Physica A 478 (2017) 188–204

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

The mutual causality analysis between the stock and futures markets

Can-Zhong Yao *, Qing-Wen Lin

School of Economics and Commerce, South China University of Technology, Guangzhou, 510006, China

HIGHLIGHTS

- Conditional Granger causality performs better in revealing asymmetric correlation.
- Direct information flows from stocks to futures are significantly greater.
- Conditional Granger causality could be used to detect the important events.
- Direct information flows are significantly less than the information flows.

ARTICLE INFO

Article history:
Received 18 October 2016
Received in revised form 20 January 2017
Available online 28 February 2017

Keywords:
Granger causality
Conditional Granger causality
Information flow
Direct information flow
Asymmetric ratio
Bootstrap technique

ABSTRACT

In this paper we employ the conditional Granger causality model to estimate the information flow, and find that the improved model outperforms the Granger causality model in revealing the asymmetric correlation between stocks and futures in the Chinese market.

First, we find that information flows estimated by Granger causality tests from futures to stocks are greater than those from stocks to futures. Additionally, average correlation coefficients capture some important characteristics between stock prices and information flows over time.

Further, we find that direct information flows estimated by conditional Granger causality tests from stocks to futures are greater than those from futures to stocks. Besides, the substantial increases of information flows and direct information flows exhibit a certain degree of synchronism with the occurrences of important events.

Finally, the comparative analysis with the asymmetric ratio and the bootstrap technique demonstrates the slight asymmetry of information flows and the significant asymmetry of direct information flows. It reveals that the information flows from futures to stocks are slightly greater than those in the reverse direction, while the direct information flows from stocks to futures are significantly greater than those in the reverse direction.

1. Introduction

Information flow is a hot topic for theoretical researches and practical applications in complex systems. By definition, information flow is characterized by interaction among different objects. Previous studies proposed numerous methodologies to estimate information flows. Haythornthwaite [1] estimated the information flows by resource exchanges

* Corresponding author.
E-mail address: ycz20120911@gmail.com (C.-Z. Yao).

http://dx.doi.org/10.1016/j.physa.2017.02.071

and revealed the directions in which information or resources flowed. Those types of social networks intuitively exhibit the information delivery, whereas they are challenged for the interaction among different objects in deeper levels.

Alternatives were proposed by scholars to estimate information flows. Cheung et al. [2] estimated information flows by Granger causality tests to analyze patterns of information flows between Eurodollar spot and futures markets, and finally uncovered the impact of information conveyed by futures data on market price movement. Gabjin et al. [3] investigated the information flows among industrial sectors in the Korean stock market by symbolic transfer entropy (STE) methods, and found that the information flows during the financial crisis were greater than those in pre-crisis and post-crisis periods. The above-mentioned two methods, the Granger causality test and the transfer entropy, both have their limitation as they are applicable to pairwise causality, and were proved equivalent by Barnett et al. [4] under Gaussian situation. Yao et al. [5] studied the industrial energy transferring paths by minimal spanning trees in pre-crisis and post-crisis periods, potentially revealing how the information flows influenced the industrial risk contagion on industrial upgrading and transformation. However, the scope of above studies is in one complex system.

In the field of finance, the price discovery between stock markets and futures markets is the stepping stone for securities investment [6,7]. In reality, the presence of futures accelerates arbitrage trading, and thereby reduces the volatility of the stock market. The high leverage of stock index futures increases the participation of funds, and boosts the liquidity of the stock market. The presences of hedging tools provided by stock index futures push agents to invest, and increase the volume of the stock market. The futures price movement is closely bound up with spot stock prices, as the agreed price of a stock index futures contract is based on the price of underlying asset. Thus it is worthwhile to study the relationship and information flows between stock markets and futures markets.

Based on the inefficiency of capital markets, many works have interpreted how stock index futures and spot goods influence each other in global markets. Chan [8] and Tse [9] successively investigated S&P 500 index and futures, and both found the leading trend of futures. Ding et al. [10] made theoretical analysis on the influence of the launch of stock index futures on stock prices in American, European, Chinese and Japanese markets. Comparably, Zakaria et al. [11] investigated the daily data of the Malaysian stock market and showed the unidirectional causal relationship from the cash market to futures market. Similar results were obtained in the studies of FTSE 100 and Nasdaq 100 [12–15]. The literatures above explored the lead–lag relationship between spots and futures from global markets, whereas they excessively focused on overall influence without considering time scales. Thus further studies set out to investigate how the lead–lag relationship between spot markets and futures markets evolved over time. Fang et al. [16] analyzed the short-term and long-term influence from stock index futures to stock market volatility. Furthermore, Gong et al. [17] used both parametric and non-parametric methods to study the relationship among China Securities Index 300 (CSI 300), Hang Seng Index (HSI), Standard and Poor 500 (S&P 500) Index and their associated futures. Their findings quantitatively supported the leading trend of index futures. In their studies, the non-parametric method showed how the lead–lag relationship changed over time. Illuminated by their works, we take time scales into consideration.

To quantify the degree of market imperfection, some scholars highlighted the concepts of information flows. Several studies focused on bidirectional mutual information flows between two objects. Based on the Granger causality test, Bose [18] discovered the information flows from the Indian futures market to stock market were slightly greater than the information flows from the spot market to futures market. Kwon [19] also found the asymmetric information flows between the stock indexes and component stocks in Australia, Canada and China markets. Despite burgeoning literatures about information flows, still few works focus on the correlation among several complex systems, in which each complex system has many objects. Thus considering different complex systems, Kim et al. [20] showed that the information flows between some fund performance indicators (FPI) and the Korean stock market increased sharply when great events happened. And information flows between the Korean fund market and the Korean stock market were significantly positively correlative with the market volatility. However, these studies were mostly based on the Granger causality test.

Motivated by these previous studies, we follow the thoughts of the information flow and causal relationship to study the interaction between the Chinese stock market and futures market. Unlike previous studies on information flows among financial objects, we employ the conditional Granger causality model to study the information flow between the stock and futures markets, and compare it to the Granger causality model [21]. Since the Granger causality model cannot detect whether the interaction is direct or mediated by other objects, the conditional Granger model serves as the supplement of the Granger causality model to distinguish the direct causality from the indirect, and hence it has been widely used in the field of neuroscience [22–24]. Concerning few applications of the conditional Granger causality model in the financial field, we think it meaningful to reveal the direct causal relationship of financial objects. As an extension of the Granger causality model, the conditional Granger causality model can serve as a benchmark to investigate their difference in relationship mining. We also try to adopt dynamic thinking to study the evolution of the causality between the stock and index futures markets. Based on Granger and conditional Granger causality models, the information flows are subdivided into direct information flows and indirect information flows.

The rest of this paper is organized as below. In the next section, we describe the time series data used in the models. In Section 3, we describe the models used in this paper. In Section 4, we present the empirical results obtained with different models. Finally, we provide conclusion of this paper.

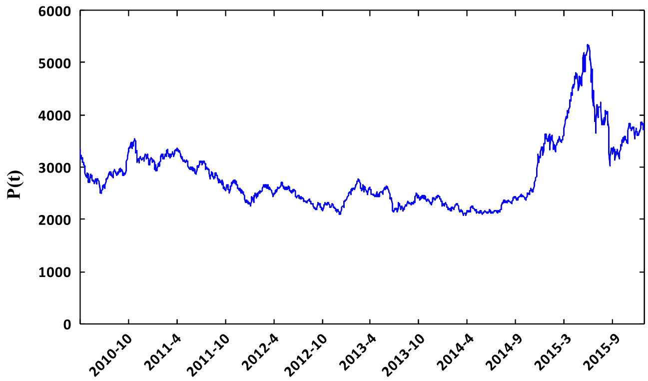

Fig. 1. Daily closing prices of Hushen 300 index futures.

2. Data and prerequisites for models

2.1. Data description

In order to study the interaction between the stock and index futures markets in China, we respectively use the daily closing price returns of 2861 stocks and stock index futures spanning from April 16, 2010 to December 31, 2015 as proxies for the stock market and the futures market. We obtain 1388 observations of each financial object in the market date from April 16, 2010 to December 31, 2015. There had been 1095 stocks on Shanghai Stock Exchange and 1766 stocks on Shenzhen Stock Exchange by December 31, 2015. As for the stocks officially listed after April 16 2010, we remove them from the studied dataset in this paper. Finally the studied stocks consist of 844 stocks on Shanghai Stock Exchange and 902 stocks on Shenzhen Stock Exchange while the stock index futures include Shanghai and Shenzhen 300 (Hushen 300) index futures, China Securities Index 500 (CSI 500) futures and Shanghai Stock 50 (SSE 50) index futures.

