DE GRUYTER Stud. Nonlinear Dyn. E. 2021; 25(3): 19–34

Can Jia, Tianmin Zhou and Handong Li*

Statistical characteristics of price impact in high-frequency trading

https://doi.org/10.1515/snde-2018-0067
Received July 8, 2018; accepted April 18, 2020; published online June 1, 2020

Abstract: Trading volume changes based on market microstructure will impact asset prices, which will lead to transaction price changes. Based on the extended Hasbrouck–Foster–Viswanathan (HFV) model, we study the statistical characteristics of daily permanent price impact and daily temporary price impact using high-frequency data from Chinese Stock Markets. We estimate this model using tick-by-tick data for 16 selected stocks that are traded on the Shanghai Stock Exchange. We find the following: (1) the time series of both the permanent price impact and temporary price impact exist in stationarity and long-term memory; (2) there is a strong correlation between the permanent price impact among assets, while the correlation coefficient of the temporary price impact is generally weak; (3) the time interval has no significant influence on the trade volume and the price change at the tick frequency, which means that it is not necessary to take into account the time interval between adjacent transaction in high-frequency trading; and (4) the bid-ask spread is an effective factor to explain trading price change, but has no significant impact on trade volume.

Keywords: HFV model; permanent price impact; price change; temporary price impact; trade volume.

1 Introduction

In financial markets, the large volume trading behavior of institutional investors will impact the transaction price of assets, resulting in price changes and transaction costs increasing. The price impact in high-frequency data reflects the price generation mechanism of the financial markets, which is the core issue of concern of the algorithmic trading in financial markets.

Price impact refers to the correlation between an incoming order (to buy or to sell) and the subsequent price change. There are two main distinct possibilities to explain why there has price impact. 1) The impact of trades reveals some private information. The arrival of new private information causes trades, which causes other agents to update their valuations, thus leading to a price change (Hopman 2007). 2) The impact is caused by illiquidity. Amihud (2002) mentions that the illiquidity measurement can serve as a rough measure of the price impacts. Chordia, Huh, and Subrahmanyam (2009) estimated illiquidity using structural formulas in line with Kyle (1985). Lespagnol and Rouchier (2018) finds that considering the liquidity as an endogenous characteristic of the market and allowing the designation of investors as bounded rational, the features such as the trend-follower expectation and the heterogeneous investment horizon are important to generate the excess volatility of asset prices.

The existing theoretical and empirical literature provides justification for typically choosing between trade size, signed trade size, and a polynomial trade variable. Theoretical models suggest that the permanent price impact is linearly related to the trade size, such as in Kyle (1985), and Huberman and Stanzl (2004), which supports using the signed trade size as the trade variable. In addition, Glosten and Harris (1988) presents an asymmetric information model in which the bid-ask spread is broken into a transitory component and an adverse-selection component. In contrast, the empirical literature shows that trade size contributes little incremental explanatory power to the trade direction when estimating the permanent price impact, such as in

***Corresponding author: Handong Li**, School of Systems Science, Beijing Normal University, No.19, Xinjiekouwai St., Haidian District, Beijing, 100875, China, E-mail: lhd@bnu.edu.cn. https://orcid.org/0000-0003-3613-7327
Can Jia and Tianmin Zhou: School of Systems Science, Beijing Normal University, No.19, Xinjiekouwai St., Haidian District, Beijing, 100875, China, E-mail: jcan@mail.bnu.edu.cn (C. Jia), lsdzdxt_28@hotmail.com (T. Zhou)

Jones, Kaul, and Lipson (1994) and Hasbrouck (2007), which supports using the trade sign as the trade variable. Bouchaud, Farmar, and Lillo (2009) find that the autocorrelation of the trade signs decays extremely slowly with time on a large variety of markets. Hasbrouck (1991) argues that the assumed linearity between the trade variable and quote updates may be a tenuous approximation and could lead to a misspecified model. To overcome this misspecification, Hasbrouck (1991) advocates to use a polynomial terms as the trade variable to capture any nonlinear relationships between quote updates and trade size. Despite these findings, polynomial terms are rarely used in practice. Brennan and Subrahmanyam (1996) combined the Hasbrouck (1991) model with the Foster and Viswanathan (1993) model, which is called the Hasbrouck–Foster–Viswanathan model (HFV), when studying the illiquidity compensation of stock returns. The model explains the price adjustment due to the unexpected transaction volume.

Furthermore, several extensions of the model have been proposed. Huberman and Stanzl (2004) give a rational explanation of the linear price impact model under the no-arbitrage framework. Almgren and Chriss (1999, 2000), Almgren (2003), Lillo, Farmer, and Mantegna (2003), and Huberman and Stanzl (2004) extend the simple linear model to the nonlinear model. Of course, this extension makes it difficult to obtain a numerical solution. Farmer and Zamani (2007) introduce a means of decomposing the total impact of trading into two components, defining the mechanical impact of a trading order as the change in the future prices in the absence of any future changes in decision making, and the informational impact as the remainder of the total impact once mechanical impact is removed, and then give the properties of the two components. Dufour and Engle (2000), and Engle and Sun (2007) introduces the effect of the duration between trades of a given stock as an important driver of the volatility of fundamental innovations. Jondeau, Lahaye, and Rockinger (2015) further discuss the impact of jump on the total variation of stock returns. Philip (2020) augments vector auto-regression estimation methods to model the nonlinear relationship between the permanent price impact and trade size and proposes an alternative technique, a flexible reinforcement learning (RL) algorithm, when capturing today's trading environment including the additional variables. Considering the transaction's timeliness and the random characteristic of the transaction price, Almgren (2003), Polimenis (2005), and Ting, Warachka, and Zhao 2007 also try to use the random function to model the effect of the transaction on the price. Almgren et al. (2005) use the transaction data of Citibank in the United States to conduct an empirical analysis on the price influence equation, and the empirical results support the hypothesis of a linear form.

The studies are mainly based on the microstructure model and using the model to measure the price impact. However, it lacks analysis of the distribution characteristics of the permanent price impact and the temporary price impact. Based on the HFV model, this paper first extends the model and then analyzes the daily permanent price impact and daily temporary price impact based on high-frequency data traded on the Shanghai Stock Exchange (SSE). In this paper, the daily permanent price impact and daily temporary price impact can well reflect the statistical characteristics and dynamics in the stock market. So we choose the HFV model instead of a more specific microstructure with lots of explanatory variables.

The remainder of the paper is organized as follows. Section 2 presents the extended HFV model. Moreover, this part explains the De-trended Fluctuation Analysis method, which is useful for analyzing time series that appear to be long-memory processes. Section 3 reports our results on the estimation of the 16 selected stocks traded on the Shanghai stock market, including the stability, distribution. and long-memory of permanent price impact and temporary price impact. Then, we discuss the effects of time intervals between transactions and bid-ask spread on the price impact. Section 4 concludes the paper.

