


## Article

# Dynamics of Information Flow between the Chinese A-Share Market and the U.S. Stock Market: From the 2008 Crisis to the COVID-19 Pandemic Period

Chun-Xiao Nie <sup>1,2,\*</sup>  and Jing Xiao <sup>1</sup><sup>1</sup> School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China<sup>2</sup> Collaborative Innovation Center of Statistical Data Engineering, Technology & Application, Zhejiang Gongshang University, Hangzhou 310018, China

\* Correspondence: niechunxiao2009@163.com

**Abstract:** The relationship between the Chinese market and the US market is widely concerned by researchers and investors. This paper uses transfer entropy and local random permutation (*LRP*) surrogates to detect the information flow dynamics between two markets. We provide a detailed analysis of the relationship between the two markets using long-term daily and weekly data. Calculations show that there is an asymmetric information flow between the two markets, in which the US market significantly affects the Chinese market. Dynamic analysis based on weekly data shows that the information flow evolves, and includes three significant periods between 2004 and 2021. We also used daily data to analyze the dynamics of information flow in detail over the three periods and found that changes in the intensity of information flow were accompanied by major events affecting the market, such as the 2008 financial crisis and the COVID-19 pandemic period. In particular, we analyzed the impact of the S&P500 index on different industry indices in the Chinese market and found that the dynamics of information flow exhibit multiple patterns. This study reveals the complex information flow between two markets from the perspective of nonlinear dynamics, thereby helping to analyze the impact of major events and providing quantitative analysis tools for investment practice.

**Keywords:** stock market; transfer entropy; information flow; surrogates

**Citation:** Nie, C.-X.; Xiao, J. Dynamics of Information Flow between the Chinese A-Share Market and the U.S. Stock Market: From the 2008 Crisis to the COVID-19 Pandemic Period. *Entropy* **2022**, *24*, 1102. <https://doi.org/10.3390/e24081102>

Academic Editors: Átila Madureira Bueno, Diego Colón, Jose Manoel Balthazar and Angelo Marcelo Tusset

Received: 13 July 2022

Accepted: 9 August 2022

Published: 10 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Both China and the United States are large economies with high influence, so analyzing the relationship between the two stock markets is of great significance for investment practice and risk management. Currently, there are many studies analyzing the relationship between the two markets based on different datasets, in which multiple empirical tools are used. For example, Goh et al. found that some economic variables in the United States can improve the forecast performance of the Chinese stock market [1]. Jian Chen et al. found that US economic variables can help predict the monthly volatility of the Chinese market [2]. In addition, previous studies have also shown that US economic policy uncertainty can significantly explain the returns of China's A-share market [3]. Researchers not only pay attention to how economic variables and policies in the US market affect the Chinese market but also directly discuss the relationship between the returns of the two markets. These studies focus on the impact of special events on the relationship between markets, such as the 2007–2009 crisis [4–10], the Eurozone crisis [11], and the COVID-19 pandemic [12–14]. Empirical analysis shows that the financial crisis has had a significant impact on the BRICS and emerging markets, such as leading to significant risk spillover effects and risk contagion [5–7]. In particular, the Chinese market was significantly affected by the US market during the crisis [4–6,8,10,11]. These studies used many types of quantitative tools, such as multivariate GARCH models [4,6,10], Granger causality test [5],

latent factor models [9], correlation contagion tests [8], etc. In particular, recently, the COVID-19 pandemic has significantly impacted the Chinese market and the US market [12]. For example, research reveals that China's stock market and the US stock market have significant spillover effects during the pandemic [13]. In addition, the pandemic has also impacted the correlation structure between market indexes [14].

Existing studies have revealed that there are extensive links between markets, and the Chinese market is significantly affected by the US market, especially under the impact of major events. However, previous studies have not fully analyzed the long-term relationship between the two markets. For example, most studies only analyze the static relationship in a specified period but do not focus on dynamics. This study analyzes the long-term relationship between the two markets from the perspective of nonlinear information flow dynamics. We use transfer entropy ( $TE$ ) to quantitatively characterize nonlinear relationships [15], and analyze the dynamics in detail by surrogate time series [16]. Surrogates is an effective technique for analyzing nonlinear time series, which generates simulated time series by keeping some characteristics of the original time series [17–19].

At present, transfer entropy has expanded into a “toolbox” that includes multiple tools. For example, Marschinski et al. introduced effective transfer entropy by shuffling the source time series, where the shuffling process is used to destroy the correlation [20]. Further, the group transfer entropy and the effective group transfer entropy can be constructed by considering multiple time series [21]. Jizba et al. constructed the Rényiian transfer entropies by extending Shannon entropy to Rényi entropy [22]. Papanas et al. extended the transfer entropy method to analyze non-stationary time series [23]. Recently, Nie proposed a method to detect local information flow [16]. This method obtains the baseline  $TE$  value from local surrogates so that it is possible to observe the contribution of observations within a small computational window to the original  $TE$  value. An advantage of this method is that changes in the intensity of the information flow can be clearly observed, so that we can identify the impact of events on the information flow [16].

Since transfer entropy is a non-parametric method that effectively extracts the causal relationship between variables, it is suitable for analyzing high-complexity financial time series. For example, methods based on transfer entropy have been widely used to analyze financial time series, such as the relationship between stock markets [24–28], the price—volume relationship [16,29–31], and the foreign exchange market [32]. In particular, early empirical research shows that the US market is a core source of information for global financial markets [28]. Kim et al. considered 10 important indexes and found that the structure of the information transfer network was affected by the crisis [33].

This study focuses on analyzing the evolutionary characteristics of the information flow between China's A-share market and the US market. We use nonlinear analysis methods to characterize the complex information flow dynamics between the two markets and, in particular, identify periods of high information flow intensity by  $TE$  and  $LRP$ . This study reveals the details of information transfer between markets based on long-term data and from multiple time scales. In the following, we first describe the data set and analyze basic descriptive statistics. Secondly, we review the concept and calculation method of transfer entropy, and the steps of  $LRP$ -based analysis. Third, we use weekly data and daily data to analyze the information flow between the two markets. Finally, we analyze the information flow between the S&P500 index and the industry indices of the Chinese A-share market.

## 2. Data

This paper uses three datasets, which include closing indices for two market composite indices and some industry indices. The composite index data includes daily and weekly closing data of S&P500 and CSI300 indices, as well as weekly closing data of some industry indices in the Chinese A-share market. All data are extracted from the iFinD database of Tonghuashun company. For the original closing index time series  $P = \{P(t)\}$ , we calculate the series  $R = \{R(t)\}$ , where  $R(t) = \log(P(t+1)) - \log(P(t))$ .

For weekly data, we use  $R_{us}$  and  $R_{cn}$  to represent the return series of the S&P500 index and the CSI300 Index, respectively. We considered the data from 31 December 2004 to 17 June 2022, a total of 890 observations, so that each preprocessed series includes 889 observations. Table 1 lists some descriptive statistics of the two series. In addition, the Pearson correlation coefficient of the two-time series is 0.2118, which implies a weak correlation.

