Journal of Management Science and Engineering 9 (2024) 53–61

Contents lists available at ScienceDirect

Journal of Management Science and Engineering

journal homepage: www.keaipublishing.com/en/journals/journal-of-management-science-and-engineering/

An MA-MRR model for transaction-level analysis of high-frequency trading processes

Qiang Zhang [a], Zudi Lu [b], Shancun Liu [c], Haijun Yang [c, d, *], Jingrui Pan [c]

[a] School of Economics and Management, Beijing University of Chemical Technology, Beijing, 100029, China
[b] Mathematical Sciences, and Southampton Statistical Sciences Research Institute, University of Southampton, Highfield, Southampton, SO17 1BJ, UK
[c] School of Economics and Management, Beihang University, Beijing, 100191, China
[d] Key Laboratory of complex System Analysis, Management and Decision (Beihang University), Ministry of Education, Beijing, 100191, China

ARTICLE INFO

Article history:
Received 8 December 2022
Received in revised form 13 May 2023
Accepted 3 August 2023
Available online 25 August 2023

JEL classification:
G10
G14
G15

Keywords:
Spread decomposition
Adverse selection risk
MA-MRR model
Information lag

ABSTRACT

The transaction-level analysis of security price changes by Madhavan, Richardson, and Roomans (1997, hereafter MRR) is a useful framework for financial analysis. The first-order Markov property of trading indicator variables is a critical assumption in the MRR model, which contradicts the information lag empirically demonstrated in high-frequency trading processes. In this study, a nonparametric test is employed, which shows that the Markov property of the trading indicator variables is rejected on most trading days. Based on the spread decomposed structure, an MA-MRR model was proposed with a moving average structure adopted to absorb the information lag as an extension. The empirical results show that the information lag plays an important role in measuring the adverse selection risk parameter and that the difference in this parameter between the original and the extension is significant. Furthermore, our analysis suggests that the information lag parameter is a useful measure of the average speed at which information is incorporated into prices.

© 2023 China Science Publishing & Media Ltd. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Transaction-level analyses of security price changes have received increasing attention in financial economics. Madhavan et al. (1997) proposed a spread decomposition model (MRR) to show that security prices change due to the arrival of new information in the trading process. This model has been widely used in empirical studies to measure adverse selection risk and liquidity costs in securities markets. For example, Bushee et al. (2018), Angelidis and Benos (2009), Sita and Westerholm (2011), Green (2004), Riordan et al. (2013), Andros (2015), Ahern (2014), Lai et al. (2014), Armstrong et al. (2011), Gregoriou and Rhodes (2017), and Sakawa and Ubukata (2014) adopted adverse selection risk as an important measure of symmetric information in equity markets. It has also been used in the futures markets (for example, Mizrach and Otsubo, 2014; Medina et al., 2014; Frijins and Tse, 2015; Ahn et al., 2002, 2008). Empirical studies on bond, fund, foreign exchange, and

* Corresponding author. School of Economics and Management, Beihang University, Beijing, 100191, China.
E-mail addresses: jqx_zhq@buaa.edu.cn (Q. Zhang), Z.Lu@soton.ac.uk (Z. Lu), Liushancun@buaa.edu.cn (S. Liu), navy@buaa.edu.cn (H. Yang), jingrui.pan@buaa.edu.cn (J. Pan).

https://doi.org/10.1016/j.jmse.2023.08.001
2096-2320/© 2023 China Science Publishing & Media Ltd. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

cryptocurrency exchange markets use adverse selection risk as a key indicator (Han and Zhou, 2014; Zhang, 2015; Chen and Gau, 2014; Fernandez-Perez et al., 2019; Dyhrberg et al., 2018).

An important assumption in the MRR model is that trade initiation follows a first-order Markov process: This implies that in the information set, only one period of lagged information is useful due to the Markov property. In particular, information innovations could be incorporated into prices after one trade. Indeed, if the trade initiation variables could be immediately adjusted to a full information level, the past initiation variables would be redundant and the current trade initiation variable would be a sufficient statistic for predicting future trade. However, Hasbrouck (1991) and Dufour and Engle (2000) use the impulse response function to test how long a unit of information innovation can be incorporated into the price. Their results show that it takes at least 20 trades to reach a fully integrated level, with the first five trades playing a crucial role. This finding contradicts the assumptions of the MRR model. Therefore, whether the asymmetric index estimated using the MRR model is trustworthy remains a key question. In this study, we have tried to answer two fundamental questions. First, is the information neglected by the MRR model essential for the estimation of adverse selection? If so, to what extent? Second, is lag information a new component in the spread decomposition structure, and what role does it play in the trading process?

An extended model is proposed. The results presented below indicate that lag information plays an important role in measuring adverse selection risk, and the lag information parameter can be used to measure the speed of incorporating information into prices. We arrive at these conclusions in three steps.

