Computational Experiments Successfully Predict the Emergence of Autocorrelations in Ultra-High-Frequency Stock Returns

Jian Zhou 1 · Gao-Feng Gu 2 · Zhi-Qiang Jiang 2 · Xiong Xiong 3 · Wei Chen 4 · Wei Zhang 3 · Wei-Xing Zhou 5

Accepted: 11 August 2016 / Published online: 24 August 2016 © Springer Science+Business Media New York 2016

Abstract Social and economic systems are complex adaptive systems, in which heterogeneous agents interact and evolve in a self-organized manner, and macroscopic laws emerge from microscopic properties. To understand the behaviors of complex systems, computational experiments based on physical and mathematical models provide a useful tool. Here, we perform computational experiments using a phenomenological order-driven model called the modified Mike–Farmer (MMF) to predict the impacts of order flows on the autocorrelations in ultra-high-frequency returns, quantified by Hurst index \( H_r \). Three possible determinants embedded in the MMF model are investigated, including the Hurst index \( H_s \) of order directions, the Hurst index \( H_x \) and the power-law tail index \( \alpha_x \) of the relative prices of placed orders. The computational experiments predict that \( H_r \) is negatively correlated with \( \alpha_x \) and \( H_x \) and positively correlated with \( H_s \). In addition, the values of \( \alpha_x \) and \( H_x \) have negligible impacts on \( H_r \), whereas \( H_s \) exhibits a dominating impact on \( H_r \). The predictions of the MMF
model on the dependence of $H_r$ upon $H_s$ and $H_x$ are verified by the empirical results obtained from the order flow data of 43 Chinese stocks.

**Keywords** Computational experiment · Order-driven model · Market efficiency · Order direction · Long memory

1 Introduction

Social and economic systems consist of interacting heterogenous agents. Macroscopic laws and collective behaviors emerge from the self-organized evolution of these complex systems. Theories in the top–down framework frequently fail in the description of complex socioeconomic phenomena and have very limited predictive power (Farmer and Foley 2009). For instance, the so-called economic men (*Homo Oeconomicus*) in economics are hypothesized to be homogenous, do not interact or interact in a linear way, and pursue to maximizing their profits rationally, like ideal gas in physics. However, agents in real societies are heterogenous and interact in a nonlinear manner (Schweitzer et al. 2009; Lux 2011). A bottom-up phenomenological framework in natural sciences should be adopted in the study of social sciences (Bouchaud 2008; Lux and Westerhoff 2009), which is now known as computational social science (Lazer et al. 2009).

Indeed, computational experiments are important and well-recognized in the fields of econophysics and sociophysics (Farmer and Foley 2009; Sornette 2014). A general framework is as follows. First, one performs phenomenological analysis to uncover important statistical laws at the microscopic level. Second, one constructs a model based on the microscopic laws. Third, one performs numerical simulations to generate macroscopic properties of the complex system. Forth, if the simulation results deviate from the stylized facts of the real system, one needs to go to the first step to find out possible missing ingredients and then improve the model (Gu and Zhou 2009; Li et al. 2014). Once the model is in store, one can perform computational experiments from the scenario-response perspective. In other words, computational experiments have predictive powers by projecting the responses of a system to external stimuli under certain scenarios and can thus provide guidance to policy makers (Farmer and Foley 2009). Alternatively, computational experiments are able to unravel uncovered properties of a system, which are testable using empirical analysis.

In this work, we perform computational experiments using an empirical order-driven model to identify the microscopic determinants of the autocorrelation structure of stock return time series. Stock markets are efficient in the weak form if no significant autocorrelations can be identified in the returns. Overwhelming evidence shows that the weak-form efficiency holds for intraday high-frequency returns and low-frequency (daily, weekly, monthly, et al.) returns in the sense that the Hurst index of returns ($H_r$) is insignificantly different from 0.5. However, it is not clear if $H_r$ differs from 0.5 or not at the transaction level. In addition, the microscopic mechanisms leading to specific $H_r$ values are unclear. We aim at understanding the impacts of order flows on the weak-form efficiency through computational experiments based on an empirical order-driven model—the modified Mike–Farmer (MMF) model. The seminal model
was proposed by Mike and Farmer (2008), mimicking the processes of order placement and cancellation. The Mike–Farmer (MF) model is able to reproduce the main stylized facts of stock returns, such as power-law tails in the return distribution and absence of long memory in return time series. However, the MF model fails to reproduce the phenomenon of volatility clustering (Mike and Farmer 2008). Gu and Zhou (2009) proposed a modified MF (MMF) model, which successfully reproduces all these stylized facts. There are three possible determinants embedded in the MMF model, including the tail heaviness of relative prices of the placed orders characterized by the tail index $\alpha$, the degree of long memory in relative prices quantified by its Hurst index $H_x$, and the strength of long memory in order directions depicted by $H_s$. We investigate the impacts of these variables on the correlation structure of return time series.