The dataset includes stocks of representative listed companies from all industries, so the scale is sufficient to reflect the whole market information. The return of a certain stock or stock index future could be calculated as the difference between the logarithmic prices at t and $t-1$, shown as follow:

$$R(t) = \ln p(t) - \ln p(t-1) \tag{1}$$

where $R(t)$ denotes the return of the stock or stock index future at time t, $p(t)$ denotes the price of the stock or stock index future at time t, and $p(t-1)$ denotes the price of the stock or stock index future at time $t-1$. The return series $R(t)$ are used to explore the interaction between the stock and futures markets.

In our analysis, the data are divided into blocks of 90 days. The data window is moved step by step to investigate the dynamic evolution of information flows and direct information flows between the stock and futures markets.

2.2. Prerequisites for models

(1) *Stationarity analysis*

Both of Granger causality and conditional Granger causality models are based on VAR models (*Vector Autoregression*), and the prerequisite for the VAR model is the stationary time series. Therefore, in order to avoid the spurious regression, it is necessary to ensure the stationarity before applying causal models. Prices and returns of Hushen 300 index futures for the period between April 16, 2010 and December 31, 2015 are shown in Figs. 1–2:

One efficient way to test the stationarity of time series is the *ADF-test* [25]. The regression for the ADF test is estimated as follow:

$$\Delta Y_t = \lambda + \phi t + \eta Y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta Y_{t-i} + \varepsilon_t \tag{2}$$

where Y represents the tested sequence, Δ is the difference operator, λ, η and ϕ are the coefficients to be estimated, p is the chosen lag value which is set to 10 in this paper, t is the time trend, and ε is the error term. The tested sequence is stationary if the null hypothesis ($\eta = 0$) is rejected. The results of *ADF-tests* are shown in Table 1:

Table 1 shows the consistent results that daily closing price sequences of index futures are non-stationary while the daily return sequences of index futures are stationary. To investigate the stationarity of 844 stocks on Shanghai Stock Exchange and 902 stocks on Shenzhen Stock Exchange, we calculate the percentages of stationary sequences. The results are shown in Table 2.

Results of *ADF-tests* show that most of price sequences are non-stationary but all the return sequences are stationary.

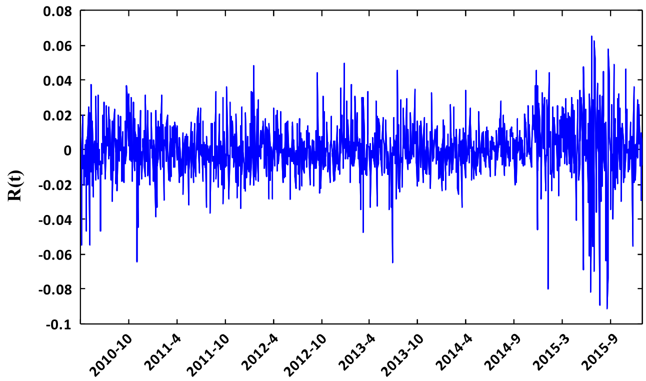

Fig. 2. Daily returns of Hushen 300 index futures.

Table 1
ADF-tests for price and return sequences of index futures.

Index futures		t-statistic	1% level	5% level	10% level	Prob.
Hushen 300	Closing price	−1.606387	−3.434882	−2.863429	−2.567825	0.4790
	Return	−35.25274	−3.434859	−2.863418	−2.567819	**0.0000**[**]
CSI 500	Closing price	−1.274862	−3.434889	−2.863432	−2.567826	0.6432
	Return	−32.96975	−3.434859	−2.863418	−2.567819	**0.0000**[**]
SSE 50	Closing price	−1.674659	−3.434869	−2.863423	−2.567821	0.4440
	Return	−36.18207	−3.434859	−2.863418	−2.567819	**0.0000**[**]

Note: prob. denotes the tail probability. The null hypothesis that the sequence is non-stationary is rejected if prob. is less than the critical values (1%, 5%, 10%).
[**] Denote significance at 1% level.

Table 2
Percentages of stationary sequences for prices and returns of stocks.

Stock		1% level	5% level	10% level
Shanghai stock exchange	Closing price	2.39%	6.27%	8.96%
	Return	100%	100%	100%
Shenzhen stock exchange	Closing price	4.13%	10.55%	16.97%
	Return	100%	100%	100%

(2) *Elimination of Heteroscedasticity*

Heteroscedasticity means different variances of random error terms. Engle et al. [26] discovered that disturbance variances of time series models were usually more instable than the assumed in macroeconomic analysis. With regard to time series data, heteroscedasticity will always appear with the feature that variances of the prediction errors depend on the previous and subsequent disturbance.

Especially in financial time series analysis, it is usual that the prediction errors of time series models are relatively small at one period, but large at the other period, namely conditional heteroscedasticity. Such fluctuations are likely caused by the volatility of financial markets vulnerable to rumors, political changes, and fiscal policies. In this paper, GARCH [27] is used to control the conditional heteroscedasticity of return sequences, expressed as:

$$\begin{cases} D_t = \mu + \sigma_t \varepsilon_t \\ \sigma_t^2 = \omega + \alpha (D_{t-1} - \mu)^2 + \beta \sigma_{t-1}^2 \end{cases} \tag{3}$$

where ε_t is an error term with zero mean and constant variance, and σ_t is the standard deviation of the model. The return sequences of stocks and index futures are adjusted to the new sequences:

$$D_t' = \frac{D_t}{\sigma_t}. \tag{4}$$

ARCH-LM [28] is used to examine the conditional heteroscedasticity of return sequences of index futures and stocks. The regression for the ARCH-LM test is estimated as follow:

$$u_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \varepsilon_t \tag{5}$$

Table 3
ARCH-LM tests of index futures before and after the adjustment.

Return sequences of index futures		F-statistic	Prob.
Hushen 300	Before	**73.66124**	**0.0000**[**]
	After	1.783087	0.1820
CSI 500	Before	**150.5550**	**0.0000**[**]
	After	0.037810	0.8459
SSE 50	Before	**41.41767**	**0.0000**[**]
	After	1.689662	0.1939

Note: If the Prob. is less than 5% (1%), there is heteroscedasticity in the corresponding sequence at the critical level of 5% (1%).
[**] Denote significance at 1% level.

Table 4
Percentages of stock sequences excluding heteroscedasticity before and after the adjustment.

Return sequences of stocks		1% level	5% level	10% level
Shanghai stock exchange	Before	26.87%	25.07%	23.88%
	After	**96.24**%	**91.46**%	**87.28**%
Shenzhen stock exchange	Before	44.04%	40.83%	38.99%
	After	**98.25**%	**95.95**%	**94.12**%

where u represents the residual error, α is the coefficient to be estimated, p represents the chosen lag value which is set to 10 in this paper, t is the time trend, and ε is the error term. The heteroscedasticity exists if the null hypothesis ($\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$) is rejected. The experimental results of ARCH-LM tests for index futures and stocks before and after the adjustment of GARCH (1, 1) are respectively shown in Tables 3 and 4.

Tables 3 and 4 show the data quality of return sequences of index futures and stocks before and after the adjustment of GARCH (1, 1). Specially, Table 3 indicates the availability of the adjustment to eliminate the heteroscedasticity of index futures sequences.

At 5% critical level, the percentage of stocks on Shanghai Stock Exchange excluding heteroscedasticity increases from 25.07% to 91.46% after the adjustment, while the percentage of stocks on Shenzhen Stock Exchange increases from 40.83% to 95.95%.

Thus, the adjustment of GARCH (1, 1) is effective for both index futures and stocks. It would be better to use the adjusted sequences instead of the original forms to investigate the information flows between the stock and futures markets.