First, we test the first-order Markov property using the conditional independence property developed by de Matos and Fernandes (2007). Our empirical examination shows that the Markov property of the trade initiation variable is rejected for most trading days. Therefore, the information structure employed by the MRR model may have neglected important trading information, which directly led to deviations in the parameter estimation.

The second step extends the MRR model to accommodate a moving-average structure to compensate for the information neglected by the original MRR model. Maximum likelihood (ML) estimation with an ARCH structure is introduced instead of the generalized moment method (GMM) used in the original MRR model. This extension allows us to estimate more parameters than in the original model (the information lag parameter), revised adverse selection risk, and revised liquidity cost. A positive estimated parameter indicates a positive information lag in the trading process, consistent with the results of Hasbrouck (1991) and Dufour and Engle (2000). There is a sharp difference between the adverse selection risk parameter in the MRR model and the revised one in our model, which is attributed to the different informational structures employed by the two models; neglected information is responsible for the deviation. Furthermore, empirical results based on the original MRR parameters are doubtful.

The third step is to explore the information lag parameters in the extended structure. Our analysis suggests that the information lag parameter can be used to measure the average speed at which information is integrated into prices. The larger the parameter value, the slower the speed. Empirically, our findings are consistent with those of Hasbrouck (1991) and Dufour and Engle (2000), and the information lag parameter provides further details on how information is integrated into prices.

The remainder of this paper is organized as follows. Section 2 introduces the high-frequency data used in the empirical results. In Section 3, we extend the MRR model by examining the Markov properties of trading processes. In Section 4, we provide an empirical comparison between the original and extended models and explore insights into the information lag parameter. Finally, Section 5 presents the conclusions.

2. Data

For the empirical studies below, we use high-frequency data of 50 stocks from the Shanghai Stock Exchange of China (SSE) constituent index and extract their transaction data from the China Center for Economic Research (CCER) database. These data are sourced from the CCER high-frequency trading database made available by the Beijing Sinofin Information Service, which includes the real-time bid-ask prices and associated limit order volumes at the top five levels on either side of the order book for each transaction, as well as instantaneous transaction price, volume, transaction amount, aggregated transaction amount, and the direction of order initiation. Our sample period is from September 1 to December 31, 2017.

The SSE is a purely order-driven market without designated market makers. It runs an electronic automated trading system and is open from Monday to Friday, with three sessions:09:15–09:25 for call auction, 09:30–11:30, and 13:00–15:00 for continuous trading double auctions. Only limit orders were allowed in the SSE. The orders are valid for one day and stored in the limit order book, of which the best five bid and ask prices and the corresponding depths of the book are revealed continuously to public investors. The tick size is 0.01 RMB, while the minimum trading quantity is 100 shares.

3. The MRR model and its extension

This section first briefly discusses the MRR model, then introduces a nonparametric density approach to test the Markov property, and finally extends the model and presents a family of ML estimators with an ARCH structure.

3.1. The MRR model

Madhavan et al. (1997) presented a spread decomposed model(MRR), in which the post-trade expected value of a stock, $\mu_i, (i = 1, 2, \cdots)$ evolves as

Table 1
Nonparametric tests of the Markov property.

Number	period	Total number of $\widehat{\lambda}_n$'s	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$	average of n
original duration	morning	2414	1993	1832	1526	15130.46
	afternoon		1924	1761	1499	13249.93
adjust duration	morning		1918	1735	1376	
	afternoon		1818	1651	1355	

$\widehat{\lambda}_n$ is the statistic that weakly converges to a standard normal distribution. Adjusted durations refer to the correction for the intra-day effects. The parameter α is the significant level of normal distribution. n is the number of trades in each period.

$$\mu_i = \mu_{i-1} + \theta(x_i - E[x_i|x_{i-1}]) + \varepsilon_i \tag{1}$$

where x_i is the trade indicator variable, equaling $+1$ if the trade is buy oriented, and -1 if it is sell oriented. The coefficient θ measures the degree of information asymmetry (or adverse selection risk) with surprise, in the order of $(x_i - E[x_i|x_{i-1}])$. Orthogonal innovation ε_i accounts for the information accumulated since the most recent trade. With transactions taking place either at the ask or bid, the transaction prices are given by Equation (2).

$$p_i = \mu_i + \varphi x_i + \xi_i \tag{2}$$

where the coefficient $\varphi \geq 0$ represents the cost per share borne by the supplier of liquidity and ξ_i is an i.i.d. mean-zero disturbance that accounts for rounding errors due to the discreteness of price changes. Madhavan et al. (1997) use the formula $E[x_i|x_{i-1}] = \rho x_{i-1}$,[1] where ρ is the first-order autocorrelation of the trade indicator x_i. By combining Equations (1) and (2), the transaction price changes are given by

$$p_i - p_{i-1} = (\varphi + \theta)x_i - (\varphi + \rho\theta)x_{i-1} + \varepsilon_i + \xi_i - \xi_{i-1} \tag{3}$$

or

$$p_i - p_{i-1} = \varphi(x_i - x_{i-1}) + \theta(x_i - \rho x_{i-1}) + \varepsilon_i + \xi_i - \xi_{i-1} \tag{4}$$

