. This finding implies that gold and silver futures may respond differently to common risk factors, especially in their long-run volatilities. Finally, in terms of crude oil futures, we find that the long-run volatility components in Brent and WTI oil react to risk factors in a high degree of consistency. Both of them increase moderately during the 2008 financial crisis but surging upwards in the recent COVID-19 epidemic.

Furthermore, to get more quantitative and robust results, we estimate the GARCH-MIDAS + ID-EMV models by setting time lags of ID-EMV through 3 to 12 months, which are presented in Tables 2–5, respectively. As we can see, Tables 2–5 show highly consistent estimation results. Firstly, almost all the estimated parameters are significant at the 1% level in both short- and long-run volatility models, indicating good fitness of the GARCH-MIDAS model with these commodity futures returns. Secondly, the parameters, α and β, for short-term volatility are all significant at the 1% level. In particular, all the β parameters are larger than 0.91, revealing strong volatility clustering features in these precious metal and crude oil futures markets. Finally, and most importantly, the estimate parameters θ are positive and significant at the 1% level across all the four futures with different time lags (from 3 to 12 months), implying that a severer infectious disease epidemic will result in greater long-run volatility in these commodity futures markets. This finding demonstrates that the infectious disease epidemic can really influence the permanent volatility in precious metal and crude oil market. Thus, it is a key factor that should be considered in volatility modelling and risk measurement for policy makers and investors.

More specific, except for the case of 6-month lags in ID-EMV, we can find that ID-EMV has larger impacts (i.e., larger θ) on the permanent volatility in crude oil markets than it does in precious metal markets, confirming the fact that crude oil is a risky asset while precious metal can be used to hedge this risk [27, 48, 58]. In addition, within precious metals, the impacts of ID-EMV on gold is greater than those on silver, proving that gold is easier than silver to be influenced by infectious disease epidemic, which is consistent with the conclusions in some extant research [58, 59]. Similarly, in terms of crude oil futures, WTI oil is more likely to be affected by ID-EMV than Brent oil. The possible reasons for this finding are twofold: on the one hand, NYMEX WTI oil price is primarily used as the benchmark for pricing U.S. oil imports and local production in North America (including Canada). However, for the vast majority of the world's crude oil production areas, such as in the North Sea, Africa, Latin America, and some Middle and Far Eastern regions, ICE Brent oil price is usually used as the pricing benchmark. Thus, when facing the same risk factors, ICE Brent oil price can be more likely to diversify these shocks through its broader oil production regions, while WTI oil price is more susceptible to shocks of uncertainty than Brent oil for its relative limited production locations. On the other hand, the U.S. shale oil revolution, which began in 2005, has increased greatly the elasticity of U.S. crude oil production and, therefore, increased the volatility in WTI oil price.

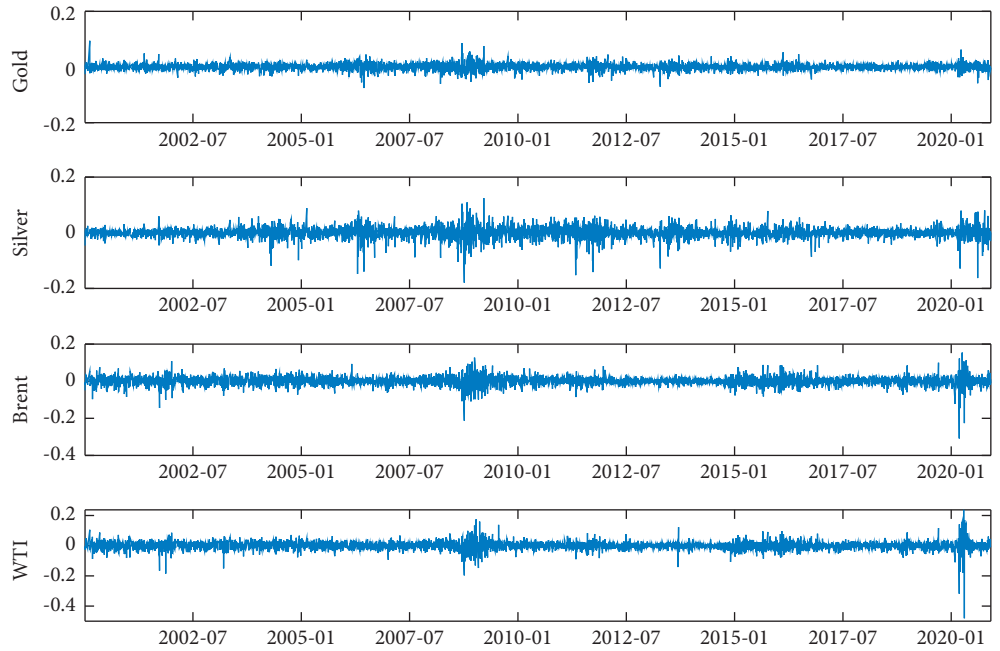


Figure 1: Time evolutions of returns for precious metal and crude oil futures.

Table 1: Descriptive statistics for ID-EMV and commodity futures returns.

	ID-EMV	Gold	Silver	Brent	WTI
Frequency	Monthly	Daily	Daily	Daily	Daily
Mean (%)	1.327	0.037	0.031	0.015	0.012
Standard deviation (%)	66.143	1.107	1.955	2.327	2.656
Skewness	0.146	−0.023	−0.912	−0.784	−1.548
Kurtosis	0.523	8.640	11.985	15.666	35.259
Jarque–Bera	3.771	6700.547***	17702.994***	34308.611***	221208.003***
Q (5)	51.601***	2.559	4.029	14.279**	15.302***
Q (10)	57.327***	16.127*	9.911	21.541**	23.993***
ADF	−16.223***	−71.327***	−71.157***	−73.889***	−73.582***
P-P	−28.514***	−71.402***	−71.168***	−73.885***	−73.577***

Notes. The Jarque–Bera statistic tests for the null hypothesis of normality in sample returns distribution. Q (n) is the Ljung–Box statistics of the return series for up to the nth-order serial correlation. ADF and P-P are statistics of the Augmented Dickey–Fuller and Phillips–Perron unit root test based on the least AIC criterion, respectively. ***, **, and * indicate rejection at the 1%, 5%, and 10% significance level, respectively.