**Table 1.** Descriptive statistics of preprocessed time series

Series	Average	Std	Skewness	Kurtosis
$R_{us}$	0.0012	0.0261	−2.0557	23.1000
$R_{cn}$	0.0016	0.0359	−0.2955	2.2400
$R_{us}^1$	$-1.2752 \times 10^{-4}$	0.0179	−0.4871	9.0371
$R_{cn}^1$	$4.3938 \times 10^{-4}$	0.0242	−0.5126	2.0974
$R_{us}^2$	$3.9919 \times 10^{-4}$	0.0077	−0.2361	4.2776
$R_{cn}^2$	$5.8243 \times 10^{-4}$	0.0161	−0.9794	6.7738
$R_{us}^3$	$6.0558 \times 10^{-4}$	0.0138	−1.0405	17.2147
$R_{cn}^3$	$2.0142 \times 10^{-4}$	0.0134	−0.3052	3.6399
$R_{ind}^1$	0.0007	0.0437	−0.2044	2.0224
$R_{ind}^2$	0.0016	0.0460	−0.3658	2.0825
$R_{ind}^3$	0.0007	0.0391	−0.3157	1.7477

In addition to the weekly data, we also analyzed the daily data of the two indexes. To clearly analyze the impact of major events, we divide the daily data into three periods. For the preprocessed index data, we use  $R_{cn}^1$  and  $R_{us}^1$  to represent the data of the first period and use similar symbols for other periods. The time intervals corresponding to the three data sets are 4 January 2007–31 December 2010, 2 January 2014–29 December 2017, and 2 January 2018–31 December 2021 respectively. Since the trading days of the two markets are not synchronized, we remove the missing values in the two series, leaving observations with the same time stamp. The time series of the three periods without missing values include 944, 948, and 942 observations, respectively. Table 1 lists the basic descriptive statistics of each return series. Here, we always use the rounding method to keep four decimal places.

We can find from Table 1 that all the averages are close to zero. In three of the four pairs of time series, the standard deviation of the series in the Chinese market is greater than that in the US market. In addition, all skewness values are less than zero, and in particular, all excess kurtosis values are greater than zero. Here, we report the excess kurtosis obtained by subtracting 3 from the original kurtosis value. The kurtosis values imply that all distributions have peak characteristics and are not normal.

In addition to the market composite index, we used 31 industry indices compiled by SWS Research Co., Ltd. Here, we used three of them for dynamic analysis. Table 1 lists the descriptive statistics of the return series for the three indices. The symbols  $R_{ind}^1$ ,  $R_{ind}^2$  and  $R_{ind}^3$  correspond to the industry index codes, 801040.SL (Steel industry), 801180.SL (Real estate) and 801960.SL (Petrochemical industry), respectively. It can be found that the average of  $R_{ind}^2$  is significantly larger than the other two indices, while the difference between the standard deviations is small. In particular, all skewness values are less than 0, and kurtosis values are greater than 0.

### 3. Method

#### 3.1. Transfer Entropy

We consider two random variables  $I$  and  $J$ , the marginal probability distribution is  $p_I(\cdot)$  is  $p_J(\cdot)$ , and the joint distribution is  $p_{I,J}(\cdot)$ . We assume that the underlying dynamic structure conforms to a stationary Markov process, where  $I$  and  $J$  correspond to orders  $k$  and  $l$ , respectively. For  $I$ , the conditional distribution of state at time  $t + 1$  ( $i_{t+1}$ ) is independent of state in  $t - k$  ( $i_{t-k}$ ), that is,  $p_I(i_{t+1}|i_t, \dots, i_{n-k+1}) = p_I(i_{t+1}|i_t, \dots, i_{n-k})$ . Similarly, for  $J$ ,  $p_J(j_{t+1}|j_t, \dots, j_{n-k+1}) = p_J(j_{t+1}|j_t, \dots, j_{n-k})$ , where  $j_t$  represents the state

at time  $t$ . The transfer entropy of  $J \rightarrow I$  is defined as Equation 1 [15], where the symbols  $i_t^k = (i_t, \dots, i_{n-k+1})$  and  $j_t^k = (j_t, \dots, j_{n-k+1})$ . In this article, we set  $l = k = 1$ .

$$TE_{J \rightarrow I} = \sum p(i_{t+1}, i_t^k, j_t^l) \log \frac{p(i_{t+1} | i_t^k, j_t^l)}{p(i_{t+1} | i_t^k)} \tag{1}$$

### 3.2. The Method Used to Estimate the Transfer Entropy

For a time series, we can encode it into a time series including states and then estimate the transfer entropy. Here, we use a method proposed in previous studies to encode the time series [25,26,34], which encodes the original time series through a group of quantiles. One of the advantages of this method is that a large weight can be assigned to the tail of the distribution, thereby focusing on the influence of extreme values in the financial market [25,26]. In addition, we use the R package RTransferEntropy developed by Behrendt et al. [34].

If the time series  $s = \{s(t)\}$  is coded into  $n$  states,  $n - 1$  quantiles  $Q = \{q_1, q_2, \dots, q_{n-1}\}$  are required, as shown in Equation (2). For example, if the observation  $s(t)$  is between the quantiles  $q_2$  and  $q_3$  of the distribution, then  $s'(t) = 2$ . The converted time series  $s' = \{s'(t)\}$  includes  $n$  states, and can be used for conditional distribution and joint distribution.

$$s'(t) = \begin{cases} 1 & s(t) \leq q_1 \\ 2 & s(t) \in (q_2, q_3) \\ \dots & \\ n - 1 & s(t) \in (q_{n-2}, q_{n-1}) \\ n & s(t) \geq q_{n-1} \end{cases} \tag{2}$$

In order to comprehensively analyze the significance of information flow at different quantile parameters, we set the following five sets of parameters:  $Q_1 = \{q(0.05), q(0.5), q(0.95)\}$ ,  $Q_2 = \{q(0.10), q(0.5), q(0.90)\}$ ,  $Q_3 = \{q(0.15), q(0.5), q(0.85)\}$ ,  $Q_4 = \{q(0.20), q(0.5), q(0.80)\}$  and  $Q_5 = \{q(0.25), q(0.5), q(0.75)\}$ , where the symbol  $q(\alpha)$  represents  $\alpha$ -quantile. From the first to the fifth set of parameters, the weight of the extreme values of the tail decreases. In the first group, extreme values are given the greatest weight, and the distribution in the fifth group is equally divided into four parts.

Here, we use the method proposed by Dimpfl et al. to test the significance of  $TE$  [25,34]. The method first estimates the dynamics between two variables through the Markov process. Secondly, based on the Markov process, a pair of simulated time series whose dependence has been eliminated can be generated, and the  $TE$  value can be estimated. Third, we repeatedly generate the simulated time series  $N_{boot}$  times to obtain a  $TE$  value distribution for comparison. Finally, the  $p$ -value is generated by comparing the original  $TE$  value with the quantile of the baseline distribution. In this article, we always set  $N_{boot} = 1000$ .

### 3.3. LRP-Based Analysis

The local random permutation (LRP) method uses  $RP$  surrogates locally to estimate the intensity of the information flow within a period [16]. Previous studies have confirmed that this method can effectively describe changes in information flow through toy models [16]. In particular, to maintain the correlation between series,  $LRP$  provides surrogates that keep the Pearson correlation unchanged.

We consider the information flow  $s_1 \rightarrow s_2$  between the series  $s_1 = \{s_1(t)\}$  and  $s_2 = \{s_2(t)\}$ , where each series includes  $N$  observations. We assume that the length of the calculation window for constructing local  $RP$  surrogates is  $L_w$ , and the  $TE_{s_1 \rightarrow s_2}$  value is significant.