Our work is directly related to the Efficient Markets Hypothesis (EMH), which is one of the cornerstones of modern finance (Fama 1970, 1991). There are three major versions of the hypothesis: weak form, semi-strong form, and strong form (Fama 1970, 1991). The weak form efficiency hypothesis suggests that asset prices are unpredictable using historical prices. The study of the weak-form EMH can be traced back to Bachelier (1900), who suggests that speculative prices follow random walks. Early empirical evidence and theoretical analysis support the random walk hypothesis (Cowles 3rd 1933; Working 1934; Kendall 1953; Osborne 1959; Cootner 1964; Samuelson 1965; Mandelbrot 1966). A classical approach to test the random walk hypothesis is to calculate the Hurst index $H$ of return time series (Mandelbrot 1971). A time series is uncorrelated if its Hurst index $H = 0.5$, antipersistent if $H < 0.5$, or persistent if $H > 0.5$ (Mandelbrot and Van Ness 1968). Different methods have been utilized to estimate Hurst indexes of financial return time series and controversial results are reported (Cajueiro and Tabak 2004; Alvarez-Ramirez et al. 2008; Cajueiro and Tabak 2008; Mishra et al. 2011; Jiang et al. 2014). We note that a well-designed statistical method is necessary to draw a conclusion on the presence of long memory in returns. It is more likely that returns are uncorrelated in long term and show inefficiency in certain short time periods (Jiang et al. 2014). However, we are not going to devote to this debate. Rather, we attempt to identify possible microscopic determinants in order flows that impact the correlation structure of financial return time series.

Our work is also related to the microstructure literature. Cont et al. (2014) find that, over short time intervals, price changes are mainly driven by the order flow imbalance, defined as the imbalance between supply and demand at the best bid and ask prices. It has also been shown that market participants who function at this very high frequency level look for this autocorrelation in order flows to design trading strategies (Clark-Joseph 2013; Fishe et al. 2015), which is actually based on herding mechanisms. Although market participants cannot make high-frequency trading due to the $T + 1$ rule, our research still provides some insights on the dynamics of Chinese stocks.
2 Description of Methods and Data Sets

2.1 Model Description

The order-driven model proposed by Mike and Farmer (2008) mimics the processes of order placement and order cancellation. An order is determined by its direction, price and size.

(1) **Order direction** We denote the order direction by “+1” if it is a buy or “−1” if it is a sell. The time series of order direction has long memory characterized by a Hurst index $H_s$ (Lillo and Farmer 2004). We can simulate the sequence of order direction using fractional Brownian motions with Hurst index $H_s$.

(2) **Order price** The logarithmic order price $\pi(t)$ at event time $t$ is determined by the relative price $x(t)$. One defines $x(t) = \pi(t) - \pi(b)(t-1)$ for buy orders and $x(t) = \pi_a(t-1) - \pi(t)$ for sell orders, where $\pi_b(t) = \ln b(t)$ and $\pi_a(t) = \ln a(t)$ with $a(t)$ and $b(t)$ being the best ask and bid prices right before the event time $t$. In the MF model, $x(t)$ follows a student distribution with the freedom degree being $\alpha_x$. In the MMF model, Gu and Zhou (2009) used an additional ingredient that the relative prices are long-term correlated with a Hurst index $H_x$.

(3) **Order size** For simplicity, all orders are set to have the same size (Mike and Farmer 2008; Gu and Zhou 2009; Meng et al. 2012).

At each simulation step, we check if the unexecuted orders on the book are cancelled or not. Following Mike and Farmer (2008), for each order on the limit order book, its conditional cancellation probability $P = A(1 - e^{-Y_i})(imb + B)/n_{tot}$ is calculated, where $A = 1.12$, $B = 0.2$, $Y_i$ is the ratio of the current distance to execution to the initial distance to execution, $imb$ is the imbalance of the order book, and $n_{tot}$ is the total number of orders on the order book. In this way, the processes of order placement and order cancellation can be simulated. Note that an excellent review of the statistical properties of limit order books is given in Gould et al. (2013).

2.2 Simulation Setting

We simulate the MMF model with different combinations of the three parameters, in which the Hurst indexes $H_s$ and $H_x$ range from 0.50 to 0.95 with a step of 0.05 and the tail exponent $\alpha_x$ varies from 1.00 to 1.65 also with a step of 0.05. For each combination $(H_s, H_x, \alpha_x)$, we simulate the MMF model for 100 repeated runs. In each run, we simulate $2 \times 10^5$ steps and record the return time series with a length of near $4 \times 10^4$ after removing the transient period, where the return is calculated as follows (Mike and Farmer 2008):

$$r(t) = \frac{\pi_a(t) + \pi_b(t)}{2} - \frac{\pi_a(t-1) + \pi_b(t-1)}{2},$$

which is the trade-by-trade price fluctuation after each transaction at time $t$ (Lillo et al. 2003; Lim and Coggins 2005; Zhou 2012a, b). In our computational experiments presented in this work, the tick size is set to be 0.01.
Computational Experiments Successfully Predict... 583

Fig. 1 Comparison of the input and output Hurst indexes for the order sign series and the relative price series. a Input and output Hurst indexes for order sign series by taking the signs of the increments of fractional Brownian motions. b Input and output Hurst indexes for the relative prices generated by the order ranking method and the iterative amplitude adjusted Fourier transform (IAAFT) approach.

