CrossMark

Identification of stock market forces in the system adaptation framework

Xiaolian Zheng, Ben M. Chen *

Department of Electrical Computer Engineering, National University of Singapore, Singapore 117576, Singapore

A R T I C L E   I N F O

A B S T R A C T

Based on the system adaptation framework which has been proposed in our previous work, this paper focuses on the input selection of this framework to identify crucial market influential factors. We first carry out an empirical research to preselect influential factors from economic and sentimental aspects. The causal relationship between each of them and the internal residue of the market is then tested. Lastly, a multicollinearity test is applied to those factors that show significant causality to the internal residue of the market to exclude the redundant indicators. As the causal relationship plays an essential role in this method, both linear time-varying and nonlinear causality tests are employed based on the predictive ability of our framework. This double selection method is applied to the US and China stock markets, and it is shown to be efficient in identifying market influential factors. We also find that these influential factors are market-dependent and frequency-dependent. Some well-tested factors in the developed market and literature may not work in the emerging market.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

System economics is a popular and promising direction of research as it has provided many powerful tools in analyzing the market as a complex system. With it, the features and structures of the market can be better explored. In our previous work [45,46], a unique system adaptation framework has been developed to model the stock market or financial markets in general from a viewpoint of system dynamics. It has been shown in our previous work that with proper input, our framework not only provides more accurate prediction results but also better understands and captures the dynamics and properties of markets. Under this framework, the information flow of the real market can be clearly reflected in the hierarchy of the identified system. The modeling process is considered as identifying a dynamical system, in which the real market is treated as an unknown plant or system and the proposed identification model is tuned by feeding back the resulting matching errors. The working scheme of the proposed structure of modeling involves a crucial component, which is to identify the input influential factors.

Unlike most physical systems, in which the system input is generally well-defined and structured to meet certain specifications and requirements, the stock market has too many input factors that might influence its internal behaviors, and hence its outcomes. The theory of nonlinear dynamic economics suggests that the economic fluctuations are caused by both internal and external forces. In the stock market, the internal force is believed to be generated by some fundamental factors, such as market mechanism, company value, profitability of a company, whereas the external force is usually constituted by the information outside the stock market including the economic, fundamental and many other influential factors. The internal and external forces induce slow and fast dynamics in the stock market respectively. Our identification model consists of

---

* Corresponding author. Tel.: +65 65162289; fax: +65 67791103.
  E-mail addresses: xlzheng919@gmail.com (X. Zheng), bmchen@nus.edu.sg (B.M. Chen).

an internal model and an adaptive filter, respectively taking the slow and fast dynamics of the market prices into consideration. The external force is much more complicated than the internal counterpart, as it represents the interaction of the whole system with its environment. Influential factors which form the external force work as the input to our framework to regulate stock prices through the adaptive filter. As their effect on the market changes from time to time, identifying the most influential factors in the given period of time is critical to yield good predicting performance. This paper focuses on the selection of appropriate input influential factors for our system adaptation framework.

There are a vast number of causality test methods reported in the literature that identify the causal relationship between the stock price and the market influential factors, of which the original and also best-established approach is the Granger causality test [21]. Generally speaking, the Granger causality test is within the linear regression context that provides useful information on whether the past knowledge of an influential factor could significantly improve the short-term prediction of stock prices, and vice versa. The so-called standard F-test was designed for stationary series. However, causality patterns may change from time to time due to various reasons, such as fast changing dynamics in the economic environment. As such, the technique has been further extended for those dynamic models with time-varying parameters. Geweke [18,19] quantified the causality in the form of linear dependence between signals based on VAR models. Such a concept can be easily extended to test dynamic models. For example, Feng et al. [16] used Bayesian networks and ant colony optimization to analyze the causal relationship among risk factors for information systems. An alternative method in testing time-varying causality is to make use of Markov regime-switching models. By using an unobservable finite Markov chain to allow for changes in causal links over the sampling period, Psaradakis et al. [35] investigated the US money–output relationship based on a VAR model with time-varying parameters. The time-varying causality test has also been frequently reported in the literature of neuroscience. Ding et al. [15] used a short-window spectral analysis to construct a time-variant Granger causality test, but it requires the stationarity of signals within the short-time window. Hesse et al. [23] loosened this requirement by recursively fitting a VAR model with time-dependent parameters, and then the time-variant Granger causality strength was calculated. Similar work could be found in Roebroeck et al. [36] as well as Bressler and Seth [10]. In our input selection process, we adopt the idea in the time-varying causality test.

The traditional Granger causality test only considers linear relationships which are incapable of attesting the nonlinear dynamics in time series. As noted by Granger [22], Hsieh [26] and many others, the nonlinearity is an intrinsic and fundamental feature in the financial time series. Baek and Brock [2] proposed a nonparametric statistical method to identify the nonlinear Granger causality. By allowing the series in testing to display short-term temporal dependence, Hiemstra and Jones [25] modified the Baek and Brock test to discover the significant bidirectional nonlinear causality between the daily returns of the Dow Jones Industrial Average (DJIA) and the percentage changes in the New York Stock Exchange trading volume, for which the linear Granger causality test fails to discover the relationship. The work popularized the research in testing nonlinear Granger causality and the Hiemstra–Jones test has become a commonly used method in economics and finance. Examples include causal relationships between international stock markets [7,20], stock returns and trading volume [12], stock price and volume [38], stock price dividend relationships [29] and so on. By developing a multivariate nonlinear causality test, Bai et al. [3] extended the Hiemstra and Jones nonlinear causality test to a multivariate setting which has also been successfully applied to the stock market [4]. We adopt the Hiemstra and Jones test in our selection process and also consider the Bai's multivariate test.

In this paper, we propose some new forms of market influential factors together with a double selection method to identify the most crucial input elements for the system adaptation framework. It involves an empirical selection process, followed by causality tests and a redundant test. Nonlinear Hiemstra and Jones test is carried out only for the periods when no significant linear type causal relationship appears. After causality tests, a multicollinearity test is used to remove some redundant factors. The US stock market, represented by the Dow Jones Industrial Average (DJIA), and China stock market, represented by Shanghai Stock Exchange Composite Index (SSE), are taken as illustrative examples to demonstrate the practicality of our selection process. This input selection part has been proven to work well.

The outline of the paper is organized as follows: Section 2 presents our system adaptation framework which serves as the basis of the input selection. Section 3 discusses the details of time-varying and nonlinear causality tests, as well as the multicollinearity test. Section 4 presents the performance of this input selection method in both US and China stock markets. Some related analyses are also given in this section. Finally, we draw some concluding remarks in Section 5.

2. The system adaptation framework

Fig. 1 presents our system adaptation framework for modeling the stock market. Inspired by ideas used in identifying physical systems, the stock market modeling process is treated as the problem of identifying a dynamic plant. The input–output behavior of the stock market is represented by the identification model $\widehat{S}$ with its output $\hat{p}$ being the estimated stock price. The actual stock price p is the output of the real stock market S. We note that both S and $\widehat{S}$ have the same input r, which consists of external influential factors of the stock market. Since both internal and external forces act upon the stock market, inducing slow and fast dynamics in the market respectively, we construct an identification model $\widehat{S}$ with an internal model I and an adaptive filter A to capture these two types of dynamics.

The internal model, I, which can be regarded as a trend generator, estimates the market trend. It is reasonable to assume that the dynamics of the internal model change at a relatively slow pace, so that it can be approximated by some time-invari-
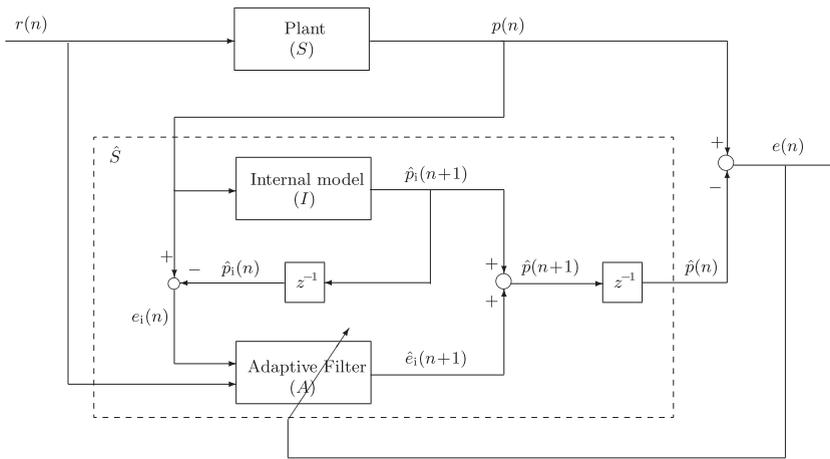
Fig. 1. Block diagram of the structure of system adaptation framework.

ant systems. Note that the internal model is independent of the adaptive filter. It uses only actual historical prices to estimate the internal stock price $\hat{p}_i$. For easy references, we define

$$e_i(n) = p(n) - \hat{p}_i(n), \tag{1}$$

as the internal residue in our framework. The historical stock price is preprocessed by the exponential moving average (EMA) model, and then an output-error (OE) model of multi-inputs and single-output (MISO) is employed to model the inherent evolution of stock prices. The parameters of OE model are estimated by the commonly used prediction error method.

In the literature, information outside the stock market always accounts for external influences to the market. The adaptive filter A is thus introduced to capture such influences. It generates the estimated error $\hat{e}_i$ of the next step by analyzing major influential factors of the stock market together with the historical information of the internal residue. Working as a cycle generator, this adaptive filter compensates the identification error e by capturing the fast dynamics of the market. It is assumed that the impact of each influential factor on the stock price movement is time-dependent, so that their significances should vary at different stages. As such, it is natural for us to choose a time-varying adaptive filter linked to the external input. We employ a time-varying state space model with instrumental variables for the adaptive filter. The parameters of the selected time-varying model are defined as state variables. Based on these state variables, a Kalman filtering technique is employed to recursively predict and update the estimated hyperparameters and instrumental variables. In this process, the identification error $e(n)$ is fed back to tune the state variables. More details about this framework including the definitions and estimations of both internal model and adaptive filter can be found in [46]. We would like to emphasize that our proposed system adaptation framework does not depend on any particular model. Besides the models we adopted, others can also be utilized as the internal model and the adaptive filter as long as they are capable of capturing the internal and external forces.