- 1 We extract the observations in the time interval  $[1, L_w]$  to obtain the sub-time series  $s_1^{[1, L_w]} = \{s_1(1), \dots, s_1(L_w)\}$  and  $s_2^{[1, L_w]} = \{s_2(1), \dots, s_2(L_w)\}$ , and shuffle the observations of the sub-sequences  $s_1^{[1, L_w]}$  and  $s_2^{[1, L_w]}$ . In the scrambling step, the cor-

- responsedence between the timestamps of the two series does not change, that is, the series  $\{(s_1(k), s_2(k))\}$  composed of pairs is scrambled.
- 2 We calculate the transfer entropy value between the time series obtained by local scrambling in step 1, and denoted it as  $TE_{s_1 \rightarrow s_2}^{[1, L_w]}(1)$ .
  - 3 We repeat steps 1 and 2 a total of  $M$  times to obtain a distribution of  $TE$  values used as a benchmark ( $TE_{ben}^1 = \{TE_{s_1 \rightarrow s_2}^{[1, L_w]}(k) | k = 1, 2, \dots, M\}$ ).
  - 4 We calculate the Z-score  $Z_{[1, L_w]} = \frac{TE - m_{[1, L_w]}}{s_{[1, L_w]}}$ .
  - 5 We move the calculation window one observation each time, and repeat steps 1–4 to get the distribution  $TE_{ben}^i = \{TE_{s_1 \rightarrow s_2}^{[i, L_w]}(k) | k = 1, 2, \dots, M\}$  and the Z-score. For example, the second Z-value is obtained by scrambling all observations in the interval  $[2, L_w + 1]$ . Finally, we get the Z-score sequence  $\{Z_{[i, i+L_w-1]} | i = 1, 2, \dots, N - L_w + 1\}$ .

The estimation of the  $TE$  value requires setting the parameter  $Q_k$ , so the parameter may affect the calculation result. Here, we use the following method to synthesize the calculation results of different parameters. For a sequence of Z-scores  $\{z_k(t)\}$  generated by a parameter  $Q_k$ , we construct a 0–1 sequence using Equation (3), where  $Z_{th}$  is the threshold. We set the threshold  $Z_{th} = 1.645$  here. In this way, we obtain a series  $I_k = \{I_k(t) | t = 1, 2, \dots, N - L_w + 1\}$  based on Z-scores, in which observations only take values in  $\{0, 1\}$ . If  $TE_{ben}^i$  is a normal distribution, then the original  $TE$  value is greater than most of the values (95%) in  $TE_{ben}^i$ , which implies that there is a significant information flow in the period  $[i, i + L_w - 1]$ . In this article, we always set  $M = 1000$ . Then, we calculate the series  $I = \{I(t)\}$  ( $I(t) = \sum_k I_k(t)$ ).

$$I_k(t) = \begin{cases} 1, & z_k(t) \geq z_{th} \\ 0, & z_k(t) < z_{th} \end{cases} \tag{3}$$

Since each Z-score corresponds to an interval, in this article, we use the following convention to plot the figure: the right end of the interval corresponding to the value is used as the time stamp in the figure. For example, for the value  $Z_{[1, L_w]}$ , the time label in the figure is the timestamp corresponding to  $L_w$ . Similarly, we plot the time series  $\{I(t)\}$ .

### 4. Results

#### 4.1. Information Flow Analysis Based on Weekly Data

##### 4.1.1. TE Values of Different Parameters

We use a group of parameters  $Q_k$  ( $k = 1, \dots, 5$ ) to calculate the information flow between  $R_{us}$  and  $R_{cn}$ . Table 2 shows the calculated results, all  $TE_{us \rightarrow cn}$  values are significant at the significance level  $\alpha = 0.05$ , and three of the  $p$ -values are less than 0.01. In addition, all  $TE_{cn \rightarrow us}$  values are not significant. Table 2 suggests that there is an asymmetric information flow between the Chinese market and the US market, and the US market is a dominant information source.

**Table 2.** Information flow analysis of long-term weekly data.

$Q_k$	$TE_{cn \rightarrow us}$	$p_{cn \rightarrow us}$	$TE_{us \rightarrow cn}$	$p_{us \rightarrow cn}$
$Q_1$	0.040537	0.054	0.048419	0.006
$Q_2$	0.0352622	0.261	0.0519378	0.005
$Q_3$	0.0381743	0.141	0.0449418	0.039
$Q_4$	0.0357750	0.183	0.0492554	0.008
$Q_5$	0.047230	0.014	0.044303	0.031

The minimum  $p$ -value of 0.005 corresponds to the parameter  $Q_2$ , which means that significant information flow can be detected when a large weight is assigned to extreme values. In addition, the  $p$ -value corresponding to the equally divided quantile interval is

also less than 0.05. This means that significant  $us \rightarrow cn$  information flow can be detected regardless of whether a greater weight is given to the extreme values.

In summary, there is sufficient evidence to support the existence of information flow  $us \rightarrow cn$ . The change of the  $p$ -value with the change of the parameter implies that the analysis of the information flow depends on the choice of the parameter, and thus the calculation results of multiple parameters need to be integrated.

#### 4.1.2. Information Flow Dynamics for Weekly Data

In the previous section, we analyzed the information flow between the two markets globally. However, the static analysis only shows the existence of information flow in the considered period, thus lacking local dynamic analysis. Below, we use *LRP* to analyze the dynamics of information flow, in particular, to detect periods with localized strong information flow.

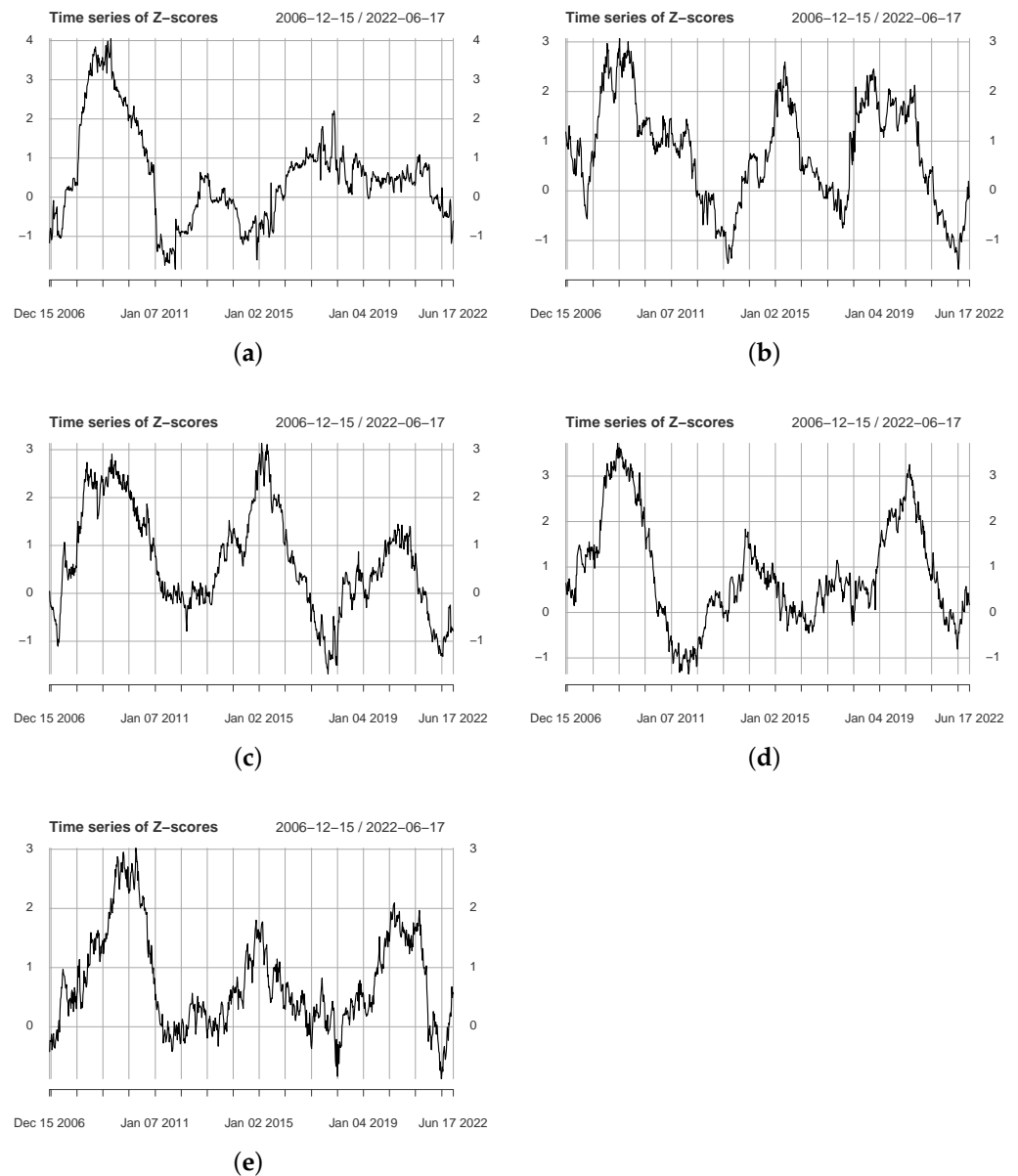
We set five sets of parameters  $Q_k$  ( $k = 1, \dots, 5$ ) and  $L_w = 96$ , where the calculation window corresponds to two years of trading weeks. The five subfigures of Figure 1 show the  $Z$ -score time series. We observe that all subfigures include local peaks larger than 1.645, implying non-trivial information flow dynamics between series. These peaks correspond to periods of localized strong information flow. For example, for Figure 1a, it can be seen that the  $Z$ -score increased rapidly in 2008 and dropped significantly thereafter. Furthermore, we can also observe that the number of peaks in all  $Z$ -score series is greater than 1. For example, three localized peaks can be observed in Figure 1b, implying the existence of three periods of strong information flow.

