

Determinants of bid and ask quotes and implications for the cost of trading[☆]

Michael Yuanjie Zhang^a, Jeffrey R. Russell^{b,*}, Ruey S. Tsay^b

^a Barr Rosenberg Research Center, AXA Rosenberg Group, 4 Orinda Way Bldg E, Orinda, CA 94563, USA

^b Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA

Received 1 September 2005; accepted 11 December 2007

Available online 11 February 2008

Abstract

Financial transaction costs are time varying. This paper proposes a model that relates transaction cost to characteristics of order flow. We obtain qualitatively consistent model results for different stocks and across different time periods. We find that an unusual excess of buyers (sellers) relative to sellers (buyers) tends to increase the ask (bid) price. Hence, the ask and bid components of spread change asymmetrically about the efficient price. For a fixed order imbalance surprise these effects are muted when unanticipated total volume is high. Unexpected high volatility in the transaction price process tends to widen the spread symmetrically about the efficient price. Our findings are consistent with predications from market microstructure theory that the cost of market making should depend on both the risk of trading with better-informed traders and inventory risk. We also find that order flow surprises have a significant impact on the efficient price and can also explain a substantial amount of persistence in the volatility of the efficient price. This dependence does not violate the efficient market hypothesis since the surprises, by definition, are not predictable.

© 2008 Published by Elsevier B.V.

JEL classification: G23; D82; C15

Keywords: Trading costs; Liquidity; Price dynamics; Market microstructure

1. Introduction

A key measure of the efficiency of a financial market making structure is the cost of trading. Most markets offer separate prices for buyers and sellers. A market's operational efficiency can therefore be measured by its bid/ask

[☆] The authors would like to thank Robert Engle, Joel Hasbrouck, Hong Li, Tee Lim, Per Mykland, Rudi Schadt, Neil Shephard, Pietro Veronesi, Kevin Wang and Li Wei for their suggestions. We specially thank Joel Hasbrouck for sharing his maximum likelihood estimation code for the benchmark Hasbrouck model in this paper. The first author is grateful to the University of Chicago Fellowship, the Oscar Mayer Fellowship and his colleagues at Barr Rosenberg Research Center.

* Corresponding author.

E-mail address: jeffrey.russell@chicagogsb.edu (J.R. Russell).

spread. As a result, there has been substantial interest in both theoretical and empirical modeling of the behavior of bid–ask spreads. The theoretical literature identifies three main factors that determine spread: inventory carrying costs (Amihud and Mendelson, 1980; Ho and Stoll, 1983; O'Hara and Oldfield, 1986), adverse selection costs (Kyle, 1985; Glosten and Milgrom, 1985; Admati and Peiderer, 1988; Easley and O'Hara, 1992) and order processing costs (Brock and Kleidon, 1992). Among those studies, researchers explore the cross-sectional and time series relationship between spread and volume (Easley and O'Hara 1992; Harris and Raviv 1993; Lee, Mucklow, and Ready 1993). In addition to its relationship with spreads and transaction costs, volume also has close link to market price variability (Gallant, Rossi, and Tauchen 1992; Campbell, Grossman, and Wang 1993; Chan and Fong 2000).

Although markets contain both buyers and sellers, an implicit assumption of “symmetry” is often imposed in the microstructure literature when order flow characteristics such as spreads and volumes are related. The impact of buyer induced and seller induced components is assumed to have a common impact on price setting. We decompose the spread into two components: the cost of buy exposure and the cost of sell exposure. We test and find evidence that the effect of volume on these components is not symmetric. Hence the classic studies which do not account for the asymmetries are missing part of the picture.

Our methodology is closely related to Hasbrouck (1999a). We model the bid and ask prices as deviations from an unobserved efficient price. The efficient price is defined as the price that would prevail in equilibrium in the absence of market frictions and is, consistent with market microstructure theory, assumed to follow a heteroskedastic random walk. The deviations of the bid and ask prices are modeled as functions of past order flows. Since the bid and ask deviations are modeled separately, we can test for symmetry in the response of the bid and ask to order flow characteristics. To the best of our knowledge, this is the first paper to perform such tests.

Information based market microstructure theory focuses on how the market evolves following the arrival of new information, especially, how new information is absorbed into market prices. Hasbrouck (1991) points out that innovations to trade order flows are the ultimate meaningful measure of information effect since innovations exclude the predictable portion of the order flow that conveys no new information. It is therefore natural to separate innovations from market variables. We decompose each market variable into anticipated and unanticipated components and test for the roles the two components play in determining trading cost and efficient price. We find that market variables such as buy/sell volumes, transaction price variance, and spread affect market making costs and the efficient price volatility only through their unanticipated components. We therefore find that surprises or shocks to these variables are more important determinants of price setting behavior than their expected components.

We have two important findings. First, by making distinctions between buyer-initiated and seller-initiated volume and decomposing the two types of volume into anticipated and unanticipated components, we find that the ask and bid components of spread change asymmetrically about the efficient price as a function of the order flow. An unusual excess of buyers relative to sellers tends to increase the ask price more than the bid price, and an unusual excess of sellers has an opposite effect. For a fixed order imbalance surprise (Pressure), these effects are muted when the unanticipated total volume (Level) is high. The impact of expected volume on trading cost is not significant for the stocks analyzed, indicating that effects of the expected component of volume is already impounded in the cost functions. Second, the dynamics of the cost of ask exposure process and the cost of bid exposure process are distinct and hence asymmetric. Except that unexpected high volatility in the transaction price process tends to widen the spread symmetrically about the efficient price, the impacts of other market variables on market making costs are asymmetric. Not surprisingly, we also find that the cost of market making is lower when the expected depth is higher. Our findings are consistent with predications from market microstructure theory that the cost of market making should depend on both the risk of trading with better-informed traders and inventory risk.

A cornerstone of market microstructure theory is that in an asymmetric information environment uninformed agents learn about private information from order flow (O'Hara, 1995). The dynamics of the efficient market price are therefore allowed to depend on the contemporaneous surprise to buyer-initiated (seller-initiated) volume, which has a significant positive (negative) impact on the mean dynamics of the efficient market price. The volatility of the underlying efficient market price is highly persistent in the estimated EGARCH model, but most of this persistence can be explained by the added market variables (bid–ask spreads, buy/sell volumes, and their lags; quoted depths). The persistence of the ARCH effects is significantly reduced after incorporating market variables. This is consistent with the story that the extent of information asymmetry in the market serves as an important source of the volatility clustering observed in the financial market.

The rest of the paper is organized as follows. Section 2 presents the specification of the proposed model. In Section 3 we describe the data used in this study, the selection of market variables, and the decomposition of market variables into anticipated and unanticipated components. We present our detailed empirical analysis for different stocks and across different time periods in Section 4. We obtain qualitatively consistent estimation results and examine closely the practical implications of the results on the cost of trading and efficient market price via linking the estimation results with existing empirical and theoretical market microstructure studies available in the literature. Section 5 concludes.

2. The model

2.1. Background issues

The focus of this paper is to understand the dynamics of intraday trading costs and the efficient market price through econometric modeling of the bid and ask quote prices in the New York Stock Exchange (NYSE). There are three fundamental issues that need to be addressed: price discreteness, intraday seasonality and time-varying parameters.

Stock prices are constrained to a discrete grid due to institutional arrangements at NYSE. The tick size over the main sampling period of this study (January 3, 2006–March 31, 2006) is one hundredth of a dollar.¹ Hence, the observed bid and ask quote data, $\{b_t, a_t\}$, are the bid and ask prices in integer values with a unit of one tick or one hundredth of a dollar. This discreteness cannot be ignored since the tick size is big relative to the quantities of interest, namely the bid–ask spreads. Predictions and maximum likelihood estimations are also affected by discreteness. In the micro-structure literature, one way to handle this discreteness issue is considering the discreteness as generated from underlying continuous variables, such as the efficient market price and market making cost(s), through some rounding mechanism.²

The deterministic intraday seasonality is a salient feature of intraday market variables such as bid–ask spreads, trading volume, and volatility.³ A model of these market variables must explicitly address this intraday seasonal pattern. We therefore not only need to consider the deterministic patterns in the latent processes driving the bid and ask quotes, but also should properly address the intraday seasonalities in the market variables used.

Lastly, we model deviations of the observed bid and ask prices from the efficient price as a stochastic process. This allows us to assess the impact of observable market characteristics on the price setting behavior of the specialist.

2.2. Model specification

Our model has three key components: (1) an asymmetric rounding mechanism that generates discrete bid and ask quotes from latent continuous quote exposure costs and efficient market price; (2) a submodel for the costs of market making which allows for both deterministic variation and stochastic variation in the costs; (3) a submodel for efficient market price that features a random walk mean dynamics with contemporaneous surprise items and an EGARCH specification of Nelson (1991) for the efficient price's volatility. In order to understand the determinants of market liquidity and price dynamics, our model also allows for adding observable market variables as regressors to the equations for the costs of market making, and the mean and volatility of efficient market price. We discuss the specification of the model along the lines of these three components.

Denote by m_t the implicit efficient price of the security under study at time t , that is, the expectation of the security's terminal value, conditional on all public information. The agent establishing the bid quote is assumed to be subject to a nonnegative cost of quote exposure $\beta_t > 0$ for small trades. The cost is assumed to be composed of fixed transaction costs and asymmetric information costs. With a one-unit tick size in the market, the agent quotes a bid price that is rounded down to the next lowest grid point: $b_t = \text{Floor}(m_t - \beta_t) = \lfloor m_t - \beta_t \rfloor$. Similarly, constrained by discreteness, the agent offers an ask quote which is rounded up to the next highest grid point: $a_t = \text{Ceiling}(m_t + \alpha_t) = \lceil m_t + \alpha_t \rceil$, where $\alpha_t > 0$ is the quote exposure cost of the agent. The asymmetric rounding of the bid and ask quotes allows the agent to avoid the

¹ The tick size has no effects on the conclusion of this paper. See Section 3.

² See Gottlieb and Kalay (1985), Ball (1988), Harris (1990) and Hasbrouck (1999b).

³ See McNish and Wood (1992), Lee et al. (1993), Jain and Joh (1988), Foster and Viswanathan (1990) and Stoll and Whaley (1990). For a related complete review, please see Goodhart and O'Hara (1997).

possibility of loss on the incoming trade. The rounding mechanism also avoids possible degenerate quotes. Hence, at time t , the observed discrete bid and ask quotes $\{b_t, a_t\}$ are generated by

$$a_t = \lceil m_t + \alpha_t \rceil \quad (1)$$

$$b_t = \lfloor m_t - \beta_t \rfloor. \quad (2)$$

We now specify the dynamics for the latent costs of quote exposure $\{\alpha_t, \beta_t\}$ and the efficient price m_t . The submodels for the latent bid and ask quote exposure costs $\{\alpha_t, \beta_t\}$ are specified as

$$\ln(\alpha_t) = \mu_t + \theta_t + \theta_t^\alpha + \phi[\ln(\alpha_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\alpha \quad (3)$$

$$\ln(\beta_t) = \mu_t + \theta_t + \theta_t^\beta + \phi[\ln(\beta_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\beta \quad (4)$$

where $\ln(\alpha_t)$ and $\ln(\beta_t)$ follow separate autoregressive processes with a common autoregressive parameter ϕ , a common time-varying deterministic mean μ_t , common and distinct market variable components θ_t , θ_t^α , and θ_t^β , and independently distributed standard normal innovations v_t^α and v_t^β .

The deterministic component μ_t of the costs is specified as a combination of exponential decay functions to allow for the intraday “U-shape” effect,

$$\mu_t = k_1 + k_2^{\text{open}} \exp(-k_3^{\text{open}} \tau_t^{\text{open}}) + k_2^{\text{close}} \exp(-k_3^{\text{close}} \tau_t^{\text{close}}) \quad (5)$$

where τ_t^{open} is the elapsed time since the opening quote of the day (in hours) and τ_t^{close} is the time remaining before the scheduled market close (in hours).

There are three components associated with market variables. $\theta_t = X_{t-1}^\theta \mathbf{d}$ is a common market variable component to α_t and β_t , where X_{t-1}^θ is a row vector of market variables observed before time t and $\mathbf{d}$ is the corresponding column vector of coefficients. $\theta_t^\alpha = X_{t-1}^\alpha \mathbf{d}_\alpha + Z_{t-1}^\alpha \mathbf{g}$ and $\theta_t^\beta = X_{t-1}^\beta \mathbf{d}_\beta + Z_{t-1}^\beta \mathbf{g}$ are an α_t -specific and a β_t -specific market variable components respectively, where X_{t-1}^α and Z_{t-1}^α are two row vectors of market variables which are informative about the ask side of the market making costs, and X_{t-1}^β and Z_{t-1}^β are that for the bid side of the market making costs. X_{t-1}^α and X_{t-1}^β can have both shared and distinct determinants. $\mathbf{d}_\alpha$ and $\mathbf{d}_\beta$ are the corresponding coefficient vectors. The market variable vectors Z_{t-1}^α and Z_{t-1}^β are constrained to have the same coefficient vector $\mathbf{g}$.

The specification allows for deterministic variation in the mean, a persistent stochastic component, and common and distinct market variable components. We will refer these submodels as the cost models.

