To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–29

Trading duration, volume and volatility: A vector multiplicative error model approach

N. Taylor*† and Y. Xu‡

†School of Economics, Finance and Management, University of Bristol, Bristol, BS8 1TU, UK.
‡Cardiff Business School, Cardiff University, Cardiff, CF10 1EU, UK.

(*Received 00 Month 20XX; in final form 00 Month 20XX*)

We develop a general form logarithmic vector multiplicative error model (log-VMEM) to examine the interdependence of trading duration, volume and return volatility. The log-VMEM improves on existing models in two ways. First, it allows the error terms to be cross-dependent and by relaxing weak exogeneity restrictions and hence it is a more general form model (cf. the recursive model of Manganelli, 2005). Second, the log-VMEM specification guarantees that the conditional means are nonnegative without any restrictions imposed on the parameters (cf. the VMEM of Cipollini el al., 2013). We further propose a multivariate lognormal distribution and a joint maximum likelihood (ML) estimation strategy. The model is applied to high frequency data associated with a number of NYSE-listed stocks. The results reveal empirical support for full interdependence among the trivariate processes and show that the log-VMEM provides a better fit to the data. Moreover, we find that unexpected duration and volume dominate observed duration and volume in terms of information content, and that volatility and volatility shocks affect duration in different directions. These results are interpreted with reference to extant microstructure theory.

Keywords: VMEM; ACD; intraday trading process; duration; volume; volatility.

JEL Classification: C32, C52, G14.

1. Introduction

Microstructure theory indicates that trading duration and volume convey information content with respect to fundamental asset prices; see, e.g., Easley and O'Hara (1992).[1] Such theory has, in turn, motivated a large number of empirical studies that investigate the information content of trading via an analysis of the dynamics of trading duration, volume and volatility; see, e.g., Manganelli (2005). We add to this body of work by specifying a flexible econometric model that avoids a number of restrictions imposed in existing models. In doing this, further insight is gained into the validity of extant microstructure theory.

The theoretical relationship between duration, volume and volatility has been the subject of considerable debate. In Diamond and Verrecchia's (1987) model, informed trader's actions are driven by the arrival of private information. If the news is bad, informed traders will wish to sell (or alternatively short sell if they do not own the stock). Given short-sale constraints there may be no trade. This is summarized as "no trade means bad news", with the empirical implication that long durations are associated with bad news that should lead to price adjustment and hence increased

*Corresponding author. Email: nick.taylor@bristol.ac.uk
[1] Trading duration is defined as the elapsed time between two consecutive transactions. For the sake of brevity we henceforth refer to trading duration as duration, trading volume as volume, and return volatility as volatility.

volatility. By contrast, Easley and O'Hara (1992) argue that informed traders only trade when there is new information (whether good or bad) arriving in the market. Consequently, variation in trading intensity is closely related to the change in the participation rate of informed traders. This is summarized as "no trade means no news", and implies that short duration is a signal of informed traders participating in the market, which manifests itself in higher volatility.[2]

The above predictions can be investigated using an augmentation of the autoregressive conditional duration (ACD) econometric model initially proposed by Engle and Russell (1998). Specifically, the trade characteristics associated with time are incorporated and modelled simultaneously, so that market microstructure predictions can be evaluated at the transaction level. For instance, Engle (2000) proposes a recursive framework to represent the dynamics of duration and volatility. In this model the joint density of duration and volatility is expressed as the product of the marginal density of duration and the conditional density of volatility (given duration). A further augmentation to the ACD model is proposed by Manganelli (2005), whereby volume is incorporated into the Engle (2000) model such that duration, volume and volatility are jointly considered. Moreover, the joint distribution of duration, volume and volatility is decomposed into the product of the marginal distribution of duration; the conditional distribution of volume (given duration); and the conditional distribution of volatility (given duration and volume). Further assumptions of weak exogeneity are made, such that the three processes are independent and can be estimated separately.

The virtue of the recursive framework of Engle (2000) and Manganelli (2005) is that it reduces the complexity of the model, since each process is estimated separately. However, recursive-type models have limitations. First, it assumes that the error terms are independent. To incorporate contemporaneous information, Manganelli (2005) specifies causality from duration to volume and from duration and volume to volatility. However, modelling volatility conditional on duration and volume is just one way to obtain the joint distribution of the trading processes. As pointed out by Engle and Sun (2007), it is also possible to model duration conditional on volatility. Other empirical studies have also assumed the interdependence of the trading processes. For instance, Hautsch (2008) has found the existence of a common unobserved component that jointly drives the dynamics of the trade and return processes. This common component explains most of the causality between these processes, even if the contemporaneous effect of the trade variable on the return variable is controlled.

A second issue concerns the weak exogeneity assumption made in previous studies. This means that the conditional expectation of one variable is a function only of its own past conditional expectation, while the past conditional expectations of other variables are not taken into consideration. This strategy has been adopted in many empirical microstructure studies; see, e.g., Engle (2000), Dufour and Engle (2000), Manganelli (2005), and Engle and Sun (2007). However, we argue that this assumption is too restrictive. When studying the information content of trading processes, various specifications of duration and volume should be considered. For example, trade innovation may be an exclusive manifestation of the private information of the informed trader. Indeed, Engle (2000) and Grammig et al. (2007) argue that it is the unexpected components of the trading process that carry informational content with respect to the fundamental asset price; also see Grammig and Wellner (2002) for empirical evidence that both volatility *and* volatility shocks have significant effects on trading intensity.

Given the above arguments it seems advisable to release weak exogeneity restrictions, and specify an interdependent model for the dynamics of duration, volume and volatility. To this end we consider a general form vector multiplicative error model (VMEM) in which duration, volume and volatility are assumed to be interdependent processes that evolve simultaneously. We contribute to

[2]A relationship between duration and volatility is also predicted by Admati and Pfleiderer (1988). They assume that frequent trading is associated with liquidity traders, and therefore low trading means that liquidity (discretionary) traders are inactive, which leaves a high proportion of informed traders in the market. This again translates into quick price adjustment and hence higher volatility.

the existing VMEM literature in two ways. First, the specification of the conditional distribution of the nonnegative-valued errors in the VMEM is open to debate. Cipollini et al. (2007) initially consider the multivariate gamma distribution, but find it too restrictive as it only allows positive error correlation. They further propose a copula-based distribution for the errors. However, they have to restrict the coefficient matrix to be diagonal in order to achieve full ML estimation. Cipollini et al. (2013) bypass the specification of the error distribution and only make use of the first two error moments. Specifically, they propose an unrestricted semiparametric VMEM and a consistent general method of moments (GMM) estimation methodology. In doing this they partially solve the above estimation issues. Within this context our contribution to the VMEM literature is to propose a multivariate lognormal distribution for the errors in combination with a full ML estimation methodology. Furthermore, we compare our existing methodology to that of Cipollini et al. (2013) via a simulation experiment, and show that the two methods are consistent. The efficiency loss of our estimation methodologies due to misspecification of the error distribution is trivial.

The second contribution relates to the nonnegativity of the conditional means of the variables under study. A sufficient condition for nonnegativity of the conditional means is that all parameters are nonnegative (He and Teräsvirta, 2004). This is a very strong restriction. Previous studies such as Manganelli (2005), Engle and Gallo (2006), and Cipollini et al. (2007, 2013) choose not to impose this restriction. However, their results show that some of their estimated parameters are negative – thus violating the nonnegativity condition and implying that predicted values could be negative. Given the nature of the data used such predictions are implausible. To avoid this issue, we consider a log-VMEM, such that the conditional means are guaranteed to be positive without any restrictions placed on the parameters. By using this model, we build a system that incorporates various causal and feedback effects among these variables.

The proposed model is applied to trade and quote NYSE stock data, and is estimated using a sample of 20 stocks observed over two different time periods. In doing this we are able to empirically study the information conveyed by trading processes over a variety of conditions. Our empirical findings are summarized as follows. First, it is recognized that duration and volume contain information content with respect to the fundamental asset value. However, the theoretical work does not draw any conclusion as to whether it is duration (volume) or the duration (volume) shock that contains information content. Based on our model, we find that both duration (volume) and duration (volume) shocks have a significant impact on volatility. Moreover, it is shocks that appear dominant – suggesting that it is unexpected components of duration or volume rather than observed duration or volume that contains information content. Second, our results show that volatility and volatility shocks affect duration in different directions – a result consistent with Hasbrouck's (1988, 1991) prediction that persistent quote changes are driven by private information, and transient quote changes are due to inventory considerations.

The remainder of this paper is organized as follows. Section 2 describes the model and methodology. Section 3 contains the application. Section 4 concludes.

2. Econometric model details

This section contains details of the proposed econometric models and how they are estimated.

2.1. *VMEM specification*

Define $\{d_t, v_t, |r_t|\}$, $t = 1, \ldots, T$, as the three-dimensional time series associated with the intraday duration, volume and volatility processes, respectively. In particular, duration is defined as the time elapsing between consecutive trades, volume is the trade size associated with each transaction, volatility is measured by the absolute return, and returns are calculated using the mid-quote price change. The trivariate trading processes are modelled as follows:

$$\{d_t, v_t, |r_t|\} \sim f(d_t, v_t, |r_t| | \Omega_t, \boldsymbol{\theta}), \tag{1}$$

where Ω_t denotes the information available up to period t, and $\boldsymbol{\theta}$ is a vector incorporating the parameters of interest.

Letting $\mathbf{x}_t = (d_t, v_t, |r_t|)'$, $\boldsymbol{\mu}_t = (\mu_{1t}, \mu_{2t}, \mu_{3t})'$, and $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t})'$, we follow Engle (2002) and Cipollini et al. (2007), and assume the following trivariate VMEM for $\mathbf{x}_t$, $\boldsymbol{\mu}_t$, and $\boldsymbol{\epsilon}_t$:

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathrm{D}(\mathbf{1}, \boldsymbol{\Sigma}), \tag{2}$$

where

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{x}_{t-i} + \sum_{i=1}^{q} \mathbf{B}_i \boldsymbol{\mu}_{t-i}. \tag{3}$$

The error vector $\boldsymbol{\epsilon}_t$ has support over $[0, +\infty)$, with a unit mean vector $\mathbf{1}$ and general variance-covariance matrix $\boldsymbol{\Sigma}$. The first two conditional moments of the VMEM are given by $\mathrm{E}(\mathbf{x}_t|\Omega_t) = \boldsymbol{\mu}_t$ and $\mathrm{var}(\mathbf{x}_t|\Omega_t) = \boldsymbol{\mu}_t\boldsymbol{\mu}_t' \odot \boldsymbol{\Sigma}$, with the latter a positive definite matrix by construction.