5. Robustness Checks

In this section, we further employ GARCH-MIDAS models with two monthly exogenous variables to test the robustness of the major findings in Section 4. As stated by Engle et al. [55], the long-run volatility in financial markets is closely related by its past realized volatility RV_{t-k}. It means that the effects of other possible risk factors on futures permanent volatility, rather than infectious disease epidemic (ID-EMV), can be captured by its past realized volatility, as pointed by Bai et al. [60]. Thus, in this section, we use a model, labelled as GARCH-MIDAS + RV + ID-EMV, defined in equation (5) to reestimate the empirical results reported in Tables 2–5.

Figure 3 further presents the estimated total and long-run volatility of the four futures assets with 6-month lags in both RV and ID-EMV. It means that the observed permanent volatilities in Figure 3 are not determined by the sole impacts of ID-EMV but by the joint effects of lagged RV and the infectious disease epidemic. In general, Figure 3 looks consistent with Figure 2 with several slight differences. Firstly, it shows that the total volatility fluctuates around its permanent component, except for silver futures. Secondly, during periods of severe market disruptions, i.e., the 2008 financial crisis and the recent COVID-19 pandemic, the long-run volatility observed in gold and silver markets are similar to those in Figure 2. However, with respect to crude oil futures, we find quite different pictures in Figure 3. For instance, the permanent volatilities in crude oil markets decrease clearly rather than increase during the period of the 2008 global financial crisis, indicating that lagged RV can really suppress the increase in the present permanent

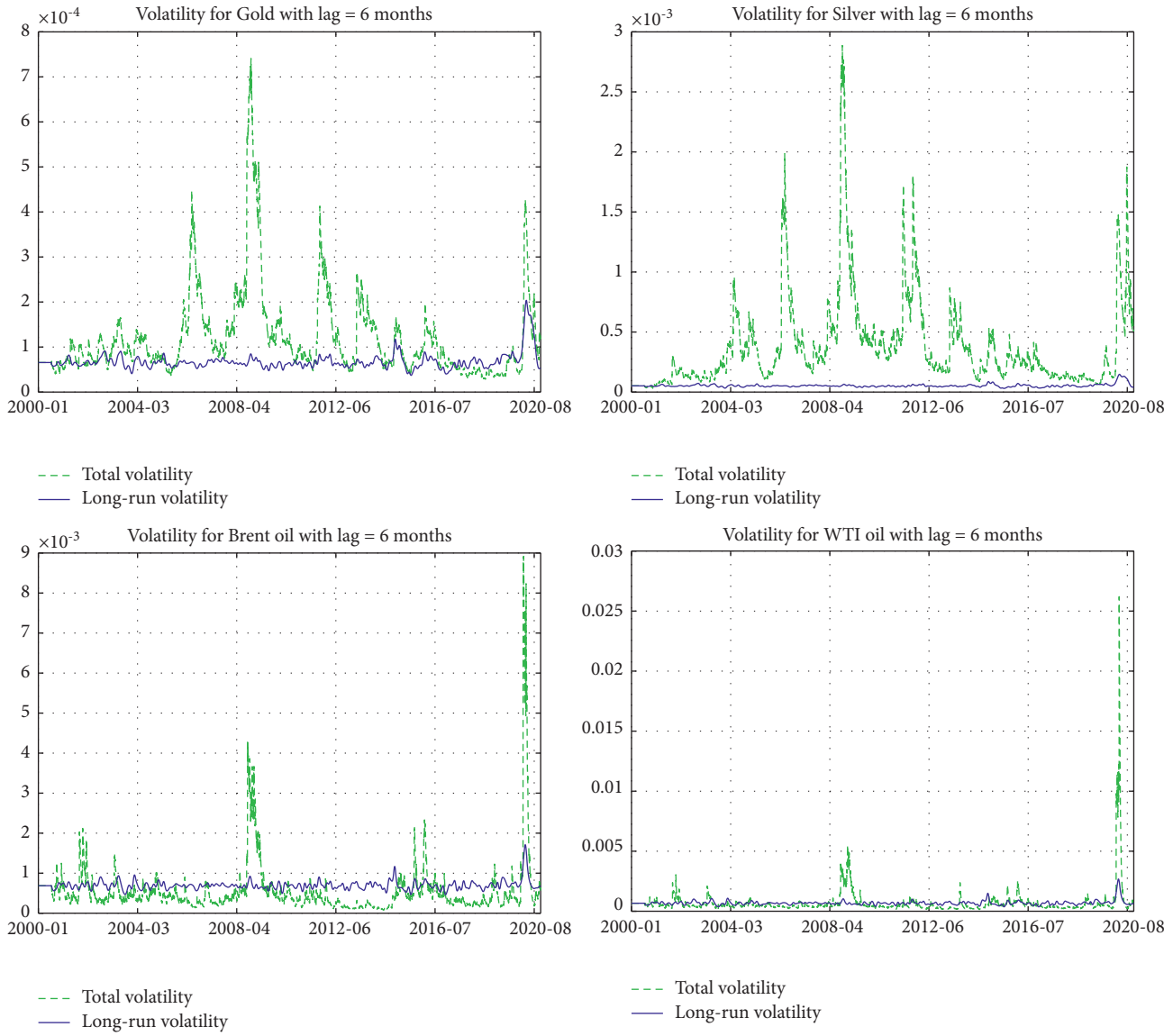


Figure 2: Estimated total and long-run volatility of precious metal and crude oil futures with GARCH-MIDAS + ID-EMV models.

