Physica A 444 (2016) 63–72

Contents lists available at ScienceDirect

Physica A

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

The lead–lag relationship between stock index and stock index futures: A thermal optimal path method

Chen-Chen Gong [a], Shen-Dan Ji [a], Li-Ling Su [a], Sai-Ping Li [b], Fei Ren [a,c,d,*]

[a] *School of Business, East China University of Science and Technology, Shanghai 200237, China*

[b] *Institute of Physics, Academia Sinica, Taipei 115, Taiwan*

[c] *Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China*

[d] *Engineering Research Center of Process Systems Engineering (Ministry of Education), East China University of Science and Technology, Shanghai 200237, China*

HIGHLIGHTS

- We use a non-parametric approach, thermal optimal path method.
- We study lead–lag relationship between stock indices and their associated futures.
- We propose an indicator to explain the surge of lead–lag function.

ARTICLE INFO

Article history:
Received 5 July 2015
Received in revised form 3 September 2015
Available online 14 October 2015

Keywords:
Econophysics
Lead–lag relationship
Stock markets

ABSTRACT

The study of lead–lag relationship between stock index and stock index futures is of great importance for its wide application in hedging and portfolio investments. Previous works mainly use conventional methods like Granger causality test, GARCH model and error correction model, and focus on the causality relation between the index and futures in a certain period. By using a non-parametric approach—thermal optimal path (TOP) method, we study the lead–lag relationship between China Securities Index 300 (CSI 300), Hang Seng Index (HSI), Standard and Poor 500 (S&P 500) Index and their associated futures to reveal the variance of their relationship over time. Our finding shows evidence of pronounced futures leadership for well established index futures, namely HSI and S&P 500 index futures, while index of developing market like CSI 300 has pronounced leadership. We offer an explanation based on the measure of an indicator which quantifies the differences between spot and futures prices for the surge of lead–lag function. Our results provide new perspectives for the understanding of the dynamical evolution of lead–lag relationship between stock index and stock index futures, which is valuable for the study of market efficiency and its applications.

1. Introduction

Concepts such as lead–lag relationship and price discovery are commonly used to study market efficiency in finance. For a capital market that is not perfectly efficient, there exists a lead–lag relationship between stock index and its associated

* Correspondence to: 130 Meilong Road, P.O. Box 114, School of Business, East China University of Science and Technology, Shanghai 200237, China. Tel.: +86 21 64253369; fax: +86 21 64253152.

E-mail address: fren@ecust.edu.cn (F. Ren).

http://dx.doi.org/10.1016/j.physa.2015.10.028
0378-4371/

futures, namely the index or its futures has price discovery function. A stock index future contract is a legally binding agreement to buy or sell a specific quantity of underlying stock market index in the future at an agreed price, which is the future price on today. Stock index futures are initially designed for hedging to prevent the risk of high fluctuations in stock markets. It can also be used for portfolio investments by exploiting the profitable opportunities between spot and futures prices. Revealing the lead–lag relationship between stock index and its associated futures is supposed to be conducive to strategy trading in hedging and portfolio investments.

Much work has been devoted to the study of the relationship between stock index and stock index futures. Diametrically conflicting conclusions are drawn for studies in different markets and time periods, and by using different methods. By analyzing intraday data, many studies have shown that futures price movements lead index movements in the short-run for a variety of stock indices such as S&P 500 [1–4], Nasdaq 100 [5], FTSE 100 [6,7], DAX [8], and S&P CNX Nifty [9]. There are also studies using daily data to show that futures price leads the spot price with daily lags, for instance Nikkei stock average [10] and Ibex 35 [11]. Other studies report the evidence of opposite opinion that the index spot price leads the futures price, for instance the intraday data analysis of French CAC index [12]. The long-run relationship between daily spot and futures prices for S&P 500 and FTSE 100 is cointegrated, and the lead role of spot returns is more pronounced [13]. Similar results are observed in Malaysian stock market [14] and Turkish stock market [15]. In addition, other studies have shown that the index and futures have a bidirectional relationship [3,16,17]. The lead and lag relationships between spot and futures prices of the Hang Seng stock average are established in the short-run [16], and the long-run relationship between the daily ISE 30 index and its futures is bidirectional [17].

An increasing number of empirical studies of the relationship between stock indices and the prices of their futures contract show that the price discovery process may evolve over time [6,18–26]. As an example, the FTSE 100 futures price leads its spot price in the mean in all observation periods, but there exists the evidence of spot returns having predictive power for the futures in some particular periods [6]. The findings in Ref. [20] show evidence of pronounced futures leadership when markets are rising. When markets are falling, futures leadership is less evident and significant feedback from the cash market is observed. Ref. [21] reported that the spot market leads the futures market during the period when structure change occurs in the spot index, and the spot market is more informationally efficient than the futures market under high variance condition in Ref. [22]. In the Warsaw Stock Exchange, price discovery occurs mainly in the spot market during the period when private investors account for the major amount of futures trading volume. After that, the futures market experiences a steep increase of trading volume shares of institutional investors and therefore the price discovery contribution of the futures market tends to rise [24]. An analysis of the daily data of HSI index and its futures indicates that the lead–lag relationship is time varying, displaying a non-linear threshold-type behavior [25]. The dynamics of price discovery process in DAX is also studied in Ref. [26].

The relationships between spot and futures prices of emerging markets have recently drawn much attention and have been extensively studied [9,14,15,27–31]. China has the largest emerging market in the world, and the CSI 300 index reflects the overall trend of Chinese stocks market. It is thus worth to study the relationship between CSI 300 index and its futures. The CSI 300 index futures was established on April 16, 2010. The underlying asset of this contract is the CSI 300 index which consists of 300 representative stocks that are most capitalized and liquid, and traded on the two stock exchanges (Shanghai Stock Exchange and Shenzhen Stock Exchange) in mainland China. To the best of our knowledge, previous studies of the relationship between CSI 300 index and its futures mainly use conventional methods like Granger causality test, GARCH model and error correction model [32]. Based on the daily data of main contracts of CSI 300 index futures from April 16, 2010 to January 14, 2014, we use a new method to study the dynamic evolution of lead–lag relationship between CSI 300 index and its futures. We also make a comparison study between the Chinese index futures with other well established index futures, namely HSI and S&P 500 index futures.