The logarithm of the efficient price m_t is specified as a random walk process. We include contemporaneous shocks to market variables in the random walk to proxy the price update due to new information.⁴ Let δ_t be a function of shocks to market variables with $\mathbb{E}_{t-1}[\delta_t] = 0$, that is, δ_t is not predictable at $t-1$. Specifically, we use a linear form that $\delta_t = X_t^m \mathbf{c}_m$, where X_t^m are a vector of shocks to market variables. Hence, we have

$$\ln(m_t) = \ln(m_{t-1}) + \delta_t + \sigma_\epsilon \epsilon_t \quad (6)$$

where the innovations $\sigma_\epsilon \epsilon_t$ represent price updates in response to the public information. The component $\delta_t = X_{t-1}^m \mathbf{c}_m$ allows price updates to respond to new and private information.

The conditional variance of the price innovation is time varying with an EGARCH specification

$$\ln(\sigma_t^2) = \eta_t + \zeta_t + \psi [\ln(\sigma_{t-1}^2) - \eta_{t-1}] + \omega \epsilon_{t-1} + \gamma [|\epsilon_{t-1}| - E|\epsilon_{t-1}|] \quad (7)$$

where ϵ_t is distributed as the generalized error distribution (GED) to allow for fat tails, and $\omega \epsilon_{t-1}$ is an asymmetric component to allow the conditional volatility to respond asymmetrically to rises and falls in efficient market price.⁵ The model also allows the tail-thickness parameter of the GED distribution to be different for intraday intervals and overnight intervals to capture the potential overnight effect on stock price.

⁴ We thank Robert Engle for the suggestion.

⁵ Hasbrouck omits the asymmetry term used in the original Nelson (1991) specification of EGARCH. We add this term back in our specification to test for the asymmetry effect in intraday data.

The deterministic variation in $\ln(\sigma_t^2)$ is represented by η_t and specified in a similar manner as that for the cost process

$$\eta_t = \eta^{\text{night}} 1_{\{\tau_t^{\text{open}}=0\}} + l_1 + l_2^{\text{open}} \exp(-l_3^{\text{open}} \tau_t^{\text{open}}) + l_2^{\text{close}} \exp(-l_3^{\text{close}} \tau_t^{\text{close}}). \quad (8)$$

For the added market variable component $\zeta_t = \mathbf{X}_{t-1}^{\zeta} \mathbf{c}$, $\mathbf{X}_{t-1}^{\zeta}$ is a vector of market variables observed before t and $\mathbf{c}$ is the corresponding vector of coefficients. We refer this submodel as the EGARCH/GED model. The EGARCH specification for the efficient price volatility can facilitate the understanding of the interaction between the changing risk environment and market behavior.

2.3. Model estimation

The model is inherently difficult to estimate since, in addition to the latent market making costs and the implicit efficient price, the volatility of the unobserved efficient price process is heteroskedastic and follows an EGARCH process. Hasbrouck proposed estimating this type of model using the non-Gaussian nonlinear state space approach of Kitagawa (1987) since the model can be cast in a state space form. But as the method relies on multi-dimensional numerical integration, it is hard to expand the model, for example, by adding exogenous variables to the cost model and/or the EGARCH/GED model.

We choose Markov chain Monte Carlo (MCMC) method to estimate our model. MCMC method easily overcomes the difficulties, such as the curse of dimensionality, faced by traditional methods such as numerical integration and state space approach and allows for adequate flexibility in a manageable manner, since the computational time using MCMC is linear in the number of parameters. We present the detailed implementations of the MCMC method in an Appendix.

Manrique and Shephard (1997) estimate a simple version of the Hasbrouck model with constant volatility using MCMC methods. Hasbrouck (1999a) also uses an MCMC method to estimate a simple version of the model, but with the added feature of allowing price clustering (the tendency of quotes to lie on ‘natural’ multiples of the tick size) in the model. The Hasbrouck model can also be cast as a limited dependent process. Manrique and Shephard (1998) discuss general issues in the application of MCMC to limited dependent variable time series analysis. Jacquier et al. (1994) and Kim et al. (1998) discuss the application of MCMC in stochastic volatility model. Ball and Chordia (2001) also estimate a model of bid and ask quotes using MCMC methods. They model the efficient market price and true spreads as a bivariate process. Our proposed model is the first to include market information beyond that of lagged dependent variables.

3. Data and market variables

3.1. Data

We use the TAQ dataset from New York Stock Exchange (NYSE). The dataset contains intraday transactions data (trades and quotes) for securities listed on NYSE. Trade price, trade size and trade time are extracted from the transaction records. Price quotes including bid price, ask price, bid depth and ask depth are obtained from the quotes record.⁶ We use the bid and ask quotes prevailing at the close of 15-minute intervals.⁷ Generally, there are 26 observations per day starting from 9:45 to 16:00. Our analysis focuses on three large cap stocks, General Electric (GE), JP Morgan Chase (JPM) and Exxon Mobil (XOM), for the three month period of January 3, 2006 to March 31, 2006. We obtain 1612 observations for each stock for the three month period.⁸

Table 1 shows summary statistics for changes in the bid and ask prices for the three stocks for the period of January 3, 2006 to March 31, 2006. The range of the bid and ask changes are large, but most of the intraday changes are within a

⁶ Only the quotes issued by the primary (NYSE) specialist are used, since the information based theory concerning quote/price revision is more appropriate in a single specialist setting.

⁷ The 15-minute interval is chosen to balance the microstructure interest and the computational tractability.

⁸ In NYSE, the tick size was changed from \$1/8 to \$1/16 on June 24 1997 then to decimal (\$1/100) on January 29, 2001. We also estimate the proposed model for GE stock using two three-month periods from November 1, 1990 to January 31, 1991 and from January 4, 1999 to March 31, 1999 for tick size regimes of \$1/8 and \$1/16 respectively and obtain qualitatively consistent model results.

Table 1

Descriptive statistics for the bid and ask changes of GE, JPM and XOM (January 3, 2006–March 31, 2006)

	Bid changes		Ask changes	
	Intraday	Overnight	Intraday	Overnight
<i>GE</i>				
Num obs	1550	61	1550	61
Min.	−21	−52	−20	−53
1st Qu.	−3	−9	−3	−10
Median	0	−1	0	−2
Mean	0	−0	0	−0
3rd Qu.	3	10	3	9
Max.	24	32	25	33
Std Dev	4.651	15.204	4.652	15.258
Skew	0.116	−0.489	0.153	−0.460
Ex kurtosis	1.796	1.033	1.789	1.231
LjungBox(15)	8.739	12.278	8.334	11.296
<i>JPM</i>				
Num obs	1550	61	1550	61
Min.	−36	−39	−35	−40
1st Qu.	−4	−13	−4	−12
Median	0	1	0	−1
Mean	0	−0	0	−1
3rd Qu.	4	13	4	12
Max.	32	46	33	46
Std Dev	7.349	18.965	7.310	19.208
Skew	0.107	0.066	0.196	0.099
Ex kurtosis	2.151	−0.236	1.981	−0.132
LjungBox(15)	23.041	11.406	23.417	10.334
<i>XOM</i>				
Num obs	1550	61	1550	61
Min.	−75	−66	−68	−65
1st Qu.	−7	−19	−7	−19
Median	0	6	0	6
Mean	−0	7	−0	6
3rd Qu.	7	26	7	26
Max.	62	200	63	203
Std Dev	13.260	40.433	13.191	41.118
Skew	−0.188	1.655	−0.162	1.675
Ex kurtosis	2.488	6.399	2.478	6.361
LjungBox(15)	23.325	21.492	22.780	22.047

This table summarizes the 15-minute bid and ask changes of GE, JPM and XOM for the three month period from January 3, 2006 to March 31, 2006. The bid and ask quote prices are in the unit of one tick (\$1/100). Only the quotes issued by the primary (NYSE) specialist are used.

couple of ticks. Changes are more dramatic for overnight intervals, as it is indicated by the significantly higher mean and standard deviation. Clearly, the intraday observations and the overnight observations have distinct distributions.

We identify each trade as buyer or seller-initiated by comparing the trade price to the prevailing bid and ask quotes at the time of the transaction. A transaction is identified as buyer-initiated and signed positively if the transaction price is above the prevailing quote midpoint and is classified as seller-initiated and signed negatively if below. Since transactions are usually recorded with a delay relative to quotes, we use a technique developed by [Lee and Ready \(1991\)](#) to match transaction with market prevailing bid and ask prices. The bid and ask prices must be at least more than 5 s before the transaction.

3.2. Market variables

For the cost model, there are two types of economic variables included: (1) variables that are related to the direction of the trade and thus potentially have different effects on the bid side and the ask side of the market making costs, for

example, signed volume, bid depth and ask depth; (2) variables that have general indication of the market sentiment without a description about the direction of trade, for example, volatility. Previous studies have only addressed the issue of market making costs in the level of bid–ask spreads. Lee et al. (1993) document a positive correlation between spread and trading volume and a negative correlation between spread and depth. High volatility is indicative of market uncertainty and possibly an increased presence of informed traders so that a risk averse specialist will widen the spread. Hence volatility and spread should be positively correlated. Since we model deviations of bid and ask quotes from the efficient price, we can study both the common dynamics and the distinct dynamics for the two components of spread, the bid side and the ask side of the market making costs.

Trading volume is arguably the most important variable to consider for the EGARCH/GED model of the efficient price since there is a robust positive relation between price volatility and trading volume. Theoretical models proposed to explain this relation include asymmetric information models (Kyle, 1985; Glosten and Milgrom, 1985; Admati and Peiderer, 1988) and mixture of distributions models (Epps and Epps, 1976; Tauchen and Pitts, 1983; Andersen, 1996). Recently, Chan and Fong (2000) provide evidence that order imbalance is also closely related to price volatility.

Market variables are constructed using all the data during each of the 15-minute interval.⁹ Since the processes for market making costs and price volatility are specified in logarithm form, it is more natural to use the log-transformations of those market variables in the model. Here is a list of the market variables we use with a brief definition:

LogSpread	log(mean bid–ask spreads over the interval)
LogTPriceVar	log(variance of trade by trade prices over the interval)
LogBuyVolume	sum of log(buyer-initiated volume) over the interval
LogSellVolume	sum of log(seller-initiated volume) over the interval
LogAskDepth	log(mean ask depth over the interval)
LogBidDepth	log(mean bid depth over the interval).

The cost equations use the variance of transaction prices (LogTPriceVar) as a proxy for the volatility of efficient price for two reasons. First, comparing to the efficient market price, transaction price potentially contains unique information about the extent of information asymmetry in the market. Second, since LogTPriceVar is observable, it simplifies the model structure and allows us to concentrate on the economic implications of the modeling results. Although the unobservable $\ln(\sigma^2)$ may be a more natural choice than LogTPriceVar, using $\ln(\sigma^2)$ in the cost model would significantly increase the complexity of the whole model.

3.3. Decomposition of market variables

The market variables are highly dependent and therefore predictable. As such, much of the current value of a market variable may already have been predicted by the market and included in the price. The surprise component of each market variable, however, will contain the genuine new information. We therefore propose decomposing the variables into two components, the predictable and the surprise. Following Hasbrouck (1991), we use the Vector Autoregression (VAR) framework to model market variables jointly and to compute the anticipated and unanticipated components. After subtracting its time of day mean for each variable, we fit an VAR model for the de-trended sequence of market variable vector. We use the conditional mean vector and residual vector from the VAR model as the corresponding anticipated and unanticipated sequences. Formally, we use $E[\cdot]$ and $e[\cdot]$ to represent the anticipated and unanticipated vector sequences respectively, where the dot represents the corresponding market variable vector.

For market variable vector

$\mathbf{X}_t = (\text{LogSpread}, \text{LogTPriceVar}, \text{LogBuyVolume}, \text{LogSellVolume}, \text{LogAskDepth}, \text{LogBidDepth})'_t$ we fit a VAR(p) model of the following form,

$$\mathbf{X}_t = \phi_0 + \Phi_1 \mathbf{X}_{t-1} + \cdots + \Phi_{t-p} \mathbf{X}_{t-p} + \mathbf{e}_t, \quad (9)$$

⁹ For the bid and ask quotes $\{a_t, b_t\}$ observed at time t , we effectively use the information vector $\mathbf{X}$ observed between t and $t-1$ (during the 15-minute interval) to understand the determinants of $\{a_t, b_t\}$. It is more natural to index $\mathbf{X}$ by $t-$, meaning *before time t* , instead of using $t-1$. Hence, for market variables, we will use the format $\mathbf{X}_{t-}$ throughout the empirical implementation.

where ϕ_0 is a 6-dimensional vector and Φ_i is a 6×6 matrix, for $i=1, \dots, p$. Hence, $E[X_t] = \phi_0 + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p}$ and $e[X_t] = e_t$ are the anticipated and unanticipated components for X_t . We use the AIC criterion to identify the best VAR model for each stock.

The six pairs of anticipated and unanticipated variables will be used in the cost model and the EGARCH model to study their relationship with market making costs and the efficient market price dynamics. We further define Level (unanticipated volume or volume shock) as $e[\text{LogBuyVolume}] + e[\text{LogSellVolume}]$ and Pressure (unanticipated order imbalance or order imbalance shock) as $e[\text{LogBuyVolume}] - e[\text{LogSellVolume}]$ and test for their impact on market making costs.¹⁰ Those are natural definitions in the refined buy volume and sell volume universe. These two variables play an important role in our empirical analysis the asymmetric impacts of volume on market making costs.