Table 2: Estimation results of GARCH-MIDAS + ID-EMV models of futures volatility by infectious disease with 3-month lags.

Futures	ω	α	β	m	θ	w	Log L
Gold	0.0003** (0.0001)	0.0281*** (0.0012)	0.9718*** (0.0014)	−9.3882*** (0.1696)	0.1439*** (0.0343)	17.7450 (11.8550)	15812.3
Silver	0.0000 (0.0002)	0.0491*** (0.0021)	0.9441*** (0.0025)	−7.8688*** (0.1109)	0.2446*** (0.0671)	3.4689*** (1.1158)	13221.1
Brent	0.0005** (0.0003)	0.0676*** (0.0033)	0.9275*** (0.0037)	−7.1387*** (0.2075)	0.7326*** (0.1164)	1.5349*** (0.2478)	12450.7
WTI	0.0004 (0.0003)	0.0658*** (0.0032)	0.9198*** (0.0036)	−7.3261*** (0.0771)	1.0272*** (0.1021)	1.6435*** (0.1686)	11925.6

Note. The bracketed numbers are the standard errors of the estimations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Log L is the logarithmic likelihood of the estimation.

volatilities in crude oil markets even with the simultaneously positive impacts of ID-EMV on them.

Tables 6–9 show the estimation results with both RV and ID-EMV lagged through 3 to 12 months. The revealed empirical findings are highly in line with those presented in Section 4. First of all, most of the parameters estimated are significant, except for several ones of ω and w_1, further indicating the nice goodness of fit of GARCH-MIDAS

Table 3: Estimation results of GARCH-MIDAS + ID-EMV models of futures volatility by infectious disease with 6-month lags.

Futures	ω	α	β	m	θ	w	Log L
Gold	0.0003** (0.0001)	0.0246*** (0.0012)	0.9753*** (0.0013)	−9.6726*** (0.2460)	1.6900*** (0.1358)	1.0947*** (0.0536)	15609.7
Silver	0.0000 (0.0002)	0.0319*** (0.0012)	0.9681*** (0.0012)	−9.9052*** (0.3591)	1.6535*** (0.1607)	1.0045*** (0.0373)	12996.7
Brent	0.0005** (0.0003)	0.0690*** (0.0034)	0.9248*** (0.0040)	−7.2990*** (0.2040)	0.6740*** (0.1748)	3.9198*** (1.2002)	12296.7
WTI	0.0004* (0.0003)	0.0696*** (0.0033)	0.9146*** (0.0037)	−7.3387*** (0.0767)	1.0803*** (0.1406)	3.4832*** (0.5312)	11782.4

Note. The bracketed numbers are the standard errors of the estimations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Log L is the logarithmic likelihood of the estimation.

Table 4: Estimation results of GARCH-MIDAS + ID-EMV models of futures volatility by infectious disease with 9-month lags.

Futures	ω	α	β	m	θ	w	Log L
Gold	0.0003** (0.0001)	0.0327*** (0.0023)	0.9581*** (0.0030)	−9.1116*** (0.0651)	1.1873*** (0.2522)	2.4184*** (0.5201)	15397.2
Silver	0.0000 (0.0002)	0.0333*** (0.0013)	0.9667*** (0.0013)	−9.2487*** (0.1578)	0.9361*** (0.2638)	2.2681*** (0.6303)	12749.8
Brent	0.0005* (0.0003)	0.0683*** (0.0035)	0.9258*** (0.0039)	−7.2736*** (0.1853)	0.7979*** (0.2370)	5.1022*** (1.7731)	12146.2
WTI	0.0005 (0.0003)	0.0680*** (0.0032)	0.9155*** (0.0036)	−7.3866*** (0.0707)	1.2215*** (0.1723)	4.8095*** (0.8107)	11634.6

Note. The bracketed numbers are the standard errors of the estimations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Log L is the logarithmic likelihood of the estimation.

Table 5: Estimation results of GARCH-MIDAS + ID-EMV models of futures volatility by infectious disease with 12-month lags.

Futures	ω	α	β	m	θ	w	Log L
Gold	0.0003** (0.001)	0.0339*** (0.0022)	0.9579*** (0.0030)	−9.0390*** (0.0751)	0.5433*** (0.1477)	8.7631** (3.4253)	15168.9
Silver	0.0000 (0.0002)	0.0347*** (0.0013)	0.9653*** (0.0013)	−9.2016*** (0.3558)	0.2437*** (0.0559)	18.0210*** (5.8944)	12530.3
Brent	0.0005** (0.0003)	0.0677*** (0.0035)	0.9255*** (0.0040)	−7.3990*** (0.1793)	0.7788*** (0.2325)	7.2045*** (2.5398)	11997.8
WTI	0.0005* (0.0003)	0.0681*** (0.0032)	0.9158*** (0.0036)	−7.3793*** (0.0731)	1.2248*** (0.1826)	6.5162*** (1.1766)	11488.7

Note. The bracketed numbers are the standard errors of the estimations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Log L is the logarithmic likelihood of the estimation.

models in these commodity futures markets. Secondly, the α and β parameters for short-term GARCH (1, 1) volatility models are all significant at the 1% level, also showing notable short-term volatility clustering in these markets. Thirdly, the majority of the estimated θ_1 coefficients for lagged RV are significantly negative with few exceptions in gold futures. This result shows that past realized volatility can adversely affect the long-term volatility of the current period. In other words, high (low) realized volatility in commodity futures markets will lead to low (high) permanent volatility in this time. Lastly, the θ_2 coefficients for ID-EMV, which we are most concerned about, are all positively significant at the 1% level, showing high agreement with the previous findings in Tables 2–6. These results reveal that even after controlling the effects of the past realized volatility, the infectious disease epidemic does have positive impacts on the long-run volatility in precious metal and crude oil futures markets. In addition, we also observe that ID-EMV has larger impacts (i.e., larger θ_2) on the permanent volatility in crude oil markets than it does in precious metal markets. Moreover, the impacts of ID-EMV on gold (WTI oil) is greater than those on silver (Brent oil), which are also highly consistent with the findings in Tables 2–5.