2 Model and method

2.1 Extended HFV model

In contrast to Hasbrouck (1991) and following Foster and Viswanathan (1993), Brennan and Subrahmanyam (1996) propose the HFV model to describe the effect of unexpected volume on price changes. They apply the model to transaction prices rather than to bid-ask quotes. The HFV model, which used transaction prices rather than bid-ask quotes, can be summarized by the following two equations.

$$q_t = \alpha_q + \sum_{j=1} \beta_j \, \Delta p_{t-j} + \sum_{j=1} \gamma_j \, q_{t-j} + \tau_t \tag{1}$$

$$\Delta p_t = \alpha_p + \psi[D_t - D_{t-1}] + \lambda \tau_t + \nu_t \tag{2}$$

where q_t is the order flow, and p_t is the transaction price. In the signed trade volume dynamics (Eq. (1)), $\Delta p_t = p_t - p_{t-1}$ denotes the transaction price change for transaction time t, and β_j measures the effect of the trading price change on the trade volume. The parameter γ_j measures the effect of the trade volume auto-regression, while τ_t measures the trade informativeness. Thus, if trades are auto-correlated or predictable from past price changes, then part of the contemporaneous order flow is predictable and should not be concluded in measuring the information content of a trade. In the trading price change dynamics (Eq. (2)), q_t denotes the signed trade quantity corresponding to the price change, and D_t denotes the indicator corresponding to the direction of a trade. ψ measures the permanent effect of the trade direction, and λ measures the temporary effect of the unexpected trade volume. ν_t denotes the error term.

In equity market, to reduce impact investors tend to split their orders into smaller pieces and execute them gradually. And for some reasons, the different investors will have the similar behavior. Both of these conditions will led to a remarkable market phenomenon that orders to buy tend to be followed by more orders to buy and orders to sell tend to be followed by more orders to sell (Toth et al. 2015). In our model, D_t is the corresponding variable to explain this situation, which means D_t has the serial correlation naturally. As illustrated in Figure 1, the scale index value defined in part 2.2 of PingAn stock all over 1, which means D_t series have a persistent long-range correlation. Corresponding, the coefficient ψ that measures the permanent effect of the trade direction will also show the characteristic of long-memory. While order flow q_t is contain the information of trade direction, and is persistent (Toth et al. 2015). Thus, the coefficient λ will also be affected by q_t.

The model focuses on the price response to unexpected volume as the measure of the adverse selection component of the price change. The rationale is that if trades are auto-correlated or predictable from past price changes, then part of the contemporaneous order flow is predictable and should not be included in measuring the information content of a trade. Since the model is formulated in tick time, observations are randomly spaced through time. We need to consider the effects of time interval between trades on the trade volume and trading price changes. Therefore, we introduce δT_t, the time interval between trade $t-1$ and t, as an explanatory variable to the basic HFV model. Here, we form the extended HFV model.

Assume that the time interval affects the signed transaction volume. That is, the time interval δT_t from the trade $t-1$ to the trade t will affect the transaction volume and the price change on trade t. The existing research results show that the bid-ask spread is a typical index to measure stock liquidity. The smaller the bid-ask spread, the less price impact there is for large orders, and thus the more liquidity is ready to be supplied in the

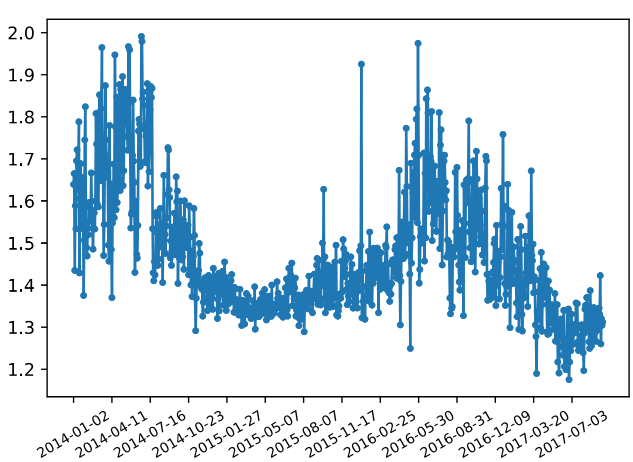

Figure 1: Long memory test results of trading direction sequence D_t of PingAn stock.

market. Then, we add the time interval effects and bid-ask spread into the basic HFV model and get the following model for estimation:

$$q_t = \alpha'_q + \sum_{j=1} \beta'_j \, \Delta p_{t-j} + \sum_{j=1} \gamma'_j q_{t-j} + \sum_{j=1} \sigma'_j \, \delta T_{t-j} + \sum_{j=1} \xi'_j \, S_{t-j} + \tau'_t \tag{3}$$

$$\Delta p_t = \alpha'_p + \psi' [D_t - D_{t-1}] + \lambda' \tau'_t + \sum_{j=1} \sigma''_j \, \delta T_{t-j} + \sum_{j=1} \xi''_j \, S_{t-j} + \nu'_t \tag{4}$$

where the definitions of the variables q_t, Δp_t, and D_t are the same as the HFV model (see Eqs. (1) and (2)). In the signed trade volume dynamics (Eq. (3)), β'_j measures the effect of a trading price change on the trade volume. The parameter γ'_j measures the effect of the trade volume auto-regression, while σ'_j and ξ'_j separate represents the trade interval and the bid-ask spread impacts on the trade volume. The residual term τ'_t represents the surprise in the signed trade volume. In the trading price change dynamics (Eq. (4)), the parameter ψ' measures the permanent effect of the trade direction. The parameter λ' measures the temporary effect of the unexpected trade volume, while σ''_j and ξ''_j respectively measures the trade interval and the bid-ask spread impacts on the trading price changes. Finally, ν'_t denotes the error term.

2.2 Detrended fluctuation analysis

The Detrended Fluctuation Analysis (DFA) is a method for analyzing time series that appear to be long-memory processes and was proposed by Peng et al. (1994) based on the DNA mechanism. It can detect long-range correlations that appear on the surface as non-stationary time series and can eliminate the pseudo-correlated phenomena in artificial non-stationary time series. Amir Bashan (2008) conducted a comparative analysis of several approaches based on the similarity of the long-range correlation of detection time series and compared the three methods of the DFA, Modified Detrended Fluctuation Analysis (MDFA), and Centered Moving Average (CMA) Methods. CMA is similar to DFA, but when the data are shorter, the result based on CMA is better than DFA, and DFA is slightly better than MDFA. In addition, when the function form of the trend in the data is not known, the DFA method is a good choice.

Take the time series $\{x_k\}$, $k = 1,2,3, \ldots ,N$, with a given length N. The general process of the DFA method is as follows.

Step 1: Construct a new sequence $y(i)$ based on $\{x_k\}$.

$$y(i) = \sum_{k=1}^{i} (x_k - x), \quad i = 1, 2, 3, \ldots, N \tag{5}$$

Step 2: Divide the new sequence into $N_s = [N/s]$ non-overlapping segments of equal length s. Consider the situation that the original sequence length N is not necessarily divisible by the subinterval length s. Therefore, in order to ensure that the original sequence information is not lost, it can be once divided from the end of the sequence. Thus, we get $2N_s$ subintervals of length s.