4. Empirical analysis

We first discuss the model estimation results and goodness of fit and then present our detailed empirical analysis. We concentrate on our model's implications for the time-varying market making costs and the efficient market price.

4.1. Estimation results and goodness of fit

The estimation results for the proposed model are given in Table 2.¹¹ Both the cost model and the EGARCH/GED model have a significant deterministic component and a significant stochastic autoregressive component coupled with an array of significant market microstructure variables as explanatory variables.

Model diagnostics are performed using posterior means of the estimated innovations, $\hat{v}_t^\alpha$, $\hat{v}_t^\beta$, and $\hat{\epsilon}_t$. The model diagnostics results shown in Table 3 indicate that the model innovations pass the Ljung–Box serial correlation tests.

4.2. Dynamics of market making costs

4.2.1. Microstructure underpinning

Understanding the dynamics of market making costs has implications for market viability and stability. Microstructure theory identifies inventory control and information asymmetry as the two major sources of risk faced by specialists or liquidity providers. The specialists are compensated for bearing these risks by charging a premium over the efficient market price. The inventory control component and the asymmetric information component of the market making costs have different characteristics and hence have different implications for the dynamics of market making costs.

In an inventory model, the specialist faces a complex balancing problem in that he must moderate random deviations in inflows and outflows where the deviations are, by assumption, unrelated to the future value of the stock. Empirical studies of inventory control by market makers in equity markets found weak intraday inventory effects.¹²

The asymmetric information component of market making costs is likely to be time varying and highly persistent driven by the arrival of new information. There are strong theoretical arguments and compelling empirical evidence that the information cost incurred by dealers is impounded in the bid and ask quotes they expose to the market.

We explore the information aspects of time-varying market making costs through selected market variables. Our model provides a unique framework to explore the sources of variation for market making costs. It allows not only common market factors that drive the market making costs, but also factors that potentially have asymmetric impact on the cost of ask quote and the cost of bid quote. The risk associated with the presence of asymmetric information is the key link between those factors and the costs of market making, as the liquidity providers will adjust the bid and ask quotes accordingly to be compensated for bearing the risk.

4.2.2. Time-varying market making costs

The estimation results for the proposed cost submodel for GE, JPM and XOM are presented in the top panel of Table 2. The cost processes, in excess of intraday deterministic mean, still have a significant persistent component.

¹⁰ Noted that $e[\text{LogBuyVolume}] \pm e[\text{LogSellVolume}] = e[\text{LogBuyVolume} \pm \text{LogSellVolume}]$.

¹¹ We also estimate the model of Hasbrouck (1999a) and obtain similar results.

¹² See Hasbrouck (1988), Madhavan and Smidt (1993) and Hasbrouck and Sofianos (1993).

Unanticipated variance of transaction prices ($e[\text{LogTPriceVar}]$) is significantly and positively associated with the exposure costs on both sides. Unanticipated buy and sell volumes ($e[\text{LogBuyVolume}]$ and $e[\text{LogSellVolume}]$) play distinct and asymmetric roles on the ask side and bid side of the cost. An increase in expected quoted depths, $E[\text{LogBidDepth}]$ and $E[\text{LogAskDepth}]$, will reduce the market making costs in the bid side and the ask side correspondingly. We also test for additional lags of market variables in the cost model and conclude that they are not significant.

4.2.2.1. Persistence. We find that the dynamics of market making costs have both a deterministic and a stochastic component. The AR ϕ coefficient for each of the three stocks is statistically different from zero. Thus, after incorporating market variable shocks into the cost model, substantial autoregressive type of persistence remains in the dynamics of market making costs. This could be attributed to the intrinsic characteristic of the market maker, that cannot be captured by the market variable shocks.

Table 2
Model parameter estimates (January 3, 2006–March 31, 2006)

Parameter	Posterior mean			t-Ratio		
	GE	JPM	XOM	GE	JPM	XOM
<i>Cost model</i>						
ϕ	0.254	0.114	0.123	3.41	8.91	6.41
σ_v	1.588	1.884	1.862	15.57	23.38	19.5
k_1	−3.872	−3.739	−3.084	−14.78	−11.51	−15.58
k_2^{open}	0.915	1.398	1.341	3.58	4.45	6.48
k_3^{open}	1.354	0.442	0.815	1.25	1.63	2.4
k_2^{close}	1.542	1.731	1.187	5.73	6.07	4.58
k_3^{close}	5.777	6.720	5.475	4.17	6.63	3.50
d_1	0.316	0.0980	0.203	4.43	1.44	3.32
d_1^{α}	0.172	0.415	0.175	1.24	3.49	1.56
d_2^{α}	−0.0472	−0.442	−0.487	−0.082	−1.31	−1.08
d_3^{α}	−0.407	−0.428	−0.215	−3.34	−2.89	−2.83
d_4^{α}	−0.0982	−0.227	−0.0517	−0.97	−1.94	−0.68
d_1^{β}	−0.144	−0.225	−0.186	−1.29	−2.05	−1.66
d_2^{β}	0.272	0.422	0.165	2.59	2.93	2.12
d_3^{β}	−0.0642	−0.175	−0.0943	−0.79	−1.46	−1.37
d_4^{β}	−0.545	−0.933	−1.117	−0.96	−1.87	−3.25
<i>Efficient price/EGARCH model</i>						
c_1^m	0.000285	0.000422	0.000338	17.17	21.21	17.16
c_2^m	−0.000328	−0.000492	−0.000459	−20	−22.88	−20.96
v^{day}	1.379	1.491	1.586	16.44	18.25	17.92
v^{night}	1.993	2.476	1.210	3.52	3.34	4.24
ψ	0.750	0.752	0.286	9.52	8.19	1.61
ω	0.0118	0.0250	0.00550	0.31	0.83	0.14
γ	0.313	0.261	0.195	4.82	4.31	2.96
η^{night}	1.269	0.764	1.0975	2.03	1.85	3.08
l_1	−13.485	−13.612	−13.396	−858.92	−67.72	−100.12
l_2^{open}	1.204	2.099	2.0746	2.03	6	8.49
l_3^{open}	3.549	1.257	0.817	1.92	2.43	4.88
l_2^{close}	0.610	0.370	0.529	2.54	1.33	2.44
l_3^{close}	3.074	3.507	3.426	2.16	1.53	1.7
c_1	1.760	0.730	0.473	5.62	4.24	3.42
c_2	0.0626	0.188	0.0224	2.59	7.4	1.27
c_3	0.0091	0.129	0.0796	0.32	4.5	4
c_4	−0.497	−0.262	−0.0618	−1.21	−1.16	−0.38
c_5	0.0276	−0.0975	−0.0397	1.01	−2.66	−2.08
c_6	−0.0127	−0.0338	0.0223	−0.44	−1.09	0.77
c_7	0.332	−0.167	−1.351	2.52	−0.84	−3.78
c_8	−0.281	0.218	0.816	−1.74	2.26	2.95

Market variables that are significant in the cost model, including unanticipated transaction price variation ($e[\text{LogTPPriceVar}]$), shocks to buy and sell volumes ($e[\text{LogBuyVolume}]$ and $e[\text{LogSellVolume}]$), absolute value of order imbalance surprise ($|e[\text{LogBuyVolume} - \text{LogSellVolume}]|$), and expected depth ($E[\text{LogAskDepth}]$ and $E[\text{LogBidDepth}]$), are closely associated with information asymmetry in the market, hence the persistence absorbed by those market variables may reflect a mean-reverting structure in the asymmetric information component of the exposure costs.

4.2.2.2. Volatility and market making costs. Market microstructure theory indicates that volatility is closely related to the extent of information asymmetry in the market. Other forces such as the leverage effect and short sale constraints can also potentially alter the ways that volatility affects market making costs. It is therefore interesting to examine whether the impact of volatility on the two sides of the market making cost is symmetric. We use as a proxy of price volatility and conduct the test by allowing the coefficients of $e[\text{LogTPPriceVar}]$ and $E[\text{LogTPPriceVar}]$ to be different in the bid and ask equations in our model specification. While the coefficients of $E[\text{LogTPPriceVar}]$ are not significant, we cannot reject the hypothesis that the two significant coefficients for $e[\text{LogTPPriceVar}]$ are equal and therefore let the coefficients to be the same across the two cost equations in our final specification.

The coefficients of $e[\text{LogTPPriceVar}]$ are all positive for the three stocks and are statistically significant for GE (t -ratio=4.432) and XOM (t -ratio=3.322). That volatility shocks as a common factor positively correlated with the costs of market making is consistent with the prediction of information based market microstructure theory. Market makers will update their bid and ask quotes more rapidly and increase the premium they charged once they learn through the sell and buy orders that the proportion of informed trader is high. Our model estimation results indicate that the common factor in the bid and ask costs to some extent also reflects the uncertainty in the underlying efficient price.

Notes to Table 2

This table reports parameter estimates of the following model.

$$\left\{ \begin{array}{l} a_t = \lceil m_t + \alpha_t \rceil \\ b_t = \lfloor m_t - \beta_t \rfloor \\ \ln(\alpha_t) = \mu_t + \theta_t + \theta_t^\alpha + \phi[\ln(\alpha_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\alpha \\ \ln(\beta_t) = \mu_t + \theta_t + \theta_t^\beta + \phi[\ln(\beta_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\beta \\ \mu_t = k_1 + k_2^{\text{open}} \exp(-k_3^{\text{open}} \tau_t^{\text{open}}) + k_2^{\text{close}} \exp(-k_3^{\text{close}} \tau_t^{\text{close}}) \\ \theta_t = d_1 e[\text{LogTPPriceVar}_{t-}] \\ \theta_t^\alpha = d_1^\alpha e[\text{LogBuyVolume}_{t-}] + d_2^\alpha e[\text{LogSellVolume}_{t-}] \\ \quad + d_3^\alpha |e[\text{LogBuyVolume}_{t-}] - e[\text{LogSellVolume}_{t-}]| + d_4^\alpha E[\text{LogAskDepth}_{t-}] \\ \theta_t^\beta = d_1^\beta e[\text{LogBuyVolume}_{t-}] + d_2^\beta e[\text{LogSellVolume}_{t-}] \\ \quad + d_3^\beta |e[\text{LogBuyVolume}_{t-}] - e[\text{LogSellVolume}_{t-}]| + d_4^\beta E[\text{LogBidDepth}_{t-}] \\ \ln(m_t) = \ln(m_{t-1}) + \delta_t + \sigma_t \epsilon_t \\ \delta_t = c_1^{\text{open}} e[\text{LogBuyVolume}_{t-}] + c_2^{\text{open}} e[\text{LogSellVolume}_{t-}] \\ \epsilon_t \sim \text{GED}(v^{\text{night}}) 1_{\{\tau_t^{\text{open}}=0\}} + \text{GED}(v^{\text{day}}) 1_{\{\tau_t^{\text{open}}>0\}} \\ \ln(\sigma_t^2) = \eta_t + \zeta_t + \psi[\ln(\sigma_{t-1}^2) - \eta_{t-1}] + \omega \epsilon_{t-1} + \gamma[|\epsilon_{t-1}| - E|\epsilon_{t-1}|] \\ \eta_t = \eta^{\text{night}} 1_{\{\tau_t^{\text{open}}=0\}} + l_1 + l_2^{\text{open}} \exp(-l_3^{\text{open}} \tau_t^{\text{open}}) + l_2^{\text{close}} \exp(-l_3^{\text{close}} \tau_t^{\text{close}}) \\ \zeta_t = c_1 e[\text{LogSpread}_{t-}] + c_2 e[\text{LogBuyVolume}_{t-}] + c_3 e[\text{LogSellVolume}_{t-}] \\ \quad + c_4 e[\text{LogSpread}_{(t-1)-}] + c_5 e[\text{LogBuyVolume}_{(t-1)-}] + c_6 e[\text{LogSellVolume}_{(t-1)-}] \\ \quad + c_7 E[\text{LogBidDepth}_{t-}] + c_8 E[\text{LogAskDepth}_{t-}] \end{array} \right.$$

The time index t represents the end of 15-minute intervals over trading day. Bid and ask quotes $\{b_t, a_t\}$ are jointly determined by the efficient price m_t , the quote exposure cost of the ask side, α_t , and the quote exposure cost of the bid side, β_t , through the rounding mechanisms and $\alpha_t = \text{Ceiling}(m_t + \alpha_t) = \lceil m_t + \alpha_t \rceil$, and $\beta_t = \text{Floor}(m_t - \beta_t) = \lfloor m_t - \beta_t \rfloor$. The logarithms of α_t and β_t follow AR(1) processes with common deterministic mean μ_t , market variable component θ_t , autoregressive coefficient ϕ and constant error variance σ_v^2 , but independent $N(0,1)$ innovations v_t^α and v_t^β , ask-side specific market variable component θ_t^α and bid-side specific market variable component θ_t^β . Market variable components θ_t , θ_t^α and θ_t^β are linear functions of expected and/or unexpected components of market variables defined below. The logarithm of the efficient market price follows a random walk with a contemporaneous surprise term δ_t in the mean and a conditional heteroskedastic volatility σ_t^2 , which governs by an EGARCH process with a deterministic mean η_t , a market variable component ζ_t , an autoregressive component with autoregressive coefficient ψ and an innovation component. The innovation of the random walk, ϵ_t , is distributed as a Generalized Error Distribution (GED) with tail-thickness parameter ν equals ν^{day} during the day and ν^{night} over night. Parameters ω and γ of the EGARCH process measure potential asymmetric effect of innovation on volatility. Market variables are defined over each 15-minute period. LogSpread is the logarithm of the mean bid–ask spreads over the interval. LogTPPriceVar is the logarithm of the variance of transaction prices over the interval. LogBuyVolume (LogSellVolume) is the sum of log buy (sell) volume over the interval. LogAskDepth (LogBidDepth) is the logarithm of mean ask (bid) depth over the interval. Anticipated and unanticipated components of a market variable x_t , denoted by $E[x_t]$ and $e[x_t]$ respectively, are constructed using the best VAR model fitted for the market variable vector $\{X_t\}$. The model is estimated using Markov Chain Monte Carlo method.