Unlike previous studies on price discovery in stock index futures, we employ a thermal optimal path (TOP) method to study the dynamic evolution of the relationship between CSI 300 index and its futures. The TOP method was initially proposed to test the dynamic time evolution of the lead–lag structure between two time series [33], and has been successfully applied to the correlation analysis of economic and financial data [33–36]. Conventional methods like cross-correlation function, Granger causality test, GARCH model, vector autoregressive model and error correction model are generally used to detect the dependence of two series in a certain period, but cannot capture the variance of correlation over time [33]. In this respect, TOP offers a new and different approach to the detection of dependences between two time series. We use this TOP method to study the lead–lag relationship between the spot and futures prices of the CSI 300, HSI and S&P 500 indices, and try to reveal the variance of their dependences over time. As a comparison study, we also include the results from the error correction model to detect the dependence between the time series of the spot and futures indices.

2. Data

The dataset used in our analysis is retrieved from the Bloomberg database. We use the daily closing prices for China Securities Index 300 (CSI 300), Hang Seng Index (HSI), and Standard and Poor 500 (S&P 500) Index, for the period between April 16, 2010 and January 14, 2014. For the futures price data, we take the closing price of the main futures contract. There are several types of futures contracts for these three indices. CSI 300 futures has four types of contracts according to different delivery month: current month, next month, and two of the months from among March, June, September, and December following the current month. HSI has four types of contracts the same as CSI 300 futures. The delivery month of S&P 500 futures includes March, June, September, and December, and it has eight contracts currently traded within a period of two years. The main contract is the most active contract among them, which has the largest trading volume on the current day. This helps to construct the time series of index futures prices more consecutive and sufficient. The closing prices of the main futures contracts during the same period are collected as the futures prices.

We denote $P_s(t)$ and $P_f(t)$ as the spot and futures prices of the stock index at time t. The returns of the stock index and stock index futures are calculated as the difference between the logarithmic prices at t and $t-1$, following the formula

$$R_s(t) = \ln P_s(t) - \ln P_s(t-1), \tag{1a}$$

$$R_f(t) = \ln P_f(t) - \ln P_f(t-1). \tag{1b}$$

The return series $R_s(t)$ and $R_f(t)$ defined above are used to quantify the lead–lag relationship between stock index and stock index futures for CSI 300, HSI, and S&P 500.

3. Preliminary analysis by error correction model

The error correction model (ECM) is a typical conventional method used to detect the dependence between the series of stock index and stock index futures. The ECM method was developed by Engle and Granger in 1987, and was used to investigate the lead–lag relationship between spot and futures series for US and UK markets in the early 1990s [13,37]. It is now widely used in econometrics as an efficient method of detecting the relationship between two non-stationary time series.

Though the spot and futures logarithmic prices $\ln P_s(t)$ and $\ln P_f(t)$ are non-stationary, their returns $R_s(t)$ and $R_f(t)$ defined in Eqs. (1a) and (1b) are shown to be stationary. We have performed the Augmented Dicky–Fuller (ADF) test to prove it. $\ln P_s(t)$ and $\ln P_f(t)$ are therefore said to be integrated. If there is a stationary linear relation between $\ln P_s(t)$ and $\ln P_f(t)$, these two time series are cointegrated. If cointegration exists between $\ln P_s(t)$ and $\ln P_f(t)$, the ECM can be expressed as

$$R_s(t) = c_s + \sum_{i=1}^{q} \alpha_{s,i} R_f(t-i) + \sum_{i=1}^{q} \beta_{s,i} R_s(t-i) + \gamma_s e_{s,t-1} + \nu_{s,t}, \tag{2a}$$

$$R_f(t) = c_f + \sum_{i=1}^{q} \alpha_{f,i} R_s(t-i) + \sum_{i=1}^{q} \beta_{f,i} R_f(t-i) + \gamma_f e_{f,t-1} + \nu_{f,t}, \tag{2b}$$

where $e_{s,t-1}$ and $e_{f,t-1}$ are cointegration errors, and $\nu_{s,t}$ and $\nu_{f,t}$ are stochastic error terms. The cointegration error $e_{s,t}$ is obtained by a linear regression between $\ln P_s(t)$ and $\ln P_f(t)$, and the general form is $\ln P_s(t) = a_s + b_s \ln P_f(t) + e_{s,t}$. In addition, we take into account the contributions of auto-regression and higher order terms, and introduce a regression formula as $\ln P_s(t) = a_s + \sum_{i=1}^{q} b_{s,i} \ln P_f(t-i) + \sum_{i=1}^{q} d_{s,i} \ln P_s(t-i) + e_{s,t}$, which has a similar form as ECM. The estimated coefficients $\hat{b}_{s,i}$ and $\hat{d}_{s,i}$ are constant and significant, indicating that $\ln P_s(t)$ and $\ln P_f(t)$ are cointegrated. Therefore $e_{s,t}$ can be obtained by $e_{s,t} = \ln P_s(t) - a_s + \sum_{i=1}^{q} \hat{b}_{s,i} \ln P_f(t-i) + \sum_{i=1}^{q} \hat{d}_{s,i} \ln P_s(t-i)$. The cointegration error $e_{f,t}$ can be obtained by the regression of $\ln P_f(t)$ on $\ln P_s(t)$ in a similar way. If the coefficient $\alpha_{s,i}$ is statistically significant, the futures leads the index; on the contrary, if the coefficient $\alpha_{f,i}$ is statistically significant, the index leads the futures.

As a comparison study of the relationship between the index and index futures, we here use the error correction model (ECM) to perform a causality analysis using all data from April 16, 2010 till January 14, 2014. In addition, Granger causality test has also been performed, and similar results are obtained (not shown in this paper). Table 1 reports the estimated coefficients of ECM and their p-values for the order $q = 3$ using data from CSI 300, HSI and S&P 500 indices and their futures. For SCI 300 and its futures, the coefficient $\alpha_{f,2}$ is statistically significant at 5% level, and none of the coefficients $\alpha_{s,i}$ $(i = 1, 2, 3)$ are statistically significant. This means that the CSI 300 index leads its futures and the change of spot price is two days ahead of the change of futures price. For HSI and S&P 500, we observe opposite result suggesting that more information flows from futures market to spot market. The coefficient $\alpha_{s,1}$ is statistically significant at 5% level, and none of the coefficients $\alpha_{f,i}$ $(i = 1, 2, 3)$ are statistically significant. This suggests that the futures of HSI and S&P 500 lead their indices and the changes of futures prices are one day ahead of the changes of their spot prices. One can conclude that in developed markets, e.g. Hong Kong and US stock markets the leading role of index futures is more significant, while in emerging markets like mainland China the leading role of stock index is more significant. Our results show that the maximum of leading order is less than 3, therefore there is no need to consider ECM with higher q order.