Step 3: Fit the local trend $y_v^{(k)}(j)$, $v = 1, 2, \ldots, N_s$, on each subinterval v using the least square method, where $y_v^{(k)}(j)$ is a k-order polynomial, $k = 1,2,3, \ldots$, denoted as DFA1, DFA2, DFA3, …. Then, we eliminate the local trend in each subinterval v and obtain the detrending sequence:

$$Z_v(j) = y_v(j) - y_v^{(k)}(j), \quad j = 1, 2, \ldots, s \tag{6}$$

Step 4: Calculate the squared mean of $2N_s$ de-intersection subinterval series.

$$F^2(s, v) = \frac{1}{s} \sum_{j=1}^{s} Z_v^2(j), \quad v = 1, 2, \ldots, 2N_s \tag{7}$$

Then we obtain the square root of the mean of $F^2(s,v)$.

$$F(s) = \sqrt{\frac{1}{2N_s}\sum_{v=1}^{2N_s} F^2(s,v)} \tag{8}$$

Step 5: Analyze the relationship $F(s) \propto s^{\varepsilon}$ between the wave function $F(s)$ and s, where ε is the scale index that reflects the relevant characteristics of the sequence. Beran et al. (1995) find that the scaling exponent ε is related to the correlation exponent ω by $\omega = 1 - \varepsilon/2$. The slope of the change in $\log F(s)$ versus $\log s$ is given as ε.

When $\varepsilon = 0.5$, there is no correlation in the sequence. That is, the sequence is random (white noise), and there is no long memory. When $0 < \varepsilon < 0.5$, the time series has an anti-persistent state, and the anti-persistent time series shows a greater reversal trend. When $0.5 < \varepsilon \leq 1$, the time series has a persistent state, and the continuous time series tends to follow the trend more. When $\varepsilon > 1$, the time series has a persistent long-range correlation. For more explanation, see Bahar et al. (2001), Bartsch et al. (2005), and Santhanam, Bandyopadhyay, and Angom 2006.

3 Empirical analysis

3.1 Data and preliminary processing

In this market, based on the SSE, liquidity is provided by the limit order book since there are no market makers. Limit orders consist of a quantity to buy or sell at that price. Orders are submitted by investors through brokers and stored in the limit order book. The matching of orders and their corresponding trades follows strict price and time priorities. Market orders (i. e., orders to buy or sell at the best price available in the book) are immediately executed and matched with the best orders on the opposite side of the book. The exchange opens at 9:30 am and closes at 3:00 pm, including a lunch break from 11:30 am to 13:00 pm.

The data consist of all intraday tick-by-tick transaction prices and quantities spanning from January 2nd, 2014, to September 25th, 2017. Since the constituents of the SSE's 50 index stocks will be changed every half year, we choose 16 financial stocks that remain as the constituent stocks from 2014 to 2017 that are traded on the SSE. They are PingAn Insurance (Group) Co. (PingAn), China Pacific Insurance (Group) Co. (CPIC), China Life Insurance (Group) Co. (ChinaLife), Shanghai Pudong Development Bank (SPDB), China Minsheng Banking Corp. (CMBC), China Merchants Bank (CMB), Industrial Bank Co. (CIB), Bank of Beijing (BOB), Agricultural Bank of China (ABC), Bank of Communications (BCM), Industrial and Commercial Bank of China (ICBC), China Everbright Bank (CEB), CITIC Securities Co. (CITIC SEC), HaiTong Securities Co. (HaiTong SEC), China Merchants Securities (CM SEC), and HuaTai Securities Co. (HuaTai SEC). There are slight differences in their quantities due to the different levels of data missing for each stock. After eliminating non-traded weekends, statutory holidays, and incomplete trading days, the trading days of all the selected stocks exceeded 860 days, and the average daily data exceeded 3400. The specific quantities of each stock are seen in Table 1.

The transaction prices were adjusted in the following way. Trades before 9:30 am and after 3 pm were eliminated. The 25 s after the opening time are also ignored because they are close to the pre-market opening session.

For specific, we obtain the intraday day structure of trade volume in every 5 min. Figure 2 shows the average changing of trade volume in one day of selected stocks. It represents a typical U shape of trading, which means trade volume is always larger at the beginning and the end than the middle of the whole exchange period. And due to the influence of lunch break time, and there exits some violent fluctuations around 1 pm. For a clearer observation, we separately give the trade volume changing of PingAn in one day.

Table 1: Basic information of the 16 selected stocks traded in the Shanghai stock market.

Ticker symbol	Trading days	Daily average data
PingAn	899	4315
CPIC	900	3431
ChinaLife	900	3305
SPDB	867	3983
CMBC	900	4013
CMB	895	3841
CIB	895	4103
BOB	862	3476
ABC	900	3954
BCM	900	3687
ICBC	900	3715
CEB	889	3671
CITIC SEC	900	4436
HaiTong SEC	898	3877
CM SEC	893	3432
HuaTai SEC	893	3807

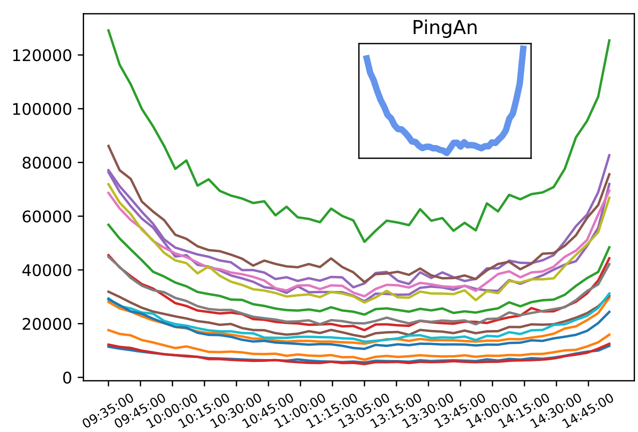

Figure 2: The intraday day structure of average trade volume of selected stocks.

Table 2: The ADF unit root test results of the price change on January 2, 2014.

Ticker symbol	ADF statistical value	p-value
PingAn	−12.854	5.3E-24
CPIC	−15.954	7.3E-29
ChinaLife	−31.347	0.0E+00
SPDB	−14.947	1.3E-27
CMBC	−32.567	0.0E+00
CMB	−33.672	0.0E+00
CIB	−32.075	0.0E+00
BOB	−20.951	0.0E+00
ABC	−16.602	1.8E-29
BCM	−23.414	0.0E+00
ICBC	−18.093	2.6E-30
CEB	−15.511	2.4E-28
CITIC SEC	−15.329	4.0E-28
HaiTong SEC	−34.256	0.0E+00
CM SEC	−12.374	5.2E-23
HuaTai SEC	−24.873	0.0E+00

3.2 Estimate the permanent and the temporary price impact coefficient

We test the time stationarity of the price change Δp_t and the signed trade volume q_t separately using the ADF unit root test. The results show that the statistical values of Δp_t and q_t of all the selected stocks are less than the test value under the 1% significance level, which indicates that all of the time series are statistically stationary and satisfy the requirement of using the least squares for estimation. To conserve space, we represent the ADF unit root test results of the price change and signed trade volume on January 2, 2014, in Table 2 and 3, respectively.