A complete parametric formulation of the VMEM requires full specification of the conditional distribution of the nonnegative random process $\boldsymbol{\epsilon}_t$. In the VMEM literature, Cipollini et al. (2007) adopt a copula-based approach. However, to enable use of ML estimation, the $\mathbf{B}$ matrix has to be diagonal. To avoid this restriction, Cipollini et al. (2013) propose a semiparametric VMEM in which GMM estimation (based on the first two moments) is employed. Our VMEM is close to Cipollini et al. (2013), in the sense that both approaches consider a general form VMEM. However, we propose to use a multivariate lognormal distribution for the errors, so that we have a full parametric VMEM.

In the univariate setting, Xu (2013) finds that the lognormal ACD model is superior to the exponential and Weibull ACD models, while its performance is similar to the Burr and generalized gamma ACD models. It is also well known that volatility is typically lognormally distributed, with Cizeau et al. (1997) and Andersen et al. (2001), among others, showing that the lognormal distribution fitted to realized volatility performs very well. Moreover, Allen et al. (2008) prove that the lognormal distribution is sufficiently flexible to provide a good approximation to a wide range of nonnegative distributions, and is also sufficiently accurate so as not to induce unnecessary numerical difficulties.

The prime limitation of the lognormal VMEM is that it assumes that the variables are positively valued. This limitation may explain why the lognormal distribution assumption is less popular in the ACD/MEM literature. Fortunately, most time series considered in extant empirical analysis (for example, duration, number of trades, volume, bid-ask spread, realised volatility, and daily range) are actually positively valued. The only exception is the absolute return. To eliminate exact zero problems in the current application, we add a small constant (10 percent of the mean) to the absolute return series.[1] This simple approach alleviates exact zero problems, but does not adversely affect the dynamics of absolute returns.

[1] In a previous version of this paper, we model the return series as an ARMA process and use the absolute value of return residuals rather than absolute returns as our proxy for volatility. This is also the approach used by Ghysels et al. (2004) to obtain return sequences that are free of bid-ask bounce effects. Results associated with this approach deliver similar results, and are available on request.

2.2. *Log-VMEM specification*

Another issue concerns the nonnegativity of the conditional mean vector $\boldsymbol{\mu}_t$. The multivariate specification has to be parameterized in a way that guarantees the nonnegativity of the conditional mean at all points in time. Since the VMEM has the same structure as the extended constant conditional correlation GARCH (ECCC-GARCH) model, the theoretical results on the positivity of this model can be applied here; see He and Teräsvirta (2004) and Conrad and Karanasos (2010). In general, a sufficient condition to guarantee nonnegativity of $\boldsymbol{\mu}_t$ is $\boldsymbol{\omega} \geq 0, \mathbf{A} \geq 0, \mathbf{B} \geq 0$ (He and Teräsvirta, 2004). Conrad and Karanasos (2010) release this restriction by allowing only one element of $\mathbf{B}$ to be negative. However, it is still highly restrictive. One could follow Manganelli (2005), Engle and Gallo (2006), and Cipollini et al. (2007, 2013) and not impose nonnegativity conditions. However, this could deliver negative estimated parameters and thus negative predicted values of the trivariate series. Given the nature of the data used such predictions are implausible. Motivated by the log-ACD model of Bauwens and Giot (2000), and Bauwens et al. (2008), we extend the VMEM into a logarithmic version, so that the conditional mean is guaranteed to be positive without any restrictions placed on the parameters. Specifically,

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathrm{D}(\mathbf{1}, \boldsymbol{\Sigma}) \tag{4}$$

$$\ln \boldsymbol{\mu}_t = \boldsymbol{\omega} + \sum_{i=1}^{p} \mathbf{A}_i \ln \mathbf{x}_{t-i} + \sum_{i=1}^{q} \mathbf{B}_i \ln \boldsymbol{\mu}_{t-i}. \tag{5}$$

This is referred to as the log-VMEM. In proposing this model we can incorporate various causal and feedback effects among the variables, while ensuring that their conditional expectations are always positive. The stationarity and invertibility conditions for the VMEM and log-VMEM are provided in Appendix A, while impulse response functions associated with first-order versions of these models are provided in Appendix B.

2.3. *The ML estimator and its asymptotic properties*

The following ML estimator is proposed:

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} -l(\boldsymbol{\theta}), \tag{6}$$

where $l(\boldsymbol{\theta})$ is the log likelihood function. To derive an expression for $l(\boldsymbol{\theta})$ we first note that by assumption the K-dimension vector $\boldsymbol{\epsilon}_t$ follows a multivariate lognormal distribution such that $\boldsymbol{\epsilon}_t \sim \ln \mathrm{N}(\mathbf{M}, \mathbf{V})$, with its density function given by

$$f(\boldsymbol{\epsilon}_t) = (2\pi)^{-K/2} |\mathbf{V}|^{-1/2} \prod_{i=1}^{K} \epsilon_{i,t} \exp(-(\ln \boldsymbol{\epsilon}_t - \mathbf{M})' \mathbf{V}^{-1} (\ln \boldsymbol{\epsilon}_t - \mathbf{M})/2). \tag{7}$$

It follows that the conditional density of $\mathbf{x}_t$ will be

$$f(\mathbf{x}_t|\boldsymbol{\theta}) = (2\pi)^{-K/2} |\mathbf{V}|^{-1/2} \prod_{i=1}^{K} x_{i,t} \exp(-(\ln \mathbf{x}_t - \ln \boldsymbol{\mu}_t - \mathbf{M})' \mathbf{V}^{-1} (\ln \mathbf{x}_t - \ln \boldsymbol{\mu}_t - \mathbf{M})/2). \tag{8}$$

The log likelihood of the model is then

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln f(\mathbf{x}_t|\boldsymbol{\theta}), \tag{9}$$

where

$$\ln f(\mathbf{x}_t|\boldsymbol{\theta}) = -\frac{K}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{V}| - \sum_{i=1}^{K}\ln x_{i,t} - \frac{1}{2}(\ln\mathbf{x}_t - \ln\boldsymbol{\mu}_t - \mathbf{M})'\mathbf{V}^{-1}(\ln\mathbf{x}_t - \ln\boldsymbol{\mu}_t - \mathbf{M}). \tag{10}$$

Note that imposing $M_i = -V_{ii}/2$ ensures that $\mathrm{E}(\boldsymbol{\epsilon}_t) = \mathbf{1}$.

Explicit expressions for the score vector and Hessian matrix in the current context are contained in the following lemmas. These assume that $\boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{\rho}]$, where $\boldsymbol{\rho} = \mathrm{vech}(\mathbf{V})$, and $\boldsymbol{\phi}_t = \ln\boldsymbol{\mu}_t + \mathbf{M}$. Here $\boldsymbol{\beta}$ contains the parameters in $\boldsymbol{\mu}_t$ and $\mathbf{M}$, and the vech operator stacks the lower triangular elements of the symmetric $(K \times K)$ $\mathbf{V}$ matrix into the $(K \times (K+1)/2)$ $\boldsymbol{\rho}$ vector.

Lemma 2.1 *The score vector associated with observation t is given by*

$$S_t(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} \\ \frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\rho}} \end{bmatrix}$$

where

$$\begin{aligned}
\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} &= -\frac{\partial\boldsymbol{\phi}_t'}{\partial\boldsymbol{\beta}}\mathbf{V}^{-1}(\ln\mathbf{x}_t - \boldsymbol{\phi}_t), \\
\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\rho}} &= -\frac{1}{2}\frac{\partial\mathrm{vec}(\mathbf{V})'}{\partial\boldsymbol{\rho}}\mathrm{vec}(\mathbf{V}^{-1} - \mathbf{V}^{-1}(\ln\mathbf{x}_t - \boldsymbol{\phi}_t)(\ln\mathbf{x}_t - \boldsymbol{\phi}_t)'\mathbf{V}^{-1}).
\end{aligned}$$

Proof. Standard vector/matrix differentiation of the log likelihood function in (9) eventually leads to the above expression. □

Lemma 2.2 *The Hessian matrix associated with observation t is given by*

$$H_t(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}} & \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\rho}'\partial\boldsymbol{\beta}} \\ \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\rho}} & \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\rho}'\partial\boldsymbol{\rho}}, \end{bmatrix}$$

where

$$\begin{aligned}
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}} &= -\Big((\mathbf{V}^{-1}(\ln\mathbf{x}_t - \boldsymbol{\phi}_t))' \otimes \mathbf{I}_K\Big)\frac{\partial^2\boldsymbol{\phi}_t'}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}} + \frac{\partial\boldsymbol{\phi}_t'}{\partial\boldsymbol{\beta}}\mathbf{V}^{-1}\frac{\partial\boldsymbol{\phi}_t}{\partial\boldsymbol{\beta}'}, \\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\rho}} &= \frac{1}{2}\frac{\partial\mathrm{vec}(\mathbf{V})'}{\partial\boldsymbol{\rho}}\Big((\mathbf{V}^{-1} \otimes \mathbf{V}^{-1})(\mathbf{I}_K \otimes (\ln\mathbf{x}_t - \boldsymbol{\phi}_t) + (\ln\mathbf{x}_t - \boldsymbol{\phi}_t) \otimes \mathbf{I}_K)\Big)\frac{\partial\boldsymbol{\phi}_t}{\partial\boldsymbol{\beta}'}, \\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\rho}'\partial\boldsymbol{\beta}} &= \frac{1}{2}\frac{\partial\boldsymbol{\phi}_t'}{\partial\boldsymbol{\beta}}\Big((\mathbf{I}_K \otimes (\ln\mathbf{x}_t - \boldsymbol{\phi}_t)' + (\ln\mathbf{x}_t - \boldsymbol{\phi}_t)' \otimes \mathbf{I}_K)(\mathbf{V}^{-1} \otimes \mathbf{V}^{-1})\Big)\frac{\partial\mathrm{vec}(\mathbf{V})}{\partial\boldsymbol{\rho}'}, \\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\rho}'\partial\boldsymbol{\rho}} &= \frac{1}{2}\frac{\partial\mathrm{vec}(\mathbf{V})'}{\partial\boldsymbol{\rho}}\Big((\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) - (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}(\ln\mathbf{x}_t - \boldsymbol{\phi}_t)(\ln\mathbf{x}_t - \boldsymbol{\phi}_t)'\mathbf{V}^{-1}) \\
&\quad - (\mathbf{V}^{-1}(\ln\mathbf{x}_t - \boldsymbol{\phi}_t)(\ln\mathbf{x}_t - \boldsymbol{\phi}_t)'\mathbf{V}^{-1} \otimes \mathbf{V}^{-1})\Big)\frac{\partial\mathrm{vec}(\mathbf{V})}{\partial\boldsymbol{\rho}'}.
\end{aligned}$$