The numbers of local peaks for the five  $Z$ -score series are 2, 3, 2, 3, and 3, respectively. Table 3 lists the periods corresponding to these peaks. We use  $T_1$ ,  $T_2$ , and  $T_3$  to represent the time intervals in three periods, where each period corresponds to two endpoints. For example, the starting time corresponding to  $T_1$  ( $Q_1$ ) is 8 June 2007, and the closing time is 24 April 2009. The first line indicates that there is a significant flow of information in the period 8 June 2007–24 April 2009. We list the  $Z$ -score at the corresponding time interval in the fourth column. The maximum value corresponds to the  $T_1$  period, while the minimum value corresponds to the  $T_2$  period.

**Table 3.** Three significant periods of the information flow of parameter.

$Q$	$T_i$	$t$	$Z$	Period
$Q_1$	$T_1$	24 April 2009	4.0550	8 June 2007–24 April 2009
$Q_1$	$T_2$	17 November 2017	2.2084	1 January 2016–17 November 2017
$Q_2$	$T_1$	9 January 2009	3.0700	2 March 2007–19 January 2009
$Q_2$	$T_2$	15 May 2015	2.5980	19 July 2013–15 May 2015
$Q_2$	$T_3$	12 October 2018	2.4534	2 December 2016–12 October 2018
$Q_3$	$T_1$	8 May 2009	2.9067	22 June 2007–8 May 2009
$Q_3$	$T_2$	17 April 2015	2.9681	21 June 2013–17 April 2015
$Q_4$	$T_1$	2 January 2009	3.6279	16 February 2007–2 January 2009
$Q_4$	$T_2$	8 November 2013	1.8349	23 December 2011–8 November 2013
$Q_4$	$T_3$	28 February 2020	3.2564	13 April 2018–28 February 2020
$Q_5$	$T_1$	9 April 2010	3.0202	23 May 2008–9 April 2010
$Q_5$	$T_2$	21 November 2014	1.8017	18 March 2013–21 November 2014
$Q_5$	$T_3$	13 March 2020	2.0941	27 April 2018–13 March 2020

We find that most of the local peaks correspond to periods that are close to each other, such as the  $T_1$  period of parameter  $Q_1$  and the  $T_1$  period of  $Q_2$  are 8 June 2007–24 April 2009 and 2 March 2007–19 January 2009 respectively. Table 3 shows that the 2008 crisis significantly affected the information flow of  $us \rightarrow cn$ . In addition, the periods corresponding to several local peaks also include major events. For example, the third peak of  $Q_5$  includes the oil crisis and the collapse of the US stock market. In summary, the *LRP*-based method identifies some local peaks, indicating that the dynamics of information flow are non-trivial.



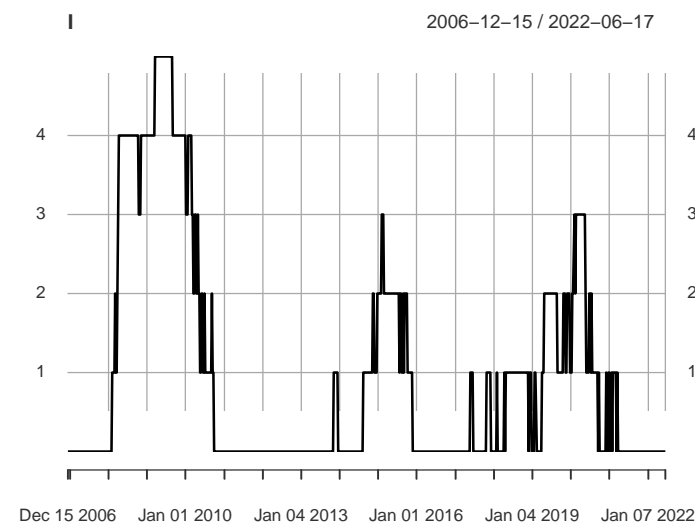
**Figure 1.** The figure shows the Z-scores for different parameters  $Q$ , where (a–e) correspond to  $Q_1$ – $Q_5$ , respectively.

#### 4.1.3. Comprehensive Analysis of Z-Score Series

We roughly identified three important periods by Z-scores in the previous subsection. Below, we calculate the  $I$  series from the  $I_k$  series of five Z-score series. The series  $I$  takes the value in the set  $\{0, 1, 2, 3, 4, 5\}$ . If  $I_k(t) = 0$ , the observations in the time period  $[t - L_w + 1, t]$  do not significantly affect the analysis results of the information flow. On the contrary, if  $I(t) = 5$ , it implies that the information flow within the period  $[t - L_w + 1, t]$  is significant and does not depend on the choice of parameters. A large  $I(t)$  value implies that there is information flow in multiple quantile intervals  $\{Q_k | k = 1, \dots, 5\}$ .

Figure 2 shows the  $I$  series. There are three periods where most observations are greater than or equal to 1. In particular, the analysis shows that there is a period where most observations are greater than or equal to 3, that is, 8 February 2008–24 September 2010. During the 2008–2009 crisis in the US market, there was strong information flow  $us \rightarrow cn$ . We use  $N_i$  to represent the number of observations equal to  $i$  in the  $I$  series, as shown in Table 4. It can be found that if the calculation results of 5 parameters are integrated, nearly

half of the observations are greater than or equal to 1. This implies that information flows widely exist between the two markets during the period under consideration. Table 5 lists the three periods identified by Figure 2. Each period includes some major events that affect the market. The first period corresponds to the 2007–2009 crisis in the US market. The second period includes the crisis in the Chinese stock market in 2015. In addition, before the stock market crash in 2015, the bubble was generated in the second half of 2014. The second period not only includes the breakage period of the bubble, but also the generation period. The third period includes a series of major events, such as the circuit breaker in the US market in March 2020, and the COVID-19 pandemic. Based on this table, we divide three periods for the analysis of daily data in the next section: 2007–2010, 2014–2017, and 2008–2021.



**Figure 2.** The  $I$  time series obtained from five sets of parameters, and takes values in  $\{0, 1, 2, 3, 4, 5\}$ . It includes three local peaks, and the maximum corresponds to the 2008 crisis.