We obtain the daily permanent price impact and daily temporary price impact for each stock and the results of the significance test of the two types of price impacts for the selected stocks. The t-test results show that the regression parameters are highly statistically significant under the 5% significance level. Take PingAn for example. Figure 3 shows the p-values of the daily permanent price impact and the daily temporary price impact using a t-test. The p-values of the daily permanent price impact are less than the test value under the 5% significance level, and for the daily temporary price impact series, the p-values of 887 days in the 899 trading

Table 3: The ADF unit root test results of the signed trade volume on January 2, 2014.

Ticker symbol	ADF statistical value	p-value
PingAn	−61.784	0.0E+00
CPIC	−13.774	9.6E-26
ChinaLife	−18.942	0.0E+00
SPDB	−61.524	0.0E+00
CMBC	−30.139	0.0E+00
CMB	−59.148	0.0E+00
CIB	−64.517	0.0E+00
BOB	−13.525	2.7E-25
ABC	−16.892	1.0E-29
BCM	−15.966	7.1E-29
ICBC	−47.920	0.0E+00
CEB	−20.104	0.0E+00
CITIC SEC	−63.373	0.0E+00
HaiTong SEC	−15.308	4.2E-28
CM SEC	−61.095	0.0E+00
HuaTai SEC	−11.534	3.8E-21

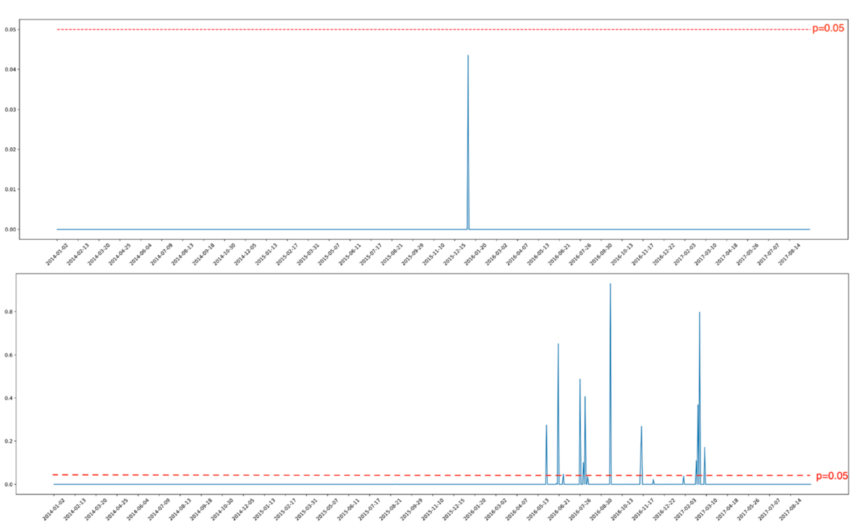

Figure 3: p-values series of the two types of price impact under the t-test of PingAn. The upper figure plots the p-values of the daily permanent price impact coefficient, and the lower figure plots the p-values of the daily temporary price impact coefficient.

data (accounting for 98.7% of the total) are less than 0.05 using a t-test. That is to say, the two types of price impact are generally significant.

Figure 4 represents the time-variance of the daily permanent price impact and the daily temporary price impact of PingAn. The regression values of the permanent price impact are persistent and have a positive impact on the price change, thus suggesting that a buy trade should push the price up, which seems obvious at first sight and is easily demonstrated empirically. The permanent price impact reflects the long-term effect on price changes of the trade direction, thus suggesting that the sign of the permanent price impact will not change over time. The temporary price impact reflects the impact of the unexpected signed trade volume on the trading price. Since the unexpected signed trade volume includes both the trade direction and trade volume, the sign of the temporary price impact is not determined to be changing with the specific circumstances.

Table 4 provides some summary statistics of the daily permanent price impact for the 16 selected stocks. The mean of the daily permanent price impact is positive, and most of the skewness is positive, thus indicating that the right tail of the probability density function is greater than the left side. That is, the long tail is on the

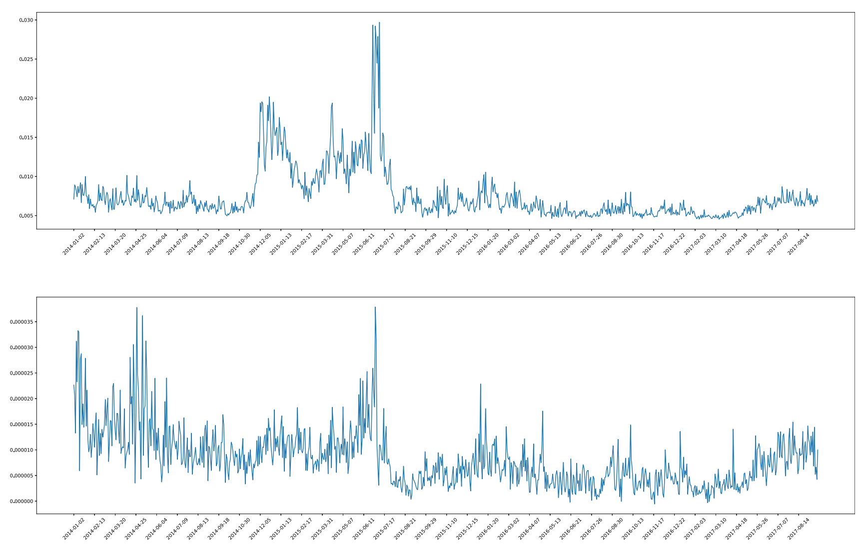

Figure 4: The evolution of the daily permanent price impact and daily temporary price impact of PingAn. The top figure displays the daily permanent price impact as the observation time proceeds, and the bottom figure plots the daily temporary price impact.

Table 4: Summary statistics. It reports the mean, standard deviation, skewness, kurtosis and L–B statistics of the daily permanent price impact.