4. Result from thermal optimal path method

4.1. Method

The thermal optimal path (TOP) method is a novel non-parametric methodology to test the dynamic evolution of the lead–lag relationship between two time series. In comparison with other conventional methods, one of the advantages of the TOP method is that there is no need to take into account extra parameters like number of regression variables and order

Table 1
Coefficients and p-values for the ECM of Eqs. (2a) and (2b).

Eq. (2a)			Eq. (2b)		
Coefficient	Value	p-value	Coefficient	Value	p-value
		CSI 300			
c_s	0.000015	0.9829	c_f	0.000013	0.9788
$\alpha_{s,1}$	−0.023610	0.6355	$\alpha_{f,1}$	0.001124	0.9598
$\alpha_{s,2}$	0.060971	0.2409	$\alpha_{f,2}$	−0.063455*	0.0045
$\alpha_{s,3}$	0.013355	0.8059	$\alpha_{f,3}$	0.041837	0.2212
$\beta_{s,1}$	0.994584	0.1152	$\beta_{f,1}$	1.048421*	0.0098
$\beta_{s,2}$	−0.011722	0.7377	$\beta_{f,2}$	−0.073733*	0.0297
$\beta_{s,3}$	−0.001771	0.9579	$\beta_{f,3}$	0.073733*	0.0297
γ_s	−0.981172	0.1206	γ_f	−1.025633*	0.0118
		HSI			
c_s	0.000011	0.9787	c_f	0.000012	0.9773
$\alpha_{s,1}$	0.760599*	0.0000	$\alpha_{f,1}$	−0.044297	0.7735
$\alpha_{s,2}$	0.070754	0.6864	$\alpha_{f,2}$	−0.270985	0.1296
$\alpha_{s,3}$	0.211645	0.1373	$\alpha_{f,3}$	−0.218526	0.1401
$\beta_{s,1}$	0.172062	0.4074	$\beta_{f,1}$	1.035235*	0.0016
$\beta_{s,2}$	−0.076682	0.6674	$\beta_{f,2}$	0.246433	0.1562
$\beta_{s,3}$	−0.224509	0.1189	$\beta_{f,3}$	0.222555	0.1286
γ_s	−0.962133*	0.0000	γ_f	−1.006243*	0.0002
		S&P 500			
c_s	0.000077	0.8490	c_f	0.000095	0.8614
$\alpha_{s,1}$	0.466722*	0.0395	$\alpha_{f,1}$	0.210157	0.3346
$\alpha_{s,2}$	0.188169	0.4455	$\alpha_{f,2}$	−0.141369	0.7218
$\alpha_{s,3}$	0.113270	0.6024	$\alpha_{f,3}$	−0.341904	0.2824
$\beta_{s,1}$	0.381633	0.3400	$\beta_{f,1}$	0.594455	0.5490
$\beta_{s,2}$	−0.071986	0.7739	$\beta_{f,2}$	0.254598	0.5762
$\beta_{s,3}$	−0.239559	0.2593	$\beta_{f,3}$	0.216085	0.4501
γ_s	−0.916021*	0.0248	γ_f	−0.874759	0.3340

Coefficients and p-values are estimated by using the data of CSI 300, HSI and S&P 500 indices and their futures during the period from April 16, 2010 till January 14, 2014.
* Coefficients are statistically significant at 5% level.

of regression, and it can also help to reveal the intermittent changes of the lead–lag relationship between the time series. The main idea of this method is to search for the optimal path in a distance matrix, which is constructed based on the matching of all sample data pairs of the two time series. In other words, the lead–lag relationship is obtained by minimizing the total mismatch between the two series. The detailed implementation of this method is introduced in the following context.

Suppose we have two time series $\{X(t_1) : t_1 = 1, \dots, N\}$ and $\{Y(t_2) : t_2 = 1, \dots, N\}$. An element of the distance matrix $E_{X,Y}$ between X and Y is defined as

$$\epsilon(t_1, t_2) = [X(t_1) - Y(t_2)]^2. \tag{3}$$

Other possible forms of distance can also be used [33], e.g., $|X(t_1) - Y(t_2)|^Q$ with $Q > 1$. In this paper, we use the form in Eq. (3).

If the two series have a steady lead–lag relationship as $Y(t_2) = X(t_1 - k)$ with $k \neq 0$, it has a mapping $t_2 = t_1 + k = \phi(t_1)$ that makes $\epsilon(t_1, t_2) = 0$. The lead–lag relationship between two time series is determined by searching for the mapping $t_1 \to t_2$ to obtain a global minimal distance as

$$Min_{\{\phi(t_1), t_1=0,1,2,\dots,N-1\}} \sum_{t_1=0}^{N-1} [X(t_1) - Y(t_2)]^2. \tag{4}$$

The distance matrix can be interpreted as an energy landscape in the plane (t_1, t_2), and the elemental distance is the energy associated with the node (t_1, t_2). Therefore, the global minimization problem can be solved by searching for the configuration with minimum energy [38,39]. It has been shown that the transfer matrix method is a useful and powerful method to solve this problem.

The transfer matrix method is implemented as follows. Fig. 1 shows the plane (t_1, t_2) with t_1 and t_2 as the horizontal and vertical coordinates. For convenience, the transfer matrix method use a rotated frame (x, t) with coordinates

$$\begin{aligned} x &= t_2 - t_1, \\ t &= t_2 + t_1. \end{aligned} \tag{5}$$

The variable x quantifies the lead–lag relation between the two series by definition, and a positive (negative) x means that the first time series X leads (lags) the second time series Y.