Equation (3) forms the basis of price movements. Price shocks consist of three parts: fundamental shock, liquidity shock, and market frictions. Meanwhile, θ represents the intensity of the fundamental shock, and φ represents the intensity of the liquidity shock. In the absence of market friction, the model reduces to the classical description of an efficient market, in which prices follow a random walk. However, in the presence of friction, transaction price movements reflect the order flow and noise induced by price discreteness, as well as information shocks.

3.2. Testing the first-order Markov property

The MRR model employs the structure $(x_i - E[x_i|x_{i-1}])$ to capture unexpected information in the trading process, or the surprise in the order flow, which follows the assumption of Glosten and Milgrom (1985). This means that the time series x_i, the trade indicator variable, follows a first-order Markov process.

With the advent of new statistical methods, some quantitative tests(Souza et al. (2018) and Chen and Hong (2012)) have enabled many fundamental assumptions in finance to be tested. The first-order Markov property is one of the most popular assumptions in continuous-time models. However, is it true? De Matos and Fernandes (2007) tested whether discretely recorded observations of a continuous-time process were consistent with the Markov property, using a smoothed nonparametric density approach. They found evidence against the Markov properties. Chen and Hong (2012) used a conditional characteristic function to check the implications of the Markov property, which is consistent with de Matos and Fernandes (2007).

We use a smoothed nonparametric density approach to check the Markov properties of the trade initiation variable. A brief introduction of this method is provided below.

Suppose X_i (which could be any trading variable, and in our test, we are only concerned with the trade indicator variable x_i) is a strictly stationary Markov time series process. Let t_i denote the observation time of the process, then the time duration between two consecutive observations $d_{i+1} = t_{i+1} - t_i$ is a measurable function of the path of X_i, and thus only depends on the information available at time t_i through X_i. This means the current duration is independent of the previous duration conditional on the previous realization. Obviously, the property of conditional independence between consecutive durations is the necessary condition of the Markov assumption X_i. So the test focuses on checking the conditional independence, in which four density functions are estimated by a nonparametric smoothing method, and a statistic, $\widehat{\lambda}_n$ is constructed (de Matos and Fernandes (2007) for the details). In this method, the statistic, $\widehat{\lambda}_n$, weakly converges to a standard normal distribution, then a

[1] It holds based on the assumption $p[x_i = 1|x_{i-1} = 1] = p[x_i = -1|x_{i-1} = -1]$ and $p[x_i = 1|x_{i-1} = -1] = p[x_i = -1|x_{i-1} = 1]$.

two-tailed test that rejects the null at the level α when $|\widehat{\lambda}_n|$ is greater or equal to the $(1-\frac{\alpha}{2})$ quantile of a standard normal distribution. We only list the descriptive statistics of $\widehat{\lambda}_n$, in which the data, x_i and d_i are given.

We adopt the intraday high-frequency data to construct the statistic $\widehat{\lambda}_n$ to test whether the first-order Markov property is satisfied in the trading process. As our samples comprise 50 stocks and each stock contains about 49 trading days, the total number of $\widehat{\lambda}_n$ is ultimately 4828.[2] Table 1 reports the number rejecting the null at level $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.01$. The results show that the Markov hypothesis is mostly rejected, only suiting less than 19% and less than 23%, respectively, when the original and adjusted durations[3] are employed in the test at the 10% level.

We clearly reject the Markov property for the trade indicator variables on most trading days, indicating that previous trades play an important role in the information integration process.

As shown above, the structure, $(x_i - E[x_i|x_{i-1}])$ used in the MRR model may have lost some important information, that is, Equation (3) could not capture the permanent impact of the order flow innovation on the price. Furthermore, the parameter θ may not capture asymmetric information accurately. Therefore, an alternative structure should be explored not only to capture the permanent impact on price but also to facilitate model estimation.

3.3. An MA-MRR model

We are proposing a new model by extending the MRR model. We use $I_i = E[x_i|\Omega_{i-1}]$ to denote the full information structure in comparison with $E[x_i|x_{i-1}]$, under the Markov structure, where Ω_{i-1} represents the full information set until trade $i-1$. In the MRR model, $\theta(x_i - E[x_i|x_{i-1}])$ represents the part of the price change from asymmetric information or the change of beliefs due to order flow, so the parameter θ measures the degree of information or the so-called permanent impact of the order flow innovation. However, as analyzed in the above section, $(x_i - E[x_i|x_{i-1}])$ represents the unexpected information conditional on the latest trade, or in some degree just an approximation, only if the neglected part is small enough. Note that the unexpected surprise of the trade should be the conditional expectation given all past information. We use $x_i - I_i$ to substitute the $(x_i - E[x_i|x_{i-1}])$. Actually, $(x_i - E[x_i|x_{i-1}])$ provides the structure, making it an easier empirical estimation. For the structure of $x_i - I_i$, a suitable approximation has to be determined to ensure that it can be estimated reasonably.