**Table 4.** Statistics of  $I$  series.

$N_i$	$N_1$	$N_2$	$N_3$	$N_4$	$N_5$	Sum
Number	142	80	29	66	24	341

**Table 5.** Significant periods identified by Figure 2.

$T_i$	Period
$T_1$	8 February 2008–24 September 2010
$T_2$	15 August 2014–20 November 2015
$T_3$	13 April 2018–19 March 2021

#### 4.2. Information Flow Analysis of Daily Data

##### 4.2.1. Global Analysis of Daily Data

In this section, we analyze the relationship between the daily data of the two indices. Table 6 lists the transfer entropy values  $TE_{us \rightarrow cn}$  of different parameters. For the first and third periods, all  $TE$  values are significant at significance level  $\alpha = 0.05$ . For the second period, the  $TE$  values of parameters  $Q_1$  and  $Q_2$  were not significant. However, the  $p$ -value for the  $TE$  value of  $Q_4$  was less than 0.01. This implies that extreme values have little effect on information flow in the second period. In summary, we still found significant information flows in the three periods of daily data. Below, we use  $LRP$ -based analysis to discuss the dynamics of information flow in each period.



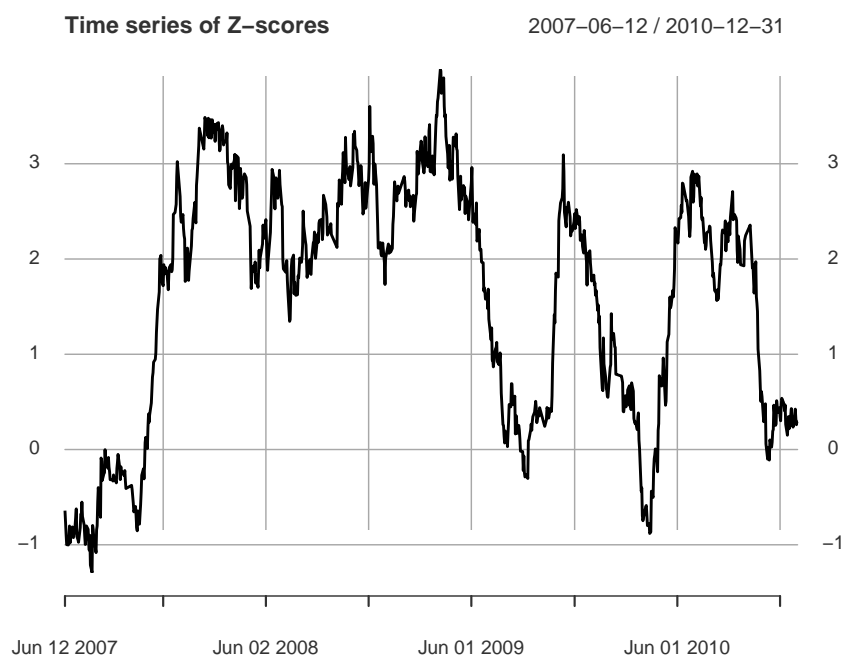
**Table 6.** Analysis of  $TE$  value of daily data.

Period	2007–2010	2014–2017	2018–2021
$Q_1$	0.0573 (<0.001)	0.0322 (0.262)	0.0587 (<0.001)
$Q_2$	0.0490 (0.006)	0.0286 (0.529)	0.0630 (<0.001)
$Q_3$	0.0510 (0.002)	0.0505 (0.002)	0.0545 (<0.001)
$Q_4$	0.0601 (<0.001)	0.0489 (0.006)	0.0519 (<0.001)
$Q_5$	0.0600 (<0.001)	0.0407 (0.041)	0.0577 (<0.001)

In the next subsection, for the first period and the second period, we choose the parameter  $Q$  corresponding to the maximum  $TE$  value in Table 6. In the analysis of the first and third periods, the parameter  $Q$  was set to  $Q_4$  and  $Q_2$ , respectively. In addition, the second period is a special period, in which both  $Q_1$  and  $Q_2$  correspond to insignificant  $TE$  values, suggesting that extreme values in this period weakly affect the level of information flow. For the second period, we set  $Q_3$  and  $Q_4$  for calculation.

#### 4.2.2. Information Flow Dynamics for 2007–2010 Data

Figure 3 shows the  $Z$ -score series for the 2007–2010 data. If the threshold is 1.645, the series includes three significant periods: 26 November 2007 (28 June 2007)–25 June 2009, 28 October 2009 (1 June 2009)–8 January 2010, and 24 May 2010 (21 December 2009)–19 October 2010. Here, the date in parentheses is the left endpoint of the calculation window for the first  $Z$ -score greater than 1.645 within the period. In particular, the maximum value (3.9880) is in the first period and the calculation window is 3 November 2008–7 April 2009.

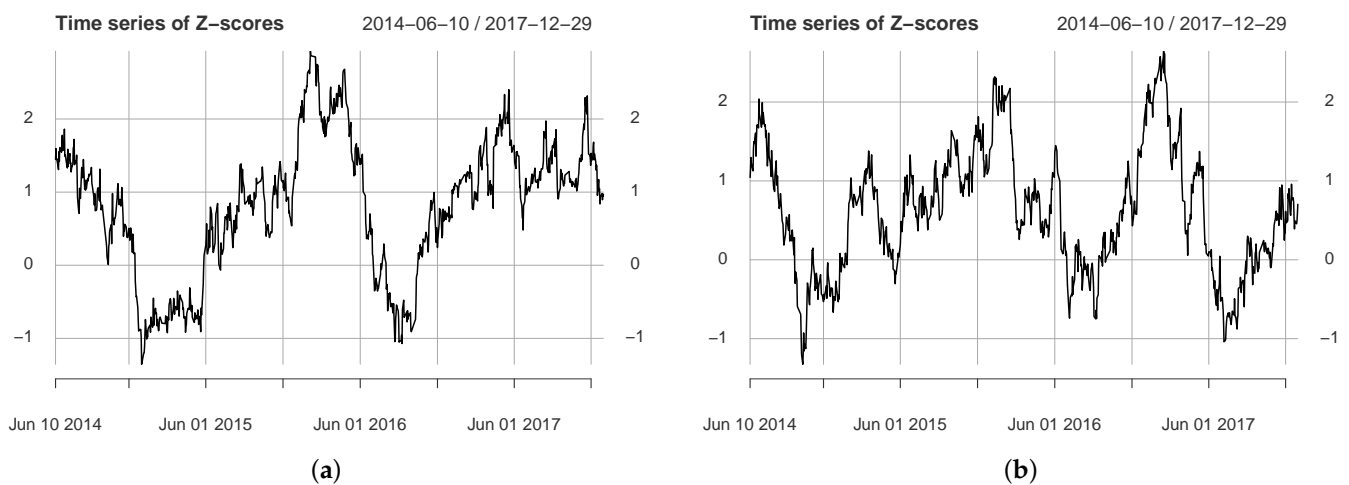


**Figure 3.** The figure shows the  $Z$ -score time series between  $R_{us}^1$  and  $R_{cn}^1$ , where  $L_w = 100$ .

The calculation results clearly show that the 2007–2009 crisis in the US market significantly affected the Chinese market. In addition, we also observe that the  $us \rightarrow cn$  information flow also exists in the post-crisis period. These results imply not one peak with strong information flow, but multiple peaks with different intensities.