All in all, the empirical outcomes in this section are in excellent conformity with the previous results, providing confirmative evidence that infectious disease does have significant impacts on precious metal and crude oil futures markets, and these effects can last through 3 to 12 months.

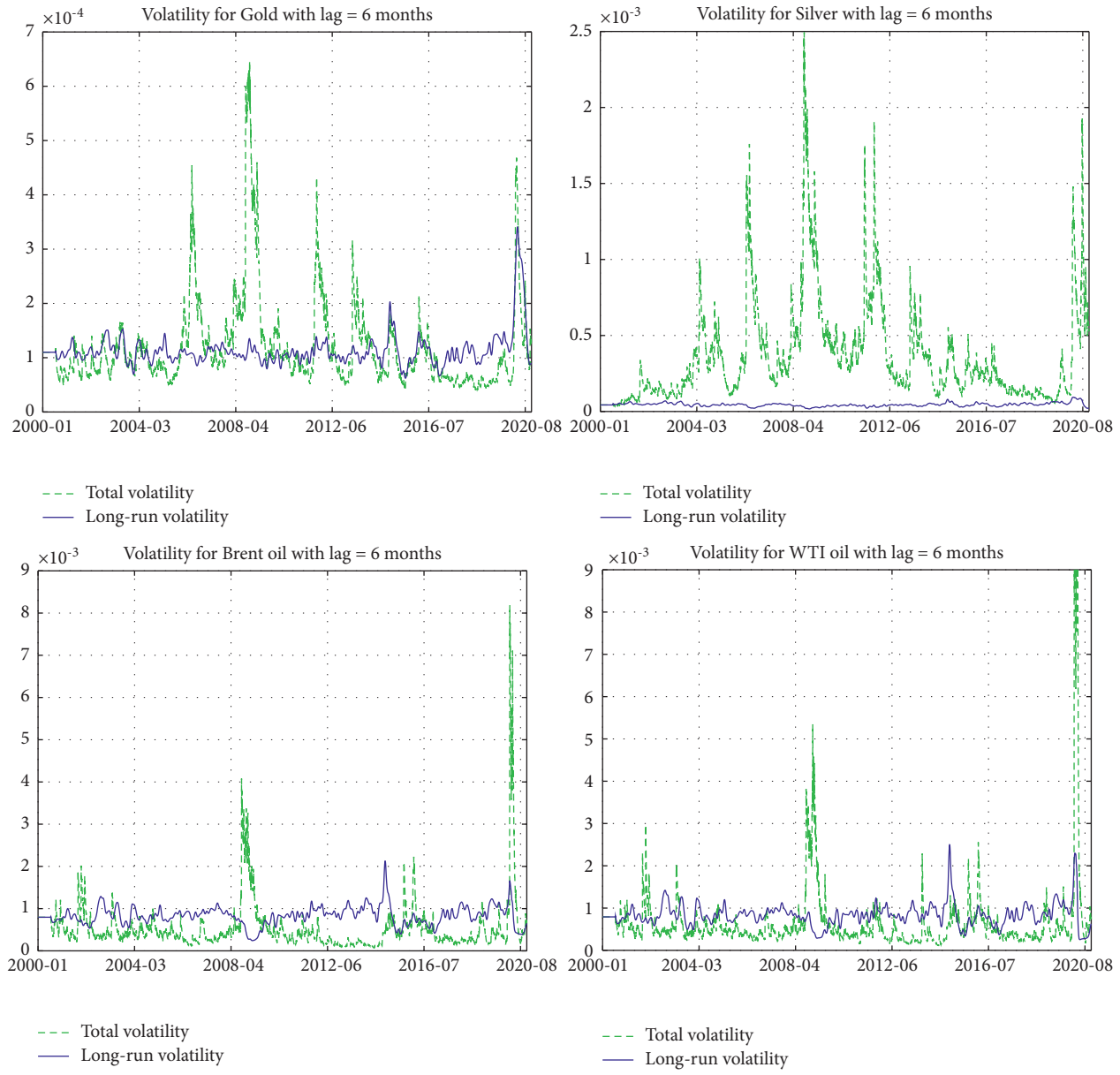


Figure 3: Estimated total and long-run volatility of precious metal and crude oil futures with GARCH-MIDAS + RV + ID-EMV models.

Table 6: Estimation results of GARCH-MIDAS + RV + ID-EMV models of futures volatility by infectious disease with 3-month lags.