Ticker symbol	Mean	Standard deviation	Skewness	Kurtosis	L–B statistics
PingAn	0.0075	0.0034	2.991	11.972	3.10E-134
CPIC	0.0075	0.0023	1.372	2.804	4.98E-73
ChinaLife	0.0069	0.0022	1.772	4.391	7.80E-104
SPDB	0.0052	0.0009	2.529	9.266	3.29E-67
CMBC	0.0049	0.0004	4.078	28.261	2.78E-64
CMB	0.0058	0.0012	2.143	6.924	2.69E-65
CIB	0.0050	0.0006	4.559	35.533	8.94E-81
BOB	0.0050	0.0005	3.008	13.707	1.96E-73
ABC	0.0049	0.0001	−2.655	8.616	5.50E-94
BCM	0.0048	0.0002	2.779	21.823	5.31E-55
ICBC	0.0049	0.0001	−0.242	9.854	6.47E-60
CEB	0.0048	0.0002	−0.113	10.500	3.98E-92
CITIC SEC	0.0050	0.0012	4.186	22.328	4.97E-123
HaiTong SEC	0.0052	0.0025	21.283	529.723	1.12E-11
CM SEC	0.0061	0.0019	2.389	7.889	1.64E-127
HuaTai SEC	0.0056	0.0014	2.339	8.154	1.74E-115

right side. The kurtosis is all positive, thus suggesting that it shows a sharp peak. The sign of the mean of the temporary price impact is unclear, and the standard deviation is very small. Almost all the skewness is positive, which means that the long tail part is on the right side of the density function, and the kurtosis is all positive, thus showing a sharp peak. The L–B statistics were constructed to test the hysteresis correlation of the obtained two types of daily price impact coefficients. The results show that the first-order lag tests of the two types of price impact coefficients are less than 0.01. That is, the two types of price impact coefficients are correlated, and the series have the ARCH effect. Table 5 reports the summary statistics of the daily temporary price impact, including the five indicators of the mean, standard deviation, skewness, kurtosis, and L–B statistics.

Furthermore, in order to find out the distribution characteristics of the two types of price impact coefficients, we obtain a histogram of 16 selected stocks from the two types of price impact coefficient series, as shown in Appendix 1.

There are differences in the distribution characteristics of each stock, but some of the distributions show similar characteristics. For example, the daily permanent shock coefficient has 12 stocks with the characteristic of left-tailing, and the distribution of four stocks shows right-tailing characteristics. Then, we integrate the daily permanent price impact coefficients and the temporary price impact coefficients of the selected stocks to obtain the whole distribution characteristics of the selected financial stocks. On the left side of Figure 5 is the distribution curve of the overall daily permanent price impact coefficient. On the right side is the distribution

Table 5: Summary statistics. It reports the mean, standard deviation, skewness, kurtosis and L–B statistics of the daily temporary price impact.

Ticker symbol	Mean	Standard deviation	Skewness	Kurtosis	L–B statistics
PingAn	8.45E-06	5.80E-06	1.543	3.809	1.53E-69
CPIC	1.22E-05	9.18E-06	1.547	4.561	1.45E-28
ChinaLife	1.29E-05	9.19E-06	3.096	20.072	9.98E-23
SPDB	1.50E-06	1.93E-06	1.977	5.704	5.12E-39
CMBC	2.68E-07	4.47E-07	1.406	3.873	3.71E-57
CMB	2.48E-06	2.40E-06	1.377	5.323	2.81E-41
CIB	7.17E-07	7.07E-07	2.390	19.872	5.58E-20
BOB	9.09E-07	9.30E-07	0.847	2.029	1.09E-36
ABC	−8.73E-09	3.11E-08	−0.438	2.710	2.32E-24
BCM	7.11E-08	2.09E-07	1.497	8.114	5.69E-16
ICBC	7.12E-09	9.02E-08	0.118	3.868	6.59E-11
CEB	−1.60E-08	1.15E-07	0.608	6.574	7.72E-15
CITIC SEC	1.22E-06	8.28E-07	1.048	2.461	1.51E-47
HaiTong SEC	2.08E-06	1.79E-06	2.111	8.412	1.65E-38
CM SEC	6.61E-06	4.05E-06	1.328	5.023	1.47E-46
HuaTai SEC	4.60E-06	3.11E-06	1.137	1.653	3.68E-63

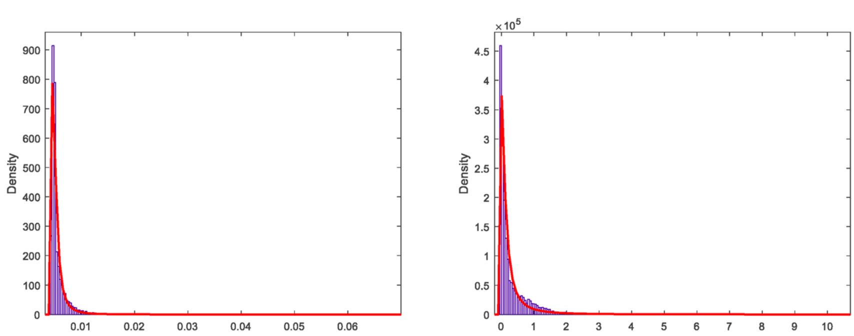

Figure 5: Distribution characteristics of the two types of price impact coefficients. The left figure plots the distribution of the daily permanent price impact, and the right figure plots the distribution of the daily temporary price impact.

curve of the overall temporary price impact coefficient. The general distribution of the two types of coefficients can be fitted by the generalized extreme value distribution.

The two types of price impact coefficients for each stock are quite different from the overall. The distribution of some permanent price impact coefficients is similar to the overall characteristics. In comparison, the distribution characteristics of the temporary price impact coefficients are different from the overall distribution characteristics.

3.3 Smoothness and long-term memory test of the price impact coefficient series

Furthermore, we discuss the statistical characteristics of the daily permanent price impact coefficient and the daily temporary price impact coefficient, and we test the stationarity and long-term memory for the two types of price impact coefficients of each stock. Table 6 shows the ADF test results for the selected stocks. The column of the permanent price impact coefficient shows the p-value of the test result by examining the unit root of the permanent impact coefficient series. The temporary price impact coefficient column shows the p-value of the test result by examining the unit root of the temporary price impact coefficient series.

From the test results in Table 6, the daily permanent price impact coefficient series of almost all stocks are time steady under the 5% significance level, and most temporary price impact coefficients are time steady under the 5% significance level.

We obtain the Hurst Index by using a non-parametric way, which is the DFA method. Table 7 gives the $\alpha_{permanent}$ of the daily permanent price impact and the $\alpha_{temporary}$ of the daily temporary price impact of the 16 selected stocks.

From the results of the above table, it can be seen that the scale index of all stocks is greater than 0.5. That is to say, the daily permanent price impact coefficient series and the daily temporary price impact coefficient series have a persistent long-term correlation, and there are more tendencies. In other words, if a stock is on an upward trend during the previous period, the next period is likely to rise, and if the stock is declining during the previous period, the next period is likely to fall.

The statistical properties of the two types of scale indices for the 16 selected stocks are shown in Table 8. The mean value of the scale index of the permanent price impact coefficient is smaller than the mean value of the temporary price impact coefficient scale index, and the variance of the permanent price impact coefficient

Table 6: The p-value of the ADF test for the price impact coefficient series.