#### 4.2.3. Information Flow Dynamics for 2014–2017 Data

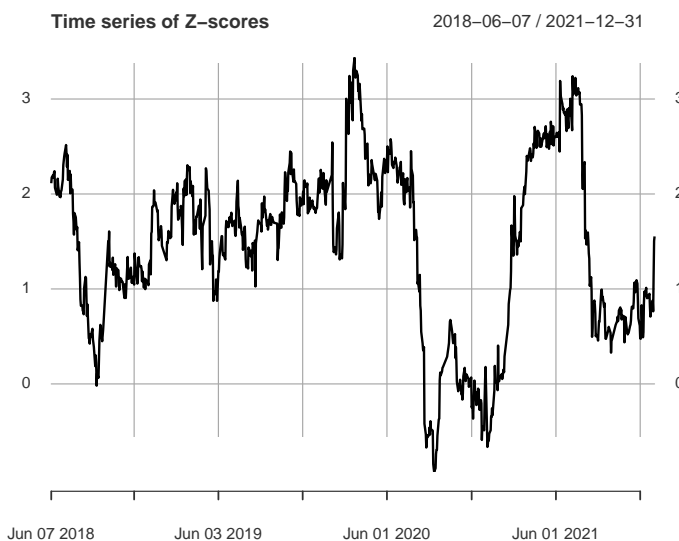
Figure 4 shows the information flow dynamics for the 2014–2017 data. Figure 4a includes a major significant period: 6 January 2016 (6 August 2015)–10 May 2016. The results show that the information flow in the  $us \rightarrow cn$  direction is enhanced after the Chinese market bubble burst. Figure 4b shows the calculation results corresponding to  $Q_4$ . We found that it also includes a major significant period: 30 December 2016 (1 August 2016)–1 March 2017. The period included an important event that affected US markets: the 2016 presidential election. A previous study showed that the event significantly affected correlation dynamics in the US stock market [35]. The analysis here suggests that there was a significant information flow from the US stock market to the Chinese stock market around this event. The two results in Figure 4 show that from 2 January 2014 to 29 December 2017, the US market significantly influenced the Chinese market in two periods, and both periods included major events.



**Figure 4.** The figure shows the Z-score time series between  $R_{us}^2$  and  $R_{cn}^2$ , where (a,b) correspond to parameters  $Q_3$  and  $Q_4$ , respectively.

#### 4.2.4. Information Flow Dynamics for 2018–2021 Data

Figure 5 shows the calculation results for the 2018–2021 data. The period includes several sub-periods greater than the critical Z-score, suggesting that from 2 January 2018 to 31 December 2021, the flow of information between the two markets was active and strong. These sub-periods are as follows: 7 June 2018 (3 January 2018)–31 July 2018 ( $T_1$ ), 22 February 2019 (12 September 2018)–15 May 2019 ( $T_2$ ), 19 June 2019 (11 January 2019)–30 July 2019 ( $T_3$ ), 27 August 2019 (29 March 2019)–5 February 2020 ( $T_4$ ), 27 February 2020 (19 September 2019)–29 July 2020 ( $T_5$ ), 3 March 2021 (24 September 2020)–2 August 2021 ( $T_6$ ). We label the different sub-periods with  $T_i$ . It can be found that  $T_2$ ,  $T_3$ ,  $T_4$  and  $T_5$  can be combined into an important period. There were several major events in this period: Sino-US trade friction, the oil crisis, the COVID-19 pandemic, the US stock market crisis in March 2020. In addition, periods  $T_1$  and  $T_6$  still belong to the period of Sino-US trade friction and the period of the COVID-19 pandemic, respectively. In particular, the calculation window corresponding to the maximum value (3.4339) is 21 October 2019–23 March 2020. Several major events are included in this period, such as the oil crisis, the COVID-19 pandemic, and the March 2020 market crisis. To sum up, it is reasonable to speculate that the impact of multiple major events led to the active information flow during the period.



**Figure 5.** The figure shows the Z-score series for  $R_{us}^3$  and  $R_{cn}^3$ , which includes two main peaks ( $Q = Q_2, L_w = 100$ ).

### 4.3. Information Flow between Industry Indices in the Chinese Market and the S&P500 Index

#### 4.3.1. Global Information Flow Analysis between Industry and Market Index

Since the composite index includes stocks belonging to various industries, in this section, we analyze the relationship between the S&P500 index and the industry index in the Chinese market. Here, we consider the case where a larger weight is given to the tail of the distribution, i.e.,  $Q = Q_1$ . Table 7 lists the  $TE$  values and  $p$ -values, where the symbol “ind” represents the industry index, and the code is listed in the second column.

The fourth and sixth columns show that there are several industry indices that are closely related to the S&P500 index. For example, there is a significant bidirectional flow of information between the petrochemical industry and the S&P500 index. In addition, some unidirectional effects can also be found, such as in real estate. It should be noted that not all industry indices are affected by the S&P500 index, such as the public utilities index and the architectural ornament index. To sum up, Table 7 shows that the information flow from S&P500 to Chinese market industry indices is heterogeneous.

**Table 7.**  $TE$  values of industry indices and S&P500 index.

Industry	Code	$TE_{ind \rightarrow us}$	$p_{ind \rightarrow us}$	$TE_{us \rightarrow ind}$	$p_{us \rightarrow ind}$
AFAF	801010.SL	0.0341	0.254	0.0444	0.011
Basic chemicals	801030.SL	0.0448	0.012	0.0394	0.061
Steel industry	801040.SL	0.0350	0.093	0.0686	0.000
Nonferrous metals	801050.SL	0.0353	0.125	0.0475	0.005
Electronics	801080.SL	0.0325	0.334	0.0413	0.034
Household Electric Appliance	801110.SL	0.0280	0.551	0.0622	0.000
Food and beverage	801120.SL	0.0293	0.491	0.0443	0.012
Textile, clothing.	801130.SL	0.0415	0.055	0.0483	0.004
Light manufacturing	801140.SL	0.0378	0.102	0.0504	0.000
Bio-pharmaceutical industry	801150.SL	0.0294	0.499	0.0418	0.036
Public utilities	801160.SL	0.0431	0.022	0.0359	0.121
Transportation industry	801170.SL	0.0338	0.257	0.0519	0.001
Real estate	801180.SL	0.0304	0.427	0.0469	0.008
Retail business	801200.SL	0.0388	0.079	0.0358	0.154
Social services	801210.SL	0.0407	0.040	0.0428	0.028
Comprehensive company	801230.SL	0.0296	0.482	0.0476	0.007
Building material	801710.SL	0.0347	0.213	0.0468	0.009

Table 7. Cont.

Industry	Code	$TE_{ind \rightarrow us}$	$p_{ind \rightarrow us}$	$TE_{us \rightarrow ind}$	$p_{us \rightarrow ind}$
Architectural ornament	801720.SL	0.0393	0.072	0.0302	0.379
Power equipment	801730.SL	0.0279	0.559	0.0347	0.160
National defense industry	801740.SL	0.0342	0.239	0.0314	0.351
Computer industry	801750.SL	0.0253	0.719	0.0383	0.096
Media industry	801760.SL	0.0309	0.334	0.0484	0.006
Communication industry	801770.SL	0.0423	0.047	0.0414	0.031
Banking	801780.SL	0.0457	0.021	0.0333	0.220
Non-bank Financial Institutions	801790.SL	0.0251	0.775	0.0372	0.087
Automobile industry	801880.SL	0.0270	0.591	0.0508	0.004
Mechanical equipment	801890.SL	0.0443	0.021	0.0447	0.017
Coal industry	801950.SL	0.0418	0.031	0.0424	0.016
Petrochemical industry	801960.SL	0.0510	0.001	0.0469	0.004
Environmental protection	801970.SL	0.0282	0.595	0.0475	0.006
Beauty care	801980.SL	0.0337	0.258	0.0436	0.017

AFAF : agriculture, forestry, animal husbandry, and fishery.