3. Methodology

3.1. Granger causality

In this section, the Granger causality test is employed to detect the causal relationship between stocks and index futures. Referring to Ref. [21], we model the Granger causality test with the following two regression equations:

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + u_t \tag{6}$$

$$X_t = \sum_{i=1}^{p} b_i X_{t-i} + \sum_{i=1}^{p} c_i Y_{t-i} + v_t \tag{7}$$

where X denotes the object needed to find the Granger cause, Y denotes the object needed to be determined whether it can Granger cause X, and residuals u_t and v_t are assumed to be mutually independent and individually with zero mean and constant variance. These equations were tested using the following hypothesis:

H_0: Y does not Granger cause X ($c_1 = c_2 = \cdots = c_p = 0$).

F-test could be expressed as follow:

$$F = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(n-2p-1)} \sim F(p, n-2p-1) \tag{8}$$

where RSS_0 is the residual sum of squares of the Eq. (6), RSS_1 is the residual sum of squares of Eq. (7), n is the number of observations, and p is a lag value. We reject the hypothesis H_0 and accept that Y is a Granger cause of X if and only if $F > F(p, n-2p-1)$. The model order p can be determined by minimizing the AIC [29] defined as:

$$AIC(p) = 2\log(|\sigma|) + \frac{2m^2 p}{n} \tag{9}$$

Table 5
Coefficient estimations and corresponding p-values.

Coefficient	Value	Prob.
a_1	0.181850	0.0000**
a_2	−0.109504	0.0001**
a_3	0.020799	0.4502
a_4	0.058902	0.0321*
a_5	−0.060237	0.0261*
b_1	0.187754	0.0000**
b_2	−0.106936	0.0002**
b_3	0.020197	0.4868
b_4	0.067934	0.0190*
b_5	−0.089929	0.0016**
c_1	−0.001195	0.6173
c_2	−0.000854	0.7210
c_3	0.000312	0.8964
c_4	−0.002116	0.3765
c_5	0.007666	0.0013*

Note:
* Denote significance at 5% level.
** Denote significance at 1% level.

where σ is the estimated noise covariance, m is the dimension of the stochastic process and n is the length of the data window used to estimate the model. In order to detect the causal relationship from stocks to index futures, Y should be set to the return sequence of certain stock while X should be set to the return sequence of index futures. Conversely, Y should be set to the return of index futures before detecting the causal relationship from index futures to stocks.

Taking Hushen 300 index futures and the first stock noted as s_1 as an example, we elaborate the processes of the Granger model estimation as follows:

(1) The maximum of the lag value p is set to a fixed number such as 10.
(2) Recursively calculating the total *AIC* of Eqs. (6)–(7) with the p value from 1 to 10, we get the corresponding p of minimum *AIC*. Experimental results show that the optimal p is 5.
(3) Eqs. (6)–(7) are estimated by *OLS* with $p = 5$. The estimations of Eqs. (6)–(7) are shown in Table 5.
(4) F and $F(p, n-2p-1)$ are calculated according to Eq. (8). Results show $F = 2.25489$ and $F(p, n-2p-1) = 2.2206$ (at 95% confidence level).
(5) Because $F > F(p, n-2p-1)$, we conclude that Hushen 300 index futures can Granger cause s_1.

3.2. Conditional Granger causality

According to the conditional Granger causality theory, the conditional Granger causality test can resolve two major problems: first, determine whether the interaction between two time series is direct or mediated by other time series; second, pin down whether the causal relationship is simply owed to different time delays in their respective driving inputs. Taking the trivariate conditional Granger causality model in time domain as an example, we express the joint autoregressive representations of two stochastic processes X_t and Z_t following Ding et al. [22] and Geweke [30]:

$$X_t = \sum_{i=1}^{p} a_{1i} X_{t-i} + \sum_{i=1}^{p} c_{1i} Z_{t-i} + \varepsilon_{1t} \tag{10}$$

$$Z_t = \sum_{i=1}^{p} b_{1i} Z_{t-i} + \sum_{i=1}^{p} d_{1i} X_{t-i} + \varepsilon_{2t} \tag{11}$$

where p in Eqs. (10)–(11) is equivalent to the p in Eqs. (6)–(7). The noise covariance matrix for the system can be represented as:

$$S = \begin{pmatrix} \mathrm{var}(\varepsilon_{1t}) & \mathrm{cov}(\varepsilon_{1t}, \varepsilon_{2t}) \\ \mathrm{cov}(\varepsilon_{1t}, \varepsilon_{2t}) & \mathrm{var}(\varepsilon_{2t}) \end{pmatrix} \tag{12}$$

where var and cov stand for variance and covariance respectively. To judge whether the relationship is mediated by Z, we extend the vector autoregressive representations for a system with three processes X, Z and Y as follows:

$$X_t = \sum_{i=1}^{p} a_{2i} X_{t-i} + \sum_{i=1}^{p} b_{2i} Y_{t-i} + \sum_{i=1}^{p} c_{2i} Z_{t-i} + \varepsilon_{3t} \tag{13}$$

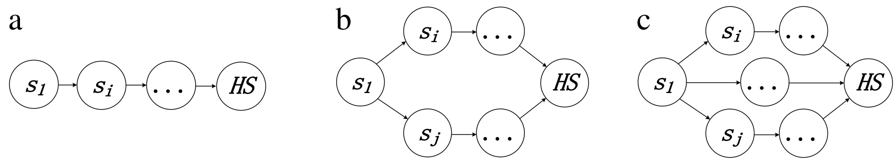

Fig. 3. The possible indirect relationship between s_1 and HS.

$$Y_t = \sum_{i=1}^{p} d_{2i}X_{t-i} + \sum_{i=1}^{p} e_{2i}Y_{t-i} + \sum_{i=1}^{p} f_{2i}Z_{t-i} + \varepsilon_{4t} \tag{14}$$

$$Z_t = \sum_{i=1}^{p} g_{2i}X_{t-i} + \sum_{i=1}^{p} h_{2i}Y_{t-i} + \sum_{i=1}^{p} k_{2i}Z_{t-i} + \varepsilon_{5t}. \tag{15}$$

The noise covariance matrix for the above system can be written as:

$$\Sigma = \begin{pmatrix} \mathrm{var}(\varepsilon_{3t}) & \mathrm{cov}(\varepsilon_{3t},\varepsilon_{4t}) & \mathrm{cov}(\varepsilon_{3t},\varepsilon_{5t}) \\ \mathrm{cov}(\varepsilon_{3t},\varepsilon_{4t}) & \mathrm{var}(\varepsilon_{4t}) & \mathrm{cov}(\varepsilon_{4t},\varepsilon_{5t}) \\ \mathrm{cov}(\varepsilon_{3t},\varepsilon_{5t}) & \mathrm{cov}(\varepsilon_{4t},\varepsilon_{5t}) & \mathrm{var}(\varepsilon_{5t}) \end{pmatrix} \tag{16}$$

where ε_{it}, $i = 1, 2, \ldots, 5$ denote the prediction errors which are mutually independent and individually with zero mean and constant variance. The conditional Granger causality from Y to X conditioned on Z is defined as follow:

$$F_{Y\to X|Z} = \ln\left(\frac{\mathrm{var}(\varepsilon_{1t})}{\mathrm{var}(\varepsilon_{3t})}\right). \tag{17}$$

If the causality from Y to X is entirely mediated by Z, the coefficients b_{2i} in Eq. (13) are uniformly zero and then $\mathrm{var}(\varepsilon_{1t}) = \mathrm{var}(\varepsilon_{3t})$, $F_{Y\to X|Z} = 0$, implying that the inclusion of Y cannot further improve the accuracy of predicting X. On the contrary, if there is direct causality from Y to X, the inclusion of Y will better predict X, implying $\mathrm{var}(\varepsilon_{1t}) > \mathrm{var}(\varepsilon_{3t})$ and $F_{Y\to X|Z} > 0$. The four variables, five variables and even N variables conditional Granger causality models can be defined in the similar ways.