3. Input selection rules

Many results can be found in the literatures concerning the influential factors of the stock market, most of which come from economic and sentimental aspects. Economic indicators are usually used to judge the well-being of the economy and predict its future performance. They determine fundamentally the movement of stock prices, as the stock market has become an increasingly important component of the economy. Interest rate, inflation rate, money supply and commodity price are all widely accepted as economic indicators [1,8,17,34,37]. Besides economic indicators, investor sentiment becomes a crucial factor in recent studies. Sentimental indicators act as a measurement of the situation of demand and supply, representing the general opinion of the investors towards the market. Their contributions to short-term variations are particularly significant [5,6,30]. Behavioral financial factors are also frequently studied in the literature (see, e.g., [13]). Therefore, in constructing indicators, we try to include more market data as it embeds the expectations of investors on the market. Our results show that these indicators are market-dependent and frequency-dependent.

Based on the predictive ability of our framework, we propose a double selection method to evaluate the effect of every selected indicator on the stock market. Based on some empirical research, we first select a set of key influential factors, which include both economic and sentiment indicators, and perform necessary preprocessing to reform these selected indicators such that the resulting data will be better fit to our proposed framework. After that, we apply a series of statistical tests, which include linear time-varying and nonlinear causality tests as well as multicollinearity tests, on each selected indicator to detect its causality relationship with the internal residue. More specifically, a linear test is first conducted on each

input–output pair of the adaptive filter, i.e., the influential factor and the internal residue. A nonlinear causality test is to be carried out if no significant causality is displayed in the linear test. Finally, a multicollinearity test is then adopted to remove redundant indicators.

3.1. Causality test

3.1.1. Time-varying causality test with threshold

Hesse et al. [23] suggests a time-varying causality test which calculates the causality strength at each time point and then compares them with related threshold values. Considering the input and output time series r and e_i, which are respectively characterized by the following AR and bivariate AR models as

$$r(n) = \sum_{i=1}^{q_r} \alpha_{1,i}(n)r(n-i) + \vartheta_1(n), \quad \Sigma_{11}(n) = var(\vartheta_1(n)), \tag{2}$$

$$e_i(n) = \sum_{i=1}^{q_e} \beta_{1,i}(n)e_i(n-i) + \upsilon_1(n), \quad \Sigma_{21}(n) = var(\upsilon_1(n)), \tag{3}$$

and

$$r(n) = \sum_{i=1}^{q_r} \alpha_{2,i}(n)r(n-i) + \sum_{i=1}^{q_e} \beta_{3,i}(n)e_i(n-i) + \vartheta_2(n), \quad \Sigma_{12}(n) = var(\vartheta_2(n)), \tag{4}$$

$$e_i(n) = \sum_{i=1}^{q_e} \beta_{2,i}(n)e_i(n-i) + \sum_{i=1}^{q_r} \alpha_{3,i}(n)r(n-i) + \upsilon_2(n), \quad \Sigma_{22}(n) = var(\upsilon_2(n)). \tag{5}$$

The time-varying strength of causality from r to e_i and from e_i to r are defined as

$$F_{r \to e_i}(n) = \ln \frac{\Sigma_{21}(n)}{\Sigma_{22}(n)}, \tag{6}$$

and

$$F_{e_i \to r}(n) = \ln \frac{\Sigma_{11}(n)}{\Sigma_{12}(n)}. \tag{7}$$

If $F_{r \to e_i}(n) > F_{e_i \to r}(n)$, we can say that r Granger causes e_i at the time n, and vice versa. Generally, we need to set an appropriate threshold to determine whether the causal effect is significant or not. If an influential factor r Granger causes e_i in certain time intervals, we randomize the order of e_i such that the causality between r and e_i is annihilated. Note that the distribution of e_i remains unchanged during such a process. This is the so-called surrogate data approach. The shuffling procedure will be repeated for a number of times, say N_s, to yield a meaningful result. After all these processes, we then calculate the resulting $\kappa\%$ quantile for each time point, which is used to represent the threshold or the significant level of the Granger causality. Values above this level have a probability of occurring chance less than $1 - \kappa\%$. A Granger causality relationship is considered to be significant when the causality strength surpasses the threshold. In this paper, we set $\kappa\% = 95\%$ and $N_s = 200$ for US and China stock markets.

3.1.2. Nonlinear causality test

For a given pair of time series $r(n)$ and $e_i(n)$, we let $e_i^m(n)$ be the m-length lead vector of e_i, $r^{Lr}(n-Lr)$ and $e_i^{Le}(n-Le)$ be the Lr-length and Le-length lag vectors of r and e_i. The nonlinear Granger noncausality is defined in the context of conditional probability [2]. Given values of m, Le, and $Lr \geqslant 1$ and for $\varrho > 0, r$ is said not to strictly Granger cause e_i if

$$P\big(\|e_i^m(n) - e_i^m(s)\| < \varrho \big| \|e_i^{Le}(n-Le) - e_i^{Le}(s-Le)\| < \varrho, \|r^{Lr}(n-Lr) - r^{Lr}(s-Lr)\| < \varrho\big)$$
$$= P\big(\|e_i^m(n) - e_i^m(s)\| < \varrho \big| \|e_i^{Le}(n-Le) - e_i^{Le}(s-Le)\| < \varrho\big), \tag{8}$$

where $P(.)$ and $\|\cdot\|$ denote probability and maximum norm respectively. The probability on the left-hand side of (8) is the conditional probability that two arbitrary m-length lead vectors of e_i are within a distance ϱ of each other, given that the corresponding Le-length lag vectors of e_i and Lr-length lag vectors of r are within the same distance ϱ of each other. We note that the condition on r is ignored on the right-hand side of (8).

Hiemstra and Jones [25] proposed an implementation of the above nonlinear Granger noncausality test by expressing the conditional probabilities in terms of the corresponding ratios of joint probability, in which (8) can be rewritten as

$$\frac{C_1(m+Le, Lr, \varrho)}{C_2(Le, Lr, \varrho)} = \frac{C_3(m+Le, \varrho)}{C_4(Le, \varrho)} \tag{9}$$

where C_1, C_2, C_3 and C_4 are the correlation-integral estimators of the joint probabilities and are given as

$$C_1(m + Le, Lr, \varrho) \equiv P(\|e_i^{m+Le}(n - Le) - e_i^{m+Le}(s - Le)\| < \varrho, \|r^{Lr}(n - Lr) - r^{Lr}(s - Lr)\| < \varrho),$$

$$C_2(Le, Lr, \varrho) \equiv P(\|e_i^{Le}(n - Le) - e_i^{Le}(s - Le)\| < \varrho, \|r^{Lr}(n - Lr) - r^{Lr}(s - Lr)\| < \varrho),$$

$$C_3(m + Le, \varrho) \equiv P(\|e_i^{m+Le}(n - Le) - e_i^{m+Le}(s - Le)\| < \varrho),$$

$$C_4(Le, \varrho) \equiv P(\|e_i^{Le}(n - Le) - e_i^{Le}(s - Le)\| < \varrho). \tag{10}$$

For e_i and r with sampling size of N, the values of C_1, C_2, C_3 and C_4 can be easily estimated by the following numerical method:

$$C_{l_1}(L_1, L_2, \varrho, n_c) = \frac{2}{n_c(n_c - 1)} \sum \sum_{n<s} I\left(\|e_i^{L_1}(n - Le) - e_i^{L_1}(s - Le)\| < \varrho\right)$$

$$\cdot I(\|r^{L_2}(n - Lr) - r^{L_2}(s - Lr)\| < \varrho),$$

$$C_{l_2}(L_1, \varrho, n_c) = \frac{2}{n_c(n_c - 1)} \sum \sum_{n<s} I\left(\|e_i^{L_1}(n - Le) - e_i^{L_1}(s - Le)\| < \varrho\right), \tag{11}$$

where $l_1 = 1, 2$; $l_2 = 3, 4$; L_1 and L_2 represent their first two parameters; $n, s = \max(Le, Lr) + 1, \ldots, N - m + 1$; $n_c = N + 1 - m - \max(Le, Lr)$; $I(\cdot)$ equals 1 when the condition inside $(\cdot)$ is fulfilled and 0 otherwise.

Given values of m, Le and $Lr \geqslant 1$, for $\varrho > 0$, the criterion of noncausality is that if r does not strictly Granger cause e_i, then

$$\sqrt{n_c}\left[\frac{C_1(m + Le, Lr, \varrho, n_c)}{C_2(Le, Lr, \varrho, n_c)} - \frac{C_3(m + Le, \varrho, n_c)}{C_4(Le, \varrho, n_c)}\right] \to N(0, \sigma^2(m, Le, Lr, \varrho)). \tag{12}$$

The estimation of $\sigma^2(m, Le, Lr, \varrho)$ can be found in [25].

Based on our framework, the nonlinear causality test is applied to series $\Sigma_{12}(n)$ of (4) and $\Sigma_{22}(n)$ of (5), which are similar to the so-called VAR residuals reported in [25]. The differences between the left- and right-hand sides of (9) and the standardized test statistic in (12) are denoted as CS and TVAL respectively. We note that there is no standard procedure in selecting the parameters Le, Lr, and ϱ. Following the Monte Carlo results in [24], for all cases in this paper, all series are standardized before conducting causality tests and ϱ is set to be 1.5σ where $\sigma = 1$ is the standard deviation of the standardized time series. We also set the lead length to be 1, i.e. $m = 1$, and set lag lengths $Le = Lr$, using common lag lengths of 1–10.