#### 4.3.2. The Dynamics of Information Flow between Industry Index and S&P500 Index

In the previous section, we found diversity in the relationship between the S&P500 index and industry indices, with some industry indices having a stronger flow of information between them and the S&P500 index. A related issue is whether the dynamics of information flow for different industry indices have similar patterns. Here, we select three industry indices and compare information flow dynamics. The three indices are 801040.SL (Steel industry), 801180.SL (Real estate) and 801960.SL (Petrochemical industry), and all have significant  $us \rightarrow ind$  information flow.

We analyze dynamics using weekly data and always set  $Q = Q_1$  and  $L_w = 96$  weeks. Figure 6 shows the dynamics of information flow for  $us \rightarrow 801040.SL$ . The Z-score series includes 794 values, and 396 values are greater than the critical threshold, implying that there is a continuous flow of information  $us \rightarrow 801040.SL$  during the period. In addition, it includes several local peaks, and the corresponding periods of the two main peaks are 18 April 2008 (2 June 2006)–15 April 2011 and 24 October 2014 (21 December 2012)–30 June 2017, respectively.

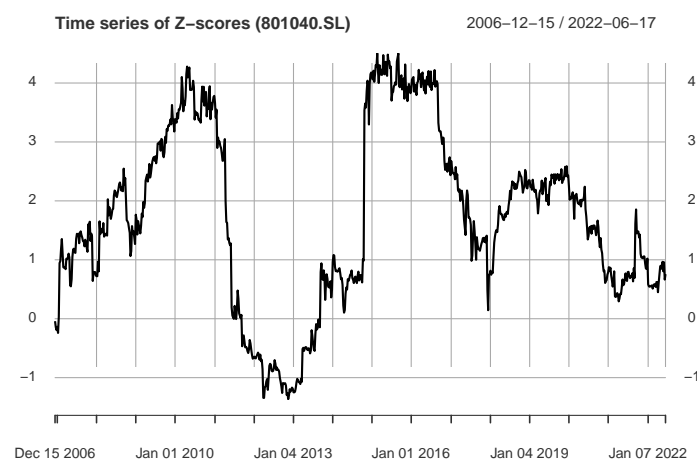
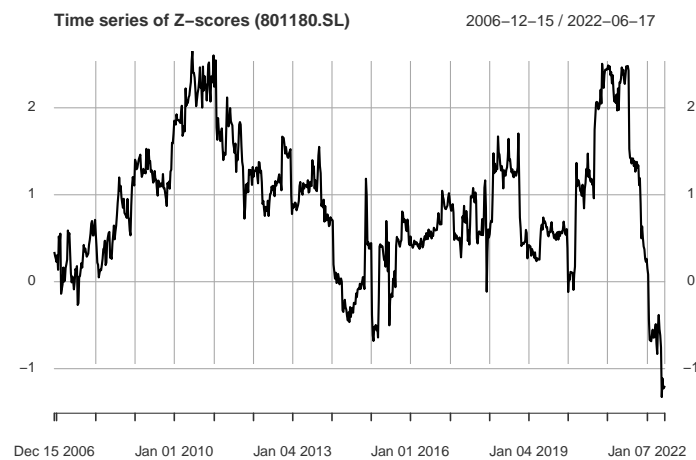


Figure 6. The figure shows the Z-score time series for the industry index 801040.SL (Steel industry) and the S&P500 index, including two clear local peaks.

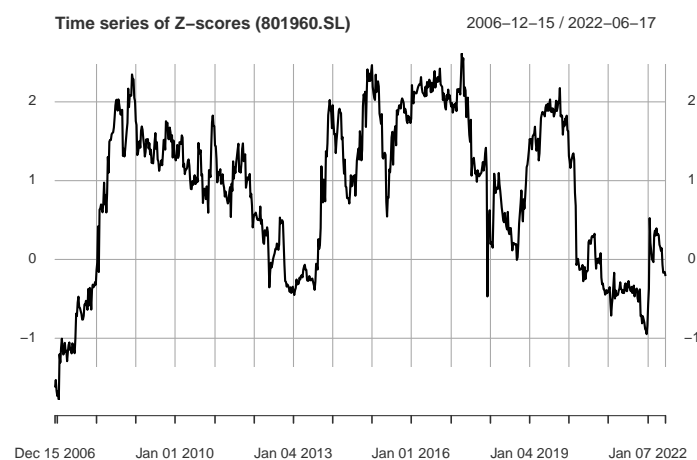
Figure 7 presents the Z-score series for the S&P500 Index  $\rightarrow$  Real estate Index. For this Z-score series, there are only 124 values greater than the critical threshold. It shows a different pattern from Figure 6, such as observations in the period 4 September 2020 (26 October 2018)–16 July 2021 contribute significantly to the information flow. In addition,

another major peak corresponds to the period 25 December 2009 (15 February 2008)–24 June 2011.



**Figure 7.** The figure shows the Z-score time series for the industry index 801180.SL (Real estate) and the S&P500 index, with only a few Z-score values above the critical threshold.

The information flow analysis result of  $us \rightarrow 801960.SL$  is shown in Figure 8. We find that it included a period with most values greater than the critical threshold: 10 October 2014–9 June 2017. This period is close to that of the second peak in Figure 6. However, for the series shown in Figure 8, it includes some values less than the critical threshold.



**Figure 8.** The figure shows the Z-score time series of the industry index 801960.SL (Petrochemical industry) and S&P500 index, and most of the Z-score values greater than the critical threshold are in the period 2014–2017.

The results in Figures 6–8 show that the information flow dynamics of  $us \rightarrow ind$  exhibit multiple patterns, thus indicating a high complexity of information flow between markets.

## 5. Discussion

This study examines periods of strong information flow between two important markets, where these periods are accompanied by major events. This implies that long-term major events may affect market relations more deeply. This provides a forward-looking analysis for analyzing the relationship between the two markets. In particular, we can analyze the dynamics of information flow in detail to identify changes in the relationship between markets when markets are impacted by major events. A related topic that can be studied is whether non-linear information flow helps predict future returns in the market.

Several previous studies have revealed that major events affect the relationship between markets [4–14]. Unlike previous studies, we analyze long-term data and focus on the dynamics of information flow. In particular, *LRP*-based analysis characterizes the strength of the local information flow, so that the time-varying characteristics of the information flow in a period can be observed in detail.

## 6. Conclusions

We discussed the long-term relationship between the Chinese market and the US market and found that there is an asymmetric information flow, where the US market is the source of information flow. Calculations show that the information flow between the Chinese market and the US market is not stable, but there are some significant periods. The *LRP*-based analysis identified three periods of significant information flow. All three periods include some major events affecting the market, such as the 2008 crisis, the oil crisis, and the COVID-19 pandemic, thus suggesting that the dynamics of information flow may be impacted by major events. In addition, we conducted a detailed dynamic analysis with daily data. Calculations show that each period includes nontrivial dynamics, some of which appear as local peaks, suggesting dramatic changes in information flow.

We also analyzed the information flow between the S&P500 index and some industry indices in the Chinese A-share market. Calculations show that the S&P500 index does not always significantly affect each industry index. Some industries are significantly affected, such as the steel industry and real estate. However, analysis based on weekly data shows that the dynamics of information flow between the industry indices and the S&P500 index exhibits multiple patterns. In addition, there are also some significant information flows from the Chinese market to the US market, thus demonstrating the heterogeneity of information flow.