To determine the Hurst index of each return time series, we adopted the detrended fluctuation analysis (DFA) proposed by Peng et al. (1994) (see also Kantelhardt et al. 2001), which is one of the most efficient methods for the estimation of Hurst index (Shao et al. 2012). However, the scaling behavior is not good for some time series. We thus turned to adopt the centred detrending moving average (DMA) method (Alessio et al. 2002; Carbone et al. 2004a,b; Xu et al. 2005; Carbone 2009; Gu and Zhou 2010; Jiang and Zhou 2011), which gives better scaling behaviors for those cases. The better performance of the centred DMA does not contradict with the numerical results in Shao et al. (2012) because the conclusions in Shao et al. (2012) are based on fractional Gaussian noises. We partition each return time series into segments of length $\ell$ and determine the average DMA fluctuation $F(\ell)$ over all segments. The Hurst index is obtained from $F(\ell) \sim \ell^H$.

To generate orders, we need to synthesize time series of order signs and relative prices with given Hurst indices $H_s$ and $H_x$. For the series of order signs, we generate a fractional Brownian motion with Hurst index $H_s$ and obtain the sign sequence of its increments. A comparison of the output Hurst index $H_x^{\text{out}}$ and the input Hurst index $H_x^{\text{in}}$ in Fig. 1a shows that the synthetic order signs are not reliable when $H_s$ is less than 0.5. For the series of relative prices, we compared the outputs of two widely used methods, the rank ordering method (Bogachev et al. 2007; Zhou 2008) and iterative amplitude adjusted Fourier transform (IAAFT) approach (Schreiber and Schmitz 1996). Figure 1b shows that the IAAFT approach provides much better results, but with deviations for small $H_x$.

2.3 Data Description

We use the order flow data of 32 A-share stocks and 11 B-share stocks traded on the Shenzhen Stock Exchange (SZSE) in 2003. The A-shares are common stocks issued by mainland Chinese companies, subscribed and traded in the Chinese currency Renminbi, listed on mainland Chinese stock exchanges, bought and sold by
Chinese nationals and approved foreign investors. The A-share market was open only to domestic investors in 2003. Note that the Chinese stock market is the largest emerging market in the world and became the second largest stock market after the USA market in 2009. Our sample stocks were part of the 40 constituent stocks included in the Shenzhen Stock Exchange component index in 2003. However, we were not able to retrieve the data for all 40 A-share constituents. The codes of stocks have six digital numbers. The codes are initiated with “0” for A-share stocks and “2” for B-share stocks. A company can be listed in both the A-share market and the B-share market. In this case, the only difference between the two codes is the first digit.

3 Computational Experiments

Figure 2 shows the average fluctuation function $F(\ell)$ with respect to the segment size $\ell$ for five typical return time series. It is evident that these curves exhibit nice power-law relationships in the sense that the average fluctuation functions fluctuate closely around the corresponding fitted straight lines. We perform linear least-squares regressions of $\ln F(\ell)$ against $\ln \ell$ within the scaling range $10 \leq \ell < 4500$. The resulting slopes are the estimates of the Hurst indexes of the return time series. It is found that the slope $H_r$ increases with $H_x$. For each combination of model parameters, the 100 Hurst indexes of the 100 return time series in repeated simulations are computed. We use all these Hurst indexes, denoted by $H_r(\alpha_x, H_x, H_s)$, for further analysis.

We calculate the three Pearson correlation coefficients of the dependent variable $H_r$ and the independent variables $\alpha_x$, $H_s$ and $H_x$. In calculating the correlation coefficient between $H_r$ and an independent variable, we ignore the influence of other two variables. We find that $\rho(H_r, \alpha_x) = -0.049$, $\rho(H_r, H_s) = -0.156$, and $\rho(H_r, H_x) = 0.953$. All these correlation coefficients are significantly different from zero at the significance level of 0.001. It suggests that the Hurst index $H_r$ of the returns is negatively correlated with $\alpha_x$ and $H_x$ and positively correlated with $H_s$. In addition, the value of $H_s$ plays the most influencing role on $H_r$, consistent with the results in Gu and Zhou (2009).

Because $\alpha_x$ seemingly has the weakest influence on $H_r$, we first study the impacts of $H_s$ and $H_x$ by fixing $\alpha_x = 1.3$. The average Hurst indexes of the return time series are presented in Table 1. The small values of the standard variations in parentheses...
Table 1  Estimated Hurst index $H_r(\alpha_x, H_x, H_s)$ of return time series generated from the MMF model for fixed $\alpha = 1.3$