Proof. Standard vector/matrix differentiation of the log likelihood function in (9) eventually leads to the above expression. □

The consistency and asymptotic normality of the ML estimator $\widehat{\boldsymbol{\theta}}$ follows from a more general maximum likelihood theory and can be found in Ling and McAleer (2003) and Lütkepohl (2005). We make use of the specific results in Nakatani and Teräsvirta (2009). Under certain regularity assumptions, the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ is given by[1]

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, J^{-1}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)J^{-1}(\boldsymbol{\theta}_0)), \tag{11}$$

where the population information matrix is given by the expectation of the outer product of the score vector evaluated at the true parameter vector $\boldsymbol{\theta}_0$, that is,

$$I(\boldsymbol{\theta}_0) = \frac{1}{T}\mathrm{E}(S(\boldsymbol{\theta}_0)S(\boldsymbol{\theta}_0)') = \mathrm{E}(S_t(\boldsymbol{\theta}_0)S_t(\boldsymbol{\theta}_0)'), \tag{12}$$

and the negative of the expected Hessian of the log likelihood function at $\boldsymbol{\theta}_0$ is given by

$$J(\boldsymbol{\theta}_0) = -\frac{1}{T}\mathrm{E}(H(\boldsymbol{\theta}_0)) = -\mathrm{E}(H_t(\boldsymbol{\theta}_0)). \tag{13}$$

The $I(\boldsymbol{\theta}_0)$ vector and $J(\boldsymbol{\theta}_0)$ matrix can be consistently estimated by their sample counterparts.

2.4. *A comparison with GMM estimation*

The general form VMEM can be estimated using the GMM approach of Cipollini et al. (2013). Specifically, they propose semiparametric GMM estimation based on the first two moment conditions. While both ML and GMM estimators have good asymptotic properties (consistent and efficient) relative to estimators based on the equation-by-equation approach, differentiating between them is an open question that ultimately depends on the nature of the data.

To compare the performance of the two approaches, we conduct a 1000-repetition Monte Carlo simulation experiment. We adopt the bivariate VMEM given by (2) and (3) with $p = q = 1$ and a sample size of 5000 observations. The sign (and size) of the parameters broadly coincide with the empirical relations between duration and volatility found by extant studies (including our own). Use of these particular values also ensures that the conditional expectations are positive (the conditions for positive definiteness of the bivariate VMEM are given in Conrad and Karanasos, 2010).

The disturbance term $\boldsymbol{\epsilon}_t$ is generated under two different distributional assumptions. First, we use the copula approach of Cipollini et al. (2013). This approach can be broken down into two steps. The marginal distribution is simulated from a gamma distribution with unit mean and standard deviation. Then we choose a Gaussian copula and assume two levels of correlation in the copula function: weak correlation ($\rho = 0.4$), and strong correlation ($\rho = 0.8$). Second, we use the multivariate lognormal distribution, with unit mean and standard deviation and the two correlation scenarios. We report the estimated parameter means and root mean square errors (RMSEs) using the GMM and ML approaches in Table 1.

Insert Table 1 here

When the disturbance terms are generated from the copula-based gamma distribution, the GMM estimator is unbiased and more efficient in all cases. However, the ML estimator (based on the lognormal distribution) is close to unbiased in most of cases, with the efficiency loss relative to the

[1]The required regularity conditions involve a stationary condition, an invertibility condition, and a positive semi-definite error variance-covariance matrix condition. Further details are available on request.

GMM approach very small. Therefore, the ML estimator appears to perform fairly well even when the error distribution is misspecified. When the disturbance term is generated from the lognormal distribution, it is no surprise that our proposed ML approach out performs the GMM approach in terms of both unbiasedness and efficiency. In particular, the average RMSE value associated with the ML approach is about half of that associated with the GMM approach. The efficiency gain achieved by using the ML approach is very large and consistent across experiments.

In general, within the context of a specific parametric model, the ML estimator is fully efficient amongst consistent and asymptotically normally distributed estimators. However, to attain this efficiency, it is necessary to make highly specific assumptions about the error distribution. By contrast, GMM estimators move away from parametric assumptions, toward estimators that are robust to alternative underlying data generating processes.[1] Thus while it seems that both have virtue we note that ML estimation is simple to implement. Moreover, for the sample sizes usually encountered in financial time series, any loss of efficiency associated with the ML estimator (encountered if the error distribution is misspecified) is relatively small.

3. An application to NYSE stock data

This section contains details of the data used, descriptions of the estimated models, and a discussion of tests of two microstructure hypotheses.

3.1. *Data*

We make use of two different datasets. The first is obtained from the NYSE-based Trades and Quotes (TAQ) dataset, while the second is obtained from Tickdata.com. Both datasets consist of time stamped trade and quote information associated with a random selection of stocks. The primary difference between the datasets relates to the time period used. The first dataset coincides with that used by Manganelli (2005), and covers the period from January 1, 1998 to June 30, 1999. The second consists of more recent data, and covers the period from January 1, 2012 to March 31, 2012. These datasets are henceforth referred to as the 1998 and 2012 datasets, respectively. By using these two datasets we are able to examine the robustness of the results to dataset design, and to investigate how the trading environment has changed.[2]

Manganelli (2005) constructs a dataset consisting of ten stocks covering the period from January 1, 1998 to June 30, 1999. Five of these stocks are randomly selected from the second decile of frequently traded stocks, while five are randomly selected from the eighth decile of stocks. We use the same raw dataset in this analysis, and prepared the data (including diurnal adjustment for intraday patterns) as in Manganelli (2005); see subsection 4.1 in Manganelli (2005) for a concise description of how the data are prepared for use in his and our paper.[3] This process of stock selection and preparation is also used in the construction of the 2012 dataset.[4] The tickers of the ten stocks in the 1998 and 2012 datasets are reported in Table 2.

Insert Table 2 here

[1] Within the context of volatility modelling, Andersen and Sørensen (1997) show that the relative merits of the ML and GMM estimators depend on the level of volatility persistence. In particular, the ML estimator is preferable when persistence is high, while the reverse holds for somewhat lower levels of volatility persistence.

[2] Since the full implementation of Regulation National Market System (Reg NMS) in 2007, the trading environment has changed dramatically with high-frequency traders providing the bulk of liquidity within an open limit order book system. These traders are proprietary traders and perform a similar role to the old specialists in the 1990s, though the former are less likely to hold inventory for more than one day.

[3] These data were kindly supplied by Simone Manganelli.

[4] The deciles associated with the 1998 and 2012 datasets are based on the number of trades of all stocks quoted on the NYSE during 1997 and 2011, respectively.

The results in Table 2 also provide summary statistics. For the frequently traded stocks, the number of observations exceeds those associated with the infrequently traded stocks. Furthermore, the latter stocks have longer durations between trades. For instance, in the 1998 dataset, the number of observations range from 46,827 to 88,918, with the average duration ranging from 99 seconds to 187 seconds. For the infrequently traded stocks, the number of observations range from 1,969 to 5,155, with the average duration ranging from 1,693 seconds to 4,441 seconds. By contrast, there is little difference between the volumes associated with the frequently and infrequently traded stocks. Comparing the 1998 and 2012 datasets, it is noticeable that the trading frequency is much higher in the latter dataset. This suggests the presence of a trend toward increased trading activity over the two samples.

The results also indicate that duration, volume and volatility show strong serial autocorrelations, and this is particularly true for the frequently traded stocks (as evinced by the Ljung-Box statistics). Thus models that are capable of allowing for such dynamics are required. The choice of which distribution to use in such models is also important. To this end, we compare the non-parametric density implied by the data with candidate parametric densities given by the exponential and lognormal distributions; see the subset of plots associated with volume in Figure 1.[1] In general, the lognormal distribution provides a more reasonable fit to the true density than the exponential distribution. Combining this result with the strong dynamic dependencies evinced in Table 2 supports use of the lognormal log-VMEM.