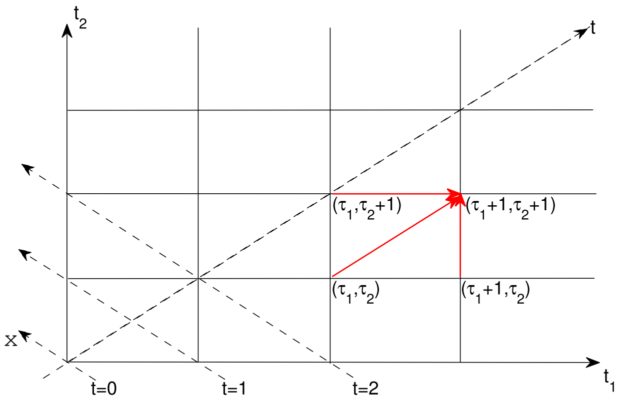

Fig. 1. Representation of the plane (t_1, t_2) and the rotated frame (t, x) as defined in Eq. (5). The three red arrows depict the three possible paths from (τ_1, τ_2) to $(\tau_1 + 1, \tau_2 + 1)$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In reality, it is very likely that both time series X and Y contain a significant amount of noise. As a result, the distance matrix $E_{X,Y}$ could also contain a possible significant amount of noise. This problem makes the optimal path sensitive to the noise, and the simple method of global minimization as described above may lead to spurious interpretation of any relationship between the two time series. To make the solution more robust and less sensitive to the existence of noise, the path configuration with energy slightly larger than the absolute minimum energy is allowed. The probability of a given path configuration with energy, quantified by $\epsilon(x, t)$ in the rotated frame (x, t), above the minimum energy is specified by a Boltzmann factor $exp[-\epsilon(x, t)/T]$, where the 'temperature' T quantifies how much deviation from the minimum energy is allowed. The introduction of this Boltzmann factor is indeed common in the Monte Carlo simulations of lattice spin models. For our purpose here, T should not be too large, otherwise one would lose all information in the lead–lag relationship between the two time series. This is the so-called 'thermal' optical path method [33–36]. In the TOP method, the optimal thermal average $\langle x(t)\rangle$ takes into account the set of neighboring (in energy) paths, which allows one to average out the noise contribution to the distance matrix. It is defined as

$$\langle x(t)\rangle = \sum_x xG(x, t)/G(t), \tag{6}$$

where $G(x, t)$ is the partition function and $G(t)$ is the sum of $G(x, t)$ over all paths emanating from $(0, 0)$ and ending at (x, t) that $G(t) = \sum_x G(x, t)$. If we treat the ratio $G(x, t)/G(t)$ as the probability for a path to be at x at time t, $\langle x(t)\rangle$ defined in Eq. (6) is the thermal average position at t and thus gives an average lead–lag relation between the two series.

We now compute the partition function $G(x, t)$ for any lattice point with coordinates x and t as shown in Fig. 1. Under the restriction of smooth and continuity mapping, the possible paths from (τ_1, τ_2) to $(\tau_1 + 1, \tau_2 + 1)$ include: either go horizontally by one step from (τ_1, τ_2) to $(\tau_1 + 1, \tau_2)$, vertically by one step from (τ_1, τ_2) to $(\tau_1, \tau_2 + 1)$; or along the diagonal from (τ_1, τ_2) to $(\tau_1 + 1, \tau_2 + 1)$. In the rotated frame (x, t), the recursive equation on $G(x, t)$ is obtained by formula

$$G(x, t + 1) = [G(x - 1, t) + G(x + 1, t) + G(x, t - 1)]e^{-\epsilon(x,t)/T}, \tag{7}$$

where $\epsilon(x, t)$ is elemental distance in Eq. (3), which can also be considered as the energy, and the 'temperature' T evaluates the deviation from the minimum energy. Eq. (6) leads to an optimal path with $\epsilon(x, t) = 0$ for $T \to 0$, and an average 'thermal optimal path' over an increasing number of path configurations with energy $\epsilon(x, t) \neq 0$ above the absolute minimum energy for an increasing T. $G(x, t) = 0$ for $t = 0$ and $t = 1$. For $t > 1$, $G(x, t)$ is obtained from $G(x, t - 1)$ and $G(x, t - 2)$ according to the recursive Eq. (7).

4.2. General result of lead–lag relationship

We now treat the return series $R_f(t)$ and $R_s(t)$ of the stock index futures and the stock index as $X(t_1)$ and $Y(t_2)$, and use the TOP method to detect the lead–lag relationship between the two series. We will present the empirical results of the TOP method for CSI 300, HSI and S&P 500. We here set the starting and ending points to be within m units from the origin and top-right point (N,N) respectively in the (t_1,t_2) plane. There are a total of $2m + 1$ possible positions for the starting point $(t_1 = 0,t_2 = 0)$, $(t_1 = 0,t_2 = j)$, and $(t_1 = j,t_2 = 0)$ with $j = 1, 2, \ldots, m$, and also for the ending point, $(t_1 = N,t_2 = N)$, $(t_1 = N,t_2 = N - j)$, and $(t_1 = N - j,t_2 = N)$ with $j = 1, 2, \ldots, m$. Consequently, there are $(2m + 1)^2$ thermal optimal paths in total. In the analysis by ECM, it has been shown that leading order of the index or its futures is less than 3. We therefore consider only situations for $m \leq 3$ here. One can of course extend the analysis to other values of m larger than 3 for the analysis.

In Fig. 2, we show the results of the average lead–lag $\langle x(t)\rangle$ between the returns of CSI 300 index and its futures with respect to different values of m. Fig. 2(a) is a plot of $\langle x(t)\rangle$ for $m = 1, 2, 3$ with a fixed temperature $T = 0.5$. $\langle x(t)\rangle$ decreases

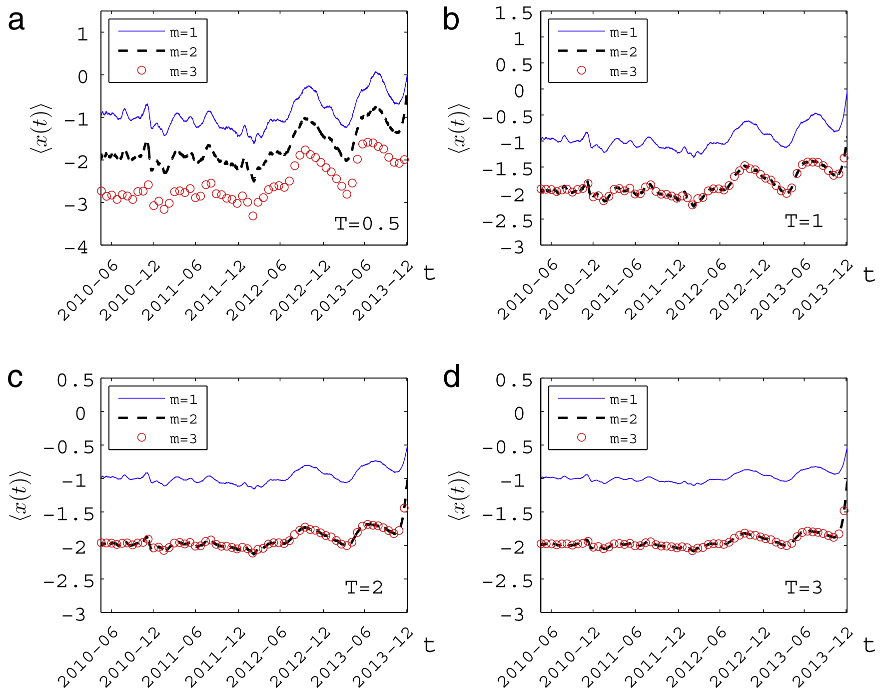

Fig. 2. (Color online) Dependence of average lead–lag $\langle x(t)\rangle$ between the returns of CSI 300 index and its futures on the starting and ending points of the paths for different temperatures T: (a) 0.5, (b) 1, (c) 2, and (d) 3.