Futures	ω	α	β	m	θ_1	θ_2	w_1	w_2	Log L
Gold	0.0003**	0.0309***	0.9609***	−9.1269***	18.3390	0.4661***	45.0800	2.1536***	15848.4
	(0.0001)	(0.0030)	(0.0041)	(0.0789)	(18.1600)	(0.0861)	(160.7300)	(0.5202)	
Silver	0.0000	0.0615***	0.9316***	−7.619***	−24.49***	0.2980***	6.5012***	2.8758***	13230.9
	(0.0002)	(0.0029)	(0.0032)	(0.1416)	(4.8943)	(0.0816)	(1.5580)	(0.8610)	
Brent	0.0001**	0.0745***	0.9207***	−7.0141***	−5.8417**	0.8973***	18.9300	1.4169***	12452.3
	(0.0000)	(0.0040)	(0.0043)	(0.2340)	(2.6558)	(0.1363)	(23.8830)	(0.1946)	
WTI	0.0005	0.0638***	0.9249***	−7.2011***	−4.6791***	1.0860***	1.0010***	1.4478***	11927.4
	(0.0003)	(0.0031)	(0.0035)	(0.1006)	(1.7145)	(0.1124)	(0.2983)	(0.1372)	

Note. The bracketed numbers are the standard errors of the estimations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Log L is the logarithmic likelihood of the estimation.

Table 7: Estimation results of GARCH-MIDAS + RV + ID-EMV models of futures volatility by infectious disease with 6-month lags.

Futures	ω	α	β	m	θ_1	θ_2	w_1	w_2	Log L
Gold	0.0003**	0.0296***	0.9626***	−9.1429***	−5.3522	1.6697***	5.0839	1.1631***	15632.3
	(0.0001)	(0.0024)	(0.0031)	(0.0899)	(24.1230)	(0.1654)	(32.9730)	(0.0718)	
Silver	0.0000	0.0369***	0.9631***	−9.8573***	−27.9310***	1.5958***	2.2015***	1.0046***	13003.8
	(0.0002)	(0.0015)	(0.0014)	(0.4055)	(6.9849)	(0.1702)	(0.5659)	(0.0398)	
Brent	0.0005**	0.0648***	0.9318***	−6.8889***	−24.7200***	1.8309***	1.5079***	1.3721***	12299.1
	(0.0003)	(0.0032)	(0.0034)	(0.3560)	(7.9097)	(0.2729)	(0.3790)	(0.1605)	
WTI	0.0005	0.0672***	0.9243***	−6.9706***	−14.8170***	2.2859***	1.6939***	1.4239***	11784.9
	(0.0003)	(0.0031)	(0.0031)	(0.1545)	(3.4674)	(0.2242)	(0.2911)	(0.1248)	

Note. The bracketed numbers are the standard errors of the estimations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Log L is the logarithmic likelihood of the estimation.

Table 8: Estimation results of GARCH-MIDAS + RV + ID-EMV models of futures volatility by infectious disease with 9-month lags.

Futures	ω	α	β	m	θ_1	θ_2	w_1	w_2	Log L
Gold	0.0003***	0.0259***	0.9741***	−9.5552***	−13.8000	1.4074***	5.6248	2.3628***	15374.3
	(0.0001)	(0.0016)	(0.0018)	(0.1906)	(25.2350)	(0.2258)	(12.8290)	(0.3957)	
Silver	0.0000	0.0403***	0.9597***	−9.1283***	−37.5240***	0.2559***	3.1900***	9.0347**	12756.8
	(0.0002)	(0.0015)	(0.0015)	(0.1605)	(8.9184)	(0.0822)	(0.8631)	(3.8279)	
Brent	0.0005*	0.0651***	0.9318***	−6.8409***	−20.6470*	1.2633***	2.2151**	3.2141**	12147.8
	(0.0003)	(0.0032)	(0.0035)	(0.3207)	(11.3380)	(0.4629)	(1.0360)	(1.2817)	
WTI	0.0005	0.0647***	0.9249***	−7.1367***	−11.2310**	1.6444***	2.3544***	3.3541***	11636.3
	(0.0003)	(0.0030)	(0.0033)	(0.1473)	(5.0179)	(0.2954)	(0.7674)	(0.6979)	

Note. The bracketed numbers are the standard errors of the estimations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Log L is the logarithmic likelihood of the estimation.

Table 9: Estimation results of GARCH-MIDAS + RV + ID-EMV models of futures volatility by infectious disease with 12-month lags.

Futures	ω	α	β	m	θ_1	θ_2	w_1	w_2	Log L
Gold	0.0003**	0.0378***	0.9540***	−8.9224***	−60.2300**	2.0407***	11.408	2.0497***	15171.8
	(0.0001)	(0.0030)	(0.0035)	(0.1138)	(27.8850)	(0.4188)	(7.8111)	(0.3817)	
Silver	0.0000	0.0406***	0.9594	−9.1784***	−37.1680***	0.2913***	4.5762***	12.369***	12537.9
	(0.0002)	(0.0015)	(0.0015)	(0.3925)	(9.3567)	(0.0828)	(1.3800)	(4.6568)	
Brent	0.0005**	0.0652***	0.9301***	−7.1647***	−14.927	1.1032***	3.2782	5.0775**	11998.7
	(0.0002)	(0.0003)	(0.0004)	(0.3600)	(13.4370)	(0.4235)	(2.6360)	(2.1906)	
WTI	0.0005	0.0655***	0.9234***	−7.1677***	−9.9499*	1.6611***	3.2375**	4.5244***	11490.1
	(0.0003)	(0.0031)	(0.0034)	(0.1528)	(5.4216)	(0.3340)	(1.3331)	(1.0596)	

Note. The bracketed numbers are the standard errors of the estimations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Log L is the logarithmic likelihood of the estimation.

6. Conclusions

Commodity futures markets are sensitive to many kinds of impacts, such as supply-demand shocks, geopolitical uncertainty, speculation surges, and public health emergencies. Especially, in the undergoing COVID-19 epidemic, commodity futures markets have experienced mind-bogglingly large swings and brought many troubles to policy makers, commodity productors and demanders, and investors. Therefore, this paper focuses on quantifying the impacts of the infectious disease epidemic on the permanent volatility of two import commodity futures, i.e., precious metal and crude oil.