Insert Figure 1 here

3.2. *Estimated models*

In the empirical analysis, we estimate two first-order models: the restricted VMEM and unrestricted log-VMEM. The VMEM is a restricted model in the sense that we have to impose restrictions to ensure that the conditional means are nonnegative. If the estimated $\mathbf{A}$ and $\mathbf{B}$ matrices were all nonnegative, then the VMEM would be unrestricted. However, if these matrices have negative elements, then we are required to restrict them to be nonnegative to ensure that predicted values are nonnegative. In doing this we may lose important information regarding the relationship between duration, volume and volatility. If this is the case, we should use the log-VMEM, as the conditional means are guaranteed to be nonnegative without any restrictions. Both models are estimated using the proposed ML estimation methodology.[2]

After estimation, we re-parameterize the model such that the impact of unexpected elements of $\mathbf{x}_{t-1}$ can be interpreted; that is, the VMEM can be written as

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + (\mathbf{A} + \mathbf{B})\mathbf{x}_{t-1} - \mathbf{B}\mathbf{e}_{t-1}, \tag{14}$$

where $\mathbf{e}_t$ is the martingale difference between $\mathbf{x}_t$ and $\boldsymbol{\mu}_t$, such that $\mathbf{e}_t = \mathbf{x}_t - \boldsymbol{\mu}_t$. This model specification is henceforth denoted M1. Similarly, the log-VMEM can be written as

$$\ln \boldsymbol{\mu}_t = \boldsymbol{\omega} + (\mathbf{A} + \mathbf{B}) \ln \mathbf{x}_{t-1} - \mathbf{B} \ln \boldsymbol{\epsilon}_{t-1}. \tag{15}$$

This model specification is henceforth denoted M2.

[1] The non-parametric density is constructed as in Grammig and Maurer (2000) and Xu (2013). First, we estimate an ACD(1,1) model with a conditional distribution given by the exponential or lognormal density function. Second, we collect the residuals from this model and plot their non-parametric density. Third, this density is compared with the parametric density implied by the exponential or lognormal density.

[2] To give the VMEM the best possible chance of success in the subsequent empirical analysis, we do not necessarily impose the restriction that *all* elements in $\mathbf{B}$ are nonnegative. Rather, we follow the approach of Conrad and Karanasos (2010), in which one off-diagonal element in $\mathbf{B}$ is permitted to be negative. If, after estimating the VMEM with no restrictions imposed, we find that more than one off-diagonal element in $\mathbf{B}$ is negative then we restrict all elements to be nonnegative and re-estimate the model.

Various causal and feedback effects among the three variables can be examined by using the above specifications. For example, $a_{31} + b_{31}$ ($a_{32} + b_{32}$) measures the impact of duration (volume) on volatility; $-b_{31}$ ($-b_{32}$) measures the impact of duration (volume) shocks on volatility; $a_{13} + b_{13}$ measures the impact of volatility on trading intensity; and $-b_{13}$ measures the impact of volatility shocks on trading intensity.

3.3. *Comparative model performance*

Measures of model fit associated with M1 and M2 applied to the 1998 and 2012 datasets are presented in the lower panel of Table 3. In particular, we present the log-likelihood, the Akaike information criterion, and the Bayesian information criterion values for each model. The results indicate that M2 almost always provides the best fit to the data. Moreover, many off-diagonal elements in the M1 **A** and **B** matrices are zero. This indicates that the nonnegativity constraints have been hit and the corresponding parameters are forced to zero. Given the inferior fit of M1 this restriction suggests a loss of useful information.

Insert Table 3 here

In comparison to the recursive model of Manganelli (2005), the proposed VMEM permits the non-diagonality of the coefficient matrix **B** and the covariance matrix **V**. Consequently, the virtue of our model can be examined by conducting Wald tests in which **B** and **V** have zero off-diagonal restrictions imposed. The results in Table 3 indicate that the tests applied to all off-diagonal elements of **B** are almost always significant for all four datasets. Moreover, the tests applied to the off-diagonal elements of **V** are universally significant, confirming the existence of cross-dependence in the error terms. These results support the general form log-VMEM specification (and joint estimation approach) proposed in this paper.

3.4. *Microstructure hypothesis tests*

The price impact of trading and the feedback effects on trading intensity have been the subject of much theoretical and empirical research; see, e.g., Dufour and Engle (2000), Engle (2000) Grammig and Wellner (2002), Engle and Lunde (2003), and Manganelli (2005). In this section, we use the results from M2 to investigate the price impact of trading and its relation to trading intensity by focusing on two hypothesis that cannot be addressed by the recursive model of Manganelli (2005).

3.4.1. **The price effect of trades.** To evaluate the price impact of trades, the vast majority of previous studies use raw duration (volume) to proxy private information. However, theoretical research does not predict whether it is duration (volume) or duration (volume) shocks that carry information content with respect to the fundamental asset value. If duration (volume) is unpredictable then there is no difference between realized values and shocks (and their impact). If, however, duration (volume) is predictable then realized values and their shocks will have different effects (cf. Hasbrouck, 1988, Engle and Russell, 1998, and Grammig et al., 2007). This motivates our first hypothesis test: duration (volume) and duration (volume) shocks contain information content on the fundamental asset value.

When M2 is applied to the data, the results in Table 3 show that the duration coefficient ($a_{31}+b_{31}$) and duration shock coefficient ($-b_{31}$) in the volatility equation are significant in most cases. This indicates that both duration and duration shocks are related to the arrival of new information, which manifests itself in higher volatility. Moreover, $a_{31} + b_{31}$ is positive and $-b_{31}$ is negative, with the latter coefficient larger in absolute terms. Hence, the overall effect is negative – a result that is consistent with Easley and O'Hara (1992). The implicit implication is that market makers will associate trading activity that is higher than the expected level as a signal of informed trading, and adjust the price accordingly.

The volume coefficient ($a_{32}+b_{32}$) and volume shock coefficient ($-b_{32}$) are also significant in most cases. Moreover, these coefficients have different signs, with volume shocks dominant in terms of overall effect. The results support the prediction of Easley and O'Hara (1987, 1992), that it is the unexpected component of volume rather than observed volume that carries information. Implicitly, market makers will only consider trade size that is larger than its expected level as a signal of private information. It is also notable that these findings are less robust for infrequently traded stocks (this is consistent with Manganelli, 2005), since many of the coefficients (a_{31}, b_{31}, a_{32}, b_{32}) are insignificant. This shows that the relevant market microstructure predictions may only be valid for frequently traded stocks.

3.4.2. **The feedback effects from volatility to trading intensity.** In terms of the feedback effect from volatility to trading intensity, previous empirical microstructure studies report apparently contradictory results. In particular, Dufour and Engle (2000) and Manganelli (2005) find that short durations follow large (squared) returns, while Grammig and Wellner (2002) find that lagged volatility significantly reduces trading intensity.

Theoretically, quote changes could either be inventory-motivated or information-motivated. Consequently, they have potentially different effects on trading intensity. If they are inventory-motivated, then large absolute quote changes (or large volumes) may attract opposite side traders, which would increase trading intensity (Dufour and Engle, 2000). If it is information-motivated, large absolute quote changes indicate a risk of informed trading such that liquidity traders may leave or slow down their trading activity to avoid adverse selection (Easley and O'Hara, 1987, and Admati and Pfleiderer, 1988). Hasbrouck (1988, 1991) has used the short-run and long-run characteristics of trading behavior to separated quote movements into short-run inventory-related effects and long-run information-related effects. Persistent quote changes could be related to private information as this information is persistent and long lived, while transient quote changes are related to inventory control as this is an inherently temporary concern. As a result, Hasbrouck predicts that persistent quote changes have a negative effect on trading intensity and transient quote changes have a positive effect on trading intensity. Our second hypothesis is to empirically evaluate these predictions.

We see from Table 3 that when M2 is applied to the frequently traded stock data, the volatility coefficient ($a_{13}+b_{13}$) is positive and significant for both the 1998 dataset and the 2012 dataset, while the volatility shock coefficient ($-b_{13}$) is negative and significant for the 10 stocks. The result is consistent with Hasbrouck's (1989, 1991) predictions. For example, information-motivated large absolute quote changes (which we measure via quote change volatility in the empirical analysis since volatility is highly persistent) indicate a risk of informed trading such that liquidity traders may leave or slow down their trading activity to avoid adverse selection. By contrast, inventory-motivated large quote changes (which we measure via shocks to quote change volatility in the empirical analysis) may attract opposite side traders and increase trading intensity. However, this result does not tend to hold for infrequently traded stocks (both datasets).

3.4.3. **Impulse response function results.** From the above analysis it is clear that shocks to the trading process contain information content with respect to asset prices. It is therefore natural to measure how long the new information takes to be impounded into prices. To answer this question we generate the impulse responses that trace the effects of one standard deviation shocks to duration, volume and volatility on future values of volatility as implied by the M1 and M2 parameter estimates. These are provided in Figure 2 for the JAX and TKF stocks, and provide a visual description of the effects up to the 70th trade.

Insert Figure 2 here

The shapes of the response functions associated with M1 and M2 are fairly similar for volatility shocks. However, for duration and volume shocks, there are clear differences. The nature of these responses can also be seen in Table 4, which contains the number of hours for a shock to return to its long-run equilibrium value.

Insert Table 4 here

The results reveal a number of interesting findings. First, it takes more time for volatility to be absorbed after a shock when M2 is used (cf. M1) for frequently traded stocks – a result that suggests that M1 may overestimate the speed of price adjustment in the market. By contrast, it takes less time for volatility to be absorbed after a shock when M2 is used (cf. M1) for infrequently traded stocks. The reason lies in the estimated system persistence, which is given by the maximum eigenvalue (Γ) of $\mathbf{A} + \mathbf{B}$ in Table 3. By noting this eigenvalue, it can be seen that M1 tends to overestimate the persistence for infrequently traded stocks, while it tends to underestimate the persistence for frequently traded stocks. Second, irrespective of the type of shock, it takes appropriately the same time for volatility to be absorbed into its long-run equilibrium value. Third, the volatility of frequently traded stocks converges much faster to its long-run equilibrium after an initial perturbation than it does for infrequently traded stocks. Fourth, volatility in the 2012 dataset converges much faster to its long-run equilibrium, as indicated by the shorter absorbtion times in this dataset (cf. the 1998 dataset results). This suggests that the speed of price adjustment is much faster during the era of high frequency trading (that is, in the 2012 dataset).