This study demonstrates the high complexity of nonlinear causal relationships between two important markets and facilitates quantitative studies of the impact of major events. In particular, empirical analysis suggests that the heterogeneity of information flow in industry indices needs to be taken into account when discussing risk spillovers between markets. In this article, we only consider the information flow between the two markets. A related and important topic is the study of how major events affect information flow networks that include multiple markets. Since *LRP*-based methods involve two variables, for the information flow network, we need to construct a method that includes multiple variables. An alternative approach is to characterize changes in network structure by shuffling observations. In addition, the relationship between industry indices within a market can be analyzed in detail through *LRP*-based methods and high-frequency data. For example, for a trading day, we can analyze how important announcements impact the information flow network constructed by multiple industry indices.

**Author Contributions:** Methodology, C.-X.N.; formal analysis, C.-X.N. and J.X.; investigation, C.-X.N. and J.X.; writing-original draft preparation, C.-X.N.; writing-review and editing, C.-X.N.; visualization, C.-X.N. and J.X.; funding acquisition, C.-X.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work is supported by the characteristic & preponderant discipline of key construction universities in Zhejiang province (Zhejiang Gongshang University—Statistics).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Goh, J.C.; Jiang, F.; Tu, J.; Wang, Y. Can US economic variables predict the Chinese stock market? *Pac.-Basin Financ. J.* **2013**, *22*, 69–87.

2. Chen, J.; Jiang, F.; Li, H.; Xu, W. Chinese stock market volatility and the role of US economic variables. *Pac.-Basin Financ. J.* **2016**, *39*, 70–83.
3. Hu, Z.; Kutan, A.M.; Sun, P.W. Is US economic policy uncertainty priced in China's A—shares market? Evidence from market, industry, and individual stocks. *Int. Rev. Financ. Anal.* **2018**, *57*, 207–220.
4. Kenourgios, D.; Samitas, A.; Paltalidis, N. Financial crises and stock market contagion in a multivariate time-varying asymmetric framework. *J. Int. Financ. Mark. Inst. Money* **2011**, *21*, 92–106.
5. Cheung, W.; Fung, S.; Tsai, S.C. Global capital market interdependence and spillover effect of credit risk: Evidence from the 2007–2009 global financial crisis. *Appl. Financ. Econ.* **2010**, *20*, 85–103.
6. Hwang, E.; Min, H.G.; Kim, B.H.; Kim, H. Determinants of stock market comovements among US and emerging economies during the US financial crisis. *Econ. Model.* **2013**, *35*, 338–348.
7. Samarakoon, L.P. Stock market interdependence, contagion, and the US financial crisis: The case of emerging and frontier markets. *J. Int. Financ. Mark. Inst. Money* **2011**, *21*, 724–742.
8. Wang, G.J.; Xie, C.; Lin, M.; Stanley, H.E. Stock market contagion during the global financial crisis: A multiscale approach. *Financ. Res. Lett.* **2017**, *22*, 163–168.
9. Dungey, M.; Gajurel, D. Equity market contagion during the global financial crisis: Evidence from the world's eight largest economies. *Econ. Syst.* **2014**, *38*, 161–177.
10. Jin, X.; An, X. Global financial crisis and emerging stock market contagion: A volatility impulse response function approach. *Res. Int. Bus. Financ.* **2016**, *36*, 179–195.
11. Ahmad, W.; Sehgal, S.; Bhanumurthy, N. Eurozone crisis and BRIICKS stock markets: Contagion or market interdependence? *Econ. Model.* **2013**, *33*, 209–225.
12. Sansa, N.A. The Impact of the COVID-19 on the Financial Markets: Evidence from China and USA. *Electron. Res. J. Soc. Sci. Humanit.* **2020**, *2*. <http://dx.doi.org/10.2139/ssrn.3567901>.
13. Nguyen, D.T.; Phan, D.H.B.; Ming, T.C. An assessment of how COVID-19 changed the global equity market. *Econ. Anal. Policy* **2021**, *69*, 480–491.
14. Zhang, D.; Hu, M.; Ji, Q. Financial markets under the global pandemic of COVID-19. *Financ. Res. Lett.* **2020**, *36*, 101528.
15. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461.
16. Nie, C.X. Dynamics of the price–Volume information flow based on surrogate time series. *Chaos Interdiscip. J. Nonlinear Sci.* **2021**, *31*, 013106.
17. Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B.; Farmer, J.D. Testing for nonlinearity in time series: The method of surrogate data. *Phys. D Nonlinear Phenom.* **1992**, *58*, 77–94.
18. Small, M. *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*; World Scientific: Singapore, 2005; pp. 115–129.
19. Lancaster, G.; Iatsenko, D.; Pidde, A.; Ticcinelli, V.; Stefanovska, A. Surrogate data for hypothesis testing of physical systems. *Phys. Rep.* **2018**, *748*, 1–60.
20. Marschinski, R.; Kantz, H. Analysing the information flow between financial time series. *Eur. Phys. J. B-Condens. Matter Complex Syst.* **2002**, *30*, 275–281.
21. Dimpfl, T.; Peter, F.J. Group transfer entropy with an application to cryptocurrencies. *Phys. A Stat. Mech. Its Appl.* **2019**, *516*, 543–551.
22. Jizba, P.; Kleinert, H.; Shefaat, M. Rényi's information transfer between financial time series. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 2971–2989.
23. Papan, A.; Kyrtsov, C.; Kugiumtzis, D.; Diks, C. Detecting Causality in Non-stationary Time Series Using Partial Symbolic Transfer Entropy: Evidence in Financial Data. *Comput. Econ.* **2016**, *47*, 341–365.
24. Kwon, O.; Yang, J.S. Information flow between composite stock index and individual stocks. *Phys. A Stat. Mech. Its Appl.* **2008**, *387*, 2851–2856.
25. Dimpfl, T.; Peter, F.J. The impact of the financial crisis on transatlantic information flows: An intraday analysis. *J. Int. Financ. Mark. Inst. Money* **2014**, *31*, 1–13.
26. Dimpfl, T.; Peter, F.J. Using transfer entropy to measure information flows between financial markets. *Stud. Nonlinear Dyn. Econom.* **2013**, *17*, 85–102.
27. Daugherty, M.S.; Jithendranathan, T. A study of linkages between frontier markets and the U.S. equity markets using multivariate GARCH and transfer entropy. *J. Multinatl. Financ. Manag.* **2015**, *32*, 95–115.
28. Kwon, O.; Yang, J.S. Information flow between stock indices. *EPL Europhys. Lett.* **2008**, *82*, 68003.
29. Ong, M.A. An information theoretic analysis of stock returns, volatility and trading volumes. *Appl. Econ.* **2015**, *47*, 3891–3906.
30. Kyrtsov, C.; Kugiumtzis, D.; Papan, A. Further insights on the relationship between SP500, VIX and volume: A new asymmetric causality test. *Eur. J. Financ.* **2019**, *25*, 1402–1419.
31. Zheng, J.; Nie, C.X. The Dynamics Of Price–Volume Information Transfer In The Cryptocurrency Markets. *Adv. Complex Syst.* **2020**, *23*, 2050014.
32. Sensoy, A.; Sobaci, C.; Sensoy, S.; Alali, F. Effective transfer entropy approach to information flow between exchange rates and stock markets. *Chaos Solitons Fractals* **2014**, *68*, 180–185.
33. Kim, Y.; Kim, J.; Yook, S.H. Information transfer network of global market indices. *Phys. A Stat. Mech. Its Appl.* **2015**, *430*, 39–45.

- 
34. Behrendt, S.; Dimpfl, T.; Peter, F.J.; Zimmermann, D.J. RTransferEntropy—Quantifying information flow between different time series using effective transfer entropy. *SoftwareX* **2019**, *10*, 100265.
  35. Nie, C.X. A network-based method for detecting critical events of correlation dynamics in financial markets. *EPL Europhys. Lett.* **2020**, *131*, 50001.