4.2.2.3. Impact of depth. Although bid and ask depths and bid and ask spreads measure different aspects of market liquidity, they are closely related to each other as both reflect the market maker's response to the market's trade flow. While the market maker will increase the spreads when he learns from the trade flow that the likelihood of informed trader is high, he will also reduce the quoted bid and ask depths accordingly. Thus we expect that past quote depths should be indicative about the extent of information asymmetry in the market, hence the costs of bid and ask quotes.

The negative coefficients of $E[\text{LogAskDepth}]$ and $E[\text{LogBidDepth}]$ for all three stocks in the cost model are consistent with our projection. Higher anticipated ask (bid) depth predicts lower cost of ask (bid) quote, indicating that there is no big buy (sell) movement caused by trading positive (negative) news.

The estimation results for the bid and ask depths are also consistent with an inventory control story that the market maker is actively managing his inventory level via his quoted prices, in order to keep his inventory in some optimal level.

4.2.3. Asymmetric cost processes

The coefficients of $e[\text{LogBuyVolume}]$ are all positive for $\ln(\alpha_i)$ and all negative for $\ln(\beta_i)$. This indicates that higher unanticipated buy volume will increase the cost of ask quote but reduce the cost of bid quote. On the other hand, the coefficients of $e[\text{LogSellVolume}]$ are all negative for $\ln(\alpha_i)$ and significantly positive for $\ln(\beta_i)$, thus, higher sell volume shocks are associated with lower cost of ask quote. Those impacts are marginal impacts and are consistent with an asymmetric information story that the market maker infers the direction of the private information from the direction of the trade.

Table 4 presents the tests of asymmetry between the cost of ask exposure process and the cost of bid exposure process. We test whether the impacts of volume on the cost of ask quote and the cost of bid quote are the same. For buy volume, we use $d_1^\alpha - d_1^\beta$ to measure the difference in impact of $e[\text{LogBuyVolume}]$ on the cost of ask exposure and the cost of bid exposure and $d_1^\alpha - d_1^\beta$ is positive for each of the three stocks and is statistically significant at 1% for JPM and 10% for XOM. Hence the impact of $e[\text{LogBuyVolume}]$ on the two sides of the market making costs is asymmetric: it has a positive impact on the cost of ask and negative impact on the cost of bid. Fig. 1(a) shows the posterior densities of d_1^α and d_1^β for JPM, and Fig. 1(b) gives the posterior densities for $d_1^\alpha - d_1^\beta$ and $d_1^\alpha + d_1^\beta$ for JPM. For $e[\text{LogSellVolume}]$, $d_2^\alpha - d_2^\beta$ is all negative and statistically significantly different from zero for each of the three stocks, hence, the impact of unanticipated sell volume on market making costs is also asymmetric.

4.2.4. Asymmetric impact of volume

We explore the asymmetric impact of volume on market liquidity and conduct formal tests in this subsection. We find the information in volume is two-dimensional and offer an intuitive interpretation.

4.2.4.1. Volume and market making costs: competing theories. There are two competing models regarding the relation between volume and quoted liquidity (hence market making costs). Easley and O'Hara (1992) suggest that abnormal high trading volume induces wider spread and lower quoted liquidity as market maker uses trading volume to infer the presence of informed traders. Thus, the model implies a negative relation between unanticipated volume and market liquidity in a time series context. An alternative theory by Harris and Raviv (1993) predicts a positive relation arguing that volume shocks indicate mainly a lack of consensus among market participants and hence increased volume primarily reflects increased liquidity trading and higher overall market liquidity (lower market making costs and narrow spreads). Lee et al. (1993) find evidence of a positive relation between volume and spread, which is consistent with the Easley and O'Hara's model.

The cost model can offer some insight on this issue from a very unique angle, since we can test directly whether the impact of volume on market making costs is asymmetric. The volume's impact on quoted liquidity is asymmetric if volume shocks have different impacts on the cost of ask quote and the cost of bid quote. The cost model can also provide a refined answer to this problem compared to previous studies. If volume affects quoted liquidity in an asymmetric fashion, then, we should reconsider the evaluation of this relationship since it is not meaningful to correlate volume with quoted liquidity directly.

4.2.4.2. Testing for asymmetry. The insignificance of anticipated volume, $E[\text{LogBuyVolume}]$ and $E[\text{LogSellVolume}]$, is consistent with both of the above mentioned theories that unanticipated volumes determine market making costs. It is intuitively appealing that the impact of volume on market making costs is asymmetric since the components of unanticipated volume can have different directional implications to the market maker and we expect the cost of ask quote and the cost of bid quote potentially evolve in different ways under different market conditions. For example, when there is a

Table 3
Model diagnostics — autocorrelations and correlations

Test statistics	Value			p-Value		
	GE	JPM	XOM	GE	JPM	XOM
<i>Panel A: Ljung–Box statistic for residuals</i>						
LjungBox(15)– $\hat{v}_t^\alpha$	16.682	20.211	9.798	0.34	0.16	0.83
LjungBox(15)– $\hat{v}_t^\beta$	33.846	24.792	13.938	0.0036	0.053	0.53
LjungBox(15)– $\hat{\epsilon}_t$	0.195	5.803	7.191	1	0.98	0.95
<i>Panel B: Correlations between residuals</i>						
corr($\hat{v}_t^\alpha, \hat{v}_t^\beta$)	–0.26	–0.0943	0.2			
corr($\hat{\epsilon}_t, \hat{v}_t^\alpha$)	–0.075	–0.19	–0.085			
corr($\hat{\epsilon}_t, \hat{v}_t^\beta$)	0.082	0.19	0.073			

This table presents diagnostic results for the model. $\hat{v}_t^\alpha$ and $\hat{v}_t^\beta$ are residuals of the two equations for α_t and β_t , respectively. $\hat{\epsilon}_t$ is the residual sequence of the equation for efficient price. Panel A shows Ljung–Box statistic for each residual series with 15 lags of autocorrelations, where the critical values are 25 at 5% level and 30.58 at 1% level. Correlations between each pair of residual series are given in Panel B.

period of big sell off of a stock, the market maker will infer that it is very likely that there is an intensive arrival of informed traders driven by private information with bad outlook for the stock, and we expect that the market maker will increase the cost of bid quote accordingly and keep the cost of ask quote more or less intact since he will incur more costs on the bid side. Similarly, for a period with a dominant number of buys, the market maker is very likely to increase the cost of ask quote. To test this potential asymmetry of volume, we link the market making costs to $e[\text{LogBuyVolume}]$ and $e[\text{LogSellVolume}]$ and allowing them to have different impact on the cost of ask quote and the cost of bid quote.

We formally test for the asymmetric impact of volume on market making costs and show the results in Table 5. The rejection of $d_1^\alpha = d_2^\alpha (d_1^\beta = d_2^\beta)$ across all three stocks shows that the impact of volume on the cost of ask quote (bid quote) is asymmetric, that is, unanticipated buy volume and unanticipated sell volume have different impacts on the market making costs.

Table 4
Tests of differential dynamics of the cost of ask exposure process and the cost of bid exposure process

Hypothesis	Test statistics	Posterior mean			t-Ratio		
		GE	JPM	XOM	GE	JPM	XOM
Panel A: Tests of same dynamics							
$d_1^\alpha = d_1^\beta$	$d_1^\alpha - d_1^\beta$	0.315	0.64	0.361	1.35	3.27	1.68
$d_2^\alpha = d_2^\beta$	$d_2^\alpha - d_2^\beta$	-0.679	-0.85	-0.38	-3.51	-3.26	-2.69
$d_3^\alpha = d_3^\beta$	$d_3^\alpha - d_3^\beta$	-0.034	-0.0527	0.0426	-0.27	-0.37	0.56
$d_4^\alpha = d_4^\beta$	$d_4^\alpha - d_4^\beta$	0.497	0.491	0.629	0.54	0.69	0.97
Panel B: Tests of reverse dynamics							
$d_1^\alpha = -d_1^\beta$	$d_1^\alpha + d_1^\beta$	0.0278	0.19	-0.0109	0.29	1.59	-0.18
$d_2^\alpha = -d_2^\beta$	$d_2^\alpha + d_2^\beta$	-0.135	-0.00599	-0.0498	-1.13	-0.045	-0.83
$d_3^\alpha = -d_3^\beta$	$d_3^\alpha + d_3^\beta$	-0.163	-0.402	-0.146	-1.21	-2.15	-1.18
$d_4^\alpha = -d_4^\beta$	$d_4^\alpha + d_4^\beta$	-0.592	-1.38	-1.6	-0.88	-2.9	-3.37

This table reports statistics of differences (sum) of corresponding coefficients of the cost of ask exposure process (α_t) and the cost of bid exposure process (β_t).

$$\begin{cases} \ln(\alpha_t) = \mu_t + \theta_t + \phi[\ln(\alpha_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\alpha \\ \ln(\beta_t) = \mu_t + \theta_t + \phi[\ln(\beta_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\beta \\ \theta_t^\alpha = d_1^\alpha e[\text{LogBuyVolume}_{t-}] + d_2^\alpha e[\text{LogSellVolume}_{t-}] \\ \quad + d_3^\alpha e[\text{LogBuyVolume}_{t-}] - e[\text{LogSellVolume}_{t-}] + d_4^\alpha E[\text{LogAskDepth}_{t-}] \\ \theta_t^\beta = d_1^\beta e[\text{LogBuyVolume}_{t-}] + d_2^\beta e[\text{LogSellVolume}_{t-}] \\ \quad + d_3^\beta e[\text{LogBuyVolume}_{t-}] - e[\text{LogSellVolume}_{t-}] + d_4^\beta E[\text{LogBidDepth}_{t-}] \end{cases}$$

Those tests are designed to test the null hypothesis that the costs of bid exposure and the cost of ask exposure have the same (reverse) dynamics in terms of the coefficients to the buy volume surprise ($e[\text{LogBuyVolume}_{t-}]$), the sell volume surprise ($e[\text{LogBuyVolume}_{t-}]$), the order imbalance surprise (Pressure) and expected depth. Posterior mean and posterior standard deviation are computed using draws from the Gibbs sampler.

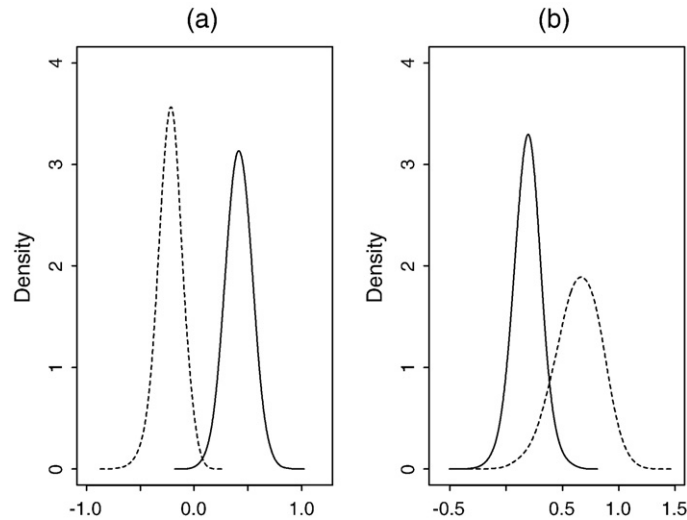


Fig. 1. Impacts of $e(\text{LogBuyVolume})$ on the Costs of Market Making for JPM. (a) Posterior densities for d_1^α (solid line) and d_1^β (dash line). (b) Posterior densities for $d_1^\alpha + d_1^\beta$ (solid line) and $d_1^\alpha - d_1^\beta$ (dash line).

One important conclusion to draw is that the information in volume utilized by the market maker in determining the quoted price is two-dimensional. The pair $e[\text{LogBuyVolume}]$ and $e[\text{LogSellVolume}]$ is a sufficient summary statistics of such information. The volume shock by itself is just an aggregate of the two pieces of information perceived by the market maker and represents a partial picture of the market conditions. Hence, the composition of the volume shock is crucial and it will be out of context and even misleading if one just uses a single dimension of the information.