Ticker symbol	Permanent price impact	Temporary price impact
PingAn	0.0149	0.0052
CPIC	0.0003	0.0534
ChinaLife	0.0002	0.0559
SPDB	0.0209	0.2229
CMBC	0.0000	0.0288
CMB	0.0029	0.0997
CIB	0.0003	0.0013
BOB	0.0000	0.0110
ABC	0.0213	0.0159
BCM	0.0000	0.0039
ICBC	0.0177	0.0004
CEB	0.0024	0.0010
CITIC SEC	0.0031	0.0006
HaiTong SEC	0.0000	0.0130
CM SEC	0.0225	0.0059
HuaTai SEC	0.0623	0.1333

Table 7: The scale index of the price impact coefficients of the selected stocks.

Ticker symbol	Scale index $\alpha_{permanent}$	Scale index $\alpha_{temporary}$
PingAn	0.959	1.055
CPIC	0.966	1.129
ChinaLife	0.995	0.992
SPDB	0.916	1.229
CMBC	0.688	0.997
CMB	0.831	1.068
CIB	0.792	0.903
BOB	0.805	0.857
ABC	0.943	0.943
BCM	0.802	0.856
ICBC	0.990	0.681
CEB	1.010	0.656
CITIC SEC	0.962	1.021
HaiTong SEC	0.761	0.889
CM SEC	0.951	1.043
HuaTai SEC	0.989	1.097

Table 8: Statistical properties of the two types of price impact scale indices.

Index	Scale index for the permanent price impact coefficient $\alpha_{permanent}$	Scale index for the temporary price impact coefficient $\alpha_{temporary}$
Mean	0.8975	0.9635
Variance	0.0096	0.0221

scale index is greater than the variance of the temporary price impact coefficient scale index. Furthermore, we use the paired t-test to analyze whether there are significant differences in the long-term memory of the two types of coefficients. The results show that at the significance level of 5%, there is no significant difference in the long-term memory of the two types of coefficients.

As we explain in Section 2.1, the two types of price impact will be affected by the auto-correlation of trading direction and order flow respectively. And the empirical results illustrate this conclusion.

3.4 Correlation analysis of the price impact coefficients

In order to discuss the price impact relationships between stocks, we eliminate the incomplete data segments for the 16 selected stocks and then obtain a complete data series of 791 days. Then, we analyze the correlation of the two types of price impact coefficients. According to the nature of the industry, the stocks are divided into the banking industry, the securities industry, and the stock industry, and we conduct a correlation analysis of the impact coefficients in the industry. The results find that there is a clear correlation between the permanent price impact coefficients of stocks, and the correlation between the temporary price impact coefficients are generally weak. Table 9 and 10 respectively gives the correlations of the permanent price impact coefficient and the temporary price impact coefficient of the stocks in the banking industry.

From the results of the correlation matrix, we can see that the overall correlation of the temporary price impact coefficient of the banking industry is relatively low with no more than 0.4. In contrast, the permanent price impact coefficient of the banking industry is relatively large. Among them, there are 23 groups where the correlation of the permanent price impact coefficient exceeds 0.5, and BCM and CEB have the highest correlations of the permanent price impact coefficient, which is 0.6483.

Table 9: Correlation of the permanent price impact coefficients of stocks in the banking industry.

Correlation	CEB	CIB	CMB	BCM	SPDB	CMBC	BOB	ABC
CEB	1							
CIB	0.0077	1						
CMB	0.0690	0.5878	1					
BCM	0.6483	0.3266	0.3413	1				
SPDB	0.1075	0.6030	0.6125	0.3128	1			
CMBC	0.2639	0.5913	0.6163	0.4624	0.5634	1		
BOB	0.0718	0.5910	0.5440	0.4252	0.5794	0.4977	1	
ABC	0.5422	−0.4819	−0.3114	0.1349	−0.1965	−0.2577	−0.3608	1

Table 10: Correlation of the temporary price impact coefficients of stocks in the banking industry.

Correlation	CEB	CIB	CMB	BCM	SPDB	CMBC	BOB	ABC
CEB	1							
CIB	0.2278	1						
CMB	−0.1184	0.2276	1					
BCM	0.2862	0.2743	0.0129	1				
SPDB	−0.0059	0.2424	0.2312	0.0929	1			
CMBC	0.2638	0.3941	0.0326	0.3172	0.1789	1		
BOB	0.3022	0.2222	−0.0319	0.2949	0.2420	0.2981	1	
ABC	0.2664	−0.0392	−0.2182	0.1624	−0.1843	0.0537	0.1659	1

The analysis results of the correlation between the daily permanent price impact coefficient and the temporary impact coefficient for the insurance industry and the securities industry are the same as those for the banking industry. See Appendix 2 for the specific results.

In summary, the paired t-test of the correlation analysis shows that the t-value is 3.678 and the p-value is 0.00035. Under the 1% significance level, the correlations between the temporary price impact coefficients of each stock are significantly lower than those for the permanent price impact coefficients. The correlation shows that the permanent price impact coefficients affect each other among the stocks, and the instantaneous impact factor has less mutual influence on the stocks.

3.5 Trade interval and bid-ask spread effect

The raw data used in this study are tick-by-tick data. Because of the randomness of transactions, the time interval of the trading data is not constant. We speculate that the length of the time delay affects the size of the price impact coefficient to a certain extent, so we add the interval item into the model.

Based on model (3) and (4), we obtain the coefficients using the least squares method. Then, the results of the significance tests show that the permanent price impact and temporary price impact coefficients are significant under the 5% significance level, and time interval's impact is not significant under the 5% significance level.

In order to know the time interval impact on the trade volume, the upper figure in Figure 6 gives the p-values of the significance test of PingAn. Under the significance level of 5%, the regression coefficient of the time interval impact is not significant, which shows that the explanation of the time interval impact is very small for trade volume. For the above reason, we remove the time interval impact as the explanatory variable of trade volume.

As for the time interval impact on the trading price change, the lower figure in Figure 6 gives the p-values of the significance test of PingAn. Under the significance level of 5%, the regression coefficient of the time interval impact is not significant, which shows that the explanation of the time interval impact is very small for trading price changes. For the above reason, we remove the time interval impact as an explanatory variable of trading price changes.

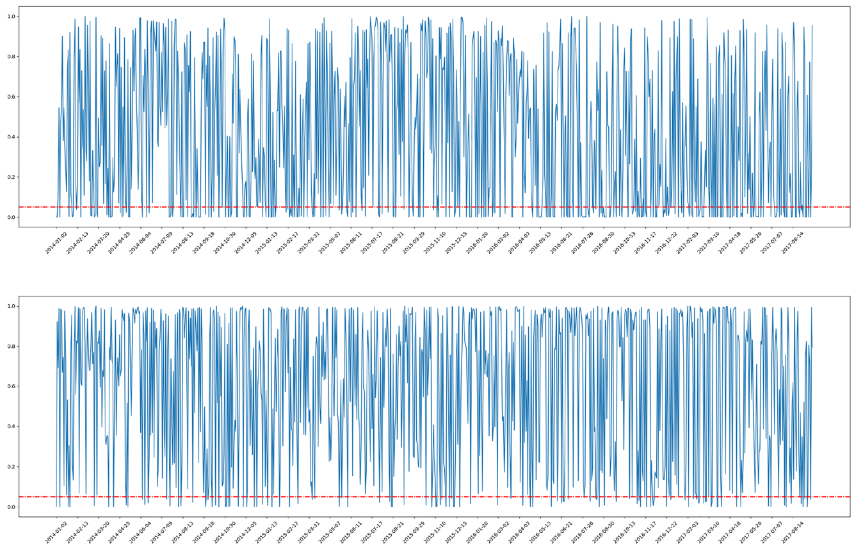

Figure 6: Significance test for the time interval impact parameters for PingAn. The upper figure shows the p-value series for the parameter of the time interval impact on the trading volume based on model (3), and the lower figure shows that the p-value series for the parameter of the time interval impact on the trading price change based on model (4).