The major empirical results show that the infectious disease epidemic does have positive impacts on the long-run volatilities of gold, silver, Brent, and WTI crude oil futures, and these effects can last from short term (3 months) to relative long term (12 months). Additionally, the influences of the infectious disease epidemic on crude oil futures are greater than those on precious metal for the reason that the former is innately tasked to be risk assets due to its tight relationships with industry and agriculture production, as well as transportation industry. Furthermore, we find that the infectious disease pandemic has larger impacts on gold (WTI oil) futures than that on silver (Brent oil) futures. All these findings are robust after controlling the effects of their own lagged realized volatilities.

The empirical results in this paper have several important policy and economic implications: firstly, infectious disease epidemic is recognized as a significant risk factor for commodity futures markets in this paper. Thus, policy makers should give it enough weight when making tailored regulatory policies in futures markets to avoid huge price volatilities and to stabilize market confidence. We also find

that these positive impacts of ID-EMV on commodity futures can last significantly as long as 12 months. In the consequence, policy makers should not withdraw their fiscal or monetary policies quickly in case it fails to meet their desired goals. Then, for commodity productors and demanders, they must pay keen attention to the trends of infectious disease epidemics and arrange raw material reserves, logistics scheduling, and production plans accordingly, which can enhance the security in their supply chain and capital chain effectively. Last, for investors in commodity futures markets, they have to know that long-run volatility in crude oil futures is easily affected by the infectious disease epidemic than precious futures. Therefore, they can tailor the risk exposure of their portfolio by rebalancing the weights in precious metal and oil assets during a pandemic. In addition, this paper shows that the impacts of ID-EMV on gold (WTI oil) are greater than those on silver (Brent oil). This result offers investors with options on different precious metal and crude oil assets to hedge the market risks stemming from the infectious disease epidemic. For investors, gold may be a better choice than silver to hedge the market risks in holding WTI crude oil, while silver is superior to gold in offsetting risks in Brent oil positions.

Data Availability

The precious metal and crude oil futures price is available at the Wind database, and the ID-EMV data are collected from the website http://www.policyuncertainty.com/infectious_EMV.html.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Y. S. proposed the original writing idea and finished the introduction section. X.-D. C. finished the empirical estimations and analyses of this paper. Y.-F. Z. prepared the methodology part. Y. W. supervised the writing process of this paper.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (71671145 and 71971191), Humanities and Social Science Fund of Ministry of Education of China (17YJA790015, 17XJA790002, 18YJC790132, and 18XJA790002), Science and Technology Innovation Team of Yunnan Provincial Universities (2019014), and Yunnan Fundamental Research Projects (202001AS070018).

References

[1] M. Ali, N. Alam, and S. A. R. Rizvi, "Coronavirus (COVID-19)—an epidemic or pandemic for financial markets," *Journal of Behavioral and Experimental Finance*, vol. 27, Article ID 100341, 2020.

[2] A. B. Amar, F. Belaid, A. B. Youssef, B. Chiao, and K. Guesmi, "The unprecedented reaction of equity and commodity markets to COVID-19," *Finance Research Letters*, vol. 38, Article ID 101853, 2020.

[3] E. Bouri, O. Cepni, D. Gabauer, and R. Gupta, "Return connectedness across asset classes around the COVID-19 outbreak," *International Review of Financial Analysis*, vol. 73, Article ID 101646, 2020.

[4] S. Corbet, Y. Hou, Y. Hu, L. Oxley, and D. Xu, "Pandemic-related financial market volatility spillovers: evidence from the Chinese COVID-19 epicentre," *International Review of Economics & Finance*, vol. 71, pp. 55–81, 2021.

[5] S. Corbet, C. Larkin, and B. Lucey, "The contagion effects of the COVID-19 pandemic: evidence from gold and cryptocurrencies," *Finance Research Letters*, vol. 35, Article ID 101554, 2020.

[6] A. P. Fassas, "Risk aversion connectedness in developed and emerging equity markets before and after the COVID-19 pandemic," *Heliyon*, vol. 6, no. 12, Article ID e05715, 2020.

[7] C. Gharib, S. Mefteh-Wali, and S. B. Jabeur, "The bubble contagion effect of COVID-19 outbreak: evidence from crude oil and gold markets," *Finance Research Letters*, vol. 38, Article ID 101703, 2020.

[8] S. Gunay, "Comparing COVID-19 with the GFC: a shockwave analysis of currency markets," *Research in International Business and Finance*, vol. 56, Article ID 101377, 2021.

[9] N. Iqbal, Z. Fareed, G. Wan, and F. Shahzad, "Asymmetric nexus between COVID-19 outbreak in the world and cryptocurrency market," *International Review of Financial Analysis*, vol. 73, Article ID 101613, 2021.

[10] T. H. Le, H. X. Do, D. K. Nguyen, and A. Sensoy, "Covid-19 pandemic and tail-dependency networks of financial assets," *Finance Research Letters*, vol. 38, Article ID 101800, 2020.

[11] N. Norouzi, G. Zarazua de Rubens, S. Choupanpiesheh, and P. Enevoldsen, "When pandemics impact economies and climate change: exploring the impacts of COVID-19 on oil and electricity demand in China," *Energy Research & Social Science*, vol. 68, Article ID 101654, 2020.

[12] M. Omane-Adjepong and I. P. Alagidede, "Exploration of safe havens for Africa's stock markets: a test case under COVID-19 crisis," *Finance Research Letters*, vol. 38, Article ID 101877, 2020.