This provides a potential direction in theoretical research to reconcile the competing theories of the Easley and O'Hara's model and the Harris and Raviv's model regarding relation between volume and quoted liquidity. The seemingly contrary predictions from the two models stem from the ambiguity about volume. When shocks to buy volume and shocks to sell volume differ dramatically and hence the extent of information asymmetry is high, we expect that there is a negative relation between volume and liquidity, which is the prediction from the Easley and O'Hara's model. If shock to buy volume and shock to sell volume are of comparable size, reflecting a lack of consensus among

Table 5

Tests of asymmetric role played by volume in determining the costs of quote exposure on the ask and bid sides

Hypothesis	Test statistics	Posterior mean			t-Ratio		
		GE	JPM	XOM	GE	JPM	XOM
<i>Panel A: Tests of same dynamics</i>							
$d_1^\alpha = d_2^\alpha$	$d_1^\alpha - d_2^\alpha$	0.579	0.843	0.39	3.01	3.48	2.25
$d_1^\beta = d_2^\beta$	$d_1^\beta - d_2^\beta$	−0.416	−0.647	−0.351	−2.74	−2.83	−2
<i>Panel B: Tests of reverse dynamics</i>							
$d_1^\alpha = -d_2^\alpha$	$d_1^\alpha + d_2^\alpha$	−0.236	−0.013	−0.04	−1.32	−0.11	−0.5
$d_1^\beta = -d_2^\beta$	$d_1^\beta + d_2^\beta$	0.128	0.197	−0.0207	0.83	1.68	−0.26

This table presents hypothesis testing results on the asymmetric role played by volume in determining the costs of quote exposures.

$$\begin{cases}
 \ln(\alpha_t) = \mu_t + \theta_t + \theta_t^\alpha + \phi[\ln(\alpha_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\alpha \\
 \ln(\beta_t) = \mu_t + \theta_t + \theta_t^\beta + \phi[\ln(\beta_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\beta \\
 \theta_t^\alpha = d_1^\alpha e[\text{LogBuyVolume}_{t-}] + d_2^\alpha e[\text{LogSellVolume}_{t-}] \\
 \quad + d_3^\alpha [e[\text{LogBuyVolume}_{t-}] - e[\text{LogSellVolume}_{t-}]] + d_4^\alpha E[\text{LogAskDepth}_{t-}] \\
 \theta_t^\beta = d_1^\beta e[\text{LogBuyVolume}_{t-}] + d_2^\beta e[\text{LogSellVolume}_{t-}] \\
 \quad + d_3^\beta [e[\text{LogBuyVolume}_{t-}] - e[\text{LogSellVolume}_{t-}]] + d_4^\beta E[\text{LogBidDepth}_{t-}]
 \end{cases}$$

Those tests are designed to test the null hypothesis that the shocks to buy volume $e[\text{LogBuyVolume}_{t-}]$ and sell volume ($e[\text{LogBuyVolume}_{t-}]$) have the same (reverse) impacts on the cost of buy exposure and cost of sell exposure. Posterior mean and posterior standard deviation are computed using draws from the Gibbs sampler.

market participants, then we expect a market with tight spread and increased depths, hence, a positive relation between volume and liquidity. This implies that both models can be right or wrong, conditionally. It suggests that the question of what is the relation between volume and market liquidity by itself is ambiguous. A complete model should utilize both dimensions of the information in volume to make unconditional and correct predictions.

The asymmetric impact of volume on market making costs also shows that there are distinct and significant dynamics in the ask and bid sides of the market making costs. Hence, prediction of the corresponding costs can be improved by using ask- and bid-specific factors. This also challenges the appropriateness of the symmetric assumption of market making costs frequently used in empirical microstructure literature.

4.2.4.3. Interpretation. Fig. 2 shows the evolution of bid–ask quote, estimated costs of market making, estimated efficient market price and the two important volume variables, Pressure and Level, for a randomly picked day, January 5, 2006 for JPM.

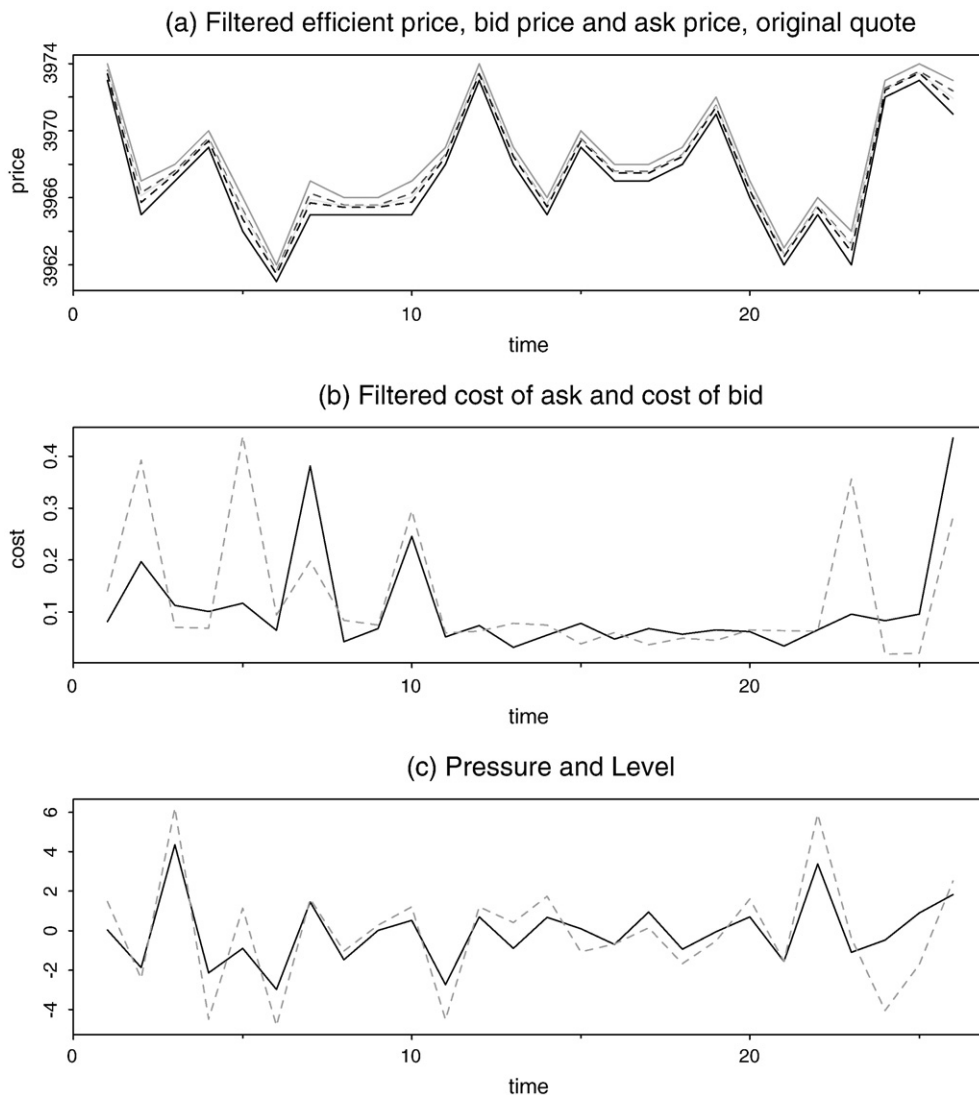


Fig. 2. Changes in price, cost, Level and Pressure for JPM for a day, January 5, 2006. The unit used for price and cost is one tick (\$1/100). Pressure and Level are standardized quantities. (a) The two outer solid lines are ask price (top) and bid price (bottom). The two dash lines are intrinsic ask price (efficient price+true cost of ask, top) and intrinsic bid price (efficient price+true cost of bid, bottom). The center solid line is the estimated efficient price. (b) The solid line is the estimated cost of ask and the dash line is the cost of bid. (c) The solid line corresponds to Pressure and the dash line corresponds to Level.

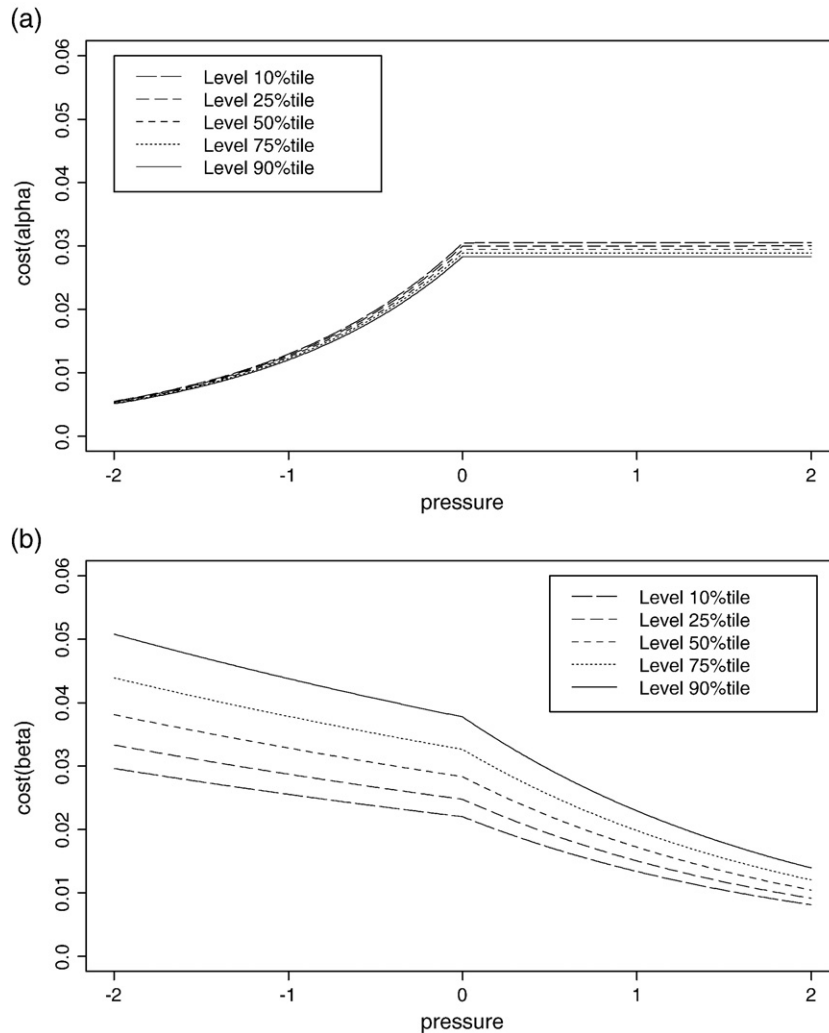


Fig. 3. Impact of Pressure on the cost of ask and cost of bid for JPM. Plots are based on the estimated dynamics for the cost of ask (α_t) and cost of bid (β_t) given in Table 2. The range for Pressure is based on its sample observations. Sample percentiles are used for level. Unconditional expectations proxied by sample means are used for other variables. Panel (a) shows the relationships between the cost of ask and Pressure conditional on Level fixed at its different percentiles. Panel (b) shows the relationships between the cost of bid and Pressure conditional on Level fixed at its different percentiles.

Using JPM as an example, we offer an intuitive interpretation of how the market maker utilizes the two pieces of volume information in the decision process. The two-dimensional information in volume can be represented by two factors, Pressure and Level. We rewrite θ_t^α (omitting the depth term for simplicity) in the following form

$$\theta_t^\alpha \approx -0.0135(\text{Level}) + 0.4285(\text{Pressure}) - 0.428|\text{Pressure}|. \quad (10)$$

Intuitively, the size of unanticipated buy volume exceeding unanticipated sell volume will put *Pressure* on the market maker to increase the cost of ask quote. Level is represented by total unanticipated volume. The negative sign, even though statistically insignificant, indicates that the market maker will discount the effect of Pressure if the Level is high.

Similarly, we can also rewrite θ_t^β (omitting the depth term) in the following form with the two factors of Pressure and Level,

$$\theta_t^\beta \approx 0.0985(\text{Level}) + 0.3235(-\text{Pressure}) - 0.175|\text{Pressure}|. \quad (11)$$

The negative Pressure represents the pressure caused by unanticipated sell volume exceeding unanticipated buy volume. The market maker will increase the cost of bid quote to maintain order balance. Similarly, when the Level is high, the market maker will discount the effect of the negative Pressure.

In order to visualize the cost function implied by the estimated model, variables other than Pressure and Level are set to their sample mean. Fig. 3 shows the costs of market making as functions of Pressure as Level fixed at its different percentiles. Fig. 4 presents a more complete picture of market making costs as functions of Pressure and Level. Figs. 3 and 4 clearly show the impact of Level and Pressure on the costs of market making. Although it is possible to interpret differently the effects of volume on the cost of trading, the message conveyed from our analysis is clear. The information in volume is two-dimensional and its effects on the cost of market making are asymmetric. It pays to consider signed volume and separate unanticipated volume out from volume.

4.3. Dynamics of efficient market price

Our specification for the efficient market price is comprehensive and informative by design. We explore the source of efficient price dynamics, both the mean and the volatility, at intraday level through market variables.