In time-series analysis, exponentially weighted moving-average rules are common technical strategies for impounding historical information that is effective in forecasting. Therefore, we propose the following structure:

$$x_i - I_i = \theta_0(x_i - E[x_i|x_{i-1}]) + \theta_1(x_{i-1} - I_{i-1}), \tag{5}$$

where θ_0 and θ_1 are the parameters. The parameter θ_0 measures the information of the current state and the parameter θ_1 measures the historical information. Notice that the formula is an iterative structure, which could represent a kind of cumulative impact or a permanent impact.

With $(x_i - E[x_i|x_{i-1}])$ replaced by $x_i - I_i$ in Equation (3), we have

$$p_i - p_{i-1} = \varphi(x_i - x_{i-1}) + \theta(x_i - I_i) + \varepsilon_i + \xi_i - \xi_{i-1} \tag{6}$$

Substituting Equation (5) into Equation (6) and simplifying the parameters, we obtain

$$p_i - p_{i-1} = \varphi(x_i - x_{i-1}) + \tilde{\theta}_0(x_i - E[x_i|x_{i-1}] + \tilde{\theta}_1(x_{i-1} - I_{i-1})) + \varepsilon_i + \xi_i - \xi_{i-1}, \tag{7}$$

where $\tilde{\theta}_0 = \theta\theta_0$ and $\tilde{\theta}_1 = \frac{\theta_1}{\theta_0}$. If $\theta_0 = 1$, then $\tilde{\theta}_0 = \theta$ and $\tilde{\theta}_1 = \theta_1$ hold. Actually, from the perspective of estimating the parameters, Equation (5) is equivalent to this form, $x_i - I_i = (x_i - E[x_i|x_{i-1}]) + \theta_1(x_{i-1} - I_{i-1})$. Compared with Equation (4), $\tilde{\theta}_0$ as a modified version of θ measures the degree of information asymmetry, resulting in a permanent impact on price. $\tilde{\theta}_1$ as an information lag parameter accounts for the direction and the amount of the information left out by the MRR model.

It is necessary to estimate the unknown parameters for the application. In the MRR model, the parameters were estimated using the GMM. Although the GMM avoids requiring a fully parametric distribution of the residual, when new market characteristics are introduced into the MRR model, the GMM moment conditions need to be added to estimate the new parameters. Note that inappropriate new moment conditions may lead to erroneous estimation results. Therefore, we introduce an ML estimator with an ARCH structure.

Let $\Delta p_i = p_i - p_{i-1} - \varphi(x_i - x_{i-1}) - \tilde{\theta}_0(x_i - E[x_i|x_{i-1}] - \tilde{\theta}_1(x_{i-1} - I_{i-1})) = \varepsilon_i + \xi_i - \xi_{i-1}$ be the residual of the innovation of the trade prices in Equation (7). Then

$$E[\Delta p_i \Delta p_{i-1}] = E[(\varepsilon_i + \xi_i - \xi_{i-1})(\varepsilon_{i-1} + \xi_{i-1} - \xi_{i-2})] = -E\xi_{i-1}^2 \tag{8}$$

It can be verified that the Δp_i follows the MA(1) model and can be written as

[2] We divide each day into two parts, so the number of the $\widehat{\lambda}_n$ is two times of that of the trading days.
[3] the adjusted durations are obtained by eliminating the intraday pattern.

Table 2
Estimation results of the two model.

Panel A: Descriptive statistics						
	GMM for MRR		MLE for MRR		MLE for MA-MRR	
	θ	φ	θ	φ	$\tilde{\theta}_0$	φ
Mean	0.001515982	0.005165877	0.001520183	0.005165197	0.001051878	0.005307157
Max	0.03000628	0.03242293	0.03002076	0.03241331	0.02173876	0.03311298
Median	0.000559031	0.004610459	0.00056368	0.004617087	0.000401217	0.004704861
Min	4.4e-05	0.003938002	4.46e-05	0.003944034	3.18e-05	0.004155266

Panel B: The difference between the corresponding parameters of the two model						
Parameter	θ			φ		
	M1 to M2	M1 to M3	M2 to M3	M1 to M2	M1 to M3	M2 to M3
Absolute Error	4.84578e-06	0.00046410	0.00046830	4.13578e-06	0.00014589	0.00014709
Relative Error	0.008086639	0.4276726	0.4342362	0.000884136	0.02734282	0.02756596