3.2. Multicollinearity test

Before finalizing the candidates for the input of our framework, we might need to eliminate the possible multicollinearity among the selected influential factors, that is, by removing some of the redundant factors. A high degree of multicollinearity might produce invalid results in predicting the market. Our adaptive filter is a time-varying linear model, thus the existence of multicollinearity of the input channels may cause the time-varying parameters to become unstable. To avoid such a problem, for each influential factor that is regressed with the others, a tolerance is calculated as $1 - R^2$, where R^2 is the coefficient of determination. A tolerance close to 1 means there is little multicollinearity, whereas a small value of tolerance suggests that multicollinearity should be noted.

4. Market input selection results

4.1. Data description

It is nature that all the indicators are market-dependent. In this section, we investigate influential factors of DJIA and SSE. The influential factors of the US and China markets are also compared to find differences between the developed market and the emerging market. The global financial crisis in 2007 is highlighted, in which the Shanghai stock market plummeted more than 70% in one year from October 2007. Almost concurrently, the US stock market began to crash, following the collapse of the housing bubble. Studying the market behavior in this period would help us understand what the major exogenous influential factors are, how the market reacts to them, and most importantly, how the market is likely to behave if it confronts similar situations in the future. We specifically investigate the DJIA from January 2008 to November 2011, the period right after the global financial crisis in 2007. For the China market, We take the daily closing prices of the SSE index from the beginning of 2007 to November 2011, including the rapid rise and crash phases, for examination. All the stock market data are obtained from the Yahoo Finance database.

4.2. US market

4.2.1. Empirical selection
Benefiting from its maturity, various derivatives are available for the US stock market, providing a valuable source of sentiment indicators. Sun [39] provides an empirical analysis on a wide range of indicators, based on which five leading economic and sentiment indicators are selected to investigate the contributions among them in generating the external market force. Of the five selected factors, three are the economic indicators, i.e., the interest rate indicator (IRI), the oil price

(OP), and the Baltic dry index (BDI); and two are sentiment indicators, i.e., the Chicago Board Options Exchange DJIA volatility index (VXD), and the exchange rate of the Euro against the Japanese Yen (EUR/JPY).

1. *Interest Rate Indicator (IRI)*

   Among the economic indicators, the interest rate is one of the most well-known factors that have a direct impact on the general trend of the stock market. The size of the federal funds rate target (FFRT) changes is often applied to investigate the role that the interest rate plays in the US stock market [40]. Since the federal funds future rate (FFFR) embodies market expectation on the monthly average of the daily effective funds rate, it is frequently used in measuring market reactions to interest rate changes [9,31]. We construct a similar indicator that takes market expectation on a daily basis into consideration. Since the daily effective federal funds rate (DEFFR)[1] is a volume-weighted average of rates on trades arranged by a group of federal funds brokers who report to the Federal Reserve Bank of New York each day, we use the difference between DEFFR and FFRT, defined as the interest rate indicator (IRI), as a gauge of the influence of interest rates on the US stock market:

$$IRI = DEFFR - FFRT. \tag{13}$$

2. *Oil Price (OP)*

   From the commodity sector, we select changes in the oil price (OP) as another key economic indicator. The oil price used in this paper is the Cushing West Texas Intermediate (WTI) spot price, obtained from Energy Information Administration. Its relative change $c(n)$ is calculated by

$$c(n) = 100 \times \frac{o(n) - o(n-1)}{o(n-1)}, \tag{14}$$

where $o(n)$ is the value of oil price.

3. *Baltic Dry Index (BDI)*

   The third economic indicator is the Baltic Dry Index (BDI), a daily average of global shipping prices for dry bulk cargoes. Functioning as an assessment of global trade and free of manipulation and speculation, it acts as an excellent leading indicator of economic activities. We also use its changes calculated with the same fashion as that in (14) to measure its influence on the stock market.

4. *Chicago Board Options Exchange DJIA Volatility Index (VXD)*

   Two sentiment indicators chosen for the DJIA are the VXD and the EUR/JPY currency pair. The VXD is a futures contract based on the prices of options on the DJIA traded at the Chicago Board Options Exchange (CBOE). It is a kind of stock fear index, reflecting the market expectation of the DJIA volatility over the next 30 days. Again, the relative changes of the VXD and the EUR/JPY exchange rate, calculated in the similar way as that in (14), are used in our process.

5. *Exchange rate of the Euro against the Japanese Yen (EUR/JPY)*

   Financial crisis would inevitably disseminate fear, resulting in investors becoming risk-averse. The currency pair EUR/JPY is a good choice for such an indicator. Besides the US dollar, the Euro and Japanese Yen are the most traded currencies. Similar as the US dollar, they are also held as reserve currencies by many countries. Among the EUR crosses and JPY crosses, the EUR/JPY is the most popular one which often has the highest trading volume. In 2013, the EUR/USD volume exceeded USD/JPY volume for the first time. It reflects the market expectation on the movement of the US economy and thus has its impact on the attractions of investments to the US market. This kind of influence is always reflected on the US stock market. In the work of Sun [39], the empirical comparison of the DJIA and the EUR/JPY exchange rate suggested that they nourish each other through interactions. Due to this reason, we choose a currency pair that does not include the US dollar.

Before conducting tests, we need to synchronize the data of the indicators and the stock prices. Specifically, if an indicator has no data available on certain samples, on which the stock market still traded, we will use the value of the indicator from the immediately preceding sample. On the other hand, if the stock price has no data available on a certain sample, but the indicator does, the latter will be removed accordingly. After synchronization, we next perform data normalization for all the series except the IRI using

$$\tilde{v}(n) = \frac{v(n) - \bar{v}}{s_v}, \tag{15}$$

where $v(n)$ denotes the original series, $\tilde{v}(n)$ denotes the normalized series, $\bar{v}$ and s_v are respectively the mean and the standard deviation of v.

The internal residue of DJIA from January 2008 to November 2011 is obtained by using the following OE model as the internal model:

---

[1] The definition is from the Federal Reserve Bank of New York.

$$H_{oe}(z) = \begin{bmatrix} \frac{0.9574z^{-1}-0.5034z^{-2}-0.321z^{-3}+0.565z^{-4}}{1-1.329z^{-1}+0.7312z^{-2}} \\ \frac{-5.127z^{-1}+2.086z^{-2}+0.3914z^{-3}+0.8506z^{-4}}{1-0.62z^{-1}-0.1239z^{-2}} \\ \frac{2.417z^{-1}+3.049z^{-2}-2.298z^{-3}-1.82z^{-4}}{1-0.3215z^{-1}-0.4643z^{-2}} \end{bmatrix}^T . \tag{16}$$

More details regarding to the OE model estimation and internal residue can be found in [45,46].

4.2.2. Causality test

When the lag length of the internal residue of the DJIA reaches 4, the corresponding Durbin–Watson statistic is 1.98 and its p value of Breusch–Godfrey test is 0.43, both showing no autocorrelation in the residuals. We thus set the lag length of the series of e_i to 4. To evaluate the effectiveness of indicators within a half month period over the internal residue, the lag lengths of all input indicators are set to 10. The resulting time-varying causal relationships between each of the five selected indicators and the internal residue of the DJIA are shown in Figs. 2–6. For each pair, we can observe that the causality strength from the indicator to the internal residue is higher than that in the opposite direction. Threshold values are then calculated by the surrogate data approach. As shown in our testing results, four out of five indicators (except the pair associated with the BDI) significantly Granger cause the internal residue as their causality strengths exceed the corresponding thresholds over the entire sampling period.

The BDI only presents significant linear causal effect over the internal residue of the DJIA from September 29, 2008 to July 21, 2010. Therefore, considering the entire period as a whole, it is necessary to further conduct a nonlinear Granger causality test on the pair associated with the BDI. Table 1 reports the results of the nonlinear Granger causality test applied to BDI-related series $\Sigma_{12}(n)$ and $\Sigma_{22}(n)$ with the same lag lengths. The p value is calculated for a one-sided test.

The null hypothesis is that the BDI does not nonlinearly cause the internal residue of the DJIA. As shown in Table 1, this null hypothesis is rejected at the 5% significance level when the lag length is larger than two. This is a strong evidence of nonlinear Granger causality from the BDI to the internal residue of DJIA, especially when the lag length is long. As such, we conclude from previous causality tests that these five indicators all Granger cause the internal residue of the DJIA, statistically supporting the rationality of our input selection.

4.2.3. Multicollinearity test

The multicollinearity test is performed among the indicators and the results are shown in Table 2. We note that the EUR/JPY exchange rate gives a relatively small tolerance, and thus could be removed. The IRI, OP, BDI, and VXD indicators are finally used as the input to our system adaptation framework to forecast the DJIA.