as m increases and generally displays large fluctuations. For other values of T, as shown in Fig. 2(b)–(d), $\langle x(t)\rangle$ has relatively small fluctuations, and is more stable as m increases. Notice that $\langle x(t)\rangle$ collapses onto a single curve for $m \geq 2$. We have also studied the dependence of $\langle x(t)\rangle$ on m for HSI and S&P 500 indices, which shows that the stability exists for $m \geq 2$ and is relatively robust for different T. Our result suggests that $m = 3$ is the most proper choice, which not only covers all possible positions for the starting and ending points for $m < 3$, but also gives stable results in agreement with the results obtained by ECM.

We next study the dependence of average lead–lag $\langle x(t)\rangle$ on temperature T for $m = 3$. In Fig. 3, the average lead–lag $\langle x(t)\rangle$ between the returns of CSI 300, HSI and S&P 500 indices and their futures for different values of T are plotted. The average lead–lag $\langle x(t)\rangle$ between CSI 300 index and its futures for $T = 0.5$, 1, 2 and 3 are shown in Fig. 3(a). The fluctuations of $\langle x(t)\rangle$ get smaller as T increases and is negative for all values of T. This suggests that the spot price of CSI 300 leads its futures price, in agreement with the results in Fig. 2. Opposite evidence that futures price leads spot price is however observed for HSI and S&P 500. As shown in Fig. 3(b)–(c), $\langle x(t)\rangle$ for HSI and S&P 500 is generally positive for different values of T. The causality between the CSI 300, HSI and S&P 500 indices and their futures is qualitatively consistent with the results of ECM.

4.3. Significance test

Our next step is to test whether the result obtained by the TOP method is statistically significant. We here use a bootstrap approach to test the statistical significance of the extracted lead–lag relationship following the work in Ref. [36]. The procedures for the bootstrap approach are as follows. For two time series $R_f(t)$ and $R_s(t)$, the lead–lag function $\langle x(t)\rangle$ is obtained by the TOP method. We first generate 1000 pairs of synthetic samples by shuffling the data of the original time series, denoted as $R_{f,i}(t)$ and $R_{s,i}(t)$, $i = 1, \ldots, 1000$. We then calculate the lead–lag function $\langle x_i(t)\rangle$ for each pair of synthetic samples. There are 1000 values of lead–lag function for each value of t. We compute the 5% quantile and 95% quantile at t denoted as $\langle x_{5\%}(t)\rangle$ and $\langle x_{95\%}(t)\rangle$ respectively. If $\langle x(t)\rangle$ of the original time series is located outside the region bounded by the 5% quantile and 95% quantile curves, the empirical result of $\langle x(t)\rangle$ is considered to be statistically significant.

Fig. 4 shows the results of the bootstrap test for the lead–lag relationship between the return series of CSI 300, HSI and S&P 500 and their futures for $m = 3$ and $T = 2$. As shown in the figure, setting $T = 2$ gives the reasonable results for all three indices. The average lead–lag $\langle x(t)\rangle$ here passes the significance test and the method offers enough deviations from the optimal path while restraining excessive noise in the computation. In Fig. 4(a), $\langle x(t)\rangle$ between the CSI 300 index and its futures is depicted by the solid line, and the upper and lower dashed lines are 95% quantile and 5% quantile curves obtained in bootstrap test approach. One can see that $\langle x(t)\rangle$ fluctuates around -2 and is below the 5% quantile curve. We conclude that $\langle x(t)\rangle$ for CSI 300 index is statistically significant and the price of CSI 300 index moves about two days ahead of its futures price. Figs. 4(b)–(c) are plots of $\langle x(t)\rangle$ for HSI and S&P 500 respectively. The curves for $\langle x(t)\rangle$ fluctuate around 1, and

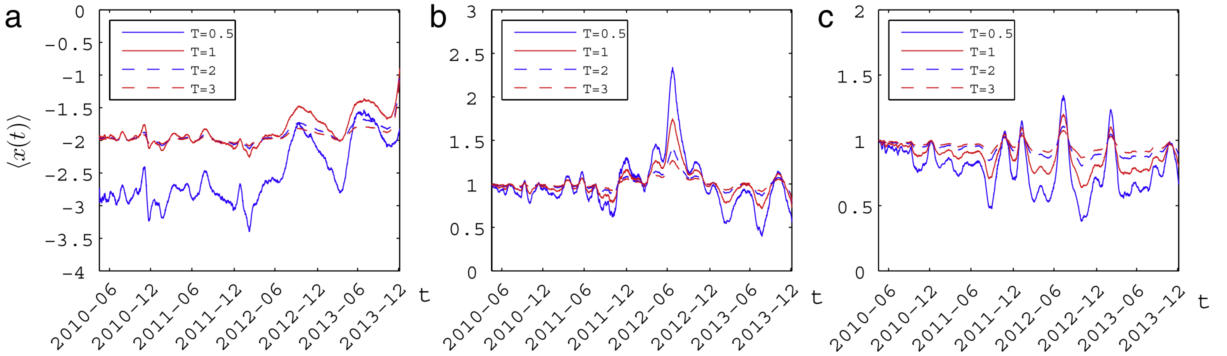

Fig. 3. (Color online) Dependence of average lead–lag $\langle x(t)\rangle$ between the returns of (a) CSI 300, (b) HSI and (c) S&P 500 indices and their futures on the temperature T for $m = 3$.