| $H_s$ | $H_x$ |
|-------|-------|
|       | 0.50  | 0.55  | 0.60  | 0.65  | 0.70  | 0.75  | 0.80  | 0.85  | 0.90  | 0.95  |
| 0.50  | 0.46 (1) | 0.45 (1) | 0.45 (1) | 0.44 (1) | 0.44 (1) | 0.43 (1) | 0.43 (2) | 0.42 (2) | 0.40 (2) |
| 0.55  | 0.47 (1) | 0.47 (1) | 0.46 (1) | 0.46 (1) | 0.45 (1) | 0.45 (1) | 0.44 (1) | 0.44 (2) | 0.43 (2) | 0.41 (2) |
| 0.60  | 0.49 (1) | 0.49 (1) | 0.48 (1) | 0.48 (1) | 0.47 (1) | 0.47 (1) | 0.46 (1) | 0.46 (1) | 0.45 (2) | 0.43 (2) |
| 0.65  | 0.52 (1) | 0.52 (1) | 0.51 (1) | 0.51 (1) | 0.50 (1) | 0.50 (1) | 0.49 (1) | 0.48 (1) | 0.47 (2) | 0.46 (2) |
| 0.70  | 0.55 (1) | 0.55 (1) | 0.55 (1) | 0.54 (1) | 0.54 (1) | 0.53 (2) | 0.52 (1) | 0.52 (1) | 0.51 (2) | 0.49 (2) |
| 0.75  | 0.58 (1) | 0.58 (1) | 0.58 (1) | 0.58 (2) | 0.57 (1) | 0.57 (1) | 0.56 (1) | 0.55 (1) | 0.54 (2) | 0.53 (2) |
| 0.80  | 0.61 (1) | 0.61 (1) | 0.61 (1) | 0.61 (1) | 0.60 (1) | 0.60 (1) | 0.60 (1) | 0.59 (1) | 0.58 (2) | 0.57 (2) |
| 0.85  | 0.63 (1) | 0.63 (1) | 0.63 (1) | 0.63 (1) | 0.63 (1) | 0.62 (1) | 0.62 (1) | 0.61 (2) | 0.61 (2) | 0.60 (2) |
| 0.90  | 0.64 (2) | 0.64 (1) | 0.64 (1) | 0.64 (1) | 0.64 (1) | 0.64 (1) | 0.64 (1) | 0.64 (2) | 0.64 (2) | 0.62 (2) |
| 0.95  | 0.64 (1) | 0.63 (1) | 0.64 (1) | 0.64 (2) | 0.64 (2) | 0.65 (1) | 0.65 (1) | 0.64 (1) | 0.65 (1) | 0.64 (2) |

The numbers in parentheses are the standard deviations multiplied by 100
show the robustness of the MMF model in generating return series. To have a better visibility, we plot the dependence of $H_r$ on $H_x$ for fixed $H_s$ in Fig. 3a and on $H_s$ for fixed $H_x$ in Fig. 3b. We find that $H_r$ decreases with increasing $H_x$ and increases with increasing $H_s$. According to Fig. 3a, $H_r$ exhibits nice linear dependence with respect to $H_x$. The dependence between $H_r$ and $H_s$ in Fig. 3b can also be roughly treated as being linear. We thus propose a linear dependence equation and obtain its coefficients as follows

$$H_r = 0.23 - 0.08H_x + 0.52H_s,$$

where all the coefficients are significantly different from 0 and the adjusted $R^2$ is 0.935. A closer scrutiny of Fig. 3b shows that the curves are nonlinear when $H_s$ is close to 0.5 and 0.95. We find that three-order polynomials can fit the data better than a linear function as depicted in Fig. 3b. This leads to the following equation:

$$H_r = 1.75 - 0.08H_x - 6.23H_s + 9.72H_s^2 - 4.55H_s^3,$$

where all the coefficients are significantly different from 0 and the adjusted $R^2$ is 0.953.

When the tail exponent $\alpha_x$ of the relative prices is taken into consideration, we propose the following three-variable linear relationship

$$H_r = 0.25 - 0.02\alpha_x - 0.08H_x + 0.52H_s,$$

where all the coefficients are significantly different from 0 and the adjusted $R^2$ is 0.935. It is interesting to notice that this equation is very similar to Eq. (2) and the coefficients for $H_x$ and $H_s$ are exactly the same in both equations. The only difference is that the intercept 0.23 in Eq. (2) is replaced by $0.25 - 0.02\alpha_x$ in Eq. (4). The positive and negative dependence between $H_r$ and the independent variables shown in Eq. (4) is consistent with the results from the Pearson correlation coefficients. When $\alpha_x = 1.3$,
$H_x = 0.80$ and $H_s = 0.75$, which are typical values for real stocks (Mike and Farmer 2008; Gu and Zhou 2009), we have $H_r = 0.55$, indicating that there is no long-term memory in the return time series.

We now investigate the robustness of the $H_r$ value with respect to the disturbance in $\alpha_x$, $H_x$ and $H_s$. According to Eq. (4), we have

$$\Delta H_r = -0.02 \Delta \alpha_x,$$

(5)

when $H_x$ and $H_s$ are fixed. It means that, if $\alpha_x$ increases by 1, $H_r$ will decrease only by 0.02. Because the variation range of $\alpha_x$ is from 1.0 to 1.65 for stocks traded on the London Stock Exchange (Mike and Farmer 2008), the variation of $H_r$ caused by $\alpha_x$ is less than 0.01. Hence, we can conclude that the tail heaviness characterized by $\alpha_x$ has a negligible impact on the Hurst index of the output return time series.

When $\alpha_x$ and $H_s$ are fixed, we have

$$\Delta H_r = -0.08 \Delta H_x,$$

(6)

We note that $H_x$ is close to 0.8 for different stocks in developed market (Zovko and Farmer 2002) and emerging market (Gu and Zhou 2009). A value of $\Delta H_x = 0.25$ will result in $\Delta H_r = -0.02$. This explains why both the MF model and the MMF model reach the same conclusion that the generated returns are not long-term correlated (Mike and Farmer 2008; Gu and Zhou 2009). It means that the strength of autocorrelations in the relative price characterized by $H_s$ also has a minor impact on the Hurst index of the return time series.