The above discussion shows that when selecting the high-frequency trading data to study the price impact coefficient characteristics, the time interval impact has no significant effect on the trade volume and trading price change. Combined with practical considerations, since market transaction information can be quickly captured by market participants and used for future trading decisions, time intervals are not the primary consideration for price impacts.

Further, we consider the impact of the bid-ask spread. The gap between the best bid and ask price level represents the bid-ask spread, which have the same fluctuation behavior with transaction price change. As an illustration, Figure 7 display the bid-ask spread curve and price change curve respectively, based on the snapshots of Limit Order Book on 6 January 2014 of PingAn.

From Figure 7, we can simply see that bid-ask spread could magnify the price change through transaction. Specifically, the bid-ask spread will affect the absolute value of price change, whether the price change is positive or negative. So we add a sign, which is the same with price change, to bid-ask spread in order to present the magnify influence in model.

Follow the above process, we obtain the p-values of the significance test of PingAn based on model (3) and (4) shown in Figure 8. Under the significance level of 5%, the regression coefficient of the bid-ask spread impact is not significant in model (3), which shows that the explanation of the bid-ask spread impact is very small for trade volume. For the above reason, we remove the bid-ask spread impact as the explanatory variable of trade volume. While under the same significance level, the regression coefficient of the bid-ask spread impact is significant in model (4), which confirm the relationship between bid-ask and price change we mentioned above.

Therefore, we remove transaction duration and bid-ask spread from model (3) and removing transaction duration from model (4) to adjust the extended HFV model. Similarly, we obtain the values of two types of price impact. The extended HFV model has better explanation on the empirical data than HFV model, which means

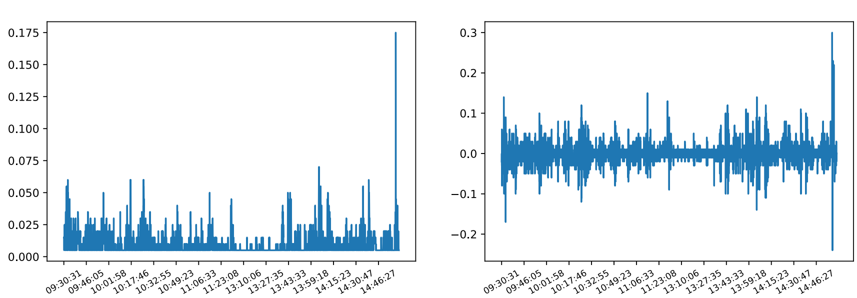

Figure 7: The left figure (bid-ask spread curve) and the right figure (price change curve) based on the snapshots of Limit Order Book on 6 January 2014 of PingAn.

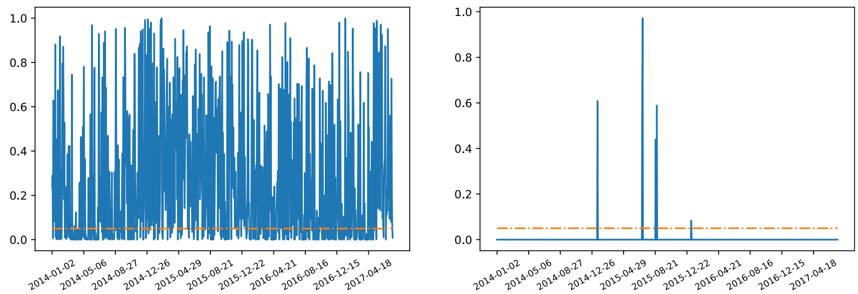

Figure 8: Significance test for the bid-ask spread impact parameters for PingAn. The upper figure shows the p-value series for the parameter of the bid-ask spread impact on the trading volume based on model (3), and the lower figure shows that the p-value series for the parameter of the bid-ask spread impact on the trading price change based on model (4).

bid-ask spread is an effective factor to explain the trading price change. For the estimation of two types of price impact coefficients, there is no significant difference between the two models in order of magnitude. Both models reflect the statistical characteristics of the above discussion.

4 Conclusion

In this paper, we discuss the statistical characteristics of the daily permanent price impact and the daily temporary price impact based on the extended HFV model. The main conclusions are as follows. (1) The permanent price impact coefficients are positive, indicating that the permanent change in price due to the transaction is the same as the trade volume sign. The temporary price impact can be positive or negative, indicating that the temporary impact of the price due to the transaction is related to both the trade volume sign and market liquidity. (2) All permanent price impact series and most temporary price impact series basically show stationarity, indicating that the price impact has a stable statistical regularity. (3) Both the permanent price impact series and the temporary price impact series have significant long-term memory, and there is no significant difference between the two types of price impacts, indicating that these two types of price impact coefficients have predictable characteristics. (4) There is a significant correlation between the permanent price impact coefficients and a weak correlation in the temporary price impact coefficients, which indicates that the permanent price impact generated by transactions will affect the price changes of other assets. However, the temporary price impact of asset prices caused by transactions mainly affects the price change of their own asset. (5) The time interval impacts of the tick-by-tick data on the trade volume and trading price changes is insignificant, indicating that the timing of consecutive orders has a negligible impact on prices. (6) The bid-ask spread is an effective factor to explain the trading price change.

The research in this paper is based on the HFV model. The core of the model is that the change in the asset price is determined by the direction of the transaction volume and the new process of the transaction volume. Essentially, the effect of changes in trade volume on the price change is linear. However, many studies confirm that the impact of trade volume on price shocks is non-linear and the price shocks generated by trading are not permanent or temporary but are elastic. Therefore, the model has its limitations. Furthermore, because we only conducted an empirical analysis of the price shocks of China's Shanghai stock market, the conclusions that are obtained have yet to be tested in other stock markets. This is also the future direction of our work.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: National Natural Science Foundation of China (no. 71671017).
Employment or leadership: None declared.
Honorarium: None declared.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

Almgren, R., and N. Chriss. 1999. "Value under liquidation." *Risk* 12: 62–4.

Almgren, R., and N. Chriss. 2000. "Optimal Execution of Portfolio Transactions." *Journal of Risk* 3: 5–39, https://doi.org/10.21314/jor.2001.041.