The absolute error is the absolute value of the average/relative difference of the corresponding parameter, take the first column of parameter θ as an example, the absolute error is calculated by $MEAN(|\theta(\mathrm{M1}) - \theta(\mathrm{M2})|)$ and the relative error is calculated by $MEAN(|\theta(\mathrm{M1}) - \theta(\mathrm{M2})|/(\theta(\mathrm{M1}) + \theta(\mathrm{M2}))$. M1 (M2) represents the estimates from the original MRR model estimated by the GMM (MLE) method. M3 represents the estimates from the extended version estimated by the MLE method.

$$\Delta p_i = (1 - \psi B)e_i \tag{9}$$

where B is the backshift operator, ψ is the coefficient, and e_i is the further error. In a general setup, following the framework of autoregressive conditional heteroscedasticity (Engle, 1982), we have the error, sharing the ARCH(1) model as the first step.

$$e_i = \sigma_i \zeta_i, \sigma_i^2 = b_0 + b_1 e_{i-1}^2, \tag{10}$$

where $b_0 > 0$, $b_1 \geq 0$, ζ_i is an i.i.d. random variable. In application, it is appropriate to assume that ζ_i follows a normal distribution for ML or quasi-ML estimation.[4] A more general ARCH structure uses the GARCH rules of Bollerslev (1986).

With ζ_i following a normal distribution and combining Equations (9) and (10), the conditional log-likelihood function is

$$l(\Delta p_1, \Delta p_2, \cdots \Delta p_T|\Theta) = -\sum_{i=2}^{T}\left[\frac{1}{2}\ln \sigma_i^2 + \frac{1}{2}\frac{(\Delta p_i - \psi_i \Delta p_{i-1})^2}{\sigma_i^2}\right], \tag{11}$$

where Θ is the parameter vector. Maximizing l leads to the MLE of Θ.

4. Empirical results

In this section, we first estimate the parameters in the two MRR models, the original MRR and the extended version, with a detailed analysis and comparison, and then explore the information lag parameter.

4.1. Model comparison

For the original MRR model, we employ two methods to compare the parameters. The MLE method can be obtained from Equations (4) and (9)–(11), and the GMM method adopts moment conditions, as in Madhavan et al. (1997). Equations (7) and (9)–(11) provide details of the MLE method for the extended model. Table 2 lists the summary statistics of the parameters of adverse selection risk and liquidity cost estimated using the two methods. Because our sample comprised 50 stocks, each parameter contained 50 values. It is worth mentioning that all the estimated parameters are significant at the 1 percent level, which is not shown in this paper.

From Tables 2 and it is clear that the values of θ and φ estimated by the MLE are approximate to those estimated by the GMM in the original MRR model. From the perspective of the model comparison, the difference between the φ s is slight while the difference between the revised adverse selection risk $\tilde{\theta}_0$ and the adverse selection risk θ is obvious. Specifically, the bottom line in Panel B of Table 3 shows the difference between θ from the perspective of relative error. It is worth noting that the relative error of parameter θ greater than 40% is not acceptable in calculating.

Furthermore, we aimed to determine the statistical properties of these differences, particularly if the distributions of these parameters were identical. Further details may be obtained by examining and testing the cumulative distribution of the estimated variables.

[4] We also adopt a heavy-tailed distribution such as a standardized Student distribution and the results are similar.

Table 3
Kruskal-Wallis Tests on parameters.

Panel A: Triple Comparison						
Parameter	θ			φ		
Test Statistic	4.9097			3.5715		
P-value	0.08588			0.1677		
Panel B: Pairwise Comparison						
Parameter	θ			φ		
	M1 to M2	M1 to M3	M2 to M3	M1 to M2	M1 to M3	M2 to M3
Test Statistic	0.0068436	3.631	3.6995	0.0007604	2.6469	2.692
P-value	0.9341	0.05671	0.05443	0.978	0.1037	0.1009

The test statistic is used to test the null hypothesis that the parameter value for all three volume samples is drawn from identical populations versus the alternative hypothesis that at least one of the populations tends to furnish greater observed values than the other populations. The parameter θ includes the adverse selection risk estimated by GMM and MLE in the original MRR model and the revised adverse selection risk estimated by MLE in the extended version. φ is the liquidity cost.
The test statistic is used to test the null hypothesis that two samples are drawn from identical populations against the alternative that their distributions are different. M1 (M2) represents the estimates from the original MRR model estimated by the GMM (MLE) method. M3 represents the estimates from the extended version estimated by the MLE method.

To compare these distributions, we used the Kruskal-Wallis test, which determines whether the compared population distribution functions are identical. Specifically, we tested whether one of the three populations differed from the others. Table 3 lists the test statistics.