4. Conclusion

In this paper, we consider a log-VMEM, in which duration, volume and volatility are interdependent. We further propose a multivariate lognormal density for this model, which allows the error terms to be contemporaneously correlated. In this way, we build a system that incorporates various causal and feedback effects among the variables. The findings are summarized as follows:

(i) We compare the proposed log-VMEM with the VMEM and show that the VMEM is a restricted model while the log-VMEM is an unrestricted model. This leads to the log-VMEM having a superior fit to the data.
(ii) We compare the proposed log-VMEM with the recursive model of Manganelli (2005). We find that the lagged (un)expected variables are widely significant in the log-VMEM, challenging the weak exogeneity assumptions used in the empirical market microstructure literature.
(iii) We highlight the importance of unexpected components of trading characteristics in that it is mostly these components that carry information with respect to asset prices. This result supports the prediction of Easley and O'Hara (1987, 1992). Furthermore, volatility and volatility shocks affect duration in different directions, confirming Hasbrouck's (1988, 1991) prediction. However, this effect is less robust for infrequently traded stocks.

The methodology used in this paper can easily be extended to model any other nonnegative-valued, highly persistent variables. An interesting application would be modeling of volatility. For example, there are different measures of volatility, but no individual one appears to be a sufficient measure on its own (see Engle and Gallo, 2006). One possibility is to consider absolute daily returns, daily high-low range and daily realized volatility within a log-VMEM framework and compare the forecasting performance with that achieved by the VMEM proposed by Cipollini et al. (2013). This proposal is left for future research.

Acknowledgements

We would like to thank Garry Phillips for his insightful comments and Simone Manganelli for providing part of the data used in this paper.

References

Admati A, Pfleiderer P. 1988. A theory of intraday patterns: Volume and price variability. Review of Financial Studies 1: 3–40.

Allen D, Chan F, McAleer M, Peiris S. 2008. Finite sample properties of the QMLE for the log-ACD model: Application to Australian stocks. Journal of Econometrics 147: 163–185.

Andersen T, Bollerslev T, Diebold F, Ebens H. 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61: 43–76.

Andersen T, Sørensen B. 1997. GMM and QML asymptotic standard deviations in stochastic volatility models: Comments on Ruiz (1994). Journal of Econometrics 76: 397–403.

Bauwens L, Giot P. 2000. The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. Annales d'Economie et de Statistique. 60: 117–49.

Bauwens L, Galli F, Giot P. 2008. The Moments of Log-ACD Models. Quant Qual Anal Soc Sci. 2: 1–28.

Bodnar T, Hautsch N. 2012. Copula-based dynamic conditional correlation multiplicative error processes. Available at SSRN 2144741.

Campbell, J, Lo A, MacKinlay A. 1997. The Econometrics of Financial Markets. Princeton University Press: Princeton, NJ.

Cipollini F, Engle R, Gallo G. 2007. A model for multivariate non-negative valued processes in financial econometrics. Unpublished paper: Stern School of Business, New York University.

Cipollini F, Engle R, Gallo G. 2013. Semiparametric vector MEM. Journal of Applied Econometrics 28: 1067–1086.

Cizeau P, Liu Y, Meyer M, Peng C, Stanley H. 1997. Volatility distribution in the S&P500 stock index. Physica A: Statistical Mechanics and its Applications 245: 441–445.

Conrad C, Karanasos M. 2010. Negative volatility spillovers in the unrestricted ECCC-GARCH model. Econometric Theory 26: 838–862.

Diamond D, Verrecchia R. 1987. Constraints on short-selling and asset price adjustment to private information. Journal of Financial Economics 18: 277-311.

Dufour A, Engle R. 2000. Time and the price impact of a trade. Journal of Finance 55: 2467–2498.

Easley D, O'Hara M. 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19: 69–90.

Easley D, O'Hara M. 1992. Time and the process of security price adjustment. Journal of Finance 47: 577–605.

Engle R. 2000. The econometrics of ultra-high-frequency data. Econometrica 68: 1–22.

Engle R. 2002. New frontiers for ARCH models. Journal of Applied Econometrics 17: 425–446.

Engle R, Gallo G. 2006. A multiple indicators model for volatility using intra-daily data. Journal of Econometrics 131: 3–27.

Engle R, Gallo G, Velucchi M. 2012. Volatility spillovers in East Asian financial markets: A MEM-based approach. Review of Economics and Statistics 94: 222–233.

Engle R, Lunde A. 2003. Trades and quotes: A bivariate point process. Journal of Financial Econometrics 1: 159–188.

Engle R, Russell J. 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66: 1127–1162.

Engle R, Sun Z. 2007. When is noise not noise – A microstructure estimate of realized volatility. Manuscript, Stern School, New York University.

Ghysels E, Gouriéroux C, Jasiak J. 2004. Stochastic volatility duration models. Journal of Econometrics 119: 413–433.

Grammig J, Maurer K. 2000. Non-monotonic hazard functions and the autoregressive conditional duration model. The Econometrics Journal 3: 16–38.

Grammig J, Theissen E, Wuensche O. 2007. Time and price impact of a trade: A structural approach. CFR Working Paper.

Grammig J, Wellner M. 2002. Modeling the interdependence of volatility and inter-transaction duration processes. Journal of Econometrics 106: 369–400.

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: 179–207.

Hautsch N. 2008. Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model. Journal of Economic Dynamics and Control 32: 3978–4015.

He C, Tersvirta T. 2004. An extended constant conditional correlation GARCH model and its fourth-moment structure. Econometric Theory. 20: 904–926.

Hosking J. 1980. The multivariate portmanteau statistic. Journal of the American Statistical Association 75: 602-608.

Ling S, McAleer M. 2003. Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory 19: 280–310.

Lütkepohl H. 2005. New Introduction to Multiple Time Series Analysis. Cambridge University Press: Cambridge, UK.

Manganelli S. 2005. Duration, volume and volatility impact of trades. Journal of Financial Markets 8: 377–399.

Nakatani T, Teräsvirta. 2009. Testing for volatility interactions in the constant conditional correlation GARCH model. The Econometrics Journal 12: 147–163.

Xu Y. 2013. The lognormal autoregressive conditional duration (LNACD) model and a comparison with alternative ACD models. Quantitative and Qualitative Analysis in Social Sciences 7.

Appendix A: Stationarity and invertibility conditions

In this section the stationarity and invertibility conditions associated with the VMEM and log-VMEM are derived.

A.1. *VMEM derivation*

Consider the following VMEM:

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \ln \mathrm{N}(\mathbf{M}, \mathbf{V}), \tag{A.1}$$

where

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{x}_{t-i} + \sum_{i=1}^{q} \mathbf{B}_i \boldsymbol{\mu}_{t-i}. \tag{A.2}$$

Taking the difference between $\mathbf{x}_t$ and $\boldsymbol{\mu}_t$, we obtain

$$\mathbf{x}_t - \boldsymbol{\mu}_t = \mathbf{e}_t, \quad \mathbf{e}_t \sim \mathrm{IID}(\mathbf{0}, \boldsymbol{\Pi}). \tag{A.3}$$

It follows that

$$\boldsymbol{\mu}_t = \mathbf{x}_t - \mathbf{e}_t, \tag{A.4a}$$

$$\sum_{i=1}^{q} \mathbf{B}_i \boldsymbol{\mu}_{t-i} = \sum_{i=1}^{q} \mathbf{B}_i \mathbf{x}_{t-i} - \sum_{i=1}^{q} \mathbf{B}_i \mathbf{e}_{t-i}. \tag{A.4b}$$

Substituting the expressions in (A.4a) and (A.4b) into (A.2) and rearranging we obtain the following vector autoregressive moving average (VARMA) representation:

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\omega} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{x}_{t-i} + \sum_{i=1}^{q} \mathbf{B}_i \mathbf{x}_{t-i} + \mathbf{e}_t - \sum_{i=1}^{q} \mathbf{B}_i \mathbf{e}_{t-i}, \\ &= \boldsymbol{\omega} + \sum_{i=1}^{r} \widetilde{\mathbf{A}}_i \mathbf{x}_{t-i} + \mathbf{e}_t - \sum_{i=1}^{q} \mathbf{B}_i \mathbf{e}_{t-i}, \end{aligned} \tag{A.5}$$

where $\widetilde{\mathbf{A}}_i = \mathbf{A}_i + \mathbf{B}_i$, and $r = \max(p, q)$. Given this VARMA(r,q) representation it follows that the process is stationary if the modulus of the roots of $|\mathbf{I} - \widetilde{\mathbf{A}}_1 z - \widetilde{\mathbf{A}}_2 z^2 \ldots \widetilde{\mathbf{A}}_r z^r| = 0$ are all greater than one, and invertible if the modulus of the roots of $|\mathbf{I} - \mathbf{B}_1 z - \mathbf{B}_2 z^2 \ldots \mathbf{B}_q z^q| = 0$ are all greater than one.

A.2. *Log-VMEM derivation*

Consider the following log-VMEM:

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \ln \mathrm{N}(\mathbf{M}, \mathbf{V}), \tag{A.6}$$

where

$$\ln \boldsymbol{\mu}_t = \boldsymbol{\omega} + \sum_{i=1}^{p} \mathbf{A}_i \ln \mathbf{x}_{t-i} + \sum_{i=1}^{q} \mathbf{B}_i \ln \boldsymbol{\mu}_{t-i}. \tag{A.7}$$

Taking logs of (A.6) we obtain

$$\ln \mathbf{x}_t = \ln \boldsymbol{\mu}_t + \ln \boldsymbol{\epsilon}_t = \mathbf{c} + \ln \boldsymbol{\mu}_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Pi}). \tag{A.8}$$

It follows that

$$\ln \boldsymbol{\mu}_t = \ln \mathbf{x}_t - \mathbf{c} - \mathbf{e}_t, \tag{A.9a}$$

$$\sum_{i=1}^{q} \mathbf{B}_i \ln \boldsymbol{\mu}_{t-i} = \sum_{i=1}^{q} \mathbf{B}_i \ln \mathbf{x}_{t-i} - \sum_{i=1}^{q} \mathbf{B}_i \mathbf{c} - \sum_{i=1}^{q} \mathbf{B}_i \mathbf{e}_{t-i}. \tag{A.9b}$$