4.3.1. Mean dynamics

We allow the efficient price dynamics to depend on the surprises of market variables. The motivation is that if trading characteristics are informative about the degree of private information then we might expect these trading

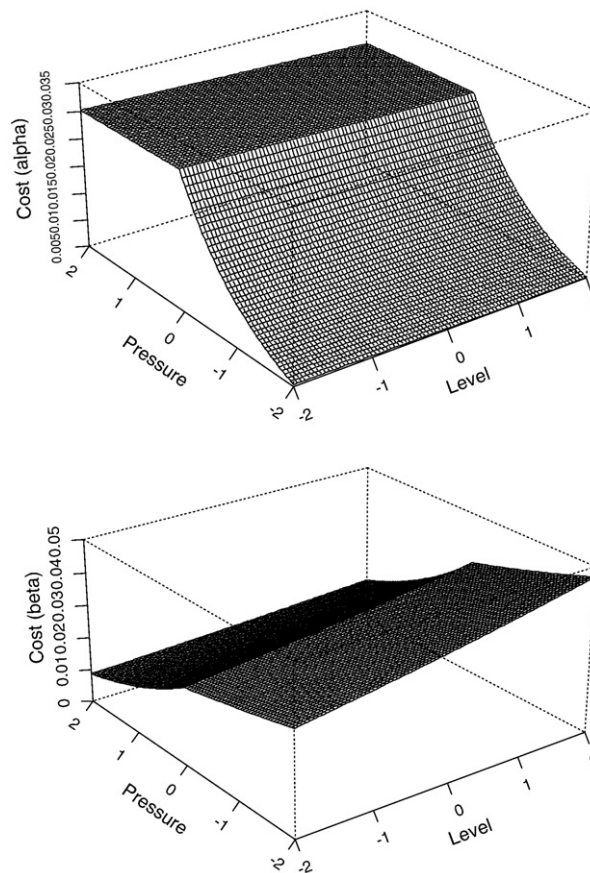


Fig. 4. Impact of Pressure and Level on the cost of ask and cost of bid for JPM. Plots are based on the estimated dynamics for the cost of ask (α_t) and cost of bid (β_t) given in Table 2. The range for Pressure and Level are based on their sample observations. Unconditional expectations proxied by sample means are used for other variables.

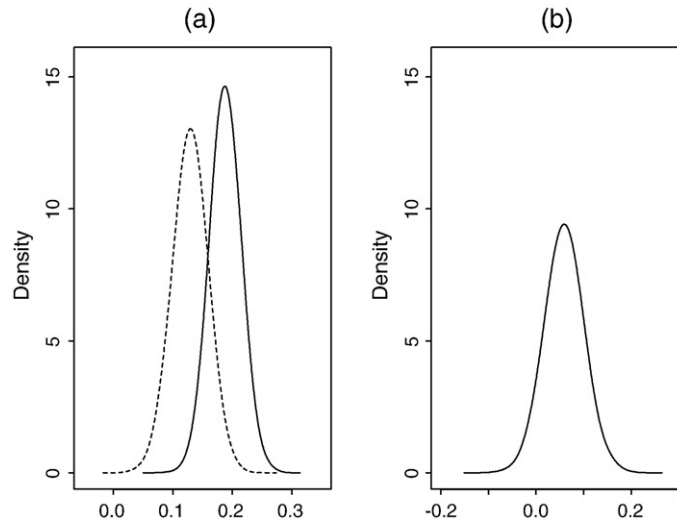


Fig. 5. Impact of $e(\text{LogBuyVolume})$ and $e(\text{LogSellVolume})$ on the volatility of the efficient market price for JPM. (a) Posterior densities for c_2 (solid line) and c_3 (dash line). (b) Posterior density for $c_2 - c_3$.

characteristics to convey information about efficient price. By putting contemporaneous surprise terms we can test the long run impact of market news on asset price directly and still preserve the random walk structure for efficient price.

The estimation results for the mean dynamics are presented in the first two rows of the lower panel of Table 2. The coefficients for $e[\text{LogBuyVolume}]$ ($e[\text{LogSellVolume}]$) are positive (negative) and statistically significant for all three stocks. Hence, unanticipated buy (sell) volume has a significant positive (negative) impact on efficient price. This is consistent with an asymmetric information story that unanticipated buy (sell) volume signals the strong likelihood that informed traders know good (bad) news about the asset traded. Therefore, the contemporaneous surprise terms have a significant long run impact on the price of the asset.

4.3.2. Volatility

The notion that the return volatility process is closely associated with information flow in the market motivates our inclusion of market microstructure variables into the volatility model. We now empirically investigate the determinants of the volatility persistence.¹³ Market microstructure theory suggests that we add five market variables closely related to the information flow of the market, LogSpread, LogBuyVolume, LogSellVolume, LogBidDepth and LogAskDepth to the volatility process as proxies to the information flows to the market. The common linkage between volatility and these market variables is the extent of information asymmetry in the market.

The estimation results of the EGARCH/GED submodel are presented in the lower panel of Table 2. The estimated persistent measure of the EGARCH process, ψ equals 0.750, 0.752 and 0.286 respectively for GE, JPM and XOM. This shows that spread, volume and depth jointly can explain a substantial portion but not all of the persistence in the volatility process. This is consistent with a information based story that the extent of information asymmetry in the market can account most of the persistence in the volatility process, where the clustering of private information arrivals drives the volatility clustering observed in equity market.

With the exception of depth, only the unanticipated components are significantly associated with volatility. The coefficients for the explanatory variables are all statistically significant with signs being consistent with the prediction

¹³ In addition to the information based market microstructure theory, an alternative framework is the Mixture of Distribution Model (MODM). Fundamental to the MODM approach is that it is new information arrival that changes the reservation prices of trades, and so induces changes in market prices.

Table 6

Tests of asymmetric effects by volume and depth on the EGARCH volatility process

Hypothesis	Test Statistics	Posterior mean			t-Ratio		
		GE	JPM	XOM	GE	JPM	XOM
<i>Panel A: Tests of same effects</i>							
$c_2 = c_3$	$c_2 - c_3$	0.0535	0.0592	−0.0572	1.49	1.48	−1.92
$c_5 = c_6$	$c_5 - c_6$	0.0405	−0.0638	−0.0621	1.05	−1.48	−1.58
$c_7 = c_8$	$c_7 - c_8$	0.613	−0.384	−2.17	2.33	−1.38	−3.61
<i>Panel B: Tests of reverse effects</i>							
$c_2 = -c_3$	$c_2 + c_3$	0.0718	0.317	0.102	1.84	8.63	4.38
$c_5 = -c_6$	$c_5 + c_6$	0.0149	−0.131	−0.0174	0.36	−2.49	−0.6
$c_7 = -c_8$	$c_7 + c_8$	0.0506	0.0507	−0.535	0.38	0.36	−2.42

This table presents tests of potential asymmetric effects by volume and depth on the EGARCH specification of the volatility process.

$$\begin{cases} \ln(\sigma_t^2) = \eta_t + \zeta_t + \psi[\ln(\sigma_{t-1}^2) - \eta_{t-1}] + \omega\epsilon_{t-1} + \gamma[|\epsilon_{t-1}| - E|\epsilon_{t-1}|] \\ \zeta_t = c_1 e[\text{LogSpread}_{t-}] + c_2 e[\text{LogBuyVolume}_{t-}] + c_3 e[\text{LogSellVolume}_{t-}] \\ \quad + c_4 e[\text{LogSpread}_{(t-1)-}] + c_5 e[\text{LogBuyVolume}_{(t-1)-}] + c_6 e[\text{LogSellVolume}_{(t-1)-}] \\ \quad + c_7 E[\text{LogBidDepth}_{t-}] + c_8 E[\text{LogAskDepth}_{t-}] \end{cases}$$

Posterior mean and posterior standard deviation are computed using draws from the Gibbs sampler.

from market microstructure theory. That is, higher volatility is associated with wider unanticipated bid–ask spreads, higher unanticipated volumes. But the coefficients for anticipated quote depths and the lag − 1 unanticipated spread and volumes are inconclusive.

We test for the potential asymmetric effect of volume on price volatility using $e[\text{LogBuyVolume}]$ and $e[\text{LogSellVolume}]$. Fig. 5(a) displays the posterior densities of c_2 and c_3 for JPM and Fig. 5(b) shows the posterior density of $c_2 - c_3$. As shown in Table 6, we cannot reject the hypothesis of $c_2 = c_3$ at 5% level. Similarly, for the lag − 1 volume shock, we cannot reject the hypothesis that $c_5 = c_6$ at 5%. Fig. 6 shows the related posterior densities.

Let $X_{t-}^1 = (e[\text{LogSpread}_{t-}], e[\text{LogBuyVolume}_{t-}], e[\text{LogSellVolume}_{t-}])$ be the row vector of shocks to spread, buy volume and sell volume, whose lag − 1 values are also significant in the model, $X_{t-}^2 = (E[\text{LogBidDepth}_{t-}], E[\text{LogAskDepth}_{t-}])$ be the row vector of shocks to quoted depths, and c_1 and c_2 be the corresponding column vector of coefficients. Given

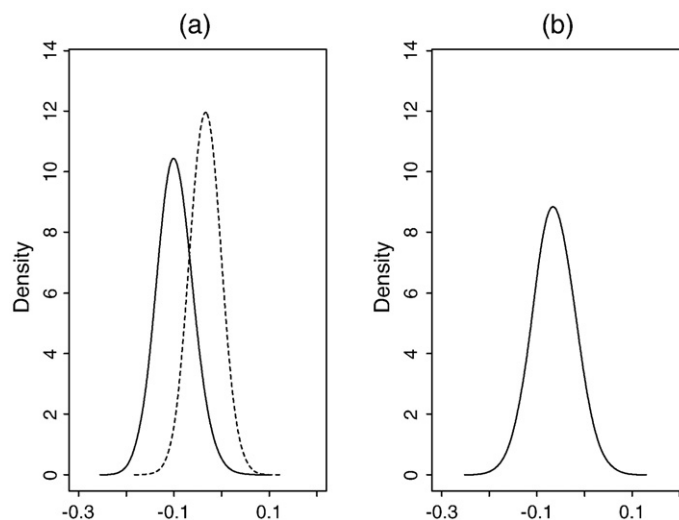


Fig. 6. Impact of Lag − 1 $e(\text{LogBuyVolume})$ and $e(\text{LogSellVolume})$ on the volatility of the efficient market price for JPM. (a) Posterior densities for c_5 (solid line) and c_6 (dash line). (b) Posterior density for $c_5 - c_6$.

Table 7

Specification tests of the EGARCH volatility process

Hypothesis	Test statistics	Posterior mean			t-Ratio		
		GE	JPM	XOM	GE	JPM	XOM
$c_1\psi + c_4 = 0$	$c_1\psi + c_4$	0.824	0.287	0.0734	2.19	1.39	0.44
$c_2\psi + c_5 = 0$	$c_2\psi + c_5$	0.0747	0.0441	−0.0334	3.2	1.34	−1.76
$c_3\psi + c_6 = 0$	$c_3\psi + c_6$	−0.00587	0.0633	0.0451	−0.26	2.33	1.56

This table presents specification tests designed to detect potential form reduction in the EGARCH specification of the volatility process.

$$\begin{cases} \ln(\sigma_t^2) = \eta_t + \zeta_t + \psi [\ln(\sigma_{t-1}^2) - \eta_{t-1}] + \omega\epsilon_{t-1} + \gamma[|\epsilon_{t-1}| - E|\epsilon_{t-1}|] \\ \zeta_t = c_1 e[\text{LogSpread}_{t-}] + c_2 e[\text{LogBuyVolume}_{t-}] + c_3 e[\text{LogSellVolume}_{t-}] \\ \quad + c_4 e[\text{LogSpread}_{(t-1)-}] + c_5 e[\text{LogBuyVolume}_{(t-1)-}] + c_6 e[\text{LogSellVolume}_{(t-1)-}] \\ \quad + c_7 E[\text{LogBidDepth}_{t-}] + c_8 E[\text{LogAskDepth}_{t-}] \end{cases}$$

If $c_1\psi + c_4 = 0$, $c_2\psi + c_5 = 0$ and $c_3\psi + c_6 = 0$, then the EGARCH process can be reduced to the form:

$$[\ln(\sigma_t^2) - \eta_t - \mathbf{X}_{t-}^1 \mathbf{c}_1] (1 - \psi B) = \mathbf{X}_{t-}^2 \mathbf{c}_2 + \omega\epsilon_{t-1} + \gamma[|\epsilon_{t-1}| - E|\epsilon_{t-1}|],$$

where $\mathbf{X}_{t-}^1 = (e[\text{LogSpread}_{t-}], e[\text{LogBuyVolume}_{t-}], e[\text{LogSellVolume}_{t-}])$ is the row vector of shocks to spread, buy volume and sell volume and $\mathbf{X}_{t-}^2 = (E[\text{LogBidDepth}_{t-}], E[\text{LogAskDepth}_{t-}])$ is the row vector of shocks to quoted depths. Posterior mean and posterior standard deviation are computed using draws from the Gibbs sampler.

the structure of the estimated model, we test whether the specification of the (log) volatility process can be reduced to the form

$$[\ln(\sigma_t^2) - \eta_t - \mathbf{X}_{t-}^1 \mathbf{c}_1] (1 - \psi B) = \mathbf{X}_{t-}^2 \mathbf{c}_2 + \omega\epsilon_{t-1} + \gamma[|\epsilon_{t-1}| - E|\epsilon_{t-1}|], \quad (12)$$

that is, whether the impact of $\mathbf{X}_{t-}^1$ on the (log) volatility is only through the mean. Formally, we need to test $c_i\psi + c_{i+3} = 0$, for $i = 1, 2, 3$. Table 7 shows the testing results. We cannot reject the specification for all three variables at the 5% level.