When $\alpha_x$ and $H_s$ are fixed, we have

$$\Delta H_r = 0.52 \Delta H_s.$$

(7)

It is clear that the strength of long-term correlations $H_s$ in order direction has a dominating impact on $H_r$. For instance, if $H_s$ increases (or decreases) by 0.2, $H_r$ will increase (or decrease) by 0.1. Consider the real case that $\alpha_x = 1.3$ and $H_x = 0.8$. If there is no memory in order direction such that $H_s$ is close to 0.5, we have $H_r = 0.42$ and the return time series is thus antipersistent. If the correlation in order direction is extremely strong such that $H_s \to 1$, we have $H_r = 0.68$ showing that the returns are persistent.

If we take the nonlinear behavior of $H_r$ against $H_s$ into consideration, we can obtain the following equation:

$$H_r = 1.75 - 0.02\alpha_x - 0.08H_x - 6.11H_s + 9.53H_s^2 - 4.45H_s^3,$$

(8)

where all the coefficients differ significantly from 0 and the adjusted $R^2$ is 0.952. This equation fits the data with a slightly smaller error than Eq. (4). We find that the coefficients for $\alpha_x$ and $H_x$ are the same as in Eq. (4). If $\alpha_x = 1.3$, $H_x = 0.8$ and $H_s$ varies from 0.5 to 1.0, $H_r$ ranges from 0.43 to 0.63.
4 Empirical Validation

We perform empirical analysis on each trading day using the order flow data of 32 A-share stocks and 11 B-share stocks traded on the Shenzhen Stock Exchange in 2003. The trading days with less than 500 transactions are discarded, because the return series is too short. We adopt the centred detrending moving average (DMA) approach to determine the Hurst indexes $H_r$, $H_s$, and $H_x$ of the time series of trade-by-trade returns, order signs, and relative prices of submitted orders for each trading day of each stock. A trading day is kept for further analysis if the estimation errors of the Hurst indexes $H_r$, $H_s$, and $H_x$ are all less than 0.02. If the number of trading days left for further analysis is less than 10, we exclude the stock from further analysis. As shown in Table 2, three B-share stocks are discarded. For a same company, the number $N_{day}$ of trading days is larger for its A-share stock than its B-share stock (say, 000002 versus 200002). This is due to the fact that A-share stocks are usually much more actively traded than B-share stocks. The correlation coefficients $\rho(H_r, H_s)$ and $\rho(H_r, H_x)$ and their corresponding $p$-values are presented in Table 2.

We first consider the correlation coefficients $\rho(H_r, H_s)$ between $H_r$ and $H_s$. For the A-shares, there are 30 stocks having positive correlation coefficients $\rho(H_r, H_s)$ and 2 stocks (000088 and 000541) with negative correlation coefficients. The two negative correlation coefficients are not significant even at the level of 10%. The positive correlation between $H_r$ and $H_s$ of stock 000488 is also not significant. The positive correlation is significant at the level of 2% for the remaining 29 stocks. For the B-shares, all the correlation coefficients are positive and 6 of them are significant at the 5% level. If we aggregate all the data for the A-share stocks, we obtain an average correlation coefficient $\rho_A(H_r, H_s) = 0.326$ with the $p$-value less than 0.1%, as also illustrated in Fig. 4a. Similarly, we obtain the average correlation coefficient $\rho_B(H_r, H_s) = 0.397$ for the B-share stocks, as illustrated in Fig. 4b. These observations indicate that $H_r$ is positively correlated with $H_s$, which verifies nicely the prediction of the computational experiments where $\rho(H_r, H_s) = 0.953$.

We now consider the correlation coefficients $\rho(H_r, H_x)$ between $H_r$ and $H_x$. For the A-shares, there are 14 stocks having positive correlation coefficients and 18 stocks having negative correlation coefficients. There are only 5 correlation coefficients are significant at the 2% level (4 negative and 1 positive), while the $p$-values of all other coefficients are greater than 10%. When we aggregate all trading days, we obtain an average correlation coefficient $\rho_A(H_r, H_x) = -0.094$ with the $p$-value less than 0.1%, as illustrated in Fig. 4c. For the B-shares, there are 3 stocks with positive correlation coefficients and 5 stocks with negative correlation coefficients. All these correlations are not significant at the 5% level. The average correlation coefficient is $\rho_B(H_r, H_x) = -0.050$, which is also not significant. The scatter plot is illustrated in Fig. 4d. Overall, we find that there is a weak negative correlation between $H_r$ and $H_x$, which supports the prediction of the computational experiments where $\rho(H_r, H_x) = -0.156$.