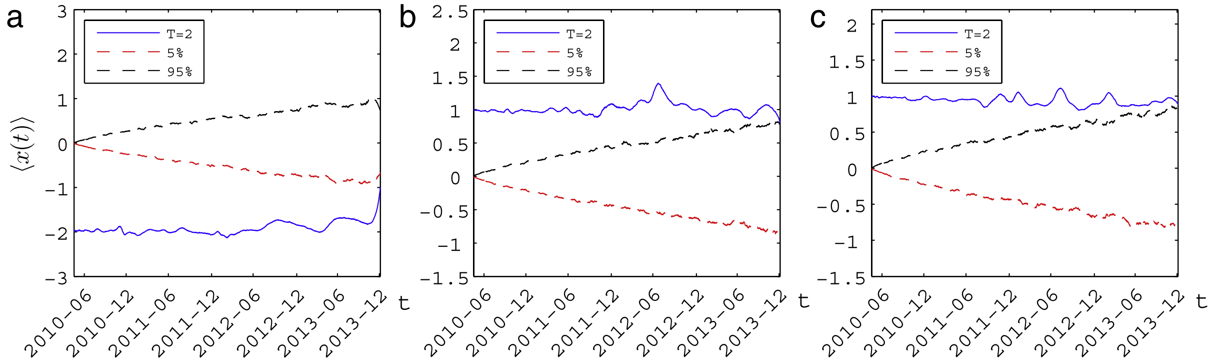

Fig. 4. (Color online) Bootstrap test for the significance of the lead–lag relationship between the return series of (a) CSI 300, (b) HSI and (c) S&P 500 indices and their futures. Solid curves are $\langle x(t)\rangle$ of the three indices for $m = 3$ and $T = 2$. Upper dashed lines are 95% quantile curves, and lower dashed lines are 5% quantile curves.

are above the 95% quantile curves. This means that both $\langle x(t)\rangle$ for HSI and S&P 500 indices can pass the significance test and their futures prices move about one day ahead of their spot prices. In general, the lead–lag relationship revealed by the TOP method is qualitatively consistent with our previous findings by ECM.

4.4. Temporal variation of the lead–lag relationship

The most interesting result by the TOP method lies in the fact that the lead–lag relationship $\langle x(t)\rangle$ evolves over time t. We can also observe the variation of $\langle x(t)\rangle$ with respect to t in Fig. 4. One can see that $\langle x(t)\rangle$ displays two local maximum around September 2012 and July 2013 for CSI 300 index. A sharp peak in $\langle x(t)\rangle$ is clearly exhibited around June 2012 for HSI index, and several peaks appear during the period of 2011–2012 for S&P 500. To better understand the variation of $\langle x(t)\rangle$ over t, we introduce an indicator to quantify the difference between the futures price and spot price of the stock index. To compare with the lead–lag $\langle x(t)\rangle$, the indicator is defined as the absolute value of the difference between the logarithmic prices of the stock index and its futures, and is further scaled by a constant factor.

Fig. 5(a) shows the comparison between $\langle x(t)\rangle$ and the indicator for CSI 300 index. The upper black curve is $\langle x(t)\rangle$ for $m = 3$ and $T = 2$. Since $\langle x(t)\rangle$ is negative for CSI 300 index, the scale factor is set to be −50. The lower blue curve is the scaled indicator $-50|\ln P_f - \ln P_s|$. It is remarkable that both $\langle x(t)\rangle$ and the indicator display extrema for certain values of t. Important events can indeed be identified in the Chinese financial market near these time points. Labels A–D in the figure correspond to the following events: A, the Shanghai Stock Exchange fell more than 5% on November 12, 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 on December 5, 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 since July 20, 2013.

The comparison between $\langle x(t)\rangle$ and the indicator for HSI index is also depicted in Fig. 5(b). The upper black curve is $\langle x(t)\rangle$, and the lower blue curve is $50|\ln P_f - \ln P_s|$. Local maxima are observed for both $\langle x(t)\rangle$ and the indicator at certain time points, where important events are also found. Labels A–E in the figure correspond to the following events: A, the Hong Kong government announced real estate policies such as raising the down payment and stamp tax on housing purchases and sales in August, 2010; B, the Chinese central bank announced that the Hong Kong offshore RMB deposit interest rates

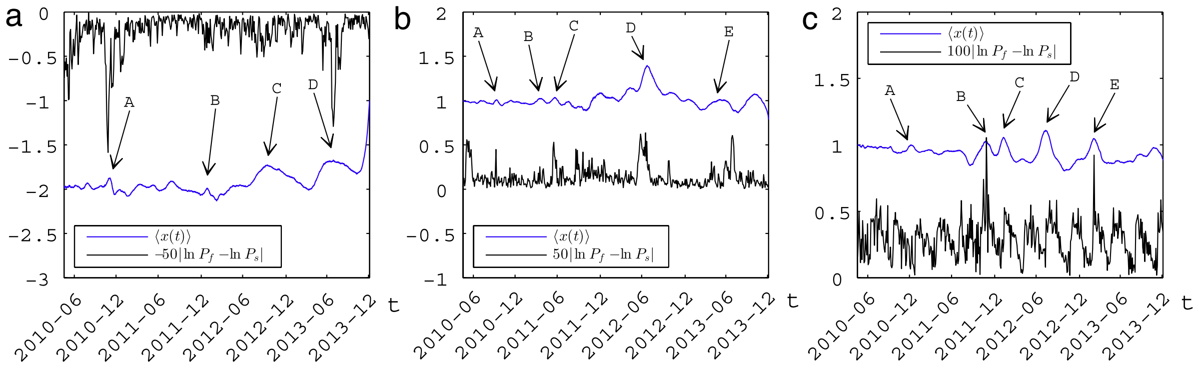

Fig. 5. Comparison between the average lead–lag $\langle x(t) \rangle$ and the indicator defined as the absolute value of the difference between the logarithmic prices of (a) CSI 300, (b) HSI and (c) S&P 500 indices and their futures. To facilitate the comparison between the two variables, the indicator is scaled by −50, 50 and 100 for CSI 300, HSI and S&P 500 respectively. Labels A–D in Fig. 5(a) refer to the following events: A, the Shanghai Stock Exchange fell more than 5% on November 12, 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 on December 5, 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 since July 20, 2013. Labels A–E in Fig. 5(b) refer to the following events: A, the Hong Kong government announced real estate policies such as raising the down payment and stamp tax on housing purchases and sales in August, 2010; B, the Chinese central bank announced that the Hongkong offshore RMB deposit interest rates would be lowered by 27 basis points on April 1, 2011; C, America ended the second round of the quantitative easing policy at the end of June, 2011; D, the Hong Kong Exchanges & Clearing (HKEx) said on June 15, 2012 in a statement that it has signed a framework agreement with the London Metal Exchange (LME) to buy the LME; E, the Treasury Markets Association (TMA) announced on April 25, 2013 its plan to launch the CNH Hong Kong Interbank Offered Rate fixing (CNH HIBOR fixing) in June 2013. Labels A–E in Fig. 5(c) refer to the following events: A, in the meeting on December 14, 2010, the Federal Reserve decided to maintain the target range for the federal funds rate between 0% and 1/4% and continue expanding its holdings of securities as announced in November (the second round of quantitative easing); B, in a meeting held on September 20–21, 2011, the Federal Open Market Committee supported a program under which the Committee would announce its intention to purchase, by the end of June 2012, $400 billion of Treasury securities with remaining maturities of 6 years to 30 years and to sell an equal amount of Treasury securities with remaining maturities of 3 years or less; C, on December 16, 2011, the US House of Representatives passed a spending bill of $1 trillion to fund the federal government until next October; D, in a meeting held on June 19–20, 2012, the Federal Open Market Committee decided to continue the maturity extension program which began in September 2012 and keep the target range for the federal funds rate at 0%–1/4%; E, the White House and Senate Republicans reached a deal to avert the fiscal cliff on December 31, 2012. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