The fact that bid–ask spreads and quoted depths are significant in the volatility equation together with volume shows that the Mixture of Distribution Hypothesis cannot be the whole story for time-varying volatility. Information based market microstructure models can provide richer insight on the determinants of this stochastically changing volatility.¹⁴

5. Conclusions

In this paper, we present an empirical microstructure model of intraday bid and ask quotes. The cost of quote exposure is allowed to depend on market information involving order flow and volatility. We find that an unusual excess of buyers relative to sellers tends to increase the ask price more than the bid price. Hence, the ask and bid components of spread change asymmetrically about the efficient price. For a fixed order imbalance surprise these effects are muted when unanticipated total volume is high. We also find that unexpected high volatility in the transaction price process tends to widen the spread symmetrically about the efficient price. Not surprisingly, our another result is that the cost of market making is lower when the expected depth is higher. Our findings are consistent with predication from market microstructure theory that the cost of market making should depend on both the risk of trading with better-informed traders and inventory risk.

We also find that the information contains in trading volume is multi-dimensional. The impacts of shocks to buy and sell volume are asymmetric on the market making costs. We offer a decomposition of the information in volume into two intuitive appealing components: a Pressure component, represented by the net order imbalance between unanticipated buy and sell volumes, and a Level component, represented by the unanticipated volume in the corresponding side (buy or sell) of the market. This provides a potential direction in theoretical research to reconcile competing theories regarding the relation between volume and quoted liquidity.

¹⁴ Lamoureux and Lastrapes (1990) find that daily trading volume as a proxy for information arrival time has significant explanatory power regarding the variance of daily returns. This supports the notion that Autoregressive Conditional Heteroskedasticity (ARCH) in daily stock return data reflects time dependence in the process generating information flow to the market.

Appendix A

The full dynamic of the proposed model is given by:

$$\left\{ \begin{array}{l} a_t = \lceil m_t + \alpha_t \rceil \\ b_t = \lfloor m_t - \beta_t \rfloor \\ \ln(\alpha_t) = \mu_t + \theta_t + \theta_t^\alpha + \phi[\ln(\alpha_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\alpha \\ \ln(\beta_t) = \mu_t + \theta_t + \theta_t^\beta + \phi[\ln(\beta_{t-1}) - \mu_{t-1}] + \sigma_v v_t^\beta \\ \mu_t = k_1 + k_2^{\text{open}} \exp(-k_3^{\text{open}} \tau_t^{\text{open}}) + k_2^{\text{close}} \exp(-k_3^{\text{close}} \tau_t^{\text{close}}) \\ \theta_t = \mathbf{X}_{t-1}^\theta \mathbf{d} \\ \theta_t^\alpha = \mathbf{X}_{t-1}^\alpha \mathbf{d}_\alpha + \mathbf{Z}_{t-1}^\alpha \mathbf{g} \\ \theta_t^\beta = \mathbf{X}_{t-1}^\beta \mathbf{d}_\beta + \mathbf{Z}_{t-1}^\beta \mathbf{g} \\ \ln(m_t) = \ln(m_{t-1}) + \delta_t + \sigma_t \epsilon_t \\ \delta_t = \mathbf{X}_t^m \mathbf{d}_m \\ \epsilon_t \sim \text{GED}(v^{\text{night}} 1_{\{\tau_t^{\text{open}}=0\}} + v^{\text{day}} 1_{\{\tau_t^{\text{open}}>0\}}) \\ \ln(\sigma_t^2) = \eta_t + \zeta_t + \psi[\ln(\sigma_{t-1}^2) - \eta_{t-1}] + \omega \epsilon_{t-1} + \gamma[|\epsilon_{t-1}| - E|\epsilon_{t-1}|] \\ \eta_t = \eta^{\text{night}} 1_{\{\tau_t^{\text{open}}=0\}} + l_1 + l_2^{\text{open}} \exp(-l_3^{\text{open}} \tau_t^{\text{open}}) + l_2^{\text{close}} \exp(-l_3^{\text{close}} \tau_t^{\text{close}}) \\ \zeta_t = \mathbf{X}_{t-1}^\zeta \mathbf{c} \end{array} \right.$$

where v_t^α and v_t^β are independent innovation distributed as $\mathcal{N}(0,1)$, and ϵ_t , the innovation for the efficient price process, is distributed as a generalized error distribution (GED) with a tail-thickness parameter v . ϵ_t is assumed to be independent with v_t^α and v_t^β , for all t .

Let $\Theta_1 = (\phi, \sigma_v^2, k_1, k_2^{\text{open}}, k_3^{\text{open}}, k_2^{\text{close}}, k_3^{\text{close}}, \mathbf{d}', \mathbf{d}_\alpha', \mathbf{d}_\beta', \mathbf{g}')'$ be the parameter vector for the cost model and $\Theta_2 = (\mathbf{c}_m', v^{\text{day}}, v^{\text{night}}, \psi, \omega, \gamma, \eta^{\text{night}}, l_1, l_2^{\text{open}}, l_3^{\text{open}}, l_2^{\text{close}}, l_3^{\text{close}}, \mathbf{c}')'$ be the parameter vector for the EGARCH/GED model. The parameter vector of the model can be written as $\Theta = (\Theta_1', \Theta_2')$. Let $s_t = (m_t, \alpha_t, \beta_t)'$ be the state variable of the model and $\mathbf{s} = (s_1, \dots, s_T)'$ be the state vector.

We use diffuse priors throughout the paper to make the statistical inference as objective as possible and to avoid any unwanted contamination from subjective priors. We draw samples from the posterior distribution $\Theta, \mathbf{s}|\mathbf{y}$ using the Gibbs sampler, which is implemented as: 1. Initialize $\mathbf{s}$; 2. Sample $\Theta \sim \Theta|\mathbf{s}, \mathbf{y}$; 3. Sample $\mathbf{s} \sim \mathbf{s}|\Theta, \mathbf{y}$; 4. Repeat 2 and 3.

Within the Gibbs steps 2 and 3, we need to decompose the conditional posterior distributions further under the Gibbs sampler framework. Some of the conditional posterior distributions are of unknown form. We sample from those distributions using more flexible methods such as Metropolis–Hasting algorithm of Metropolis et al. (1953) and Hastings (1970), Adaptive Rejection Metropolis Sampling (ARMS) method of Gilks et al. (1995) and the Griddy–Gibbs sampler of Ritter and Tanner (1992).

The draws from the algorithm will converge to unconditional draws from the distribution $\Theta, \mathbf{s}|\mathbf{y}$, under some weak regularity conditions (Tierney 1994). For the models in this paper, we run the algorithm for 10,000 iterations and drop the first 2000 runs as initial burn-ins. Formal inferences are based on the 8000 iterations left.¹⁵

We next present the details of the implementation of the Gibbs sampler. To facilitate the presentation, define $\boldsymbol{\mu} = (\mu_1, \dots, \mu_T)'$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_T)'$, $\boldsymbol{\theta}_\alpha = (\theta_1^\alpha, \dots, \theta_T^\alpha)'$ and $\boldsymbol{\theta}_\beta = (\theta_1^\beta, \dots, \theta_T^\beta)'$.

I. Sample Θ_1

We run the simulation according to the following sequence:

1. Sample $\mathbf{g}, \phi, \sigma_v^2$. Let

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{Z}_1^\alpha & \ln(\alpha_1) - \mu_1 \\ \vdots & \vdots \\ \mathbf{Z}_{T-1}^\alpha & \ln(\alpha_{T-1}) - \mu_{T-1} \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} \mathbf{Z}_1^\beta & \ln(\beta_1) - \mu_1 \\ \vdots & \vdots \\ \mathbf{Z}_{T-1}^\beta & \ln(\beta_{T-1}) - \mu_{T-1} \end{bmatrix}$$

¹⁵ Convergence checks are conducted using CODA, the Convergence Diagnosis and Output Analysis software for Gibbs sampler, ©MRC Biostatistics Unit.

and

$$Y_1 = \begin{bmatrix} \ln(\alpha_2) - \mu_2 - \theta_2 - \mathbf{X}_1^\alpha \mathbf{d}_\alpha \\ \vdots \\ \ln(\alpha_T) - \mu_T - \theta_T - \mathbf{X}_{T-1}^\alpha \mathbf{d}_\alpha \end{bmatrix}, \quad Y_2 = \begin{bmatrix} \ln(\beta_2) - \mu_2 - \theta_2 - \mathbf{X}_1^\beta \mathbf{d}_\beta \\ \vdots \\ \ln(\beta_T) - \mu_T - \theta_T - \mathbf{X}_{T-1}^\beta \mathbf{d}_\beta \end{bmatrix}.$$

Denote $\mathbf{h} = (\mathbf{g}', \phi)'$, we have $\mathbf{h} | \sigma_v^2, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{d}_\alpha, \mathbf{d}_\beta \sim \mathcal{N}(\bar{\mathbf{h}}, \sigma_h^2)$ where

$$\bar{\mathbf{h}} = (\mathbf{X}_1' \mathbf{X}_1 + \mathbf{X}_2' \mathbf{X}_2)^{-1} (\mathbf{X}_1' Y_1 + \mathbf{X}_2' Y_2), \sigma_h^2 = \sigma_v^2 (\mathbf{X}_1' \mathbf{X}_1 + \mathbf{X}_2' \mathbf{X}_2)^{-1},$$

and $\sigma_v^2 | \mathbf{h}, \boldsymbol{\mu}, \boldsymbol{\theta}, \mathbf{d}_\alpha, \mathbf{d}_\beta \sim \text{SSQ} \cdot \chi_{2(T-1)}^2 = \mathcal{IG}(T-1, \frac{\text{SSQ}}{2})$ where

$$\text{SSQ} = (Y_1 - \mathbf{X}_1 \mathbf{h})' (Y_1 - \mathbf{X}_1 \mathbf{h}) + (Y_2 - \mathbf{X}_2 \mathbf{h})' (Y_2 - \mathbf{X}_2 \mathbf{h}).$$

2. Sample $\boldsymbol{\kappa}_1 = (k_1, k_2^{\text{open}}, k_2^{\text{close}}, \mathbf{d}')'$. Let

$$X = \begin{bmatrix} 1 - \phi & e^{(-k_3^{\text{open}} \tau_2^{\text{open}})} - \phi e^{(-k_3^{\text{open}} \tau_1^{\text{open}})} & e^{(-k_3^{\text{close}} \tau_2^{\text{close}})} - \phi e^{(-k_3^{\text{close}} \tau_1^{\text{close}})} & \mathbf{X}_1^\theta \\ \vdots & \vdots & \vdots & \vdots \\ 1 - \phi & e^{(-k_3^{\text{open}} \tau_T^{\text{open}})} - \phi e^{(-k_3^{\text{open}} \tau_{T-1}^{\text{open}})} & e^{(-k_3^{\text{close}} \tau_T^{\text{close}})} - \phi e^{(-k_3^{\text{close}} \tau_{T-1}^{\text{close}})} & \mathbf{X}_{T-1}^\theta \end{bmatrix}$$

and

$$Y_1 = \begin{bmatrix} \ln(\alpha_2) - \theta_2^\alpha - \phi \ln(\alpha_1) \\ \vdots \\ \ln(\alpha_T) - \theta_T^\alpha - \phi \ln(\alpha_{T-1}) \end{bmatrix}, \quad Y_2 = \begin{bmatrix} \ln(\beta_2) - \theta_2^\beta - \phi \ln(\beta_1) \\ \vdots \\ \ln(\beta_T) - \theta_T^\beta - \phi \ln(\beta_{T-1}) \end{bmatrix}.$$

We have $\boldsymbol{\kappa}_1 | \phi, \sigma_v^2, k_3^{\text{open}}, k_3^{\text{close}}, \mathbf{d}_\alpha, \mathbf{d}_\beta, \mathbf{g} \sim \mathcal{N}(\bar{\boldsymbol{\kappa}}_1, \sum_{\boldsymbol{\kappa}_1})$ where $\bar{\boldsymbol{\kappa}}_1 = \frac{1}{2} (X'X)^{-1} X' (Y_1 + Y_2)$, and $\sum_{\boldsymbol{\kappa}_1} = \frac{1}{2} \sigma_v^2 (X'X)^{-1}$.

3. Sample $\boldsymbol{\kappa}_2 = (k_3^{\text{open}}, k_3^{\text{close}})'$. Since the posterior density of $\boldsymbol{\kappa}_2$ is not of a known form, we first use Adaptive Rejection Metropolis Sampling (ARMS) for a pilot run and then use a customized Metropolis–Hastings algorithm to formally sample from the posterior. We choose two log-normal distributions as the candidate-generating densities of the Metropolis–Hastings algorithm.