We further perform linear regressions for the 32 A-share stocks and 8 B-share stocks using the following equation

$$H_r = \beta_0 + \beta_x H_x + \beta_s H_s.$$  \hfill (9)
Table 2 Empirical analysis of the cross-correlations between the Hurst index $H_r$ of the trade-by-trade returns and the Hurst indexes $H_s$ and $H_x$ of the order signs and the relative prices

| Stock   | $N_{\text{day}}$ | $\rho(H_r, H_s)$ | $p$-value | $\rho(H_r, H_x)$ | $p$-value |
|---------|------------------|------------------|-----------|------------------|-----------|
| **Panel A: A-share stocks** | | | | | |
| 000001  | 234              | 0.600            | 0.000     | 0.170            | 0.009     |
| 000002  | 221              | 0.391            | 0.000     | −0.268           | 0.000     |
| 000009  | 218              | 0.643            | 0.000     | −0.012           | 0.860     |
| 000012  | 144              | 0.440            | 0.000     | −0.209           | 0.012     |
| 000016  | 146              | 0.487            | 0.000     | −0.113           | 0.174     |
| 000021  | 216              | 0.390            | 0.000     | −0.017           | 0.806     |
| 000024  | 99               | 0.423            | 0.000     | 0.130            | 0.200     |
| 000027  | 164              | 0.476            | 0.000     | 0.026            | 0.739     |
| 000063  | 167              | 0.422            | 0.000     | −0.108           | 0.166     |
| 000066  | 175              | 0.356            | 0.000     | −0.048           | 0.525     |
| 000088  | 56               | −0.028           | 0.839     | −0.207           | 0.125     |
| 000089  | 131              | 0.320            | 0.000     | −0.021           | 0.811     |
| 000406  | 180              | 0.464            | 0.000     | 0.096            | 0.200     |
| 000429  | 59               | 0.533            | 0.000     | −0.105           | 0.427     |
| 000488  | 87               | 0.153            | 0.157     | 0.084            | 0.439     |
| 000539  | 66               | 0.313            | 0.011     | 0.065            | 0.607     |
| 000541  | 24               | −0.197           | 0.356     | −0.489           | 0.015     |
| 000550  | 176              | 0.363            | 0.000     | 0.052            | 0.493     |
| 000581  | 60               | 0.363            | 0.004     | 0.212            | 0.105     |
| 000625  | 192              | 0.257            | 0.000     | −0.048           | 0.513     |
| 000709  | 137              | 0.413            | 0.000     | 0.078            | 0.364     |
| 000720  | 96               | 0.248            | 0.015     | 0.076            | 0.462     |
| 000778  | 106              | 0.320            | 0.001     | −0.067           | 0.497     |
| 000800  | 231              | 0.332            | 0.000     | −0.256           | 0.000     |
| 000825  | 193              | 0.527            | 0.000     | −0.031           | 0.665     |
| 000839  | 237              | 0.277            | 0.000     | −0.062           | 0.338     |
| 000858  | 192              | 0.455            | 0.000     | −0.036           | 0.625     |
| 000898  | 205              | 0.477            | 0.000     | 0.066            | 0.346     |
| 000917  | 119              | 0.404            | 0.000     | 0.032            | 0.733     |
| 000932  | 188              | 0.504            | 0.000     | 0.069            | 0.348     |
| 000956  | 227              | 0.373            | 0.000     | −0.009           | 0.895     |
| 000983  | 130              | 0.310            | 0.000     | 0.020            | 0.818     |
| **Aggregate** | **4876** | **0.326** | **0.000** | **−0.094** | **0.000** |
| **Panel B: B-share stocks** | | | | | |
| 200002  | 18               | 0.502            | 0.034     | −0.326           | 0.187     |
| 200012  | 20               | 0.184            | 0.439     | −0.058           | 0.807     |
| 200024  | 12               | 0.555            | 0.061     | −0.016           | 0.961     |
Panel A and Panel B show respectively the results of A-share stocks and B-share stocks with the codes given in the first column. The second column presents the number of trading days $N_{\text{day}}$ included in the analysis for each stock.

For the A-share stocks, we obtain that $\beta_0 = 0.21$, $\beta_x = -0.33$ and $\beta_s = 0.66$. The adjusted $R^2$ is 0.156 and all the coefficients are significant at the 1% level. For the B-shares, we obtain that $\beta_0 = 0.13$, $\beta_x = -0.10$ and $\beta_s = 0.57$ with the $p$-values.
being 0.052, 0.057 and 0.000, respectively. The adjusted $R^2$ is 0.140. These results are again in good agreement with Eq. 2 of the computational experiments. Note that we do not investigate the impact of the relative price distribution since there is no evidence showing that the relative prices are distributed as a Student function for the real LOB data (Gu and Zhou 2009).

5 Discussion

In this work, we have performed computational experiments to investigate the impacts of order flow characteristics on the correlation structure in return time series. The statistical properties of the relative prices of placed orders are found to have statistically significant but quantitatively negligible impacts on the Hurst index of returns. A fatter distribution of relative prices means more order are market orders with the prices larger than the opposite best prices. Since the order size is 1 in the MMF model, an order will not have persistent impacts on price movements. Hence, the parameter $\alpha_x$ only have negligible impacts on the persistence of returns quantified by $H_r$.

In contrast, the long memory effect in order direction plays a dominating role on the Hurst index of returns. The remarkable impact of the long memory feature of order direction can be explained. The Hurst index of the time series of order direction quantifies the degree of herding in order placement. Stronger herding incurs stronger persistence in the changes of prices. Moreover, the persistence of price changes is more evident because the order size is fixed to 1 and each order will be very likely to move the mid-price. These predictions from the computational experiments are in excellent agreement with empirical analysis on real order flow data.