To study the direct relationship from a certain stock in the stock market to index futures, we define s_i as the ith stock in the stock market, HS as Hushen 300 index futures. The processes to determine the conditional Granger causality from s_1 to HS are elaborated as follows:

(1) Granger causality: Granger causality model is used to determine whether s_1 correlates with HS. If s_1 cannot Granger cause HS, s_1 has no causality on HS. Otherwise, we turn to the next step.
(2) Trivariate conditional causality: the conditional Granger causality from s_1 to HS conditioned on s_i $(i \neq 1)$ is calculated according to Eq. (17). If there is a s_i $(i \neq 1)$ subject to $F_{s_1\to HS|s_i} \leq 0$, the causality from s_1 to HS is indirect. Otherwise, we turn to the next step.
(3) Four variables conditional causality: the conditional Granger causality from s_1 to HS conditioned on s_i and s_j $(i \neq j \neq 1)$ is calculated according to Eq. (17). If there are s_i and s_j $(i \neq j \neq 1)$ subject to $F_{s_1\to HS|s_i,s_j} \leq 0$, the causality from s_1 to HS is indirect. Otherwise, we turn to the next step.

……

(k) N variables conditional causality: if the value of causality from s_1 to HS conditioned on $(N-2)$ different stocks (except s_1) is less than or equal to zero, the causality could be considered indirect. Otherwise, causality from s_1 to HS is direct. The maximum number of mediated variables can be defined according to the number of stocks Granger caused by s_1.

Step (1) aims to find out the pairs of stocks and index futures that have Granger relationship and remove the pairs which have no causality with each other. Whereas it leaves three questions: first, whether the causality is direct or mediated by other stocks is indeterminate; second, the relationship from s_1 to HS can be shaped through one or more paths; third, each path can have one or more mediated stocks.

The above procedures are illustrated in Fig. 3. In order to address the problem shown in Fig. 3(a), we apply the trivariate conditional causality model in step (2). Step (2) aims to remove the relationship between s_1 and HS mediated by other stocks through one path by recurrently traversing other stocks. The total number of cycles is C_{n-1}^{1} (supposing the total number of stocks is n). If there is a certain stock enhancing the predictive power of HS, the causality from s_1 to HS is mediated by others. It reveals their relationship is indirect.

After step (2), there is still a left problem shown in Fig. 3(b) that the causality from s_1 to HS is mediated by others through two paths, so the four variables conditional Granger causality model is proposed in step (3). The similar traversal method is applied in step (3) and the total number of cycles is C_{n-1}^{2}, because the number of combinations between s_i and s_j is C_{n-1}^{2}. Analogously, steps $(4-k)$ aim to deal with the problem shown in Fig. 3(c) that causality from s_1 to HS is mediated by others through three or more paths.

Table 6
Correlation coefficients between $IFF(F \to S)$ and $IFF(S \to F)$.

	Hushen 300	CSI 500	SSE 50
Correlation coefficient	0.2380	0.3452	0.1667

3.3. Information flows and direct information flows

In this paper, the information flow is estimated by the Granger causality test while the direct information flow is estimated by the conditional Granger causality test. The degree of Granger causality and degree of conditional Granger causality can be defined with the following methods proposed by Billio et al. [31]:

$$(F \to S)_{GC} = \begin{cases} 1 & \text{if } F \text{ Granger causes } S \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

$$(S \to F)_{GC} = \begin{cases} 1 & \text{if } S \text{ Granger causes } F \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

$$(F \to S)_{CGC} = \begin{cases} 1 & \text{if } F \text{ conditional Granger causes } S \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

$$(S \to F)_{CGC} = \begin{cases} 1 & \text{if } S \text{ conditional Granger causes } F \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

where $(S \to F)_{GC}$ denotes the degree of Granger causality from a certain stock to index futures, and $(S \to F)_{CGC}$ denotes the degree of conditional Granger causality from a certain stock to index futures. The inverse is defined similarly. Therefore the information flow and direct information flow between the stock and futures markets at time t could be defined as follows:

$$IFF(F \to S, t) = \frac{\sum_{i=1}^{n} (F_t \to S_{it})_{GC}}{n} \tag{22}$$

$$IFF(S \to F, t) = \frac{\sum_{i=1}^{n} (S_{it} \to F_t)_{GC}}{n} \tag{23}$$

$$DIFF(F \to S, t) = \frac{\sum_{i=1}^{n} (F_t \to S_{it})_{CGC}}{n} \tag{24}$$

$$DIFF(S \to F, t) = \frac{\sum_{i=1}^{n} (S_{it} \to F_t)_{CGC}}{n} \tag{25}$$

where *IFF* denotes the information flow, *DIFF* denotes the direct information flow, and n represents the total number of stocks.

4. Empirical results

4.1. Results of Granger causality

Figs. 4–6 show the evolution of Granger causality between the stock and index futures markets over time. The confidence level of the Granger model in Eq. (8) is set to 95%.

As shown in Table 6, the correlation coefficient between $IFF(F \to S)$ and $IFF(S \to F)$ of CSI 500 is the largest, followed by Hushen 300 and SSE 50. Their correlation coefficients reflect that the relationship between stocks and CSI 500 futures is most connected.

To further study the symmetry between $IFF(F \to S)$ and $IFF(S \to F)$, we summarize the statistical indicators of the *IFF* of Hushen 300, CSI 500 and SSE 50, (Table 7).

From the mean values shown in Table 7, we demonstrate that the number of Granger causality links from futures to stocks is slightly greater than that from stocks to futures. And mean values of SSE 50 in two opposite directions are very close. Additionally, the values of skewness and kurtosis from stocks to futures are greater than those from futures to stocks. It means that the distribution of the former is more uneven and the degree is more concentrated. Furthermore it reveals that the impact from external factors on the causality from stocks to futures is more pronounced.

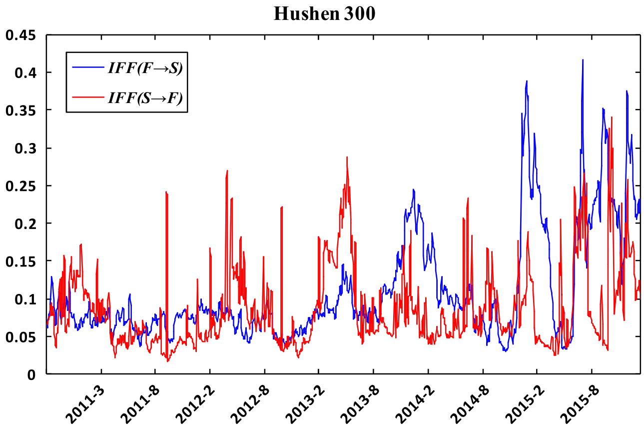

Fig. 4. Granger causality between stocks and Hushen 300 index futures.

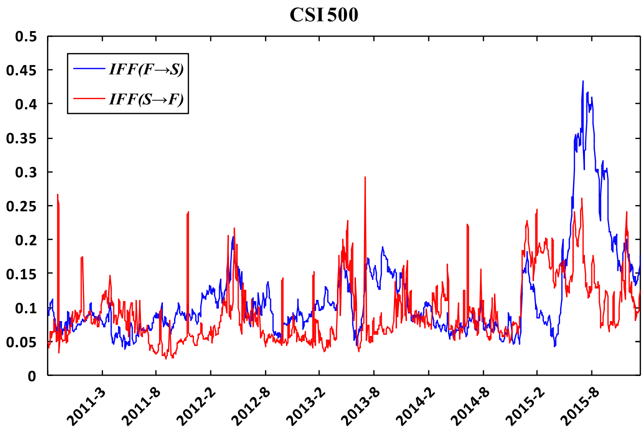

Fig. 5. Granger causality between stocks and CSI 500 index futures.

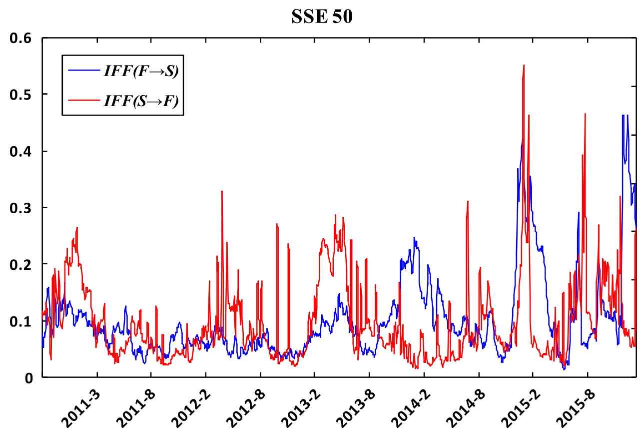

Fig. 6. Granger causality between stocks and SSE 50 index futures.