Substituting the expressions in (A.9a) and (A.9b) into (A.7) and rearranging we obtain the following vector autoregressive moving average (VARMA) representation:

$$\begin{aligned}
\ln \mathbf{x}_t &= \overline{\mathbf{c}} + \sum_{i=1}^{p} \mathbf{A}_i \ln \mathbf{x}_{t-i} + \sum_{i=1}^{q} \mathbf{B}_i \ln \mathbf{x}_{t-i} + \mathbf{e}_t - \sum_{i=1}^{q} \mathbf{B}_i \mathbf{e}_{t-i}, \\
&= \overline{\mathbf{c}} + \sum_{i=1}^{r} \widetilde{\mathbf{A}}_i \ln \mathbf{x}_{t-i} + \mathbf{e}_t - \sum_{i=1}^{q} \mathbf{B}_i \mathbf{e}_{t-i},
\end{aligned} \tag{A.10}$$

where $\overline{\mathbf{c}} = \mathbf{c} + \boldsymbol{\omega} - \sum_{i=1}^{q} \mathbf{B}_i \mathbf{c}$, $\widetilde{\mathbf{A}}_i = \mathbf{A}_i + \mathbf{B}_i$, and $r = \max(p, q)$. Given this VARMA(r,q) representation it follows that the process is stationary if the modulus of the roots of $|\mathbf{I} - \widetilde{\mathbf{A}}_1 z - \widetilde{\mathbf{A}}_2 z^2 \ldots \widetilde{\mathbf{A}}_r z^r| = 0$ are all greater than one, and invertible if the modulus of the roots of $|\mathbf{I} - \mathbf{B}_1 z - \mathbf{B}_2 z^2 \ldots \mathbf{B}_q z^q| = 0$ are all greater than one.

Appendix B: Impulse response functions

Under stationary conditions, the impulse response functions associated with the first-order VMEM and first-order log-VMEM are derived.

B.1. *VMEM derivation*

Consider the following first-order VMEM:

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \ln \mathrm{N}(\mathbf{M}, \mathbf{V}), \tag{B.1}$$

where

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\boldsymbol{\mu}_{t-1}. \tag{B.2}$$

Suppose the system is in steady state up to time $t = 0$. That is, all the errors before $t = 0$ equal unity. It follows that

$$\mathbf{x}_t = \boldsymbol{\mu}_t, \tag{B.3}$$

$$\boldsymbol{\mu}_t = (\mathbf{I} - \mathbf{A} - \mathbf{B})^{-1}\boldsymbol{\omega} \quad \forall t < 0. \tag{B.4}$$

If a shock occurs to $\mathbf{x}_t$ at time $t = 0$ then the impulse response function for $t > 0$ is given by

$$\frac{\partial \mathrm{E}(\boldsymbol{\mu}_t|\Omega_t)}{\partial \mathbf{x}_0'} = \boldsymbol{\Lambda}^t \tag{B.5}$$

where $\boldsymbol{\Lambda} = \mathbf{A} + \mathbf{B}$.

The next step is to investigate the effect of a shock occurring to $\boldsymbol{\epsilon}_0$ on $\mathbf{x}_0$. Suppose now that at time $t = 0$ a shock occurs to the ith element of $\boldsymbol{\epsilon}_0$ (denoted ϵ_{i0}). The size of the shock is assumed to equal the unconditional standard deviation of the ith element of $\boldsymbol{\epsilon}_t$. The effect of this shock on $\mathbf{x}_0$ is given by

$$\frac{\partial \mathbf{x}_0}{\partial \epsilon_{i0}} = \boldsymbol{\mu}_0 \odot \mathbf{s}_i, \tag{B.6}$$

where

$$\mathbf{s}_i = \left(\sigma_i \frac{\sigma_{i1}}{\sigma_i^2}, \sigma_i \frac{\sigma_{i2}}{\sigma_i^2}, \sigma_i \frac{\sigma_{i3}}{\sigma_i^2}\right)'. \tag{B.7}$$

Here σ_{ij} is the unconditional covariance between ϵ_{it} and ϵ_{jt}, and σ_i is the unconditional standard deviation of ϵ_{it}. This result relies on Engle et al. (2012), who show that $\mathrm{E}(\epsilon_{jt}|\epsilon_{it} = 1 + \sigma_i) = 1 + \sigma_i\sigma_{ij}/\sigma_i^2$.

It follows that the impulse response function for $t > 0$ given a one standard deviation shock to $\boldsymbol{\epsilon}_0$ is given by

$$\frac{\partial \mathrm{E}(\boldsymbol{\mu}_t|\Omega_t)}{\partial \boldsymbol{\epsilon}_0'} = \boldsymbol{\Lambda}^t \mathrm{dg}(\boldsymbol{\mu}_0)\boldsymbol{\Xi}, \tag{B.8}$$

where $\mathrm{dg}(\boldsymbol{\mu}_0)$ is a diagonal matrix with $\boldsymbol{\mu}_0$ along the diagonal, and

$$\boldsymbol{\Xi} = \begin{bmatrix} \sigma_1 & \frac{\sigma_{21}}{\sigma_2} & \frac{\sigma_{31}}{\sigma_3} \\ \frac{\sigma_{12}}{\sigma_1} & \sigma_2 & \frac{\sigma_{32}}{\sigma_3} \\ \frac{\sigma_{13}}{\sigma_1} & \frac{\sigma_{23}}{\sigma_2} & \sigma_3 \end{bmatrix}. \tag{B.9}$$

This impulse response function is analogous to the Manganelli (2005) result, if we restrict $\boldsymbol{\Xi}$ to be an identity matrix. Our result differs in that we look at the impulse response of one (unconditional) standard deviation shocks of $\boldsymbol{\epsilon}_0$, while Manganelli (2005) concentrates on the impulse response of unit shocks of $\boldsymbol{\epsilon}_0$.

B.2. *Log-VMEM derivation*

Consider the following first-order log-VMEM:

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \ln \mathrm{N}(\mathbf{M}, \mathbf{V}), \tag{B.10}$$

where

$$\ln \boldsymbol{\mu}_t = \boldsymbol{\omega} + \mathbf{A} \ln \mathbf{x}_{t-1} + \mathbf{B} \ln \boldsymbol{\mu}_{t-1}. \tag{B.11}$$

Suppose the system is in steady state up to time $t = 0$. That is, all the errors before $t = 0$ equal unity. It follows that

$$\ln \mathbf{x}_t = \ln \boldsymbol{\mu}_t, \tag{B.12}$$

$$\ln \boldsymbol{\mu}_t = (\mathbf{I} - \mathbf{A} - \mathbf{B})^{-1} \boldsymbol{\omega} \quad \forall t < 0. \tag{B.13}$$

If a shock occurs to $\mathbf{x}_t$ at time $t = 0$ then the impulse response function for $t > 0$ is given by

$$\frac{\partial \mathrm{E}(\ln \boldsymbol{\mu}_t | \Omega_t)}{\partial \ln \mathbf{x}_0'} = \boldsymbol{\Lambda}^t, \tag{B.14}$$

where $\boldsymbol{\Lambda}$ is as previously defined.

The next step is to investigate the effect of a shock occurring to $\boldsymbol{\epsilon}_0$ on $\ln \mathbf{x}_0$. Suppose now that at time $t = 0$ a shock occurs to the ith element of $\boldsymbol{\epsilon}_0$ (denoted ϵ_{i0}). The size of the shock is assumed to equal the unconditional standard deviation of the ith element of $\boldsymbol{\epsilon}_t$. The effect of this shock on $\ln \mathbf{x}_0$ is given by

$$\frac{\partial \ln \mathbf{x}_0}{\partial \epsilon_{i0}} = \boldsymbol{\mu}_0 \odot \frac{1}{\mathbf{x}_0} \odot \mathbf{s}_i, \tag{B.15}$$

where $\mathbf{s}_i$ is as previously defined. It follows that the impulse response function for $t > 0$ given a one standard deviation shock to $\boldsymbol{\epsilon}_0$ is given by

$$\frac{\partial \mathrm{E}(\ln \boldsymbol{\mu}_t | \Omega_t)}{\partial \boldsymbol{\epsilon}_0'} = \boldsymbol{\Lambda}^t \boldsymbol{\Xi}, \tag{B.16}$$

where $\boldsymbol{\Xi}$ is as previously defined. Moreover,

$$\frac{\partial \mathrm{E}(\boldsymbol{\mu}_t | \Omega_t)}{\partial \boldsymbol{\epsilon}_0'} = \mathrm{dg}(1/\boldsymbol{\mu}_t) \boldsymbol{\Lambda}^t \boldsymbol{\Xi}, \tag{B.17}$$

where $\mathrm{dg}(\boldsymbol{\mu}_t)$ is a diagonal matrix with $\boldsymbol{\mu}_t$ along the diagonal. $\boldsymbol{\mu}_t$ can be derived recursively from initial values in $\boldsymbol{\mu}_0$, such that,

$$\boldsymbol{\mu}_t = \exp(\boldsymbol{\omega} + \mathbf{A} \ln \mathbf{x}_{t-1} + \mathbf{B} \ln \boldsymbol{\mu}_{t-1}) \quad \forall t > 0. \tag{B.18}$$

where $\boldsymbol{\mu}_0 = (\mathbf{I} - \mathbf{A} - \mathbf{B})^{-1} \boldsymbol{\omega}$.