Both computational experiments and empirical analysis show that the Hurst index of trade-by-trade return time series can deviate significantly from $H = 0.5$. The intraday returns can be either antipersistent, persistent, or uncorrelated. This finding shows the evolutionary feature of traders’ heterogenous behaviors. The deviation of $H_r$ from 0.5 does not necessarily imply that there are arbitrage opportunities at the micro scales and that the market is inefficient. Market friction such as transaction costs may eliminate the possibility to gain positive profits from such a predictability. In addition, the $T + 1$ trading rule makes it more unlikely to make profits based on this finding of autocorrelations in the returns. The $T + 1$ trading rule also provides a possible interpretation of the fact that the deviation is more severe in real stocks than in artificial stocks. The deviation of $H_r$ can also be understood in the sense that it takes time for the market to digest information.

Our findings imply that the correlation structure of return time series is not necessary to be stable over time. In particular, at small timescales, the Hurst index of returns can fluctuate following the fluctuations in the herding strength of order directions. We can conjecture that the return time series may exhibit long memory if the directions of placed orders are strongly correlated and become antipersistent if order directions are less correlated. These findings provide latent supportive evidence to the Adaptive Markets Hypothesis proposed by Lo (2004).

We would like to note that the MMF model is very useful in the cost benefit analysis of market policies, as already done in Hayes et al. (2012). For instance, there is
a price limit rule in the Chinese stock market. The daily price limits are ±10% for common stocks and ±5% for specially treated stocks. A further modified MMF model shows that the price dynamics will be very different for asymmetric price limit setting. However, a detailed analysis of the topic is beyond the scope of this paper, which will be reported in a future work.

Acknowledgements Zhi-Qiang Jiang, Gao-Feng Gu and Wei-Xing Zhou received support from the National Natural Science Foundation of China (71501072) and the Fundamental Research Funds for the Central Universities. Xiong Xiong and Wei Zhang received support from the National Natural Science Foundation of China (71532009, 71131007) and the Program for Changjiang Scholars and Innovative Research Team in University (IRT1028). Wei Chen received support from the National Natural Science Foundation of China (71571121).

References
Alessio, E., Carbone, A., Castelli, G., & Frappietro, V. (2002). Second-order moving average and scaling of stochastic time series. European Physical Journal B, 27(2), 197–200. doi:10.1140/epjb/e20020150.
Alvarez-Ramirez, J., Alvarez, J., Rodriguez, E., & Fernandez-Anaya, G. (2008). Time-varying Hurst exponent for US stock markets. Physica A, 387(24), 6159–6169. doi:10.1016/j.physa.2008.06.056.
Bachelier, L. (1900). Théorie de la Spéculation. Ph.D. thesis, University of Paris, Paris.
Bogachev, M. I., Eichner, J. F., & Bunde, A. (2007). Effect of nonlinear correlations on the statistics of return intervals in multifractal data sets. Physical Review Letters, 99, 240601. doi:10.1103/PhysRevLett.99.240601.
Bouchaud, J. P. (2008). Economics needs a scientific revolution. Nature, 455, 1181. doi:10.1038/4551181a.
Cajueiro, D. O., & Tabak, B. M. (2004). The Hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient. Physica A, 336(3–4), 521–537. doi:10.1016/j.physa.2003.12.031.
Cajueiro, D. O., & Tabak, B. M. (2008). Testing for long-range dependence in world stock markets. Chaos, Solitons & Fractals, 37(3), 918–927. doi:10.1016/j.chaos.2006.09.090.
Carbone, A. (2009). Detrending moving average algorithm: A brief review. Science and Technology for Humanity (TIC-STH) IEEE pp 691–696. doi:10.1109/TIC-STH.2009.5444412
Carbone, A., Castelli, G., & Stanley, H. E. (2004a). Analysis of clusters formed by the moving average of a long-range correlated time series. Physical Review E, 69, 026105. doi:10.1103/PhysRevE.69.026105.
Carbone, A., Castelli, G., & Stanley, H. E. (2004). Time-dependent Hurst exponent in financial time series. Physica A, 344, 267–271. doi:10.1016/j.physa.2004.06.130.
Clark-Joseph, A. D. (2013). Exploratory trading, job market paper
Cont, R., Kukanov, A., & Stoikov, S. (2014). The price impact of order book events. Journal of Financial Econometrics, 12(1), 47–88. doi:10.1093/jjfinec/nbt003.
Cootner, P. H. (1964). The random character of stock market prices. Cambridge: MIT Press.
Cowles, A. 3rd. (1933). Can stock market forecasters forecast? Econometrica, 1, 309–324.
Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. Journal of Finance, 25(2), 383–417. doi:10.1111/j.1540-6261.1970.tb00518.x.
Fama, E. F. (1991). Efficient capital markets: II. Journal of Finance, 46(5), 1575–1617. doi:10.1111/j.1540-6261.1991.tb04636.x.
Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. Nature, 460, 685–686. doi:10.1038/460685a.
Fishe, R. P. H., Haynes, R., & Onur, E. (2015). Anticipatory traders and trading speed, http://ssrn.com/abstract=2606949
Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., & Howison, S. D. (2013). Limit order books. Quantitative Finance, 13(11), 1709–1742. doi:10.1080/14697688.2013.803148.
Gu, G. F., & Zhou, W. X. (2009). Emergence of long memory in stock volatilities from a modified Mike-Farmer model. EPL (Europhysics Letters), 86(4), 48002. doi:10.1209/0295-5075/86/48002.
Gu, G. F., & Zhou, W. X. (2010). Detrending moving average algorithm for multifractals. Physical Review E, 82, 011136. doi:10.1103/PhysRevE.82.011136.
Hayes, R., Paddrik, M., Todd, A., Yang, S., Beling, P., & Scherer, W. (2012). Agent based model of the e-mini future: Application for policy making. In C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, & A. M. Uhrmacher (Eds.), Proceedings of the 2012 Winter Simulation Conference. pp. (1–12) Berlin: IEEE. doi:10.1109/WSC.2012.6465037