4.2.4. Prediction results with selected input

Having four selected influential factors, we proceed to estimate the hyperparameters for each influential indicator, i.e. the diagonal entries of Noise Variance Ratio matrix Q_r [45] which determines the variations of the state variables. Initially, we set all the hyperparameters as 0.002 and the covariance matrix $\hat{P}$ as a diagonal matrix with all its diagonal entries equal to 10^5. Note that the same initial values will be adopted when combining all indicators together as the input for the prediction in the next section. In this part, we have 14 state variables and 14 hyperparameters in the Q_r matrix. With the estimated
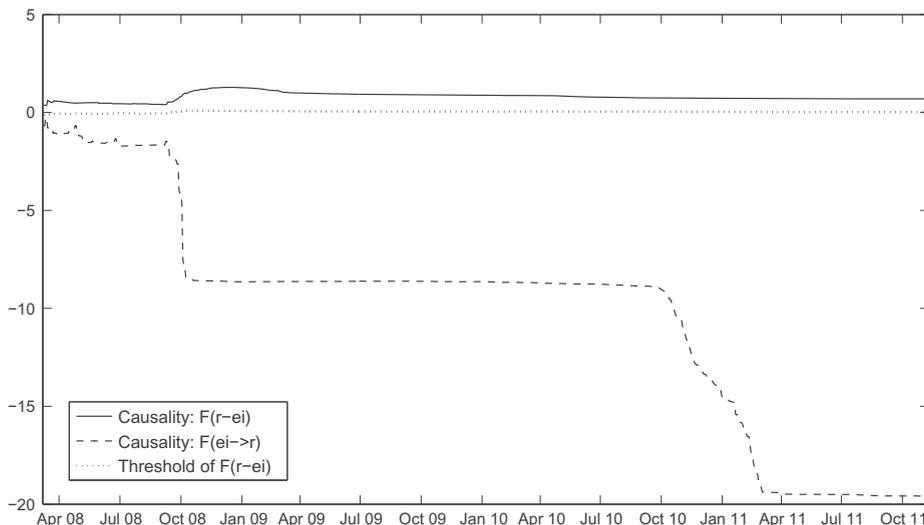

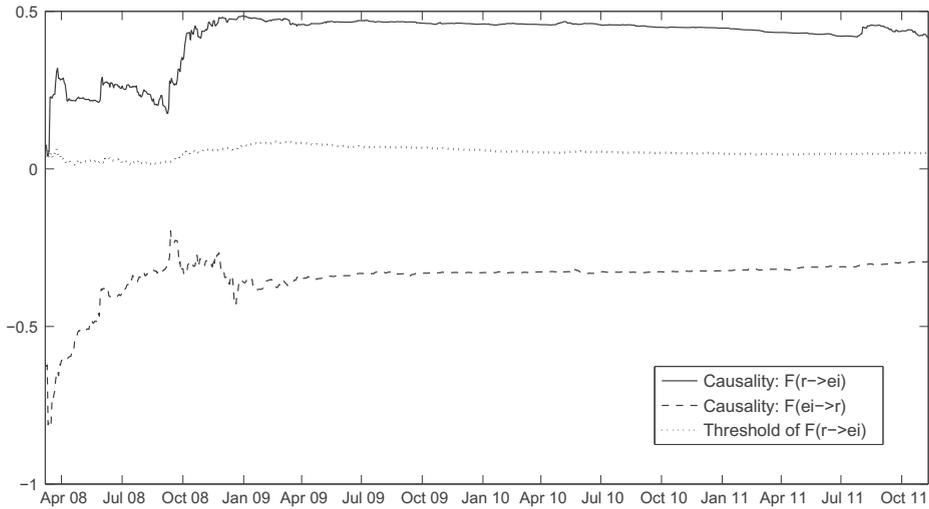**Fig. 2.** Time-varying causality between the internal residue of the DJIA and the IRI.

Fig. 3. Time-varying causality between the internal residue of the DJIA and the OP.
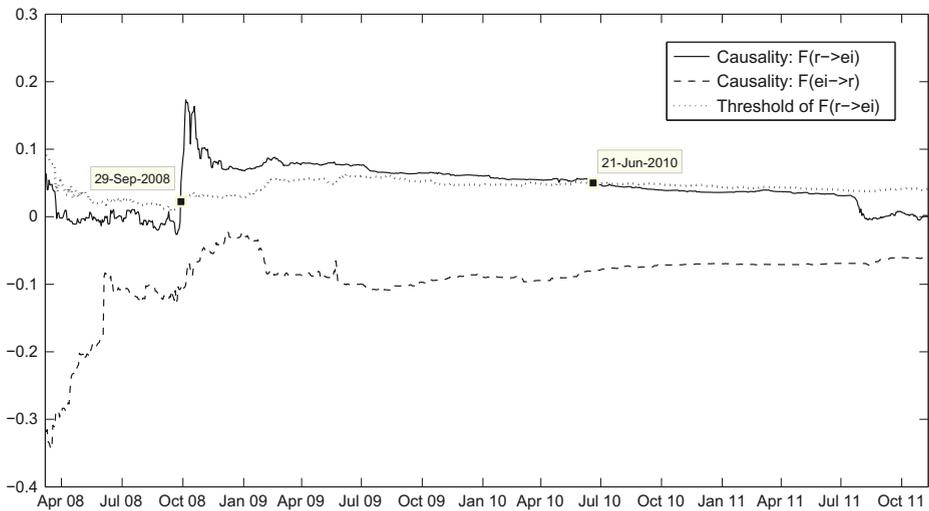

Fig. 4. Time-varying causality between the internal residue of the DJIA and the BDI.

hyperparameters, one-step-ahead prediction results can be obtained. Related definitions and other details please refer to [45].

For the period of the DJIA data under studies, we divide the market data into four subperiods by observing the corresponding internal residue, and to highlight the performance of our proposed framework during the 2007 US sub-prime financial crisis. Subperiod S1 is referred to the time interval from September to December 2008; Subperiod S2 from January 2009 to April 2010; Subperiod S3 from May to December 2010; and lastly, Subperiod S4 lasts from January to November 2011. It is easy to observe that in Subperiod S1, which starts from the month when the financial crisis hit its most critical stage, the variance of the corresponding internal residue is extremely large. After becoming relatively small in Subperiod S2, the variance of the internal residue increases again in May 2010, which is the beginning of Subperiod S3. It becomes much larger during Subperiod S4, corresponding to the downturn of the market in 2011.

We compare the predicting ability of our framework with the commonly adopted autoregressive moving average model with exogenous input (ARMAX). The lag lengths for the ARMAX model are set similarly as those used in our framework, i.e., 4 for the AR and MA (moving average) terms and 10 for all the exogenous inputs. Summarized in Table 3 are the prediction error results of our proposed framework and the ARMAX method. From the mean average error (MAE) and the root mean squared error (RMSE), it is obvious that the system adaptation framework significantly outperforms the ARMAX approach, especially in Subperiods S1 and S4. The effectiveness of the system adaptation framework structure, the ability of the dynamical design of the adaptive filter and the distinguished function of the internal model are comprehensively testified.
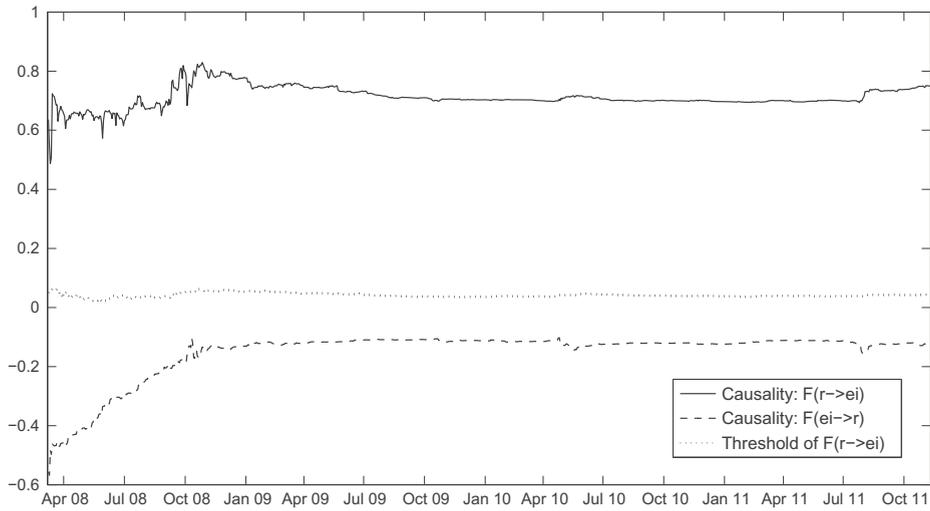
Fig. 5. Time-varying causality between the internal residue of the DJIA and the VXD.
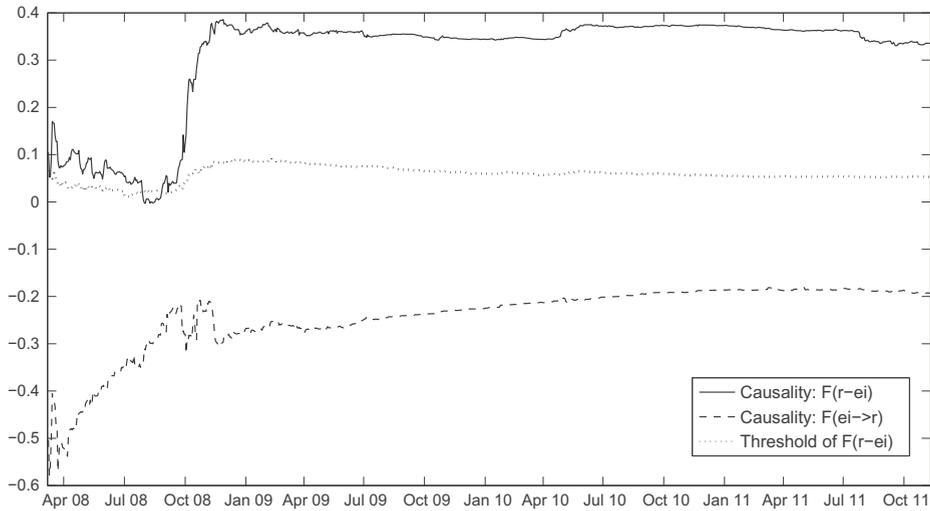

Fig. 6. Time-varying causality between the internal residue of the DJIA and the EUR/JPY.

Table 1
Nonlinear Granger causality test results in the US stock market.

$Le = Lr$	BDI does not cause the external force of DJIA		
	CS	TVAL	p-Value
1	9.1884×10^{-5}	0.0435	0.4826
2	0.0040	1.0691	0.1425
3	0.0116	1.7759	0.0379**
4	0.0200	2.0952	0.0181**
5	0.0300	2.3977	0.0082**
6	0.0467	2.9158	0.0018**
7	0.0625	3.0650	0.0011**
8	0.0690	2.8864	0.0019**
9	0.0731	2.7644	0.0029**
10	0.0790	2.4545	0.0071**

** Significance at 5% level for a one-sided test.

Table 2
Multicollinearity test results in the US stock market.

	IRI	OP	BDI	VXD	EUR/JPY
Tolerance	0.971	0.820	0.992	0.785	0.626

Table 3
Comparison of the prediction results between the ARMAX and the system adaptation framework approaches for the DJIA index.