In order to investigate whether the substantial increase of causality is interrelated with the market price movement, we divided the observations of all stocks and information flows into blocks of 240 days (nearly a year of the market date), and investigate the dynamic evolution of average correlation coefficients between stock prices and information flows over time. The average correlation coefficient at time t is defined as the following formula:

$$\bar{\rho}_t = \frac{\sum_{i=1}^{n} \rho_{it}}{n} \tag{26}$$

Table 7
Statistical indicators of *IFF*.

	IFF of Hushen 300		*IFF* of CSI 500		*IFF* of SSE 50	
	$(F \rightarrow S)$	$(S \rightarrow F)$	$(F \rightarrow S)$	$(S \rightarrow F)$	$(F \rightarrow S)$	$(S \rightarrow F)$
Mean	0.1102	0.0878	0.1147	0.0947	0.1022	0.0991
Median	0.0814	0.0633	0.0922	0.0759	0.0814	0.0705
Maximum	0.4575	0.6835	0.4593	0.6872	0.5045	0.7432
Minimum	0.0253	0.0145	0.0362	0.0217	0.0127	0.0109
Std. Dev.	0.0731	0.0702	0.0685	0.06	0.0759	0.0898
Skewness	1.7057	2.9213	2.2796	3.1567	2.2608	2.8745
Kurtosis	5.6159	15.4529	8.7923	22.4749	8.7753	15.0677

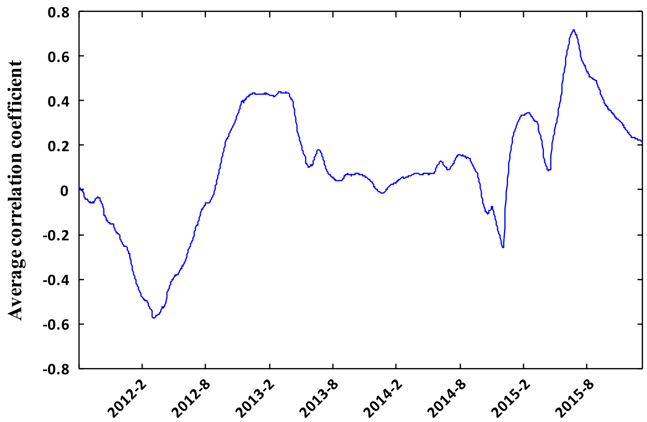

Fig. 7a. Yearly average correlation coefficients between stock prices and information flows from stocks to Hushen 300 index futures. The value at each temporal point means the average correlation coefficient in the past year.

where $\bar{\rho}_t$ denotes the average correlation coefficient at time t, ρ_{it} denotes the correlation coefficient between the price of the ith stock and the information flow at time t, and n denotes the number of stocks. The evolution of $\bar{\rho}_t$ of Hushen 300 index futures over time t is shown in Fig. 7a.

It is remarkable that the average correlation coefficient also achieved several local maximums. It indicates that the fluctuations of stock prices are closely relative to the Granger causality and could possibly affect the Granger causality between the stock and futures markets.

One could see that the information flows experienced several substantial increases around November 2010, December 2011, October 2012, July 2013, October 2014, and June 2015, and respectively reached local maximums around January 2011, April 2012, May 2013, January 2014, January 2015, and August 2015. As known to all, fluctuations of stock prices are closely relative to the rumors of financial market, political changes, and fiscal policies. This motivated us to consider the occurrences of market events.

Fig. 7a shows that the average correlation coefficient between the stock price and the information flow was negative before August 2012 and reached a minimum value near −0.6 in March 2012. It indicated that the drop of stock prices would lead to the increases of information flows between stocks and futures during this period, combined with the plot showing of mean stock prices (Fig. 7b). Moreover, the average correlation coefficient reached a maximum value over 0.7 in June 2015, uncovering the closely positive correlation during this period: increases of stock prices would increase the correlation between stocks and futures.

Based on Figs. 4–6 and 7c which shows the corresponding important events, we can conclude two points: first, the increases of information flows estimated by Granger causality tests are relative to the market volatility; second, the Chinese stock market's 'stock disaster' in 2015 had the greatest impact on the relationship between the stock and futures markets.

4.2. Results of conditional Granger causality

Similar to the Granger causality, the conditional Granger causality also frequently fluctuates over time (Figs. 8–10). Nevertheless, the fluctuations of $DIFF(S \rightarrow F)$ are greater than $DIFF(F \rightarrow S)$ in contrast to Granger causality. It means that the direct information flows from stocks to futures are more vulnerable to external factors than those in the reverse direction.

Different from information flows, the fluctuation magnitude of direct information flows around the Chinese stock market's 'stock disaster' in 2015 was smaller, especially for SSE 50. Moreover, for Hushen 300 and SSE 50, the fluctuation magnitude around 2011 and 2013 was greater than the magnitude in 2015.

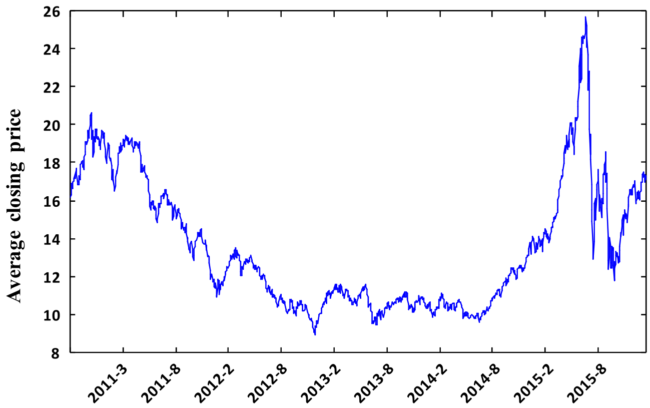

Fig. 7b. The average daily closing prices of stocks.

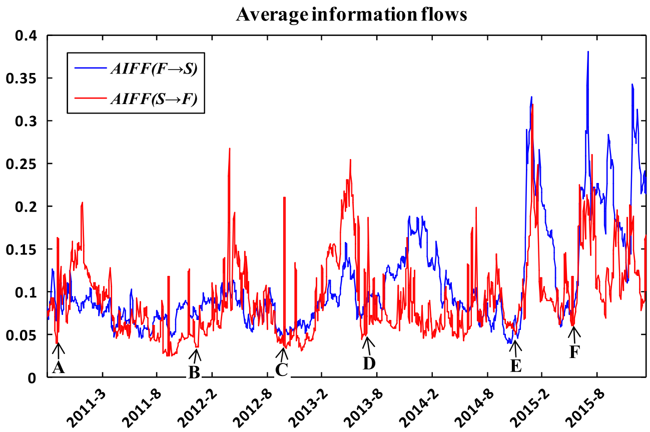

Fig. 7c. Average information flows, which are the average of flows between stocks and 3 index futures. $AIFF(F \rightarrow S)$ denotes average information flows from futures to stocks. $AIFF(S \rightarrow F)$ denotes average information flows from stocks to futures. The following important events happened before the substantial increases of information flows: (A) Shanghai Stock Exchange fell more than 5% in November 2010, which was caused by the expectation of raising interest rates on bank savings; (B) The central bank lowered the reserve deposit ratio of banks by 0.5 percentage points in December 2011; (C) China Development Bank started China credit assets securitization with CNY 10.16 billion products in September 2012; (D) China fully liberalized the financial institutions lending rate control on July 20, 2013; the Chinese stock market's 'stock disaster' in 2015, which experienced two periods: (E) The distinguishable increases of stock prices in the stock market for the period between June 2014 and February 2015; (F) The sharp drop of stock prices and Shanghai composite index in June 2015.