Almgren, R., C. Thum, E. Hauptman, and L. Hong. 2005. "Direct Estimation of Equity Market Impact." *Risk* 18.

Almgren, R. F. 2003. "Optimal Execution with Nonlinear Impact Functions and Trading-enhanced Risk." *Applied Mathematical Finance* 10: 1–18, https://doi.org/10.1080/135048602100056.

Amihud, Y. 2002. "Illiquidity and Stock Returns: Cross-section and Time-series Effects." *Journal of Financial Markets* 5 (1): 31–56, https://doi.org/10.1016/s1386-4181(01)00024-6.

Bahar, S., J. W. Kantelhardt, A. Neiman, H. H. A. Rego, D. F. Russell, L. Wilkens, A. Bunde, and F. Moss. 2001. "Long-range Temporal Anti-correlations in Paddlefish Electroreceptors." *Europhysics Letters* (*EPL*) 56 (3): 454–60, https://doi.org/10.1209/epl/i2001-00540-7.

Bartsch, R., T. Henning, A. Heinen, and P. Maass. 2005. "Statistical Analysis of Fluctuations in the ECG Morphology." *Physica A: Statistical Mechanics and its Applications* 354 (1): 415–31, https://doi.org/10.1016/j.physa.2005.03.019.

Bashan, A., R. Bartsch, J. W. Kantelhardt, and S. Havlin. 2008. "Comparison of Detrending Methods for Fluctuation Analysis." *Physica A Statistical Mechanics its Applications* 387 (21): 5080–90, https://doi.org/10.1016/j.physa.2008.04.023.

Beran, J., R. Sherman, M. S. Taqqu, and W. Willinger 1995. "Long-range Dependence in Variable-Bit-Rate Video Traffic." *IEEE Transactions on Communications* 43 (2): 1566–79, https://doi.org/10.1109/26.380206.

Bouchaud, J. P., J. D. Farmer, and F. Lillo. 2009. "How Markets Slowly Digest Changes in Supply and Demand," In *Handbook of Financial Markets: Dynamics and Evolution*. North-Holland: Elsevier.

Brennan, M. J., and A. Subrahmanyam. 1996. "Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns." *Journal of Financial Economics* 41 (3): 441–64, https://doi.org/10.1016/0304-405x(95)00870-k.

Hopman, C. 2007. "Do Supply and Demand Drive Stock Prices?." *Quantitative Finance* 7 (1): 37–53, https://doi.org/10.1080/14697680600987216.

Chordia, T., S. W. Huh, and A. Subrahmanyam. 2009. "Theory-Based Illiquidity and Asset Pricing." *Review of Financial Studies* 22 (9): 3629–68, https://doi.org/10.1093/rfs/hhn121.

Dufour, A., and R. F. Engle. 2000. "Time and the Price Impact of a Trade." *Journal of Finance* 55, 2467–98, https://doi.org/10.1111/0022-1082.00297.

Engle, R. F., and Z. Sun. 2007. "When is Noise not Noise a Microstructure Estimate of Realized Volatility." NYU Working Paper No. FIN-07-047.

Farmer, J. D., and N. Zamani. 2007. "Mechanical vs. Informational Components of Price Impact." *European Physical Journal B* 55 (2), 189–200, https://doi.org/10.1140/epjb/e2006-00384-5.

Foster, F. D., and S. Viswanathan. 1993, "Variations in Trading Volume, Return Volatility, and Trading Costs: Evidence on Recent Price Formation Models." *Journal of Finance* 48, 187–211, https://doi.org/10.1111/j.1540-6261.1993.tb04706.x.

Glosten, L. R., and L. E. Harris. 1988. "Estimating the Components of the Bid/ask Spread." *Journal of Financial Economics* 21 (1): 123–42, https://doi.org/10.1016/0304-405x(88)90034-7.

Hasbrouck, J. 1991. "Measuring the information content of stock trades." *The Journal of Finance* 46 (1): 179–207, https://doi.org/10.1111/j.1540-6261.1991.tb03749.x.

Hasbrouck, J. 2007. "Empirical Market Microstructure." *The Institutions, Economics, and Econometrics of Securities Trading*. USA: Oxford University Press.

Huberman, G., and W. Stanzl. 2004. "Price Manipulation and Quasi-Arbitrage." *Econometrica* 72 (4): 1247–75, https://doi.org/10.1111/j.1468-0262.2004.00531.x.

Jondeau, E., J. Lahaye, and M. Rockinger. 2015. "Estimating the Price Impact of Trades in a High-Frequency Microstructure Model with Jumps." *Journal of Banking Finance* 61, S205–24, https://doi.org/10.1016/j.jbankfin.2015.09.005.

Jones, C. M., G. Kaul, and M. L. Lipson 1994. "Transactions, Volume, and Volatility." *Review of Financial Studies* 7 (4): 631–51, https://doi.org/10.1093/rfs/7.4.631.

Kyle, A. S. 1985. "Continuous Auctions and Insider Trading." *Finance Stochastics* 53 (6): 1315–35, https://doi.org/10.2307/1913210.

Lespagnol, V., and J. Rouchier. 2018. "Trading Volume and Price Distortion: An Agent-Based Model with Heterogenous Knowledge of Fundamentals." *Computational Economics* 4: 1–30, https://doi.org/10.1007/s10614-017-9655-y.

Lillo, F., J. D. Farmer, and R. N. Mantegna. 2003. "Econophysics: Master Curve for Price-Impact Function." *Nature* 421 (6919): 129, https://doi.org/10.1038/421129a.

Peng, C. K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. 1994. "Mosaic Organization of DNA Nucleotides." *Physical Review E* 49 (2): 1685, https://doi.org/10.1103/physreve.49.1685.

Philip, R., 2020. Estimating permanent price impact via machine learning. Journal of Econometrics 215 (2), 414–49. https://doi.org/10.1016/j.jeconom.2019.10.002.

Polimenis, V. 2005. "A Realistic Model of Market Liquidity and Depth." *Journal of Futures Markets* 25 (5): 443–64, https://doi.org/10.1002/fut.20147.

Santhanam, M. S., J. N. Bandyopadhyay, and D. Angom 2006. "Quantum Spectrum as a Time Series: Fluctuation Measures." *Physical Review E Statistical Nonlinear Soft Matter Physics* 73 (1) Pt 2: 015201, https://doi.org/10.1103/physreve.73.015201.

Ting, C., M. Warachka, and Y. Zhao 2007. "Optimal Liquidation Strategies and their Implications." *Journal of Economic Dynamics Control* 31 (4): 1431–50, https://doi.org/10.1016/j.jedc.2006.07.003.

Toth, B., I. Palit, F. Lillo, and D. Farmer 2015. "Why is Equity Order Flow so Persistent?." *Journal of Economic Dynamics and Control* 51: 218–39, https://doi.org/10.1016/j.jedc.2014.10.007.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/snde-2018-0067).