The Kruskal-Wallis test in Panel A of Table 3 shows that the null hypothesis about the adverse selection parameter θ is rejected with a test statistic of 0.08588, which is well above the 0.1 confidence level. The test statistic φ is 0.1677, suggesting no significant differences in the liquidity cost estimates. In Table 3, panel B, the test statistics of θ and φ in the comparison of "M1 to M2" indicate that there is not a significant difference between the two estimate methods. However, the test statistic of the adverse selection parameter θ in the comparisons of "M1 to M3" and "M2 to M3" verifies the fact that the difference between the revised adverse selection risk $\tilde{\theta}_0$ and the adverse selection risk θ is significant, attributed to the different structures of the models. Meanwhile, the effect of different structures of the models on the estimates φ is not significant with the test statistics of values 0.1037 and 0.1009 in the comparisons of "M1 to M3" and "M2 to M3," respectively.

Based on the analysis of the non-Markov property of the trade indicator variables and the Kruskal-Wallis tests, the variability in the estimates suggests that the information lag component plays an important role in revealing significant differences in the estimates of adverse selection risk.

4.2. Economic implication of the model

Our model emphasizes the contradiction between MRR model assumptions and the non-Markovian nature of real-world data. We then account for delayed information in the MA-MRR model and find that the liquidity parameter barely changes, whereas the adverse selection cost parameter exhibits significant differences.

Currently, there is no universally accepted metric for assessing adverse selection risks. The MRR model deviated from the Markov structure, and Zhang et al. (2015) established that this inflates the estimates of adverse selection risk, lending substantiation to our conclusions. Moreover, from a statistical standpoint, we validated the notable dissimilarities in the estimation of parameters for adverse selection risk.

Empirical evidence confirms a significant discrepancy in adverse selection risk between the two models based on the consideration of a delayed information term. In our study, quarterly data were used for parameter estimation, and the discrepancies between the two models were analyzed. We confirm that significant discrepancies (or errors) may arise in parameter estimation when the actual data fail to satisfy the Markov property.

4.3. Insight into the information lag parameter

We now focus on the $\tilde{\theta}_1$ in Equation (7), which is the information lag parameter. The empirical results show that all $\tilde{\theta}_1$ values are positive, ranging from 0.1069351 to 0.459482. Fig. 1 shows detailed estimates for the 50 stocks. It is notable that all estimates are significant at the 1 percent level. The maximum was 0.459482, occurring at the 44th stock, and the minimum was 0.1069351, occurring at the 8th stock. The mean and median are 0.2275045 and 0.2339847, respectively, as represented by the dot-dashed and solid lines in Fig. 1.

Our concern in this study is to gain insight into the information lag parameter rather than why the estimates differ according to stock. We remove the information part $\tilde{\theta}_0(x_i - E[x_i|x_{i-1}] + \tilde{\theta}_1(x_{i-1} - I_{i-1}))$ from Equation (7)and explore the meaning of the parameter $\tilde{\theta}_1$ from the perspective of information integration into the price.

Consider one unit information shock in the first trade, that is $x_1 - I_1 = x_1 - E[x_1|x_0] = x_1 - E[x_1]$. Then the information shock in the second trade can be expressed as $x_2 - I_2 = (x_2 - E[x_2|x_1]) + \tilde{\theta}_1(x_1 - I_1)$, from which it can be deduced that a

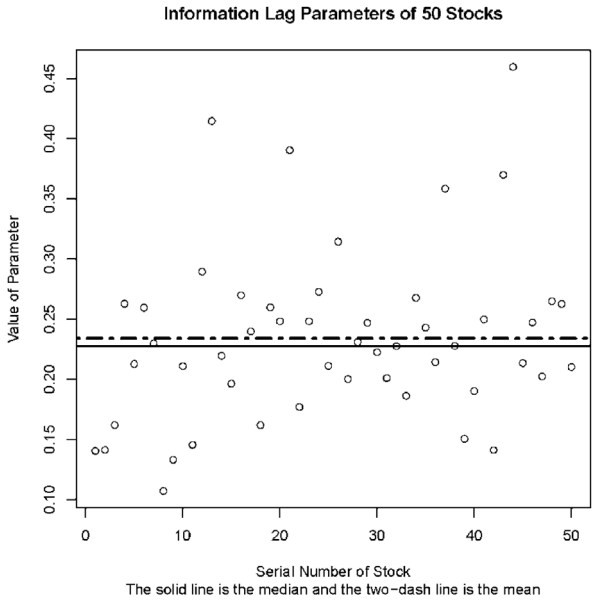

Fig. 1. The values of information lag parameters of 50 stocks.

proportion of the information in the first trade, about $\tilde{\theta}_1$, lags behind the second trade. In other words, only a proportion about $1-\tilde{\theta}_1$ of the first shock is integrated into the price at the first trade. From this, it is easy to see that the cumulative proportion of the first shock integrated into the price at the n th trade is $1-\tilde{\theta}_1{}^n$. Obviously, if $|\tilde{\theta}_1|<1$, the shock information will be integrated to a full level when n approximates to infinity. So the value of $\tilde{\theta}_1$ can be seen as an index to measure the speed of information integration. The smaller the value is, the faster the information integrates into the price in each trade. Fig. 2 intuitively illustrates the information integration processes according to the corresponding $\tilde{\theta}_1$ values.