[13] J. Sadefo Kamdem, R. Bandolo Essomba, and J. Njong Berinyuy, "Deep learning models for forecasting and analyzing the implications of COVID-19 spread on some commodities markets volatilities," *Chaos, Solitons & Fractals*, vol. 140, Article ID 110215, 2020.

[14] Y. Sakurai and T. Kurosaki, "How has the relationship between oil and the US stock market changed after the COVID-19 crisis?" *Finance Research Letters*, vol. 37, Article ID 101773, 2020.

[15] A. H. Samadi, S. Owjimehr, and Z. Nezhad Halafi, "The cross-impact between financial markets, Covid-19 pandemic, and economic sanctions: the case of Iran," *Journal of Policy Modeling*, vol. 43, pp. 34–55, 2020.

[16] A. Sharif, C. Aloui, and L. Yarovaya, "COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: fresh evidence from the wavelet-based approach," *International Review of Financial Analysis*, vol. 70, Article ID 101496, 2020.

[17] J. Wang and X. Wang, "COVID-19 and financial market efficiency: evidence from an entropy-based analysis," *Finance Research Letters*, Article ID 101888, 2021.

[18] O. B. Adekoya, J. A. Oliyide, and G. O. Oduyemi, "How COVID-19 upturns the hedging potentials of gold against oil and stock markets risks: nonlinear evidences through threshold regression and Markov-regime switching models," *Resources Policy*, vol. 70, Article ID 101926, 2020.

[19] A. Dutta, D. Das, R. K. Jana, and X. V. Vo, "COVID-19 and oil market crash: revisiting the safe haven property of gold and bitcoin," *Resources Policy*, vol. 69, Article ID 101816, 2020.

[20] W. Mensi, A. Sensoy, X. V. Vo, and S. H. Kang, "Impact of COVID-19 outbreak on asymmetric multifractality of gold and oil prices," *Resources Policy*, vol. 69, Article ID 101829, 2020.

[21] A. A. Salisu, X. V. Vo, and A. Lawal, "Hedging oil price risk with gold during COVID-19 pandemic," *Resources Policy*, vol. 70, Article ID 101897, 2020.

[22] P. Li and Z. Dong, "Time-varying network analysis of fluctuations between crude oil and Chinese and U.S. gold prices in different periods," *Resources Policy*, vol. 68, Article ID 101749, 2020.

[23] K. Mokni, S. Hammoudeh, A. N. Ajmi, and M. Youssef, "Does economic policy uncertainty drive the dynamic connectedness between oil price shocks and gold price?" *Resources Policy*, vol. 69, Article ID 101819, 2020.

[24] K. Morema and L. Bonga-Bonga, "The impact of oil and gold price fluctuations on the South African equity market: volatility spillovers and financial policy implications," *Resources Policy*, vol. 68, Article ID 101740, 2020.

[25] A. A. Salisu and I. Adediran, "Gold as a hedge against oil shocks: evidence from new datasets for oil shocks," *Resources Policy*, vol. 66, Article ID 101606, 2020.

[26] K. H. Al-Yahyaee, W. Mensi, A. Sensoy, and S. H. Kang, "Energy, precious metals, and GCC stock markets: is there any risk spillover?" *Pacific-Basin Finance Journal*, vol. 56, pp. 45–70, 2019.

[27] K. H. Al-Yahyaee, M. U. Rehman, I. M. Wanas Al-Jarrah, W. Mensi, and X. V. Vo, "Co-movements and spillovers between prices of precious metals and non-ferrous metals: a multiscale analysis," *Resources Policy*, vol. 67, Article ID 101680, 2020.

[28] M. Balcilar and Z. A. Ozdemir, "The volatility effect on precious metals price returns in a stochastic volatility in mean model with time-varying parameters," *Physica A: Statistical Mechanics and Its Applications*, vol. 534, Article ID 122329, 2019.

[29] K. Guhathakurta, S. R. Dash, and D. Maitra, "Period specific volatility spillover based connectedness between oil and other commodity prices and their portfolio implications," *Energy Economics*, vol. 85, Article ID 104566, 2020.

[30] L. Hau, H. Zhu, R. Huang, and X. Ma, "Heterogeneous dependence between crude oil price volatility and China's agriculture commodity futures: evidence from quantile-on-quantile regression," *Energy*, vol. 213, Article ID 118781, 2020.

[31] H. Hu, D. Chen, B. Sui, L. Zhang, and Y. Wang, "Price volatility spillovers between supply chain and innovation of financial pledges in China," *Economic Modelling*, vol. 89, pp. 397–413, 2020.

[32] S. Husain, A. K. Tiwari, K. Sohag, and M. Shahbaz, "Connectedness among crude oil prices, stock index and metal prices: an application of network approach in the USA," *Resources Policy*, vol. 62, pp. 57–65, 2019.

[33] Y. Jiang, Y. Fu, and W. Ruan, "Risk spillovers and portfolio management between precious metal and BRICS stock markets," *Physica A: Statistical Mechanics and Its Applications*, vol. 534, Article ID 120993, 2019.

[34] M. C. K. Lau, S. A. Vigne, S. Wang, and L. Yarovaya, "Return spillovers between white precious metal ETFs: the role of oil, gold, and global equity," *International Review of Financial Analysis*, vol. 52, pp. 316–332, 2017.

[35] Y. Li, J. Huang, and J. Chen, "Dynamic spillovers of geopolitical risks and gold prices: new evidence from 18 emerging economies," *Resources Policy*, vol. 70, Article ID 101938, 2021.

[36] T. Liu and X. Gong, "Analyzing time-varying volatility spillovers between the crude oil markets using a new method," *Energy Economics*, vol. 87, Article ID 104711, 2020.