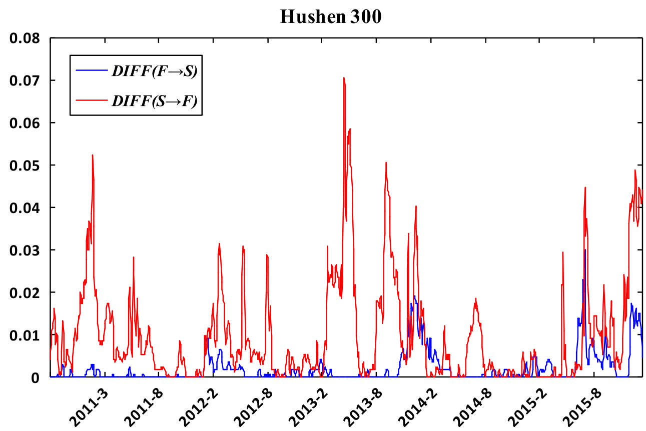

Fig. 8. Conditional Granger causality between stocks and Hushen 300 index futures.

To further study the symmetry between $DIFF(F \rightarrow S)$ and $DIFF(S \rightarrow F)$, we summarize the statistical indicators of *DIFF* of Hushen 300, CSI 500 and SSE 50, and show them in Table 8. Table 8 shows consistent results that mean values of

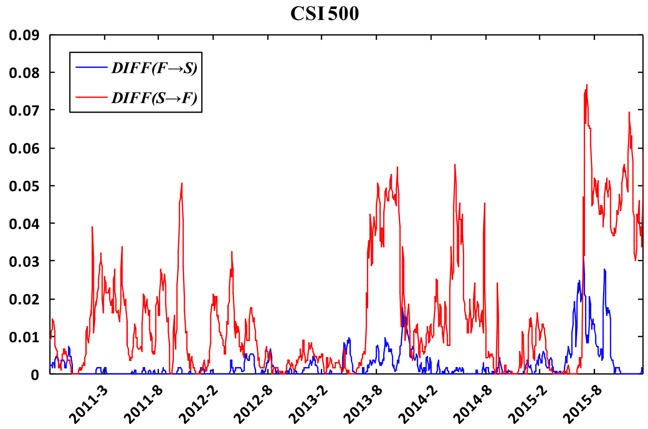

Fig. 9. Conditional Granger causality between stocks and CSI 500 index futures.

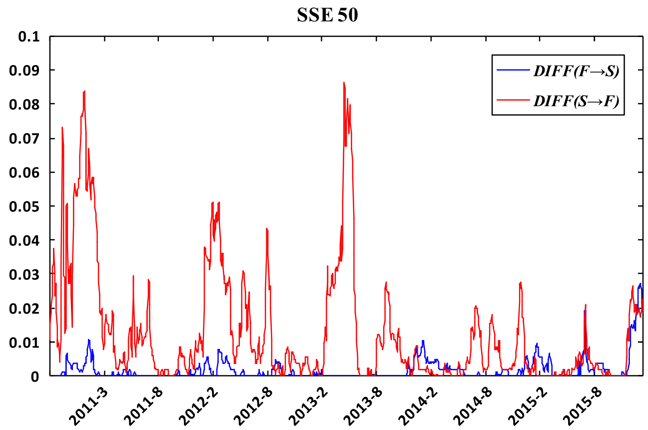

Fig. 10. Conditional Granger causality between stocks and SSE 50 index futures.

Table 8
Statistical indicators of *DIFF*.

	DIFF of Hushen 300		*DIFF* of CSI 500		*DIFF* of SSE 50	
	$(F \to S)$	$(S \to F)$	$(F \to S)$	$(S \to F)$	$(F \to S)$	$(S \to F)$
Mean	0.002	0.0099	0.0025	0.0154	0.0018	0.0127
Median	0	0.0054	0	0.009	0	0.0054
Maximum	0.038	0.0796	0.038	0.0958	0.0289	0.1157
Minimum	0	0	0	0	0	0
Std. Dev.	0.004	0.0125	0.0046	0.0171	0.0036	0.0176
Skewness	3.2259	1.8719	3.2277	1.3685	3.9333	2.0901
Kurtosis	16.552	6.5792	15.7549	4.2628	22.9736	7.5109

DIFF from stocks to futures are greater than those in the reverse direction. Additionally, the values of skewness and kurtosis from futures to stocks are both greater than those in the reverse direction. The different performance between the Granger causality test and the conditional Granger causality test in relationship mining is revealed with Tables 7 and 8.

4.3. Analysis of information flows and direct information flows

The above empirical results of Granger and conditional Granger causality models respectively show characteristics of information flows and direct information flows. To better understand the similarity and difference between the information flows and direct information flows, we show them together in Figs. 11–16:

Figs. 11–16 intuitively show that the direct information flows are less than information flows, especially for the direction from futures to stocks.

Table 9 shows that the correlation coefficients between *IFF* and *DIFF* from futures to stocks are greater than the correlation coefficients from stocks to futures. It is remarkable that for CSI 500, the correlation coefficient between *IFF* and *DIFF* from

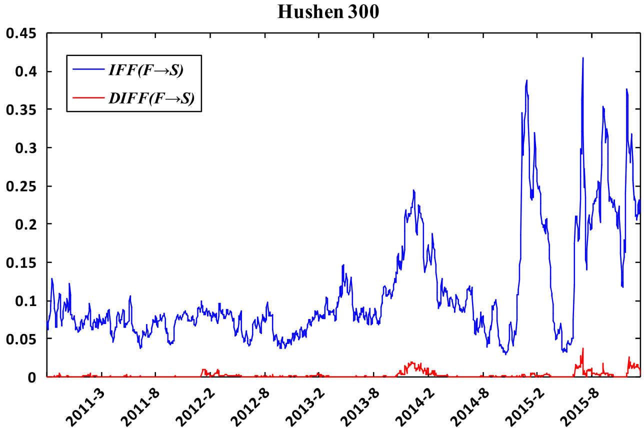

Fig. 11. Granger and conditional Granger causality from Hushen 300 index futures to stocks.

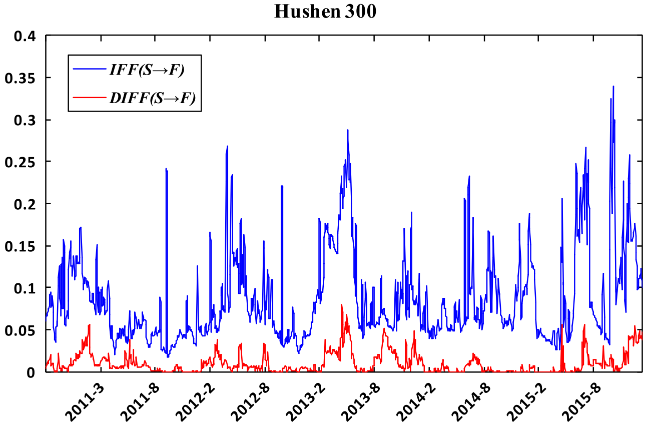

Fig. 12. Granger and conditional Granger causality from stocks to Hushen 300 index futures.

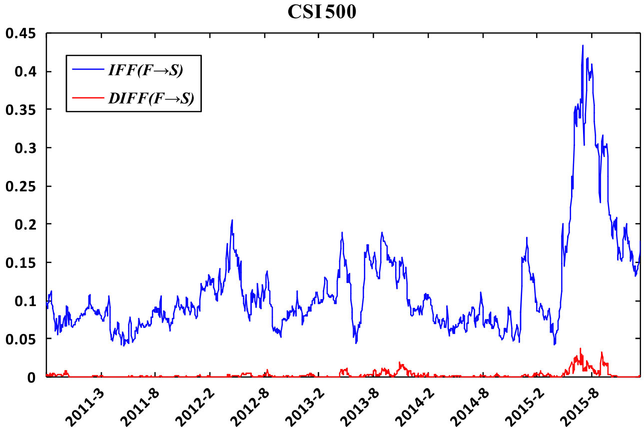

Fig. 13. Granger and conditional Granger causality from CSI 500 index futures to stocks.

Table 9
Correlation coefficients between *IFF* and *DIFF*.