	Subperiod 1 September 2008–December 2008		Subperiod 2 January 2009–April 2010	
	MAE	RMSE	MAE	RMSE
ARMAX	215.61	272.47	88.09	115.96
System adaptation framework	29.82	38.69	29.81	39.50
Improvement	86.17%	85.80%	66.16%	65.94%
	Subperiod 3 May 2010–December. 2010		Subperiod 4 January 2011–November 2011	
ARMAX	85.60	119.27	110.15	147.87
System adaptation framework	36.18	48.77	35.48	45.83
Improvement	57.73%	59.11%	67.79%	69.01%

It is clear that the predicting or forecasting ability of the system adaptation framework is far superior especially for complicated economic situations when the market is highly volatile. For example, in Subperiod S1, our framework gives amazingly accurate prediction results whereas the ARMAX model totally fails to measure the dynamics of the market. Nevertheless, we should also note that under the framework structure, the predictions in Subperiods S1 and S2 are better than those in Subperiods S3 and S4. Such a phenomenon could probably be due to the weakening of the determinant effect of the input sources. For references, we note that the average daily changes of the closing prices of the DJIA from Subperiod S1 to S4 are $263.82, 85.61, 79.29$ and 110.19, respectively. Compared with the MAE generated by the ARMAX which are slightly within the range of this average daily price changes, our framework provides much smaller and reasonable MAE in all subperiods. Hence, all these results prove that this double selection method works well in selecting the appropriate input to our framework. With these input, our framework shows its excellent ability in understanding the dynamics of complex systems especially in the complicated situations. In the next subsection, we apply the same selection method to the China stock market.

4.3. China market

4.3.1. Empirical selection

The Shanghai Stock Exchange (SSE) is the largest market in China and the third largest in the world by market capitalization. However, it is still not entirely open to foreign investors and is significantly influenced by the Chinese government. Many characteristics of the China stock market are very unique. In this section, we once again focus on the impact of the 2007 global financial crisis on the China market. We take the daily closing prices of the SSE index from the beginning of 2007 to November 2011, including the rapid rise and crash phases, for examination. The influential factors of the China and US markets are also compared to find differences between the developed market and the emerging market.

The internal residue of the SSE is obtained through the same procedure as that for the DJIA studied previously. The internal OE model for the SSE under the system adaptation framework is estimated based on its daily closing prices from 2002 to 2006, and the resulting OE model is given by

$$H_{\text{SSE}}(z) = \begin{bmatrix} \frac{0.5799z^{-1}-0.7365z^{-2}+0.506z^{-3}-0.2967z^{-4}}{1-0.7929z^{-1}-0.1169z^{-2}} \\ \frac{-2.234z^{-1}+1.558z^{-2}-1.158z^{-3}-0.3087z^{-4}}{1-0.8926z^{-1}+0.6835z^{-2}} \\ \frac{2.646z^{-1}-1.092z^{-2}-1.061z^{-3}-0.1253z^{-4}}{1-0.6328z^{-1}-0.2496z^{-2}} \end{bmatrix}^{\text{T}}. \tag{17}$$

Subsequently we proceed to identify potential economic and sentiment influential factors for the SSE. Various influential factors of the China stock market have been studied in the literature. For example, Zheng and Wong [47] adopted a two-stage bivariate GARCH model to analyze the conditional dependence between the so-called A-type and B-type Shares in the China stock market, and the impacts of the US and Hong Kong markets over the Chinese counterpart. Yao et al. [44] analyzed the relationships between the SSE and ten banking stocks listed in the market. The indicators related to the interest rate, money supply and inflation have also been investigated and reported.

Generally, there is a lack of sentiment indicators for the China markets. Since Chinese derivatives market is still relatively small and some market data are manipulated by the Chinese government, it is difficult to find influential factors that reflect

the true sentiment of investors. Even for the commodities like the oil, China has a totally different price schedule, which is almost completely controlled by the government, regardless of the global price trend. As a result, we cannot make use of many common influential factors that work well in the developed countries. After some intensive search and testing, we finally select the following indicators that are suitable for the SSE.

1. Interest Rate Indicator (IRI):

   For a stock market, the interest rate is always the primary influential factor under consideration. The Shanghai Interbank Offered Rate (SHIBOR) aims to become a new interest rate benchmark in the China market, and to provide the similar functions and roles as those of the Federal Funds Rate (FFR) in the US market and the London Inter Bank Offered Rate (LIBOR) in the UK market. Its maturity is important in making interest rates more market-based. Thus, researchers usually adopt SHIBOR to study the interaction between the interest rate and the China stock market. We use the changes between the daily SHIBOR overnight rate as the corresponding Interest Rate Indicator (IRI), which is defined similarly as that in (14):

   $$\text{SHIBOR}(n) = 100 \times \frac{i_c(n) - i_c(n-1)}{i_c(n-1)}, \tag{18}$$

   where i_c is the value of the SHIBOR overnight rate. Since China has just experimented the SHIBOR trial from October 2006, the IRI data are only available from then.

2. International Stock Market Indicator (ISMI):

   The interactions among stock markets around the world have become more and more intensified nowadays. The dynamic relationships between the China stock market and the stock markets in other countries have been extensively studied in the literature, among which the influence of the US stock market is pervasive. In terms of daily stock returns, Laurencec et al. [32] found that the US stock market has a strong causal effect to both the China and Hong Kong stock markets. In Chen et al. [11], it was reported that the Standard & Poor's 500 Index (S&P 500) led the SSE with respect to return transmission. Hung also did a similar study on several stock market indices [28]. Studies had also been conducted on other markets, which include the Hong Kong market [32,47], the Japan market [43], and the India market [11]. In this paper, we use the internal residues of the S&P 500 and Hong Kong Hang Seng Index (HSI), generated by our system adaptation framework, to represent the influences of the international stock markets on the SSE. More specifically, the internal residues of the S&P 500 and the HSI are generated by the system adaptation framework with the following respective internal OE models

   $$H_{\text{S\&P500}}(z) = \begin{bmatrix} \frac{-45.94z^{-1}-14.02z^{-2}+9.758z^{-3}-0.098z^{-4}}{1+0.2986z^{-1}-0.2124z^{-2}} \\ \frac{12.62z^{-1}+34.94z^{-2}-37.14z^{-3}+12.06z^{-4}}{1-0.8782z^{-1}+0.2605z^{-2}} \\ \frac{2.489z^{-1}-13.79z^{-2}-3.118z^{-3}+7.741z^{-4}}{1+0.1254z^{-1}-0.5448z^{-2}} \end{bmatrix}^{\text{T}}, \tag{19}$$

   and

   $$H_{\text{HSI}}(z) = \begin{bmatrix} \frac{2.418z^{-1}+2.45z^{-2}-0.108z^{-3}-0.6531z^{-4}}{1+1.158z^{-1}+0.1604z^{-2}} \\ \frac{0.0552z^{-1}-1.573z^{-2}+0.0189z^{-3}+1.519z^{-4}}{1+0.0662z^{-1}-0.9176z^{-2}} \\ \frac{2.573z^{-1}-1.364z^{-2}-3.253z^{-3}+1.912z^{-4}}{1+0.0694z^{-1}-0.9244z^{-2}} \end{bmatrix}^{\text{T}}. \tag{20}$$

   These OE models are respectively obtained by using their corresponding historical data for a period from 2001 to 2005.

3. Exchange Rate of the US Dollar against the Chinese Yuan (USD/CNY):

   China has reformed its currency policies in recent years, shifting to a flexible exchange rate regime and pegging the Chinese Yuan to a basket of foreign currencies rather than strictly tying to the US dollar. Since then, the Chinese Yuan has been appreciated a lot against the US dollar. It is allowed to float within a daily band of 0.5% around the central parity. This revaluation of the CNY/USD rate marked a new era of a managed floating exchange rate and influenced significantly the China stock market. Nieh and Yau [33] proved the existence of an asymmetric causal relationship between the appreciation of CNY/USD and the SSE. Similar causal relationships have also been confirmed by Yang [42] and Tian and Ma [41]. In our study, we select the change of the USD/CNY exchange rate as one of the influential factors for testing the Shanghai stock market.

4. *Inflation Rate Indicator (IFRI):*

   The relationship between the inflation rate and the stock market is still debatable. Chow and Lawler [14] found that the higher mean rate of return in the SSE than that in the New York Stock Exchange Composite Index was partially the result of a higher rate of inflation in China. Huang *et al.* [27] employed multiresolution wavelets to investigate three influential factors in different time scales. In their studies, the inflation was found to have an impact on the SSE in the 16-to-32-month trend, but it vanished in the 2-to-4-month cycles. The inflation rate $\pi(n)$ is defined based on the logarithmic changes of Consumer Price Index (CPI) $C(n)$ (from the EIU country database), i.e.,

   $$\pi(n) = 100 \times [\ln(C(n)) - \ln(C(n-1))]. \tag{21}$$

Since only the monthly CPI data are available, a cubic spline interpolation is used to increase its frequency. Inspired by the work of Huang *et al.* [27], we investigate the IFRI indicator at different frequencies, i.e., at the daily, weekly and monthly, respectively. It indeed turns out that the causal relationship changes at different frequencies.