would be lowered by 27 basis points on April 1, 2011; C, America ended the second round of the quantitative easing policy at the end of June, 2011; D, the Hong Kong Exchanges & Clearing (HKEx) announced on June 15, 2012 in a statement that it had signed a framework agreement with the London Metal Exchange (LME) to buy the LME; E, the Treasury Markets Association (TMA) announced on April 25, 2013 its plan to launch the CNH Hong Kong Interbank Offered Rate fixing (CNH HIBOR fixing) in June 2013.

Fig. 5(c) is a plot of $\langle x(t) \rangle$ and the scaled indicator $100|\ln P_f - \ln P_s|$ for S&P 500 index, depicted by upper black curve and lower blue curve respectively. The indicator for S&P 500 index shows a periodic behavior with a period of 3-months which matches the period of futures contract. The deviation between futures price and spot price is generally large at the beginning of the contract period, and the futures price converges to spot price as the delivery date approaches, which can explain the periodic behavior of the indicator. Such a periodic behavior is not clear for CSI 300 and HSI indices, which may due to variation of contract periods for the four types of futures contracts. Besides the periodic fluctuations, the indicator displays several maximal values at certain time t where $\langle x(t) \rangle$ also shows local maxima. Important events are also identified in the US financial market near these time points. Labels A–E refer to the following events: A, in the meeting on December 14, 2010, the Federal Reserve decided to maintain the target range for the federal funds rate between 0% and 1/4% and continue expanding its holdings of securities as announced in November (the second round of quantitative easing); B, in a meeting held on September 20–21, 2011, the Federal Open Market Committee supported a program under which the Committee would announce its intention to purchase, by the end of June 2012, $400 billion of Treasury securities with remaining maturities of 6 years to 30 years and to sell an equal amount of Treasury securities with remaining maturities of 3 years or less; C, on December 16, 2011, the US House of Representatives passed a spending bill of $1 trillion to fund the federal government until next October; D, in a meeting held on June 19–20, 2012, the Federal Open Market Committee decided to continue the maturity extension program which began in September 2012 and keep the target range for the federal funds rate between 0% and 1/4%; E, the White House and Senate Republicans reached a deal to avert the fiscal cliff on December 31, 2012.

In general, there are large gaps between the futures prices and spot prices when important events occur in financial markets. In developed markets like Hang Kong and US stock markets, the futures markets first react to the information, and the information flows from the futures markets to the stock markets which leads to a delayed response of the stock prices. Therefore, the lead–lag relationships between the stock indices and their futures reach local maxima around these important events. In emerging markets like mainland China, though the movements of stocks markets are ahead of the movements of futures markets when important financial events occur, the rapid increase of lead–lag structure $\langle x(t) \rangle$ can still be observed.

This point of view is supported by the results of both TOP method and the indicator quantifying the difference between future price and spot price.

5. Conclusion

Based on the thermal optimal path (TOP) method, we study the dynamic evolution of the lead–lag relationship of CSI 300, HSI and S&P 500 indices and their associated futures respectively using the daily data in the period between April 16, 2010 and January 14, 2014. The causality analysis by the error correction model (ECM) shows that the leading order of CSI 300 is two days while the leading order of the futures of HSI and S&P 500 is one day. This result is in agreement with that obtained by the optimal path in the TOP method. By calculating the lead–lag function, we confirm that CSI 300 has more pronounced leadership while the futures of HSI and S&P 500 have more pronounced leadership, and the result is proven to be statistically significant by a bootstrap test. We further show that the lead–lag relationship evolves over time, and find some local extrema for the average lead–lag $\langle x(t) \rangle$ at some time points. To better understand the dynamic evolution of the lead–lag relationship, an indicator is introduced to quantify the difference between the futures price and spot price of the stock index. We find that both the lead–lag function and the indicator display extrema when important events occur in financial markets. Based on this observation, we offer an explanation for the surge of lead–lag relationship between stock index and stock index futures.

In the analysis by ECM, we note that the lead–lag relationship depends on the specific form of the cointegration error term and the ECM in Eq. (2). The causality behavior persists as we introduce different terms in the formula, but the lead–lag relationship shows quantitative differences. In the TOP method, there is no need to take into account extra factors similar to the formula and order of the regression equation. Moreover, it can help to reveal the intermittent changes of the lead–lag relationship between two time series. Our study demonstrates the capability of the TOP method, to uncover the dynamic evolution of the lead–lag relationship between stock index and stock index futures for three typical markets: mainland China, Hong Kong and US. This work is valuable for the understanding of lead–lag relationship between stock index and stock index futures, and can be applied to strategy trading in hedging and portfolio investments.

Acknowledgments

The authors would like to thank Wei-Xing Zhou and Hao Meng for helpful discussion and suggestions. This work was partially supported by the National Natural Science Foundation (Nos. 10905023, 11075054 and 71131007), Humanities and Social Sciences Fund sponsored by Ministry of Education of the Peoples Republic of China (No. 09YJCZH042), the Shanghai (Follow-up) Rising Star Program Grant 11QH1400800, and the Fundamental Research Funds for the Central Universities (2015).