	$(F \rightarrow S)$	$(S \rightarrow F)$
Hushen 300	0.5972	0.3675
CSI 500	0.7281	0.0869
SSE 50	0.6476	0.3542

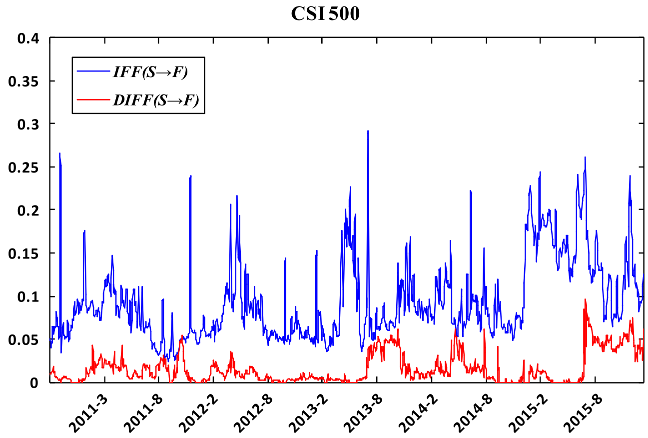

Fig. 14. Granger and conditional Granger causality from stocks to CSI 500 index futures.

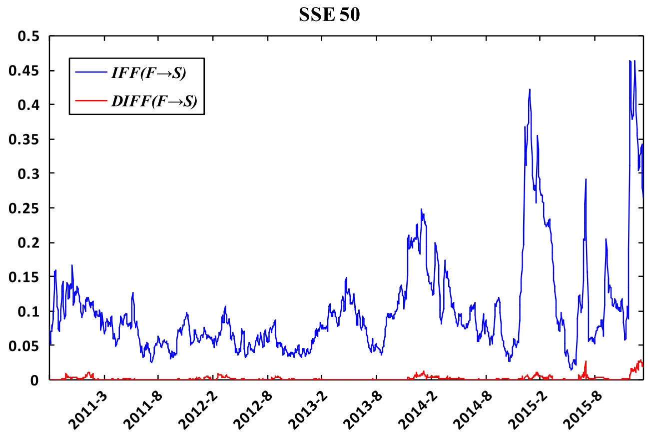

Fig. 15. Granger and conditional Granger causality from SSE 50 index futures to stocks.

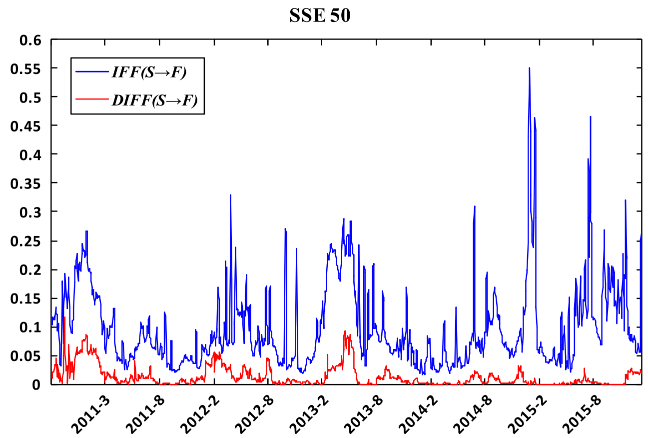

Fig. 16. Granger and conditional Granger causality from stocks to SSE 50 index futures.

stocks to futures is the smallest, while the correlation coefficient between *IFF* and *DIFF* in the reverse direction is oppositely the greatest in all mentioned index futures. Fig. 7c shows the correlation between the occurrences of important events and the dramatic increases of information flows. Thus from Fig. 7c and Table 9, we can conclude that the direct information flows correspond to important events.

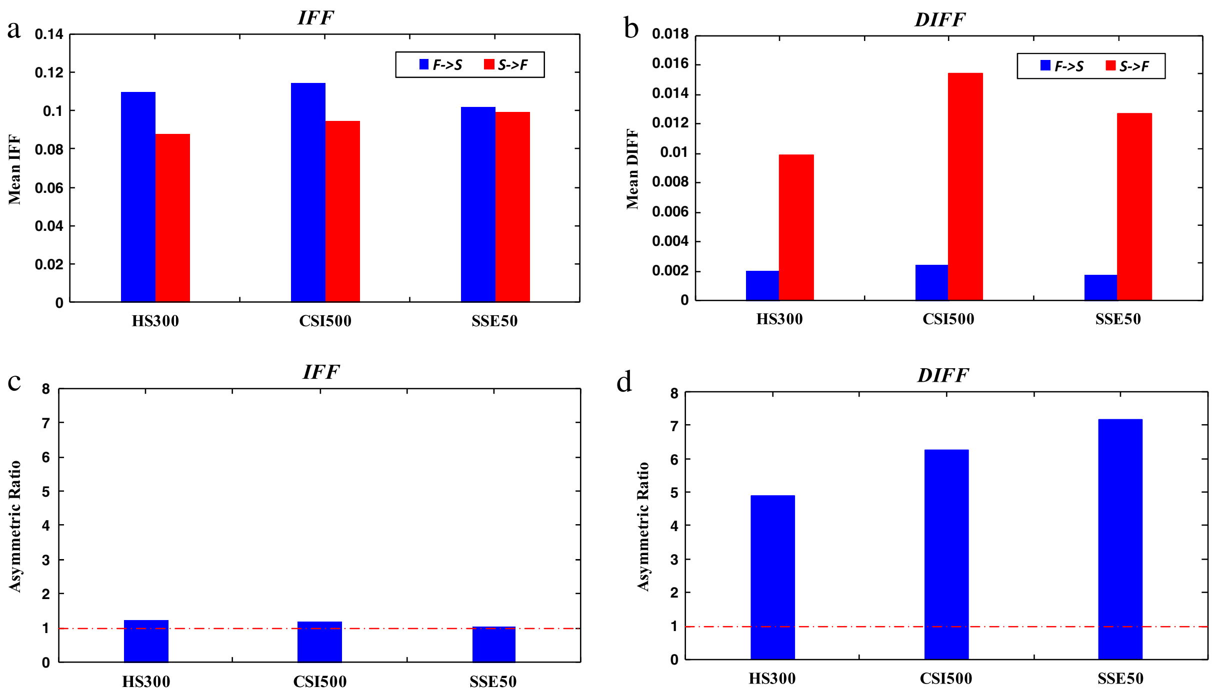

Fig. 17. Mean values and asymmetric ratios of *IFF* and *DIFF*. The asymmetric ratio more than 1 reflects the asymmetry of information flows.

Table 10
Bootstrap values of asymmetric ratios for *IFF* and *DIFF*.

	IFF	*DIFF*
Hushen300	76.2%	98.3%
CSI500	88.7%	99.6%
SSE50	45.6%	96.4%

4.4. Bootstrap tests for information flows and direct information flows

Bootstrap technique is widely used for estimating statistical significance [32–34]. [32-34] Before the application of the bootstrap technique, we quantify some indicators. In this paper, the symmetry of *IFF* and *DIFF* is quantified by asymmetric ratios. Specifically, the asymmetric ratio of *IFF* is calculated by dividing the mean value of $IFF(F \to S)$ by the mean value of $IFF(S \to F)$, and the asymmetric ratio of *DIFF* is calculated by dividing the mean value of $DIFF(S \to F)$ by the mean value of $DIFF(F \to S)$. Results are shown in Fig. 17.

The significance value of the asymmetric ratio is obtained with the bootstrap technique. Taking *IFF* as example, we show the bootstrap processes as follows:

(1) Generate a sample $B = \{X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_p\}$ according to the tested sequences in Eqs. (6)–(7).
(2) Generate a bootstrap sample $B^* = \{X_1^*, X_2^*, \ldots, X_p^*, Y_1^*, Y_2^*, \ldots, Y_p^*\}$ by randomly sampling half of the observations of B without replacement.
(3) Compute $(S \to F)_{GC}$ and $(F \to S)_{GC}$ based on the bootstrap sample B^*.
(4) Compute *IFF* by repeating steps (1–3) for all the stocks.
(5) Compute the asymmetric ratio (noted as AR_i $i = 1, 2, \ldots, 1000$).
(6) Repeat steps (1–5) 1000 times and obtain an asymmetric ratio collection $ARC = \{AR_1, AR_2, \ldots, AR_{1000}\}$. Typically in the systemic analysis 1000 is thought of as an enough number of replicas.
(7) Compute the percentage of AR_i $i = 1, 2, \ldots, 1000$ greater than 1 and get the bootstrap value of the asymmetric ratio. The bootstrap processes of *DIFF* are defined in similar ways. The results of bootstrap processes are shown in Table 10.

As shown in Table 10, the asymmetric ratios of *DIFF* are all significantly greater than 1 at the confidence level of 95%. However, the bootstrap values of asymmetric ratios of *IFF* are less than 95%.