In Fig. 2, the horizontal axis represents the transaction time from the first trade to the 20th trade. The vertical axis represents the proportion of cumulative information incorporated into the prices. Although the velocities corresponding to $\tilde{\theta}_1$ are dramatically different, all the curves of the cumulative information integrated in the price cross the 99 percent line, approximating a full level after seven trades. Furthermore, from the perspective of the average effect (the mean of $\tilde{\theta}_1$ is 0.23), the curve with $\tilde{\theta}_1=0.30$ shows that more than 99 percent of the information about the shock information is integrated into the price after the first five trades. Hasbrouck (1991) employs an impulse response function to explore how long one unit of information shock is integrated into a price. Their results indicate that it takes approximately 20 trades to reach a fully

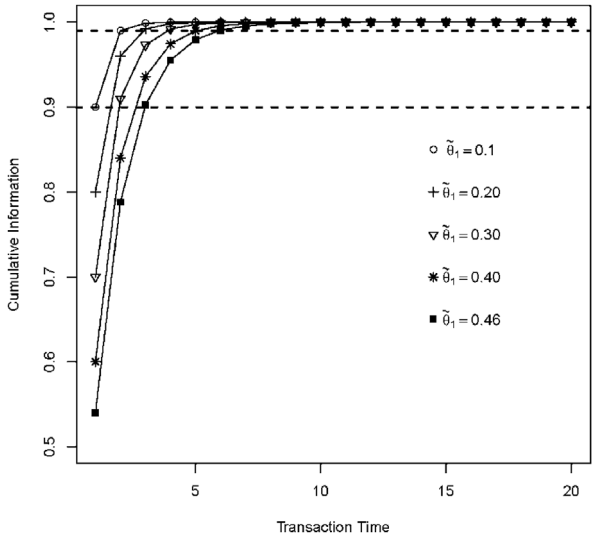

Fig. 2. Information integration processes according to the corresponding $\tilde{\theta}_1$ values.

incorporated level, and that the first five trades are critical, in which most of the information shock is integrated into the price. He suggests that a VAR structure with five lags is well suited for empirical studies. From the perspective of the total transaction time and average information integration speed of all stocks, our empirical results are consistent with those of Hasbrouck (1991) and Dufour and Engle (2000). However, from the perspective of individual stocks, a VAR structure with five lags may be insufficient or redundant. For example, the bottom curve in Fig. 2, in the case of $\tilde{\theta}_1 = 0.46$, shows that seven trades are required to reach a 99 percent level, whereas the top curve in Fig. 2, in the case of $\tilde{\theta}_1 = 0.10$, shows that only two trades are required to reach a 99 percent level. Our data come from the most frequent stock; otherwise, it will take at least 18 trades to reach a 99% confidence level when the estimate $\tilde{\theta}_1$ is 0.80 as example. Therefore, the information lag component can provide a useful reference for empirical studies on transaction-level analysis.

5. Conclusion

This study presents an MA-MRR model that provides a new approach to measuring adverse selection risk that seems particularly well suited to financial data. The lag component in the model captures the speed of information integration as an indicator of market quality, which is closely related to market liquidity and price informativeness.

When the data do not satisfy the Markov property, a moving average process often makes the model more reasonable. Our tests reject the assumption of the first-order Markov property of the trading indicator variables for most trading days. Therefore, the empirical application of the adverse selection risk estimated by the MRR is questionable because the first-order Markov property assumed in this model is doubtful. We further show that the difference between the adverse selection risk of the MRR model and the corresponding parameter of the extended version is significant because there is an information lag in the trading process. This new estimate of the adverse selection risk of the extended model provides a more accurate measure of asymmetric information. Moreover, the speed of information integration by the information lag component in the decomposed structure can provide useful clues for other empirical studies.

Ethics statement

Not applicable because this work does not involve the use of animal or human subjects.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant number: 71771008), Science and Technology Support Plan of Guizhou (Grant No. 2023-221) and the Funds for the First-class Discipline Construction (XK 1802-5).