References

[1] R. Zeckhauser, V. Niederhoffer, The performance of market index futures contracts, Financ. Anal. J. 39 (1983) 59–65.
[2] I.G. Kawaller, P.D. Koch, T.W. Koch, The temporal price relationship between S&P500 futures and the S&P500 index, J. Finance 42 (1987) 1309–1329.
[3] M.A. Pizzi, A.J. Economopoulos, H.M. O'Neill, An examination of the relationship between stock index cash and futures markets: A cointegration approach, J. Futures Markets 18 (1998) 297–305.
[4] Q.C. Chu, W.G. Hsieh, Y.K. Tse, Price discovery on the S&P 500 index markets: An analysis of spot index, index futures, and SPDRs, Int. Rev. Financ. Anal. 8 (1999) 21–34.
[5] J. Hasbrouck, Intraday price formation in US equity index markets, J. Finance 58 (2003) 2375–2399.
[6] A.H. Abhyankar, Return and volatility dynamics in the FT-SE 100 stock index and stock index futures markets, J. Futures Markets 15 (1995) 457–488.
[7] C. Brooks, A.G. Rew, S. Ritson, A trading strategy based on the lead–lag relationship between the spot index and futures contract for the FTSE 100, Int. J. Forecast. 17 (2001) 31–44.
[8] G.G. Booth, R.W. So, Y.K. Tse, Price discovery in the German equity index derivatives markets, J. Futures Markets 19 (1999) 619–643.
[9] P.C. Pati, P. Rajib, Intraday return dynamics and volatility spillovers between NSE S&P CNX Nifty stock index and stock index futures, Appl. Econ. Lett. 18 (2011) 567–574.
[10] Y.K. Tse, Lead-lag relationship between spot index and futures price of the Nikkei stock average, J. Forecast. 14 (1995) 553–563.
[11] M.L. Nieto, A. Fernandez, M.J. Munoz, Market efficiency in the spanish derivatives markets: An empirical analysis, Int. Adv. Econ. Res. 4 (1998) 349–355.
[12] G. Shyy, V. Vijayraghavan, B. Scott-Quinn, A further investigation on the lead–lag relationship between the cash market and stock index futures market with the use of bid/ask quotes: The case of France, J. Futures Markets 16 (1996) 405–420.
[13] M. Wahab, M. Lashgari, Price dynamics and error correction in stock index and stock index futures markets: A cointegration approach, J. Futures Markets 13 (1993) 711–742.
[14] Z. Zakaria, K.L. Shamsuddin, Relationship between stock futures index and cash prices index: Empirical evidence based on Malaysia data, J. Bus. Stud. Q. 4 (2012) 103–112.
[15] A. Kasman, S. Kasman, The impact of futures trading on volatility of the underlying asset in the Turkish stock market, Physica A 387 (2008) 2837–2845.
[16] G. Rajaguru, S.S. Pattnayak, Investigation of a lead–lag relationship between spot and futures indices of the hang seng stock average, Int. J. Bus. Stud. 15 (2007) 69–82.
[17] E. Özen, T. Bozdogan, M. Zügül, The relationship of causality between the price of futures transactions underlying stock exchange and price of cash market: The case of Turkey, Middle East. Finance Econ. 4 (2009) 28–37.
[18] Y. Iihara, K. Kato, T. Tokunaga, Intraday return dynamics between the cash and the futures markets in Japan, J. Futures Markets 16 (1996) 147–162.
[19] A. Frino, A. West, The lead–lag relationship between stock indices and stock index futures contrasts: Futher Australian evidence, ABACUS 35 (1999) 333–341.
[20] A. Chatrath, R. Christie-David, K.K. Dhanda, T.M. Koch, Index futures leadership, basis behavior, and trader selectivity, J. Futures Markets 22 (2002) 649–677.

[21] D. Lien, Y.K. Tse, X. Zhang, Structural change and lead–lag relationship between the Nikkei spot index and futures price: a genetic programming approach, Quant. Finance 3 (2003) 136–144.
[22] M.Y. Li, The dynamics of the relationship between spot and futures markets under high and low variance regimes, Appl. Stoch. Models Bus. Ind. 25 (2009) 696–718.
[23] Y.K. Tse, W.S. Chan, The lead–lag relationship between the S&P spot and futures markets: An intraday-data analysis using a threshold regression model, Jpn. Econ. Rev. 61 (2010) 133–144.
[24] M.T. Bohl, C.A. Salm, M. Schuppli, Price discovery and investor structure in stock index futures, J. Futures Markets 31 (2011) 282–306.
[25] H.L. Chung, W.S. Chan, J.A. Batten, Threshold non-linear dynamics between Hang Seng stock index and futures returns, Eur. J. Finance 17 (2011) 471–486.
[26] E. Theissen, Price discovery in spot and futures markets: a reconsideration, Eur. J. Finance 18 (2012) 969–987.
[27] M. Zhong, A.F. Darrat, R. Otero, Price discovery and volatility spillovers in index futures markets: Some evidence from Mexico, J. Bank. Financ. 28 (2004) 3037–3054.
[28] J.K. Kang, C.J.L. Lee, S. Lee, An empirical investigation of the lead–lag relations of returns and volatilities among the KOSPI200 spot, futures and options markets and their explanations, J. Emerg. Mark. Finance 5 (2006) 235–261.
[29] M.G. Kavussanos, L.D. Visvikis, P.D. Alexakis, The lead–lag relationship between cash and stock index futures in a new market, Eur. Financ. Manag. 14 (2008) 1007–1025.
[30] H. Ullah, Lead-lag relationship in spot and future marke evidence from Pakistani stock market KSE-100 index, Bus. Rev. 8 (2013) 135–148.
[31] E.C. Cagli, P.E. Mandaci, The long-run relationship between the spot and futures markets under multiple regime-shifts: Evidence from Turkish derivatives exchange, Expert Syst. Appl. 40 (2013) 4206–4212.
[32] J.-Y. Zhu, J.-W. Diao, Price discovery in index futures and spot market in China: Based on HS300 stock index futures, IEEE 5 (2011) 424–427.
[33] D. Sornette, W.-X. Zhou, Non-parametric determination of real-time lag structure between two time series: The optimal thermal causal path method, Quant. Finance 5 (2005) 577–591.
[34] W.-X. Zhou, D. Sornette, Non-parametric determination of real-time lag structure between two time series: The optimal thermal causal path method with application to economic data, J. Macroecon. 28 (2006) 195–224.
[35] W.-X. Zhou, D. Sornette, Lead-lag cross-sectional structure and detection of correlated-anticorrelated regime shifts: Application to the volatilities of inflation and economic growth rates, Physica A 380 (2007) 287–296.
[36] K. Guo, W.-X. Zhou, S.-W. Cheng, D. Sornette, The US stock market leads the Federal funds rate and Treasury bond yields, PLoS One 6 (2011) 22794.
[37] A. Ghosh, Cointegration and error correction models: intertemporal causaliry between index and futures prices, J. Futures Markets 13 (1993) 193–198.
[38] B. Derrida, J. Vannimenus, Y. Pomeau, Simple frustrated systems: Chains, strips and squares, J. Phys. C 11 (1978) 4749–4765.
[39] B. Derrida, J. Vannimenus, Interface energy in random systems, Phys. Rev. B 27 (1983) 4401–4411.