4.3.2. Causality test

We first conduct a series of causality tests for the inflation rate indicator (IFRI) by testing the indicator at the daily, weekly and monthly frequencies, and the results are shown in Figs. 7–9. It is clear that the linear causal relationship is only significant at the weekly frequency. With the monthly data, although the inflation rate has significant causal effect to the SSE at the beginning, the causal relationship changes its direction after 2008. We further proceed to identify the nonlinear causal relationship using daily and monthly data. As shown in Table 4, there is no significant result found in the nonlinear tests. As such, the IFRI is not considered to be a final candidate for the input to the system adaptation framework of the SSE for daily prediction. Nonetheless, it could be used for weekly forecasting. These results suggest whether an indicator affecting the market is frequency-dependent. Similar study regarding to the influence of inflation rate on the US market has also been done at the daily, weekly and monthly frequencies. The weekly and daily data are created by a cubic spline interpolation which is the same as it is used in the China stock market. Testing results show that the inflation rate has linear causal relationships to the DJIA during the whole investigated period at all three frequency levels. However, since we focus on the daily prediction, we would not use the daily inflation rate as it is manipulated data by the interpolation. This kind of manipulated
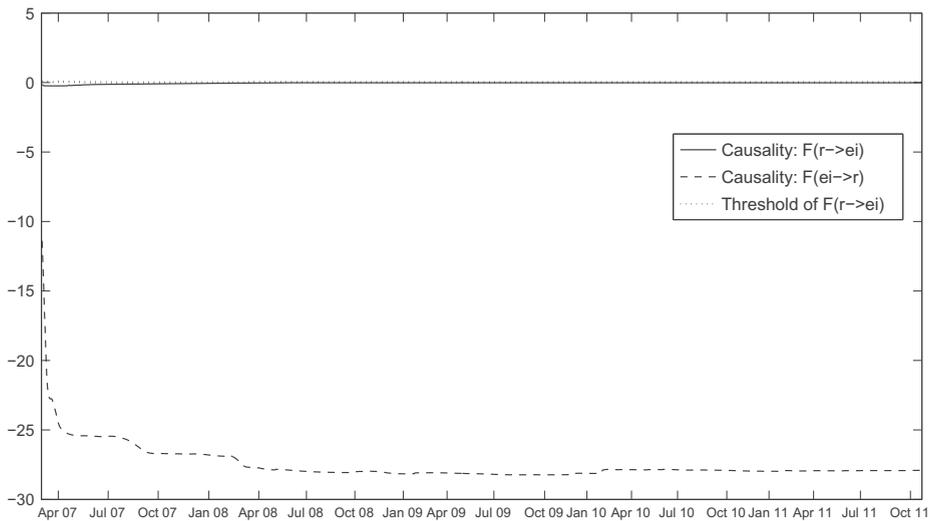

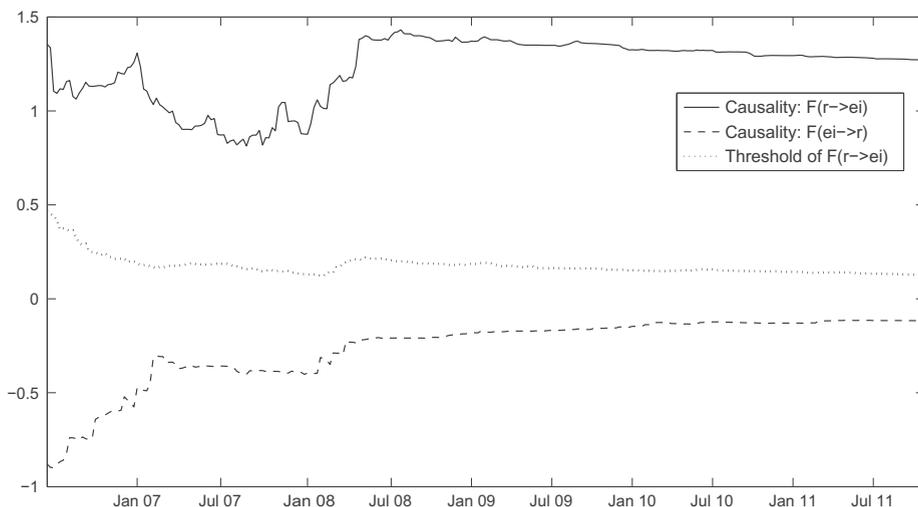**Fig. 7.** Time-varying causality between the internal residue of the SSE and IFRI (daily data).

Fig. 8. Time-varying causality between the internal residue of the SSE and IFRI (weekly data).
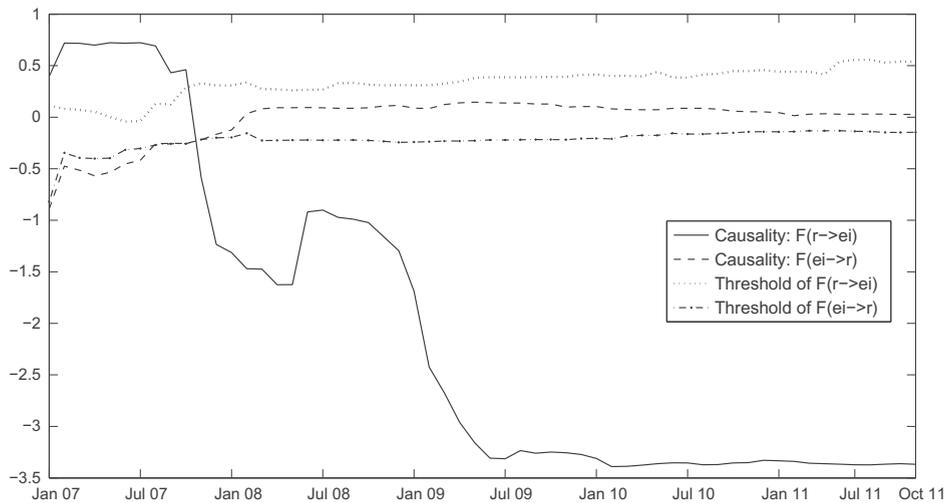
Fig. 9. Time-varying causality between the internal residue of the SSE and IFRI (monthly data).

Table 4
Nonlinear Granger causality test results between the internal residue of the SSE and IFRI at different frequencies.

$Le = Lr$	Daily IFRI does not cause SSE external force		Monthly IFRI does not cause SSE external force	
	CS	TVAL	CS	TVAL
1	−0.0749	−1.4569	−0.0084	−0.6020
2	−0.1293	−2.0022	−0.0229	−1.5541
3	−0.2297	−2.9022	−0.0445	−1.6982
4	−0.2577	−2.9222	−0.0546	−1.0853
5	−0.2661	−3.7184	−0.1252	−1.8803
6	−0.3202	−3.6521	−0.2701	−2.3571
7	−0.5006	−2.9899	−0.1941	−1.1481
8	−0.3514	−1.7628	−0.1789	−0.6418
9	−0.6746	−40.6963	0.0765	1.1455
10	NA	NA	NA	NA

* Significance at 10% level for a one-sided test.
** Significance at 5% level for a one-sided test.

data will only be considered when the available market data are significantly insufficient. The US stock market is famous for its maturity that sufficient market data are available. In this way, we exclude the inflation rate from the input indicator set of the DJIA.

Data adjustment and data normalization are applied to the remaining four indicators before conducting causality tests. The lag length is selected in a similar way as it is in the DJIA. For the internal residue of the SSE, when the order reaches 4, the Durbin–Watson statistic is 2.00 and the p value of Breusch–Godfrey test is 0.44, both showing no autocorrelation in the residuals. As such, the lag lengths of the internal residues of the SSE, S&P 500 and HSI are all set to be 4, and for other influential factors, it is set to 10. Figs. 10–13 present the time-varying causality between each of these indicators and the internal residue of the SSE. Only the S&P 500 after September 25, 2007 and the currency pair USD/CNY after January 30, 2008 significantly Granger cause the internal residue.

We next apply the nonlinear causality test to the IRI and the HSI in the whole period; the S&P 500 from January 4, 2007 to September 24, 2007; and the USD/CNY from January 4, 2007 to January 29, 2008, and the results are given in Table 5, in which 'NA' denotes that results not available in the nonlinear causality tests. It is caused by $C_2 = 0$ or $C_4 = 0$ in (10) so that the condition in (12) cannot be tested. We note that $C_2 = 0$ or $C_4 = 0$ means that with selected lag length and given conditions, there is no vectors of e_i and r, whose distance is within ϱ or less.

It can be observed from the obtained results that significant nonlinear causality is only evidenced in the HSI with the lag length from 1 to 9. The nonlinear causal relationship is not found in the direction from the IRI, or the beginning parts of the S&P 500 and the USD/CNY, to the internal residue of the SSE. Considering both the time-varying and nonlinear causality test results, we can conclude that the HSI over the whole testing period, the S&P 500 after September 25, 2007, and the USD/CNY after January 30, 2008, significantly Granger cause the internal residue of the SSE. We also apply the multivariate nonlinear causality test [3,4] to these input variables but no significant causal relationship is found.
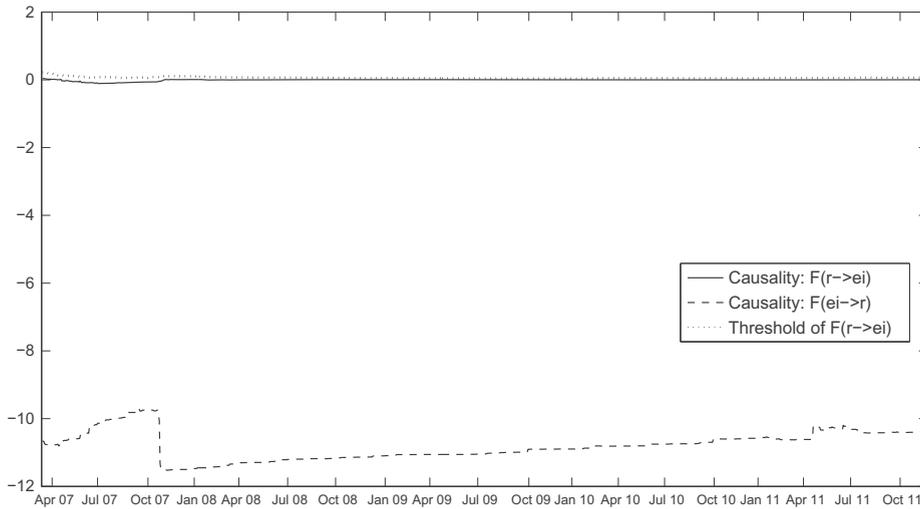
Fig. 10. Time-varying causality between the internal residue of the SSE and the IRI.
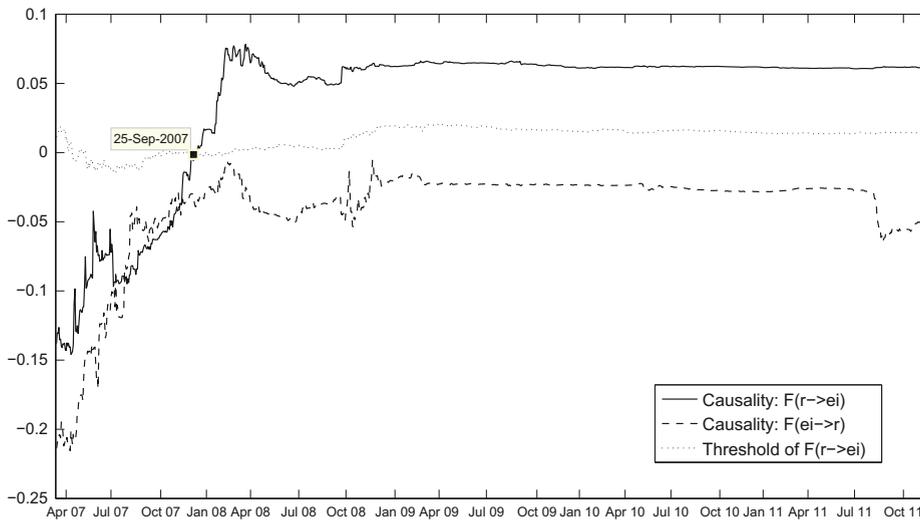