References

Ahern, K. R. (2014). Do common stock have perfect substitutes? Product market competition and the elasticity of demand for stock. *Rev. Econ. Stat., 96*(4), 756–766.
Ahn, H. J., Cai, Jun, Hamao, Yasushi, & Ho, Richard Y. K. (2002). The components of the bid-ask spread in a limit-order market: evidence from the Tokyo stock exchange. *J. Empir. Finance, 9*, 399–430.
Ahn, H. J., Kang, J., & Ryu, D. (2008). Informed trading in the index option market: the case of KOSPI 200 options. *J. Futures Mark., 28*, 1118–1146.
Andros, G. (2015). Market quality of dealer versus hybrid markets for illiquid securities: new evidence from the FTSE AIM Index. *Eur. J. Finance, 21*(6), 466–485.
Angelidis, T., & Benos, A. (2009). The components of the bid-ask spread: the case of the athens stock exchange. *Eur. Financ. Manag., 15*, 112–144.
Armstrong, C., Core, J., Taylor, D., & Verrecchia, R. (2011). When does information asymmetry affect the cost of capital? *J. Account. Res., 49*(1), 1–40.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econom., 31*(3), 307–327.
Bushee, B. J., Gow, I. D., & Taylor, D. J. (2018). Linguistic complexity in firm disclosures: obfuscation or information? *J. Account. Res., 56*(1), 85–121.
Chen, Y.-L., & Gau, Y.-F. (2014). Asymmetric responses of ask and bid quotes to information in the foreign exchange market. *J. Bank. Finance, 38*, 194–204.
Chen, B., & Hong, Y. (2012). Testing for the Markov property in time series. *Econom. Theor., 28*, 130–178.
de Matos, J. A., & Fernandes, M. (2007). Testing the Markov property with high frequency data. *J. Econom., 141*(1), 44–64.
Dufour, A., & Engle, R. F. (2000). Time and the price impact of a trade. *J. Finance, 55*, 2467–2498.
Dyhrberg, A. H., Foley, S., & Svec, J. (2018). How investible is Bitcoin? Analyzing the liquidity and transaction costs of Bitcoin markets. *Econ. Lett., 171*, 140–143.
Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*, 987–1008.
Fernandez-Perez, A., Frijns, B., Indriawan, I., & Tourani-Rad, A. (2019). Surprise and dispersion: informational impact of USDA announcements. *Agric. Econ., 50*(1), 113–126.
Frijns, B., & Tse, Y. M. (2015). The informativeness of trades and quotes in the FTSE 100 index futures market. *J. Futures Mark., 35*(2), 105–126.
Glosten, L., & Milgrom, P. (1985). Bid, ask, and transaction prices in a specialist market with heterogeneously informed agents. *J. Financ. Econ., 14*, 71–100.
Green, T. C. (2004). Economic news and the impact of trading on bond prices. *J. Finance, 59*(3), 1201–1233.
Gregoriou, A., & Rhodes, M. (2017). The accuracy of spread decomposition models in capturing informed trades. *Rev. Behav. Finance, 9*(1), 2–13.

Han, S., & Zhou, X. (2014). Informed bond trading, corporate yield spreads, and corporate default prediction. *Manag. Sci., 60*(3), 675–694.
Hasbrouck, J. (1991). The summary informativeness of stock trades: an econometric analysis. *Rev. Financ. Stud., 4*, 571–595.
Lai, S., Ng, L., & Zhang, B. (2014). Does PIN affect equity prices around the world? *J. Financ. Econ., 114*(1), 178–195.
Madhavan, A., Richardson, M., & Roomans, M. (1997). Why do security prices change? A transaction-level analysis of nyse stocks. *Rev. Financ. Stud., 10*, 1035–1064.
Medina, V., Pardo, A., & Pascual, R. (2014). The timeline of trading frictions in the European carbon market. *Energy Econ., 42*, 378–394.
Mizrach, B., & Otsubo, Y. (2014). The market microstructure of the European climate exchange. *J. Bank. Finance, 39*, 107–116.
Riordan, R., Storkenmarier, A., & Wagener, M. (2013). Public information arrival: price discovery and liquidity in electronic limit order markets. *J. Bank. Finance, 37*, 1148–1159.
Sakawa, H., & Ubukata, M. (2014). Watanabel, Naoki, Market liquidity and bank-dominated corporate governance: evidence from Japan. *Int. Rev. Econ. Finance, 31*, 1–11.
Sita, B. B., & Westerholm, P. J. (2011). The role of trading intensity estimating the implicit bid-ask spread and determining transitory effects. *Int. Rev. Financ. Anal., 20*, 306–310.
Souza, I. V. M., Reisen, V. A., & Franco, G. D. C. (2018). The estimation and testing of the cointegration order based on the frequency domain. *J. Bus. Econ. Stat., 36*, 695–704.
Zhang, Y. (2015). The securitization of gold and its potential impact on gold stocks. *J. Bank. Finance, 58*, 309–326.
Zhang, Q., Liu, S. C., & Qiu, W. H. (2015). Does MRR model overestimate the information risk: empirical results from Chinese data of SSE. *J.Manag. Sci. China, 18*, 62–72.