4. Sample $\mathbf{d}_\alpha, \mathbf{d}_\beta$. Let

$$X_1 = \begin{bmatrix} \mathbf{X}_1^\alpha \\ \vdots \\ \mathbf{X}_{T-1}^\alpha \end{bmatrix}, \quad Y_1 = \begin{bmatrix} \ln(\alpha_2) - \mu_2 - \theta_2 - \mathbf{Z}_1^\alpha \mathbf{g} - \phi(\ln(\alpha_1) - \mu_1) \\ \vdots \\ \ln(\alpha_T) - \mu_T - \theta_T - \mathbf{Z}_{T-1}^\alpha \mathbf{g} - \phi(\ln(\alpha_{T-1}) - \mu_{T-1}) \end{bmatrix}$$

and

$$X_2 = \begin{bmatrix} \mathbf{X}_1^\beta \\ \vdots \\ \mathbf{X}_{T-1}^\beta \end{bmatrix}, \quad Y_2 = \begin{bmatrix} \ln(\beta_2) - \mu_2 - \theta_2 - \mathbf{Z}_1^\beta \mathbf{g} - \phi(\ln(\beta_1) - \mu_1) \\ \vdots \\ \ln(\beta_T) - \mu_T - \theta_T - \mathbf{Z}_{T-1}^\beta \mathbf{g} - \phi(\ln(\beta_{T-1}) - \mu_{T-1}) \end{bmatrix}.$$

We have $\mathbf{d}_\alpha | \phi, \sigma_v^2, \boldsymbol{\mu}, \mathbf{d}, \mathbf{g} \sim \mathcal{N}(\bar{\mathbf{d}}_\alpha, \sum_{\mathbf{d}_\alpha})$ where $\bar{\mathbf{d}}_\alpha = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' Y_1$ and $\sum_{\mathbf{d}_\alpha} = \sigma_v^2 (\mathbf{X}_1' \mathbf{X}_1)^{-1}$, and $\mathbf{d}_\beta | \phi, \sigma_v^2, \boldsymbol{\mu}, \mathbf{d}, \mathbf{g} \sim \mathcal{N}(\bar{\mathbf{d}}_\beta, \sum_{\mathbf{d}_\beta})$ where $\bar{\mathbf{d}}_\beta = (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' Y_2$ and $\sum_{\mathbf{d}_\beta} = \sigma_v^2 (\mathbf{X}_2' \mathbf{X}_2)^{-1}$.

II. Sample $\boldsymbol{\Theta}_2$

Given the efficient price sequence $\{m_t\}_{t=1}^T$, the volatilities are observable conditional on the parameters of the EGARCH/GED submodel. Assuming $\ln(\sigma_1^2) = \eta_1$, we obtain the volatility series using the following recursive equations:

$$\begin{cases} \epsilon_{t-1} &= \xi_{t-1} / \sigma_{t-1} \\ \ln(\sigma_t^2) &= \eta_t + \xi_t + \psi [\ln(\sigma_{t-1}^2) - \eta_{t-1}] + \omega \epsilon_{t-1} + \gamma [|\epsilon_{t-1}| - E|\epsilon_{t-1}|] \end{cases}$$

where $\xi_t = \ln(m_t) - \ln(m_{t-1}) - \delta_t$. Hence, the log likelihood function of the EGARCH/GED model can be written as

$$\ln L(\Theta_2 | m_1, \dots, m_T) = \sum_{t=2}^T \left[\ln f_{\text{night}} \left(\frac{\xi_t}{\sigma_t} \right) 1_{\{\tau_t^{\text{open}}=0\}} + \ln f_{\text{day}} \left(\frac{\xi_t}{\sigma_t} \right) 1_{\{\tau_t^{\text{open}}>0\}} - \ln(\sigma_t) \right].$$

We use Adaptive Rejection Metropolis Sampling (ARMS) to draw from the posterior distribution of each element of in the pilot run. We use a specially tuned Metropolis–Hastings algorithm in our formal estimation.

III. Sample s

To sample from $s|\Theta; \mathbf{y}$, we draw from conditional posterior distribution $s_t|s_{\setminus t}, \Theta; \mathbf{y}$ for $t=1, \dots, T$, where $s_{\setminus t} = (s_1, \dots, s_{t-1}, s_{t+1}, \dots, s_T)$. The feasible set of s_t conditional on the observed bid–ask quote $y_t = (b_t, \alpha_t)'$ is given by

$$\mathcal{Q}(b_t, \alpha_t) = \{(m_t, \alpha_t, \beta_t) : \alpha_t > 0; \beta_t > 0; b_t \leq m_t - \beta_t < b_t + 1; \alpha_t - 1 < m_t + \alpha_t \leq a_t\}.$$

This feasible set effectively imposes a truncation to the distribution $s_t|s_{\setminus t}, \Theta$. Following Hasbrouck (1999b) we sample s_t according to the following sequence,

$$[m_t | \alpha_t, \beta_t, s_{t-1}, s_{t+1}, \Theta; y_t], \quad [\alpha_t | m_t, s_{t-1}, s_{t+1}, \Theta; y_t], \quad [\beta_t | m_t, s_{t-1}, s_{t+1}, \Theta; y_t].$$

We still have a truncation set for m_t conditional on y_t and (α_t, β_t) . The truncation set is given by $M_t = (\underline{m}_t, \bar{m}_t)$, where $\underline{m}_t = \max\{b_t + \beta_t, \alpha_t - 1 - \alpha_t\}$ and $\bar{m}_t = \min\{b_t + 1 + \beta_t, \alpha_t - \alpha_t\}$. Thus, we can draw m_t using the kernel $f(m_t | m_{t-1}, m_{t+1}, \Theta)$. $1M_t$, where $f(m_t | m_{t-1}, m_{t+1}, \Theta)$ is the conditional density of m_t given m_{t-1} and m_{t+1} .

Similarly, the truncation sets for α_t and β_t conditional on y_t and m_t are given by $A_t = (\underline{\alpha}_t, \bar{\alpha}_t]$, where $\underline{\alpha}_t = \max\{0, \alpha_t - 1 - m_t\}$ and $\bar{\alpha}_t = \alpha_t - m_t$, and $B_t = (\underline{\beta}_t, \bar{\beta}_t]$, where $\underline{\beta}_t = \max\{0, m_t - b_t - 1\}$ and $\bar{\beta}_t = m_t - b_t$.

1. Sample $\mathbf{m}$. We use Griddy–Gibbs to sample from the posterior distribution of each m_t .

$$\begin{aligned} \ln(m_1) | \ln(m_2), \Theta &\propto \frac{1}{\sigma_2(m_1)} f_2 \left(\frac{\ln(m_2) - \ln(m_1)}{\sigma_2(m_1)} \right) \\ \ln(m_t) | \ln(m_{t-1}), \ln(m_{t+1}), \Theta &\propto \frac{1}{\sigma_{t+1}(m_t)} f_{t+1} \left(\frac{\ln(m_{t+1}) - \ln(m_t)}{\sigma_{t+1}(m_t)} \right) \\ &\quad f_t \left(\frac{\ln(m_t) - \ln(m_{t-1})}{\sigma_t} \right), \quad t = 2, \dots, T-1 \\ \ln(m_T) | \ln(m_{T-1}) &\propto f_T \left(\frac{\ln(m_T) - \ln(m_{T-1})}{\sigma_T} \right) \end{aligned}$$

Where

$$\begin{cases} \ln(\sigma_{t+1}^2(m_t)) = \eta_{t+1} + \zeta_{t+1} + \psi[\ln(\sigma_t^2) - \eta_t] + \omega\epsilon_t + \gamma[|\epsilon_t| - E|\epsilon_t|] \\ \epsilon_t = [\ln(m_t) - \ln(m_{t-1})]/\sigma_t \end{cases}$$

2. Sample α .

$$\begin{aligned} \ln(\alpha_1) | \ln(\alpha_2) &\sim \mathcal{N} \left(\mu_1 + \frac{1}{\phi} (\ln(\alpha_2) - \mu_2 - \theta_2^{\alpha} - \theta_2), \sigma_v^2 \right), \\ \ln(\alpha_t) | \ln(\alpha_{t-1}), \ln(\alpha_{t+1}) &\sim \mathcal{N} \left(\mu_t + \frac{\mu_t^{\alpha}}{1 + \phi^2}, \frac{\sigma_v^2}{1 + \phi^2} \right), \quad t = 2, \dots, T-1, \\ \ln(\alpha_T) | \ln(\alpha_{T-1}) &\sim \mathcal{N} (\mu_T + \theta_T^{\alpha} + \theta_T + \phi(\ln(\alpha_{T-1}) - \mu_{T-1}), \sigma_v^2), \end{aligned}$$

where $\mu_t^{\alpha} = \phi[(\ln(\alpha_{t-1}) - \mu_{t-1}) + (\ln(\alpha_{t+1}) - \mu_{t+1})] + \theta_t^{\alpha} + \theta_t - \phi(\theta_{t+1}^{\alpha} + \theta_{t+1})$.

3. Sample β . It is symmetric to α .

References

- Admati, A.R., Peiderer, P., 1988. A theory of intraday patterns: volume and price variability. *Review of Financial Studies* 1, 3–40.
- Amihud, Y., Mendelson, H., 1980. Dealership markets: market making with uncertainty. *Journal of Financial Economics* 8, 31–54.
- Andersen, T.G., 1996. Return volatility and trading volume: an information of interpretation of stochastic volatility. *Journal of Finance* 51 (1), 169–204.
- Ball, C.A., 1988. Estimation bias introduced by discrete security prices. *Journal of Finance* 43, 841–865.
- Ball, C.A., Chordia, T., 2001. True spreads and equilibrium prices. *Journal of Finance* 56, 1801–1835.
- Brock, W.A., Kleidon, A.W., 1992. Periodic market closure and trading volume. *Journal of Economic Dynamics and Control* 16, 451–489.
- Campbell, J.Y., Grossman, S.J., Wang, J., 1993. Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics* 108, 905–934.
- Chan, K., Fong, W.M., 2000. Trade size, order imbalance, and the volatility–volume relation. *Journal of Financial Economics* 57, 247–273.
- Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. *Journal of Finance* 47, 577–606.
- Epps, T.W., Epps, M.L., 1976. The stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distribution hypothesis. *Econometrica* 44 (2), 305–321.
- Foster, F.D., Viswanathan, S., 1990. A theory of interday variations in volumes, variances, and trading costs in securities markets. *Review of Financial Studies* 3 (4), 593–624.
- Gallant, A.R., Rossi, P.E., Tauchen, G., 1992. Stock prices and volume. *Review of Financial Studies* 5, 199–242.
- Gilks, W.R., Best, N.G., Tan, K.K.C., 1995. Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* 44, 155–173.
- Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.
- Goodhart, C.A.E., O'Hara, M., 1997. High frequency data in financial markets: issues and applications. *Journal of Empirical Finance* 4, 73–114.
- Gottlieb, G., Kalay, A., 1985. Implications of the discreteness of observed stock prices. *Journal of Finance* 40, 135–153.
- Harris, L.E., 1990. Estimation of stock price variance and serial covariances from discrete observations. *Journal of Financial and Quantitative Analysis* 25 (3), 291–306.
- Harris, M., Raviv, A., 1993. Difference of opinion make a horse race. *Review of Financial Studies* 6, 473–506.
- Hasbrouck, J., 1988. Trades, quotes, inventories and information. *Journal of Financial Economics* 22, 229–252.
- Hasbrouck, J., 1991. Measuring the information content of stock trades. *Journal of Finance* 46 (1), 179–207.
- Hasbrouck, J., 1999a. The dynamics of discrete bid and ask quotes. *Journal of Finance* 54 (6), 2109–2142.
- Hasbrouck, J., 1999b. Security bid/ask dynamics with discreteness and clustering: simple strategies for modeling and estimation. *Journal of Financial Markets* 2, 1–28.
- Hasbrouck, J., Sofianos, G., 1993. The trades of market makers: an empirical analysis of NYSE specialists. *Journal of Finance* 48, 1565–1593.
- Hastings, W.K., 1970. Monte-Carlo sampling methods using Markov chain and their applications. *Biometrika* 57, 97–109.
- Ho, T.S.Y., Stoll, H.R., 1983. The dynamics of dealer markets under competition. *Journal of Finance* 38, 1053–1074.
- Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* 12, 371–417.
- Jain, P.C., Joh, G.H., 1988. The dependence between hourly prices and trading volume. *Journal of Financial and Quantitative Analysis* 23, 269–283.
- Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.
- Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series. *Journal of American Statistical Association* 82, 1032–1041.
- Kyle, A., 1985. Continuous time auctions and insider trading. *Econometrica* 53, 1315–1336.
- Lamoureux, C.G., Lastrapes, W.D., 1990. Heteroskedasticity in stock return data: volume versus GARCH effects. *Journal of Finance* 45 (1), 221–229.
- Lee, C.M., Mucklow, B., Ready, M.J., 1993. Spreads, depths and the impact of earnings information: an intraday analysis. *Review of Financial Studies* 6 (2), 345–374.
- Lee, C.M., Ready, M.J., 1991. Inferring trade direction from intraday data. *Journal of Finance* 46 (6), 733–746.
- Madhavan, A., Smidt, S., 1993. An analysis of changes in specialist inventories and quotations. *Journal of Finance* 48, 1595–1628.
- Manrique, A., Shephard, N., 1997. Likelihood analysis of a discrete bid/ask price model for a common stock. Working Paper. Nuffield College, Oxford University.
- Manrique, A., Shephard, N., 1998. Simulation-based likelihood inference for limited dependent processes. *Econometrics Journal* 1, C174–C202.
- McInish, T.H., Wood, R.A., 1992. An analysis of intraday patterns in bid/ask spreads for NYSE stocks. *Journal of Finance* 47, 753–764.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21, 1087–1092.
- Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.
- O'Hara, M., 1995. *Market Microstructure Theory*. Basil Blackwell.
- O'Hara, M., Oldfield, G.S., 1986. The microeconomics of market making. *Journal of Financial and Quantitative Analysis* 21, 361–376.
- Ritter, C., Tanner, M.A., 1992. Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy–Gibbs sampler. *Journal of American Statistical Association* 87, 861–868.
- Stoll, H.R., Whaley, R.E., 1990. Stock market structure and volatility. *Review of Financial Studies* 3, 37–71.
- Tauchen, G.E., Pitts, M., 1983. The price variability–volume relationship on speculative markets. *Econometrica* 51 (2), 485–505.
- Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* 22, 1701–1762.