Fig. 11. Time-varying causality between the internal residues of the SSE and S&P 500.

4.3.3. Multicollinearity test

The tolerance is calculated for each of the indicators that significantly Granger cause the external force of SSE, i.e. S&P 500, HSI and USD/CNY. In Table 6, all the tolerances are all close to 1, which means little multicollinearity exists. In this way, all these three indicators are used as the input to the SSE from January 30, 2008 to November of 2011.

4.3.4. Prediction results with selected input

As in the DJIA case, we partition the time interval of interest for the SSE into four subperiods according to different phases of the market and the variance of its internal residue. The training period for estimating the adaptive filter is from January 30, 2008 to the end of April 2008. Subperiod S1 is from May to December 2008, a period characterized as a steep decline. From January 2009, the China stock market began to recover until July 2009. Then, it came into an oscillation period. As such, Subperiod S2 is set to be from January to July 2009. The oscillation period, from August 2009 to December 2010, is Subperiod S3. In 2011, the market turned down again. We therefore define Subperiod S4 as from January 2011 to November 2011.

The initial hyperparameters in the adaptive filter (the diagonal entries of Q_r) are also set to be 5.5×10^{-4} and other initial conditions are set to be the same as those for the DJIA. The resulting estimations of the hyperparameters in the adaptive filter, i.e., the coefficients associated with autoregressive part, are

$$4.6958 \times 10^{-5}, \quad 3.1908 \times 10^{-5}, \quad 6.2916 \times 10^{-7}, \quad 7.7498 \times 10^{-6}, \tag{22}$$
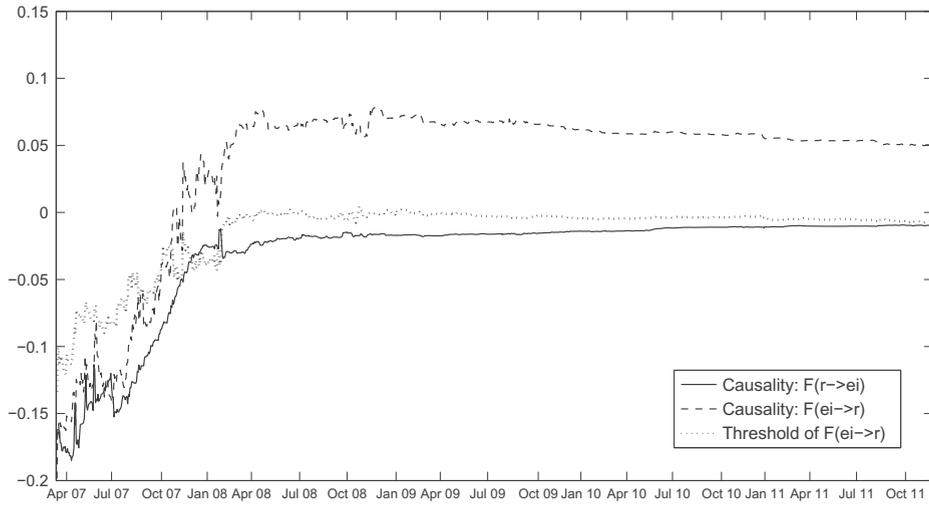
Fig. 12. Time-varying causality between the internal residues of the SSE and HSI.
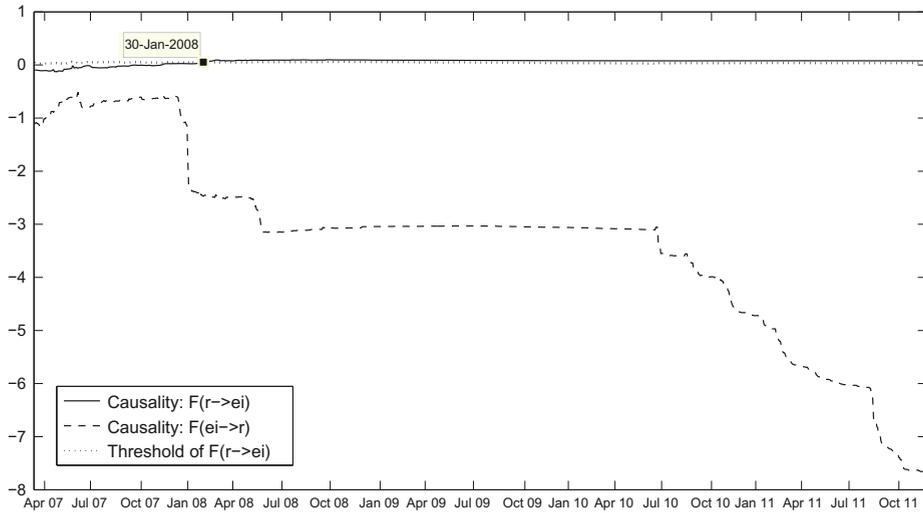

Fig. 13. Time-varying causality between the internal residues of the SSE and USD/CNY.

the estimated coefficients associated with the S&P 500 input are

$$5.1311 \times 10^{-4}, \quad 0.0980, \quad 0.2599, \quad 0.3537, \tag{23}$$

the estimated coefficients associated with the HSI input are

$$0.1398, \quad 4.2367 \times 10^{-4}, \quad 3.7451 \times 10^{-4}, \quad 0.1066, \tag{24}$$

and finally, the estimated coefficients associated with the USD/CNY input are

$$8.1674 \times 10^{-5}, \quad 0.2927, \quad 5.1603 \times 10^{-7}, \quad 7.5472 \times 10^{-4}, \quad 7.9529 \times 10^{-5}, \tag{25}$$

$$0.0114, \quad 6.6762 \times 10^{-6}, \quad 8.6714 \times 10^{-5}, \quad 6.5181 \times 10^{-5}, \quad 6.2095 \times 10^{-8}. \tag{26}$$

With the estimated hyperparameters, the one-day-ahead prediction results are obtained and shown in Table 7. It is once again confirmed that our framework gives more accurate one-step-ahead predictions than the conventional ARMAX approach. For illustration, we calculate the average daily changes of the SSE closing prices in the four subperiods, which are $53.60, 35.63, 36.26$ and 23.30, respectively. The MAE of the ARMAX model is about the same as the average daily changes, whereas the MAE resulting from our framework is much smaller and much more reasonable. Nevertheless, we should also note that the prediction results for the China market are not as good as those for the US market. One possible reason could be

Table 5
Nonlinear Granger causality test results in the China stock market.

$Le = Lr$	IRI does not cause SSE external force		S&P 500 does not cause SSE external force (January 4, 2007–September 24, 2007)	
	CS	TVAL	CS	TVAL
1	0.0121	0.9849	−0.0017	−0.2824
2	0.0021	0.0572	−0.0037	−0.3319
3	−0.1590	−1.8365	0.0014	0.0721
4	−0.1892	−1.0867	−0.0042	−0.1917
5	−0.3318	−1.5631	0.0007	0.0205
6	−0.2217	−0.9902	−0.0360	−0.7195
7	−0.7098	−41.6938	−0.0747	−0.8306
8	NA	NA	−0.0816	−0.3760
9	NA	NA	−0.0287	−0.1039
10	NA	NA	−0.2381	−1.3421
	HSI does not cause SSE external force		USD/CNY does not cause SSE external force (January 4, 2007–January 29, 2008)	
1	0.0091	3.0486**	−0.0024	−0.1740
2	0.0193	4.2702**	−0.0101	−0.4492
3	0.0261	4.1049**	−0.0204	−0.6314
4	0.0344	4.0747**	0.0130	0.2560
5	0.0329	3.2191**	0.0708	1.0750
6	0.0205	1.7307**	0.1196	0.9968
7	0.0274	2.0246**	0.5032	6.6323**
8	0.0221	1.3425*	0.5279	6.6009**
9	0.0255	1.3359*	NA	NA
10	0.0255	1.1054	NA	NA

* Significance at 10% level for a one-sided test.
** Significance at 5% level for a one-sided test.

Table 6
Multicollinearity test results in the China stock market.

	S&P 500	HSI	USD/CNY
Tolerance	0.863	0.977	0.867

Table 7
Comparison of the prediction results between the ARMAX approach and the proposed framework for the China stock market.