Jiang, Z. Q., & Zhou, W. X. (2011). Multifractal detrending moving-average cross-correlation analysis. *Physical Review E*, 84, 016106. doi:10.1103/PhysRevE.84.016106.

Jiang, Z. Q., Xie, W. J., & Zhou, W. X. (2014). Testing the weak-form efficiency of the WTI crude oil futures market. *Physica A*, 405, 235–244. doi:10.1016/j.physa.2014.02.042.

Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. A. H., Havlin, S., & Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A*, 295(3–4), 441–454. doi:10.1016/S0378-4371(01)00144-3.

Kendall, M. G. (1953). The analysis of economic time-series: Part I: Prices. *Journal of the Royal Statistical Society A*, 116(1), 11–34. doi:10.2307/2980947.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., et al. (2009). Social science: Computational social science. *Science*, 323(5915), 721–723. doi:10.1126/science.1167742.

Li, Y. L., Zhang, W., Zhang, Y. J., Zhang, X. T., & Xiong, X. (2014). Calibration of the agent-based continuous double auction stock market by scaling analysis. *Information Science*, 256, 46–56. doi:10.1016/j.ins.2012.06.019.

Lillo, F., & Farmer, J. D. (2004). The long memory of the efficient market. *Studies in Nonlinear Dynamics and Econometrics*, 8(3), 1–33. doi:10.2202/1558-3708.1226.

Lillo, F., Farmer, J. D., & Mantegna, R. (2003). Master curve for price impact function. *Nature*, 421, 129–130. doi:10.1038/421129a.

Lo, A. W. (2004). The adaptive markets hypothesis. *Journal of Portfolio Management*, 30(5), 15–29. doi:10.3905/jpm.2004.442611.

Lux, T. (2011). Network theory is sorely required. *Nature*, 469, 303. doi:10.1038/469302a.

Mishra, R. K., Sehgal, S., & Bhanumurthy, N. R. (2011). A search for long-range dependence and chaotic structure in Indian stock market. *Chaos, Solitons & Fractals*, 20(2), 96–104. doi:10.1016/j.rfe.2011.04.002.

Osborne, M. F. M. (1959). Brownian motion in the stock market. *Operations Research*, 7(2), 145–173. doi:10.1287/opre.7.2.145.

Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49(2), 1685–1689. doi:10.1103/PhysRevE.49.1685.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6(2), 41–49. doi:10.1177/0019786814566926.0002.

Schreiber, T., & Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical Review Letters*, 77(4), 635–638. doi:10.1103/PhysRevLett.77.635.

Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., & White, D. R. (2009). Economic networks: The new challenges. *Science*, 325, 422–425. doi:10.1126/science.1173644.

Shao, Y. H., Gu, G. F., Jiang, Z. Q., Zhou, W. X., & Sornette, D. (2012). Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series. *Scientific Reports*, 2, 835. doi:10.1038/srep00835.
Sornette, D. (2014). Physics and financial economics (1776–2014): Puzzles, Ising and agent-based models. Reports on Progress in Physics, 77(6), 062001. doi: 10.1088/0034-4885/77/6/062001.

Working, H. (1934). A random-difference series for use in the analysis of time series. Journal of the American Statistical Association, 29(185), 11–24. doi: 10.2307/2278456.

Xu, L. M., Ivanov, P. C., Hu, K., Chen, Z., Carbone, A., & Stanley, H. E. (2005). Quantifying signals with power-law correlations: A comparative study of detrended fluctuation analysis and detrended moving average techniques. Physical Review E, 71, 051101. doi: 10.1103/PhysRevE.71.051101.

Zhou, W. X. (2008). Multifractal detrended cross-correlation analysis for two nonstationary signals. Physical Review E, 77, 066211. doi: 10.1103/PhysRevE.77.066211.

Zhou, W. X. (2012a). Determinants of immediate price impacts at the trade level in an emerging order-driven market. New Journal of Physics, 14(9), 023055. doi: 10.1088/1367-2630/14/2/023055.

Zhou, W. X. (2012b). Universal price impact functions of individual trades in an order-driven market. Quantitative Finance, 12(8), 1253–1263. doi: 10.1080/14697688.2010.504733.

Zovko, I., & Farmer, J. D. (2002). The power of patience: A behavioural regularity in limit-order placement. Quantitative Finance, 2(5), 387–392. doi: 10.1088/1469-7688/2/5/308.