From Fig. 17 and Table 10, we can draw 3 conclusions:

(1) The direct information flows from stocks to futures are significantly greater than the flows from futures to stock.
(2) The information flows from futures to stocks are slightly greater than the flows from stocks to futures.
(3) For SSE 50, the causal relationship between stocks and futures is unstable in the sense that the bootstrap value is less than 50%.

5. Conclusion

Based on Granger and conditional Granger causality models, we investigate the information flows and direct information flows between the Chinese stock market and futures market. In our studies, the information flow is estimated by the Granger causality test while the direct information flow is estimated by the conditional Granger causality test. We also study the evolution of causality links over time to reveal the relationship between market behaviors and the intermittent changes of information flows.

First, we respectively investigate the bidirectional Granger causality and conditional Granger causality both from stocks to index futures and from index futures to stocks, and sum up the number of causality links at each sliding window to quantify information flows and direct information flows at the corresponding time.

Further, we study the quantitative characteristics of information flows and direct information flows by self-comparison and cross-comparison of Granger and conditional Granger causality from stocks to futures and the reverse direction. In addition, the introductions of occurrence events help us better understand the characteristics of information flows and direct information flows in the analysis of empirical results. Results show the synchronism between occurrences of important events and sudden increases of information flows. We also reveal that the correlation between stock prices and information flows evolved from negative to positive during the observation period.

Finally, the comparative analysis with the asymmetric ratio and the bootstrap technique demonstrates the difference between information flows and direct information flows in relationship mining. Results show that the information flows from futures to stocks are slightly greater than those from stocks to futures while the direct information flows turn out the opposite result that the flows from stocks to futures are significantly greater than those from futures to stocks. Our work potentially reveals the influence of important events or market volatility on fluctuations of information flows and direct information flows. The empirical results of our study would be valuable for marketing policy and investor decision making.

Acknowledgments

The authors declare that there is no conflict of interest regarding the publication of this paper. This research was supported by the National Natural Science Foundation of China (Grant 171201060), Research Fund for the Doctoral Program of Higher Education (Grant 20120172120051), and the Fundamental Research Funds for the Central Universities (Grant 2015ZZ059, 2015ZDXM04).

References

[1] C. Haythornthwaite, Social network analysis: An approach and technique for the study of information exchange, Libr. Inf. Sci. Res. 18 (4) (1996) 323–342.
[2] Y.W. Cheung, H.G. Fung, Information flows between Eurodollar spot and futures markets, Multinatl. Financ. J. 1 (1997) 255–271.
[3] O. Gabjin, O. Tamina, K. Hoyong, K. Okyu, An information flow among industry sectors in the Korean stock market, J. Korean Phys. Soc. 65 (2014) 2140–2146.
[4] L. Barnett, A.B. Barrett, A.K. Seth, Granger causality and transfer entropy are equivalent for Gaussian variables, Phys. Rev. Lett. 103 (23) (2009) 4652–4657.
[5] C.-Z. Yao, J.-N. Lin, X.-F. Liu, A study of hierarchical structure on South China industrial electricity-consumption correlation, Physica A 444 (2016) 129–145.
[6] K.D. Garbade, W.L. Silber, Price movements and price discovery in futures and cash markets, Rev. Econ. Stat. 65 (2) (1983) 289–297.
[7] J. Fleming, B. Ostdiek, R.E. Whaley, Trading costs and the relative rates of price discovery in stock, futures, and option markets, J. Futures Mark. 16 (4) (1996) 353–387.
[8] K. Chan, A further analysis of the lead-lag relationship between the cash market and stock index futures market, Rev. Financ. Stud. 5 (1992) 123–152.
[9] Y.K. Tse, W.S. Chan, The lead-lag relation between S&P500 spot and futures markets: An intraday-data analysis using threshold regression model, Jpn. Econ. Rev. 6 (2010) 134–144.
[10] H. Ding, S. Zeng, X.J. Guo, Influence of stock index futures on stock market price: Theoretical analysis and experiences of the Chinese and overseas markets, Int. J. Econ. Finance 3 (4) (2011) 113–118.
[11] Z. Zakaria, S. Shamsuddin, Relationship between stock futures index and cash prices index: Empirical evidence based on Malaysia data, J. Bus. Stud. Quart. 4 (2) (2012) 103–112.
[12] O. Gwilym, M. Buckle, The lead-lag relationship between the FTSE100 stock index and its derivative contracts, Appl. Financ. Econ. 11 (2001) 385–393.
[13] A.H. Abhyankar, Return and volatility dynamics in the FT-SE 100 stock index and stock index futures markets, J. Futures Mark. 15 (1995) 457–488.
[14] A. Antoniou, P. Holmes, Futures trading, information and spot price volatility: evidence for the FTSE-100 Stock Index Futures contract using GARCH, J. Bank. Finance 19 (1995) 117–129.
[15] J. Li, J.K.W. Fung, L.T.W. Cheng, The lead-lag relation between spot and futures markets under different short-selling regimes, Financ. Rev. 36 (3) (2001) 63–88.
[16] Z. Fang, R.D. Chen, The influence from stock index futures to stock market volatility, in: 2011 Fourth International Conference on Business Intelligence and Financial Engineering, BIFE, October, Wuhan, China 4, 2011, pp. 676–678.
[17] C.-C. Gong, S.-D. Ji, L.-L. Su, S.-P. Li, F. Ren, The lead–lag relationship between stock index and stock index futures: A thermal optimal path method, Physica A 444 (2016) 63–72.
[18] S. Bose, Contribution of India index futures to price formation in the stock market, Money Finance 3 (1) (2007) 39–55.
[19] O. Kwon, G. Oh, Asymmetric information flow between market index and individual stocks in several stock markets, EPL (Europhys. Lett.) 97 (2012) 28007.
[20] H.Y. Kim, O. Kwon, G. Oh, A causality between fund performance and stock market, Physica A 443 (2016) 439–450.
[21] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (3) (1969) 424–438.
[22] M. Ding, Y. Chen, S. Bressler, Granger causality: basic theory and application to neuroscience, in: S. Schelter, M. Winterhalder, J. Timmer (Eds.), Handbook of Time Series Analysis, Wiley, Wienheim, 2006, pp. 438–460.
[23] Z. Zhou, Y. Chen, M. Ding, P. Wright, Z. Lu, Y. Liu, Analyzing brain networks with PCA and conditional Granger causality, Hum. Brain Mapp. 30 (2008) 2197–2206.

[24] Q. Gao, X. Duan, H. Chen, Evaluation of effective connectivity of motor areas during motor imagery and execution using conditional Granger causality, NeuroImage 54 (2) (2011) 1280–1288.
[25] J.D. Hamilton, Time Series Analysis, Princeton University, Chichester, West Sussex, 1994, pp. 528–547.
[26] R.F. Engle, D.F. Kraft, Multiperiod Forecast Error Variances of Inflation Estimated from Arch Models, in: Applied Time Series Analysis of Economic Data, 1983.
[27] R.F. Engle, An Introduction to the Use of ARCH/GARCH models in Applied Econometrics, NYU Working Paper, 2001.
[28] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation, Econometrica 50 (1982) 987–1008.
[29] H. Akaike, Fitting autoregressive models for regression, Ann. Inst. Statist. Math. 21 (1969) 243–247.
[30] J. Geweke, Measurement of linear dependence and feedback between multiple time series, J. Amer. Statist. Assoc. 77 (1982) 304–313.
[31] M. Billio, M. Getmansky, A.W. Lo, L. Pelizzon, Econometric measures of connectedness and systemic risk in the finance and insurance sectors, J. Financ. Econ. 104 (2012) 535–559.
[32] C. Krajewski, A.W. Dickerman, Bootstrap analysis of phylogenetic trees derived from DNA hybridization distances, Syst. Biol. 39 (4) (1990) 383–390.
[33] A. Zharkikh, W.H. Li, Estimation of confidence in phylogeny: The complete and partial bootstrap technique, Mol. Phylogenet. Evol. 4 (1) (1995) 44–63.
[34] B.Q. Minh, M.A.T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic bootstrap, Mol. Biol. Evol. 30 (5) (2013) 1188–1195.