Table 1. Simulation results

		Weak Correlation ($\rho = 0.4$)				Strong Correlation ($\rho = 0.8$)			
		GMM		ML		GMM		ML	
Parameter	True	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
Panel A: Copula-based gamma distribution									
a_{11}	0.10	0.10	0.18	0.13	0.33	0.10	0.15	0.13	0.34
a_{12}	0.01	0.01	0.04	0.01	0.07	0.01	0.05	0.01	0.07
a_{21}	0.01	0.01	0.12	0.02	0.17	0.01	0.18	0.02	0.22
a_{22}	0.08	0.08	0.18	0.10	0.25	0.08	0.12	0.10	0.24
b_{11}	0.85	0.85	0.31	0.85	0.19	0.85	0.20	0.85	0.20
b_{12}	0.02	0.02	0.05	0.02	0.06	0.02	0.06	0.02	0.07
b_{21}	−0.04	−0.04	0.18	−0.04	0.20	−0.04	0.25	−0.05	0.26
b_{22}	0.90	0.89	0.34	0.90	0.12	0.90	0.18	0.90	0.13
Mean RMSE			0.18		0.17		0.15		0.19
Panel B: Lognormal distribution									
a_{11}	0.10	0.10	0.18	0.10	0.09	0.10	0.19	0.10	0.10
a_{12}	0.01	0.01	0.04	0.01	0.03	0.01	0.05	0.01	0.04
a_{21}	0.01	0.01	0.13	0.01	0.09	0.01	0.18	0.01	0.12
a_{22}	0.08	0.08	0.17	0.08	0.07	0.08	0.17	0.08	0.07
b_{11}	0.85	0.85	0.28	0.85	0.13	0.85	0.26	0.85	0.14
b_{12}	0.02	0.02	0.05	0.02	0.04	0.02	0.07	0.02	0.05
b_{21}	−0.04	−0.04	0.19	−0.04	0.14	−0.04	0.24	−0.04	0.17
b_{22}	0.90	0.90	0.27	0.90	0.08	0.90	0.25	0.90	0.09
Mean RMSE			0.16		0.08		0.18		0.10

Notes: Results in this table are based on 1000-repetition Monte Carlo simulations each with a sample size of 5000 observations. The true parameter values are reported in the first column. Within each correlation scenario, we report the estimated parameter mean and RMSE values associated with the GMM and ML approaches. The last row reports the mean RMSE values across all parameters. For presentation purposes the RMSE values are multiplied by 10.

Table 2. Summary statistics

	S1	S2	S3	S4	S5
Panel A: 1998 dataset (infrequently traded stocks)					
Ticker	DTC	FTD	GBX	GSE	JAX
# Observations	4162	3925	5155	1969	2766
Mean-Duration	2093.91	2417.11	1693.45	4441.14	3164.67
Mean-Volume	2136.83	736.25	1434.04	1523.77	1000.04
Mean-Volatility	2.25	0.53	1.82	1.69	2.63
LB-Duration	155.35	19.29	1047.46	99.65	109.03
LB-Volume	142.41	69.24	170.84	117.59	104.80
LB-Volatility	143.27	217.45	593.94	149.22	214.31
MLB	531.28	710.52	1967.80	484.34	575.40
Panel B: 1998 dataset (frequently traded stocks)					
Ticker	AVT	COX	CP	DLP	GAP
# Observations	58390	88918	71673	65305	46827
Mean-Duration	150.02	98.60	122.42	134.15	187.30
Mean-Volume	1070.01	2678.86	2892.23	1486.35	824.63
Mean-Volatility	4.57	5.93	5.44	7.42	5.57
LB-Duration	6772.70	29905.30	3713.70	19672.60	4962.56
LB-Volume	413.04	364.91	769.86	1775.60	174.97
LB-Volatility	1258.44	2697.83	3374.14	3253.61	1975.02
MLB	9490.56	33636.59	10160.16	25091.35	7558.88
Panel C: 2012 dataset (infrequently traded stocks)					
Ticker	CXE	KEF	MYC	NCA	TKF
# Observations	6231	2784	5296	6115	2835
Mean-Duration	232.84	521.12	273.94	237.25	511.55
Mean-Volume	859.23	712.58	590.79	664.10	459.96
Mean-Volatility	2.88	2.13	1.49	2.65	1.95
LB-Duration	919.03	205.47	579.17	510.81	470.39
LB-Volume	248.25	302.17	232.56	276.43	122.54
LB-Volatility	337.11	175.99	643.68	207.31	260.50
MLB	1956.66	813.47	1829.47	951.85	1068.52
Panel D: 2012 dataset (frequently traded stocks)					
Ticker	ARW	DLR	HLS	NBL	SPR
# Observations	103468	146810	100634	214021	134762
Mean-Duration	14.02	9.58	14.00	6.78	10.77
Mean-Volume	380.14	390.63	467.00	389.33	467.87
Mean-Volatility	13.27	10.32	15.12	17.70	13.74
LB-Duration	12307.34	12920.46	14672.92	32444.20	32175.59
LB-Volume	302.58	568.94	379.95	872.55	933.73
LB-Volatility	4999.48	9159.89	2837.55	9811.20	6183.91
MLB	20058.15	26314.20	18891.64	51713.23	40663.70

Notes: This table contains summary statistics. The mean of volatility is given in per second terms and is obtained by dividing the mean of absolute returns by mean duration and multiplying by 10^6. The univariate Ljung-Box (LB) test statistic is based on 15 lags of duration, volume, or volatility (given by absolute return). The multivariate Ljung-Box (MLB) statistics are computed using the method described in Hosking (1980). The 95% critical value associated with the LB test statistic equals 25.00 and the value associated with MLB test statistic equals 61.66. Mean statistics pertain to the series before diurnal adjustment, while the (M)LB statistics pertain to the series after diurnal adjustment.

Table 3. Estimated model parameters

Parameter	S1 M1	S1 M2	S2 M1	S2 M2	S3 M1	S3 M2	S4 M1	S4 M2	S5 M1	S5 M2
	Panel A: 1998 dataset (infrequently traded stocks)									
$a_{11}+b_{11}$	0.99**	0.65**	0.92**	0.84**	0.99**	0.96**	0.99**	0.33**	0.99**	0.33
$a_{12}+b_{12}$	0.00	−0.37	0.07	0.07	0.00	0.14**	0.00	0.63	0.00	0.67
$a_{13}+b_{13}$	0.00	0.24	0.00	0.02	0.00	0.11**	0.00	−0.57	0.00	−0.87
$a_{21}+b_{21}$	0.00	−0.27*	0.01	0.00	0.00	0.00	0.00	0.17**	0.00	0.12
$a_{22}+b_{22}$	0.20	0.19	0.67**	0.99**	0.55**	0.72**	0.96**	0.78**	0.61**	0.66*
$a_{23}+b_{23}$	0.08	0.11	0.00	−0.11**	0.00	−0.12**	0.01	0.18	0.14	0.23
$a_{31}+b_{31}$	0.02	−0.25*	0.00	−0.01	0.01**	0.01	0.00	0.18*	0.00	0.07
$a_{32}+b_{32}$	−0.39	−0.63**	−0.32	0.04	−0.18**	−0.19**	0.00	−0.14	0.00	−0.13
$a_{33}+b_{33}$	0.93**	1.10**	0.97**	0.56**	0.98**	0.88**	0.98**	1.09**	0.96**	1.03**
$-b_{11}$	−0.83**	−0.58**	−0.77**	−0.81**	−0.84**	−0.90**	−0.74**	−0.27*	−0.48**	−0.25
$-b_{12}$	0.00	0.27	−0.07	−0.14	0.00	−0.24**	0.00	−0.85*	0.00	−0.98*
$-b_{13}$	0.00	−0.29	0.00	−0.03	0.00	−0.14**	0.00	0.53	0.00	0.88
$-b_{21}$	0.00	0.22*	0.00	−0.01	0.00	−0.02	0.00	−0.18**	0.00	−0.13
$-b_{22}$	−0.07	−0.03	−0.51**	−0.90**	−0.45**	−0.56**	−0.94**	−0.68**	−0.51**	−0.48
$-b_{23}$	−0.07	−0.08	0.00	0.12**	0.00	0.14**	−0.01	−0.17	−0.14	−0.24
$-b_{31}$	−0.02	0.22*	0.00	−0.04	−0.01	−0.04	0.00	−0.21**	0.00	−0.10
$-b_{32}$	0.47	0.73**	0.49	0.27	0.22**	0.30**	0.00	0.21	0.01	0.23
$-b_{33}$	−0.85**	−1.06**	−0.93**	−0.49**	−0.92**	−0.82**	−0.92**	−1.05**	−0.91**	−0.99**
Wald ($b_{ij,i\neq j}=0$)		R		R		R		R		R
σ_{11}	6.32**	6.26**	4.35**	4.34**	4.56**	4.73**	7.90**	8.02**	8.88**	8.80**
σ_{21}	0.01	0.07	−0.11**	−0.12**	−0.08*	−0.02	−0.19*	−0.17*	−0.08	−0.02
σ_{22}	1.61**	1.56**	0.93**	0.91**	1.55**	1.50**	1.41**	1.37**	1.03**	1.01**
σ_{31}	1.23**	1.26**	0.66**	0.45**	0.88**	0.95**	1.34**	1.41**	1.57**	1.66**
σ_{32}	0.08**	0.06*	0.05*	0.08**	0.07**	0.06*	0.04	0.01	0.12**	0.11**
σ_{33}	2.15**	2.14**	2.08**	2.18**	2.05**	2.05**	2.25**	2.21**	2.20**	2.19**
Wald ($\sigma_{ij,i\neq j}=0$)		R		R		R		R		R
Γ	0.99	0.98	0.97	0.98	0.99	0.99	0.99	0.97	0.99	0.89
−LL	23.90	23.77	19.23	19.19	28.78	28.59	11.45	11.41	15.74	15.66
AIC	47.85	47.59	38.51	38.44	57.62	57.23	22.95	22.87	31.54	31.38
BIC	48.02	47.76	38.68	38.60	57.80	57.40	23.10	23.02	31.70	31.54

Notes: This table contains estimated parameters associated with the VMEM (M1) and log-VMEM (M2) specifications. $\mathbf{\Gamma}$ denotes the largest eigenvalue of $\mathbf{A}+\mathbf{B}$, LL is the system log-likelihood, AIC is the system Akaike information criterion, and BIC is the system Bayesian information criterion. Wald ($b_{ij,i\neq j}=0$) and Wald ($\sigma_{ij,i\neq j}=0$) provide an indication of Wald test rejection (R) or non-rejection (N) at the 5% level of the null of zero off-diagonal elements in $\mathbf{B}$ and $\mathbf{V}$, respectively. Significance at the 1% level is indicated by $**$, and at the 5% level by $*$.