[37] W. Mensi, K. H. Al-Yahyaee, and S. Hoon Kang, "Time-varying volatility spillovers between stock and precious metal markets with portfolio implications," *Resources Policy*, vol. 53, pp. 88–102, 2017.

[38] W. Mensi, S. Hammoudeh, M. U. Rehman, A. A. S. Al-Maadid, and S. Hoon Kang, "Dynamic risk spillovers and portfolio risk management between precious metals and global foreign exchange markets," *The North American Journal of Economics and Finance*, vol. 51, Article ID 101086, 2020.

[39] W. Mensi, J. A. Hernandez, S. M. Yoon, X. V. Vo, and S. H. Kang, "Spillovers and connectedness between major precious metals and major currency markets: the role of frequency factor," *International Review of Financial Analysis*, vol. 74, Article ID 101672, 2021.

[40] W. Mensi, M. U. Rehman, and X. V. Vo, "Spillovers and co-movements between precious metals and energy markets: implications on portfolio management," *Resources Policy*, vol. 69, Article ID 101836, 2020.

[41] W. Mensi, A. Sensoy, A. Aslan, and S. H. Kang, "High-frequency asymmetric volatility connectedness between Bitcoin and major precious metals markets," *The North American Journal of Economics and Finance*, vol. 50, Article ID 101031, 2019.

[42] J. C. Reboredo and A. Ugolini, "Price spillovers between rare earth stocks and financial markets," *Resources Policy*, vol. 66, Article ID 101647, 2020.

[43] M. U. Rehman, "Do bitcoin and precious metals do any good together? An extreme dependence and risk spillover analysis," *Resources Policy*, vol. 68, Article ID 101737, 2020.

[44] S. J. H. Shahzad, M. U. Rehman, and R. Jammazi, "Spillovers from oil to precious metals: quantile approaches," *Resources Policy*, vol. 61, pp. 508–521, 2019.

[45] A. K. Tiwari, B. R. Mishra, and S. A. Solarin, "Analysing the spillovers between crude oil prices, stock prices and metal prices: the importance of frequency domain in USA," *Energy*, vol. 220, Article ID 119732, 2021.

[46] G. S. Uddin, J. A. Hernandez, S. J. H. Shahzad, and S. H. Kang, "Characteristics of spillovers between the US stock market and precious metals and oil," *Resources Policy*, vol. 66, Article ID 101601, 2020.

[47] G. S. Uddin, S. J. H. Shahzad, G. Boako, J. A. Hernandez, and B. M. Lucey, "Heterogeneous interconnections between precious metals: evidence from asymmetric and frequency-domain spillover analysis," *Resources Policy*, vol. 64, Article ID 101509, 2019.

[48] D. C. Yildirim, E. I. Cevik, and O. Esen, "Time-varying volatility spillovers between oil prices and precious metal prices," *Resources Policy*, vol. 68, Article ID 101783, 2020.

[49] C. Zhang, X. Shi, and D. Yu, "The effect of global oil price shocks on China's precious metals market: a comparative analysis of gold and platinum," *Journal of Cleaner Production*, vol. 186, pp. 652–661, 2018.

[50] F. X. Diebold and K. Yilmaz, "Better to give than to receive: predictive directional measurement of volatility spillovers," *International Journal of Forecasting*, vol. 28, no. 1, pp. 57–66, 2012.

[51] R. Seyedeh Fatemeh, B. Mehdi, B. Bahareh Ramezanian, and S. M. J. Razmi, "The impact of US monetary policy uncertainties on oil and gas return volatility in the futures and spot markets," *Journal of Petroleum Science and Engineering*, vol. 191, Article ID 107232, 2020.

[52] Y. Wei, S. Qin, X. Li, S. Zhu, and G. Wei, "Oil price fluctuation, stock market and macroeconomic fundamentals: evidence from China before and after the financial crisis," *Finance Research Letters*, vol. 30, pp. 23–29, 2019.

[53] Y. Zhang, Y. Wei, Y. Zhang, and D. Jin, "Forecasting oil price volatility: forecast combination versus shrinkage method," *Energy Economics*, vol. 80, pp. 423–433, 2019.

[54] R. F. Engle and J. G. Rangel, "The spline-GARCH model for low-frequency volatility and its global macroeconomic causes," *Review of Financial Studies*, vol. 21, no. 3, pp. 1187–1222, 2008.

[55] R. F. Engle, E. Ghysels, and B. Sohn, "Stock market volatility and macroeconomic fundamentals," *Review of Economics and Statistics*, vol. 95, no. 3, pp. 776–797, 2013.

[56] F. Corsi, "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, vol. 7, pp. 174–196, 2009.

[57] S. Baker, B. Nicholas, J. D. Steven, K. Kyle, S. Marco, and V. Tasaneeya, "The unprecedented stock market reaction to COVID-19," *Covid Economics: Vetted and Real-Time Papers*, vol. 1, 2020.

[58] A. K. Pradhan, B. R. Mishra, A. K. Tiwari, and S. Hammoudeh, "Macroeconomic factors and frequency domain causality between gold and silver returns in India," *Resources Policy*, vol. 68, Article ID 101744, 2020.

[59] T. L. D. Huynh, M. A. Nasir, V. X. Vo, and T. T. Nguyen, ""Small things matter most": the spillover effects in the cryptocurrency market and gold as a silver bullet," *North American Journal of Economics & Finance*, vol. 54, Article ID 101277, 2020.

[60] L. Bai, Y. Wei, G. Wei, X. Li, and S. Zhang, "Infectious disease pandemic and permanent volatility of international stock markets: a long-term perspective," *Finance Research Letters*, vol. 40, Article ID 101709, 2021.