	Subperiod 1 May 2008–December 2008		Subperiod 2 January 2009–July 2009	
	MAE	RMSE	MAE	RMSE
ARMAX	52.00	67.47	34.44	42.12
System adaptation framework	19.63	30.30	16.96	24.56
Improvement	62.26%	55.09%	50.75%	46.75%
	Subperiod 3 August 2009–December 2010		Subperiod 4 January 2011–November 2011	
ARMAX	36.30	49.60	23.70	30.41
System adaptation framework	26.56	38.13	12.79	18.22
Improvement	26.82%	23.13%	46.05%	40.07%

due to the inaccurate data that we gather from the open sources. Another possible way to enhance the prediction performance is to incorporate the influence of the Chinese Government, if possible, as an input to the system adaptation framework. The China stock market is clearly interfered by some factors beyond the common market influential sources. For example, the interest rate is believed to be the most important and direct factor that affects a stock market, but the SHIBOR, a market-based interest rate that reflects the tightness of market liquidity, does not Granger cause the internal residue of the SSE at all. Last but not least, the China stock market is still not well regulated and is too immature to have sentiment indicators as which have seriously affected the developed markets. This is also a common feature of emerging markets.

5. Conclusion

In this paper, we have proposed a complete double selection method in identifying external influential factors for a particular stock market. Both time-varying and nonlinear causality tests play a decisive role in the selection procedure, while the multicollinearity test is mainly used to exclude some redundant indicators. These tests are carried out between the influential factors and the internal residue of a stock market index. The influential factors are preselected by some empirical research and tested by this double selection method.

We have applied this selection method to the US and China stock markets. Our work shows that influential factors are market-dependent and frequency-dependent. The interest rate, oil price, VXD, BDI and EUR/JPY are found to Granger cause the external force of DJIA while SSE is only affected by USD/CNY and international stock markets which are represented by S&P 500 and HSI. We have also investigated inflation rate at different frequencies including interpolated weekly and daily data, but in the China stock market, a significant causality relationship is only testified based on weekly data. With appropriate influential factors as the input, our system adaptation framework provides accurate one-step-ahead prediction results. Comparing the US market with the China market, we also found the differences between a developed market and an emerging market. Influential factors tend to be complicated and hard to find in the emerging market due to its immaturity.

References

[1] G. Armano, M. Marchesi, A. Murru, A hybrid genetic-neural architecture for stock indexes forecasting, Inform. Sci. 170 (1) (2005) 3–33.
[2] E. Baek, W. Brock, A General Test for Non-Linear Granger Causality: Bivariate Model, Working paper, Iowa State University and University of Wisconsin, Madison, WI, 1992.
[3] Z. Bai, W.-K. Wong, B. Zhang, Multivariate linear and non-linear causality tests, Math. Comput. Simul. 81 (2010) 5–17.
[4] Z. Bai, H. Li, W.-K. Wong, B. Zhang, Multivariate causality tests with simulation and application, Stat. Probab. Lett. 81 (8) (2011) 1063–1071.
[5] M. Baker, J. Wurgler, Investor sentiment and the cross-section of stock returns, J. Finan. 61 (2006) 1645–1680.
[6] M. Baker, J. Wurgler, Investor sentiment in the stock market, J. Econ. Perspect. 21 (2007) 129–151.
[7] M. Beine, G. Capelle-Blancard, H. Raymond, International nonlinear causality between stock markets, Euro. J. Finan. 14 (2008) 663–686.
[8] G. Bekaert, E. Engstrom, Inflation and the stock market: understanding the "Fed Model", J. Monet. Econ. 57 (3) (2010) 278–294.
[9] B.S. Bernanke, K.N. Kuttner, What explains the stock market's reaction to federal reserve policy?, J Finan. 60 (3) (2005) 1221–1257.
[10] S.L. Bressler, A.K. Seth, Wiener–Granger causality: a well established methodology, NeuroImage 58 (2) (2011) 323–329.
[11] H. Chen, B.J. Lobo, W.-K. Wong, Links Between the Indian, US and Chinese Stock Markets, Working paper, No. 0602, National University of Singapore, 2006.
[12] S.-H. Chen, C.-C. Liao, Agent-based computational modeling of the stock price-volume relation, Inform. Sci. 170 (1) (2005) 75–100.
[13] O. Choustova, Quantum probability and financial market, Inform. Sci. 179 (5) (2009) 478–484.
[14] G.C. Chow, C.C. Lawler, A time series analysis of the Shanghai and New York stock price indices, Ann. Econ. Finan. 4 (2003) 17–35.
[15] M. Ding, S.L. Bressler, W. Yang, H. Liang, Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment, Biol. Cybern. 83 (2000) 35–45.
[16] N. Feng, H.J. Wang, M. Li, A security risk analysis model for information systems: causal relationships of risk factors and vulnerability propagation analysis, Inform. Sci. 256 (2014) 57–73.
[17] G. Filis, Macro economy, stock market and oil prices: do meaningful relationships exist among their cyclical fluctuations?, Energy Econ 32 (4) (2010) 877–886.
[18] J.F. Geweke, Measurement of linear dependence and feedback between multiple time series, J. Am. Stat. Assoc. 77 (1982) 304–324.
[19] J.F. Geweke, Measures of conditional linear dependence and feedback between time series, J. Am. Stat. Assoc. 79 (1984) 907–915.
[20] J.G. De Gooijera, S. Sivarajasingham, Parametric and nonparametric Granger causality testing: linkages between international stock markets, Phys. A: Statist. Mech. Appl. 387 (2008) 2547–2560.
[21] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (3) (1969) 424–438.
[22] C.W.J. Granger, Forecasting in Business and Economics, second ed., Academic Press, San Diego, 1989.
[23] W. Hesse, E. Moller, M. Arnold, B. Schack, The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies, J. Neurosci. Meth. 124 (2003) 27–44.
[24] C. Hiemstra, J.D. Jones, Some Information Relating to the Finite Sample Properties of the Modified Baek and Brock Nonlinear Granger Causality Test, Working Paper, University of Strathclyde and Securities and Exchange Commission, 1993.
[25] C. Hiemstra, J.D. Jones, Testing for linear and nonlinear Granger causality in the stock price-volume relation, J. Finan. 49 (1994) 1639–1664.
[26] D. Hsieh, Chaos and nonlinearity dynamics: application to financial market, J. Finan. 46 (1991) 1839–1877.
[27] L. Huang, C. Wang, S. Qin, A multiresolution wavelet based analysis of the Chinese stock market, in: Proceedings of 2010 Third International Conference on Business Intelligence and Financial Engineering, 2010, pp. 305–309.
[28] J.-C. Hung, A fuzzy asymmetric GARCH model applied to stock markets, Inform. Sci. 179 (22) (2009) 3930–3943.
[29] A. Kanas, Nonlinearity in the stock price-dividend relation, J. Int. Money Finan. 24 (2005) 583–606.
[30] S.P. Kothari, J. Lewellen, J.B. Warner, Stock returns, aggregate earnings surprises, and behavioral finance, J. Finan. Econ. 79 (3) (2006) 537–568.
[31] K.N. Kuttner, Monetary policy surprises and interest rates: evidence from the Fed funds futures market, J. Monet. Econ. 47 (3) (2001) 523–544.
[32] M. Laurencec, F. Cai, S. Qian, Weak-form efficiency and causality tests in Chinese stock markets, Multinat. Finan. J. 1 (4) (1997) 291–307.
[33] C.-C. Nieh, H.-Y. Yau, The impact of renminbi appreciation on stock prices in China, Emerg. Mark. Finan. Trade 46 (1) (2010) 16–26.
[34] H.V. Phama, E.W. Cooperb, T. Caob, K. Kameib, Hybrid Kansei-SOM model using risk management and company assessment for stock trading, Inform. Sci. 256 (2014) 8–24.
[35] Z. Psaradakis, M.O. Ravn, M. Sola, Markov switching causality and the money–output relationship, J. Appl. Econom. 20 (2005) 665–683.
[36] A. Roebroeck, E. Formisano, R. Goebel, Mapping directed influence over the brain using Granger causality and fMRI, Neuroimage 25 (2005) 230–242.
[37] J. Shanken, M.I. Weinstein, Economic forces and the stock market revisited, J. Empir. Finan. 13 (2) (2006) 129–144.
[38] P. Silvapulle, J.-S. Choi, Testing for linear and nonlinear Granger causality in the stock price-volume relation: Korean evidence, Quart. Rev. Econ. Finan. 39 (1999) 59–76.
[39] W. Sun, A Comprehensive Approach to Predicting Market Bottoms, Bachelor of Engineering Thesis, Department of Electrical & Computer Engineering, National University of Singapore, 2009.
[40] D.L. Thornton, Tests of the market's reaction to federal funds rate target changes, Fed. Res. Bank St. Louis Rev. (1998) 25–36.
[41] G.G. Tian, S. Ma, The relationship between stock returns and the foreign exchange rate: the ARDL approach, J. Asia Pac. Econ. 15 (2010) 490–508.
[42] Y. Yang, The Relationship Between Stock and Foreign Exchange Markets: Evidence From Periods of Exchange-Rate-Regime Shifts, Master of Philosophy, Department of Economics and Finance, City University of Hong Kong, 2009.

[43] F. Yao, Y. Ying, L. Zhang, Analysis of stock market information – a new financial engineering approach, in: Proceedings of IEEE International Conference on Information and Automation (ICIA), 2010, pp. 430–435.
[44] S. Yao, D. Luo, M. Stephen, Shanghai Stock Exchange Composite Index and Bank Stock Prices in China: A Causality Analysis, Research Papers, University of Nottingham, No. 08/25, 2008.
[45] X. Zheng, B.M. Chen, Identification of stock market forces in the system adaptation framework, in: Proceedings of the 9th IEEE International Conference on Control and Automation, 2011, pp. 340–347.
[46] X. Zheng, B.M. Chen, Modeling and forecasting of stock markets under a system adaptation framework, J. Syst. Sci. Complex. 25 (2012) 641–674.
[47] Y. Zheng, W.-K. Wong, Mean, Volatility Spillover and Time-Varying Conditional Dependence in Chinese Stock Markets, RMI Working Paper, No. 07/02, National University of Singapore, 2007.