Table 3. Estimated model parameters (cont.)

	S1		S2		S3		S4		S5	
Parameter	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Panel B: 1998 dataset (frequently traded stocks)										
$a_{11}+b_{11}$	0.99**	0.97**	0.99**	0.93**	0.99**	1.04**	0.99**	0.96**	0.99**	0.97**
$a_{12}+b_{12}$	0.00	0.09**	0.00	0.05**	0.00	0.02	0.00	0.03**	0.00	0.07**
$a_{13}+b_{13}$	0.00	0.29**	0.00	0.22**	0.00	0.24**	0.00	0.13**	0.00	0.09**
$a_{21}+b_{21}$	0.00	−0.01	0.00	0.04**	0.00	−0.06*	0.00	0.03**	0.00	0.00
$a_{22}+b_{22}$	0.88**	0.88**	0.99**	0.93**	0.53**	0.96**	0.93**	0.92**	0.88**	0.77**
$a_{23}+b_{23}$	0.00	−0.19**	0.00**	−0.10**	0.01	−0.24**	0.00	−0.14**	0.00	−0.16**
$a_{31}+b_{31}$	0.01	0.02	0.01	0.08**	0.00	−0.13*	0.01**	0.06**	0.01	0.02
$a_{32}+b_{32}$	−0.06	−0.19**	0.00	−0.06**	−0.37	−0.06	−0.02**	−0.09**	−0.06	−0.24**
$a_{33}+b_{33}$	0.93**	0.53**	0.96**	0.75**	0.93**	0.44**	0.97**	0.73**	0.97**	0.76**
$-b_{11}$	−0.85**	−0.87**	−0.86**	−0.82**	−0.94**	−0.98**	−0.91**	−0.89**	−0.88**	−0.91**
$-b_{12}$	0.00	−0.20**	0.00	−0.16**	0.00	−0.11**	0.00	−0.09**	0.00	−0.13**
$-b_{13}$	0.00	−0.33**	0.00	−0.26**	0.00	−0.27**	0.00	−0.15**	0.00	−0.11**
$-b_{21}$	0.00	0.00	0.00	−0.06**	0.00	0.03	0.00	−0.05**	0.00	−0.02*
$-b_{22}$	−0.85**	−0.75**	−0.99**	−0.83**	−0.41**	−0.85**	−0.88**	−0.79**	−0.85**	−0.63**
$-b_{23}$	0.00	0.19**	0.00	0.10**	0.00	0.25**	0.00	0.15**	0.00	0.17**
$-b_{31}$	−0.01**	−0.10**	−0.01**	−0.18**	0.00	0.05	0.00	−0.12**	−0.01	−0.07**
$-b_{32}$	0.08*	0.36**	0.00	0.20**	0.49	0.27**	0.04**	0.22**	0.08	0.37**
$-b_{33}$	−0.87**	−0.48**	−0.92**	−0.69**	−0.87**	−0.38**	−0.94**	−0.68**	−0.92**	−0.69**
Wald ($b_{ij,i\neq j}=0$)		R		R		R		R		R
σ_{11}	2.59**	2.57**	2.41**	2.43**	2.05**	2.01**	2.09**	2.11**	2.89**	2.93**
σ_{21}	−0.09**	−0.07**	−0.04**	−0.01	0.06**	0.06**	0.02**	0.03**	−0.08**	−0.05**
σ_{22}	1.39**	1.34**	1.58**	1.52**	2.32**	2.24**	1.36**	1.32**	1.29**	1.25**
σ_{31}	0.79**	0.82**	0.77**	0.83**	0.54**	0.58**	0.75**	0.78**	0.82**	0.85**
σ_{32}	0.01	−0.01*	0.07**	0.04**	0.10**	0.08**	0.02**	0.00	0.01	0.00
σ_{33}	2.24**	2.19**	2.32**	2.27**	2.36**	2.32**	2.22**	2.20**	2.23**	2.22**
Wald ($\sigma_{ij,i\neq \jmath}=0$)		R		R		R		R		R
Γ	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99	0.99	0.99
−LL	30.65	30.38	47.14	46.72	38.94	38.66	33.45	33.27	24.70	24.55
AIC	61.30	60.77	94.29	93.44	77.88	77.32	66.90	66.54	49.40	49.11
BIC	61.32	60.80	94.31	93.46	77.91	77.34	66.93	66.57	49.42	49.13

Notes: This table contains estimated parameters associated with the VMEM (M1) and log-VMEM (M2) specifications. $\mathbf{\Gamma}$ denotes the largest eigenvalue of $\mathbf{A}+\mathbf{B}$, LL is the system log-likelihood, AIC is the system Akaike information criterion, and BIC is the system Bayesian information criterion. Wald ($b_{ij,i\neq j}=0$) and Wald ($\sigma_{ij,i\neq j}=0$) provide an indication of Wald test rejection (R) or non-rejection (N) at the 5% level of the null of zero off-diagonal elements in $\mathbf{B}$ and $\mathbf{V}$, respectively. Significance at the 1% level is indicated by $**$, and at the 5% level by $*$.

Table 3. Estimated model parameters (cont.)

	S1		S2		S3		S4		S5	
Parameter	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Panel C: 2012 dataset (infrequently traded stocks)										
$a_{11}+b_{11}$	0.99**	0.79**	0.97**	0.70**	0.99**	0.48**	0.93**	0.74**	0.99**	0.78**
$a_{12}+b_{12}$	0.00	0.06	0.00	−0.10	0.00	0.15	0.00	0.16	0.00	0.47**
$a_{13}+b_{13}$	0.00	0.06	0.02	0.29**	0.00	0.00	0.06	0.02	0.00	−0.14
$a_{21}+b_{21}$	0.02**	0.01	0.04	−0.02	0.04	0.10**	0.05**	0.07**	0.04	0.02
$a_{22}+b_{22}$	0.92**	0.94**	0.86**	0.90**	0.80**	0.77**	0.67**	0.83**	0.77	0.75**
$a_{23}+b_{23}$	0.00	−0.09**	0.00	0.02	0.00	0.05	0.00	−0.01	0.02	0.13*
$a_{31}+b_{31}$	0.00	0.01	0.06**	0.10**	0.03	0.12**	0.00	0.09**	0.03	0.06*
$a_{32}+b_{32}$	0.04	−0.06*	0.10	0.10*	0.21	0.08	0.29	0.19	0.11	0.36**
$a_{33}+b_{33}$	0.69**	0.72**	0.70**	0.69**	0.57	0.76**	0.45*	0.58**	0.77	0.56**
$-b_{11}$	−0.43**	−0.55**	−0.34**	−0.46**	−0.34**	−0.33**	−0.17**	−0.58**	−0.52**	−0.59**
$-b_{12}$	0.00	−0.14**	0.00	0.00	0.00	−0.34**	0.00	−0.37*	0.00	−0.71**
$-b_{13}$	0.00	−0.03	0.00	−0.29*	0.00	−0.01	0.00	0.01	0.00	0.12
$-b_{21}$	−0.01	−0.02	0.00	0.06*	−0.02	−0.10*	−0.04**	−0.09**	−0.01	−0.01
$-b_{22}$	−0.85**	−0.79**	−0.74**	−0.76**	−0.71**	−0.57**	−0.54**	−0.65**	−0.69	−0.60**
$-b_{23}$	0.00	0.10**	0.00	−0.01	0.00	−0.06	0.00	0.01	−0.02	−0.14*
$-b_{31}$	0.00	−0.10**	−0.06**	−0.15**	−0.03	−0.19**	0.00	−0.19**	−0.03	−0.10**
$-b_{32}$	0.11**	0.23**	−0.09	−0.04	−0.07	0.05	−0.12	−0.07	−0.04	−0.35**
$-b_{33}$	−0.56**	−0.59**	−0.59**	−0.53**	−0.43	−0.64**	−0.33	−0.44**	−0.63	−0.38**
Wald ($b_{ij,i\neq j}=0$)		R		R		R		R		R
σ_{11}	3.88**	3.99**	5.13**	5.18**	4.60**	4.57**	4.59**	4.53**	4.91**	5.04**
σ_{21}	−0.16**	−0.13**	0.14**	0.16**	−0.18**	−0.14**	−0.22**	−0.18**	0.05	0.08*
σ_{22}	1.43**	1.37**	1.16**	1.14**	1.19**	1.15**	1.20**	1.17**	0.93**	0.91**
σ_{31}	0.61**	0.76**	1.07**	1.17**	0.99**	1.03**	0.97**	1.01**	1.17**	1.27**
σ_{32}	0.04	0.02	0.08**	0.08**	0.13**	0.11**	0.09**	0.07**	0.05*	0.06*
σ_{33}	2.27**	2.23**	1.98**	1.97**	2.11**	2.08**	2.18**	2.12**	1.76**	1.77**
Wald ($\sigma_{ij,i\neq j}=0$)		R		R		R		R		R
Γ	0.99	0.96	0.97	0.91	0.99	0.87	0.94	0.90	0.99	0.94
−LL	34.56	34.10	15.20	15.05	28.72	28.51	33.39	33.05	14.88	14.74
AIC	69.18	68.24	30.46	30.16	57.49	57.08	66.83	66.15	29.81	29.53
BIC	69.36	68.43	30.62	30.32	57.66	57.26	67.01	66.33	29.97	29.70

Notes: This table contains estimated parameters associated with the VMEM (M1) and log-VMEM (M2) specifications. $\mathbf{\Gamma}$ denotes the largest eigenvalue of $\mathbf{A}+\mathbf{B}$, LL is the system log-likelihood, AIC is the system Akaike information criterion, and BIC is the system Bayesian information criterion. Wald ($b_{ij,i\neq j}=0$) and Wald ($\sigma_{ij,i\neq j}=0$) provide an indication of Wald test rejection (R) or non-rejection (N) at the 5% level of the null of zero off-diagonal elements in $\mathbf{B}$ and $\mathbf{V}$, respectively. Significance at the 1% level is indicated by $**$, and at the 5% level by $*$.

Table 3. Estimated model parameters (cont.)

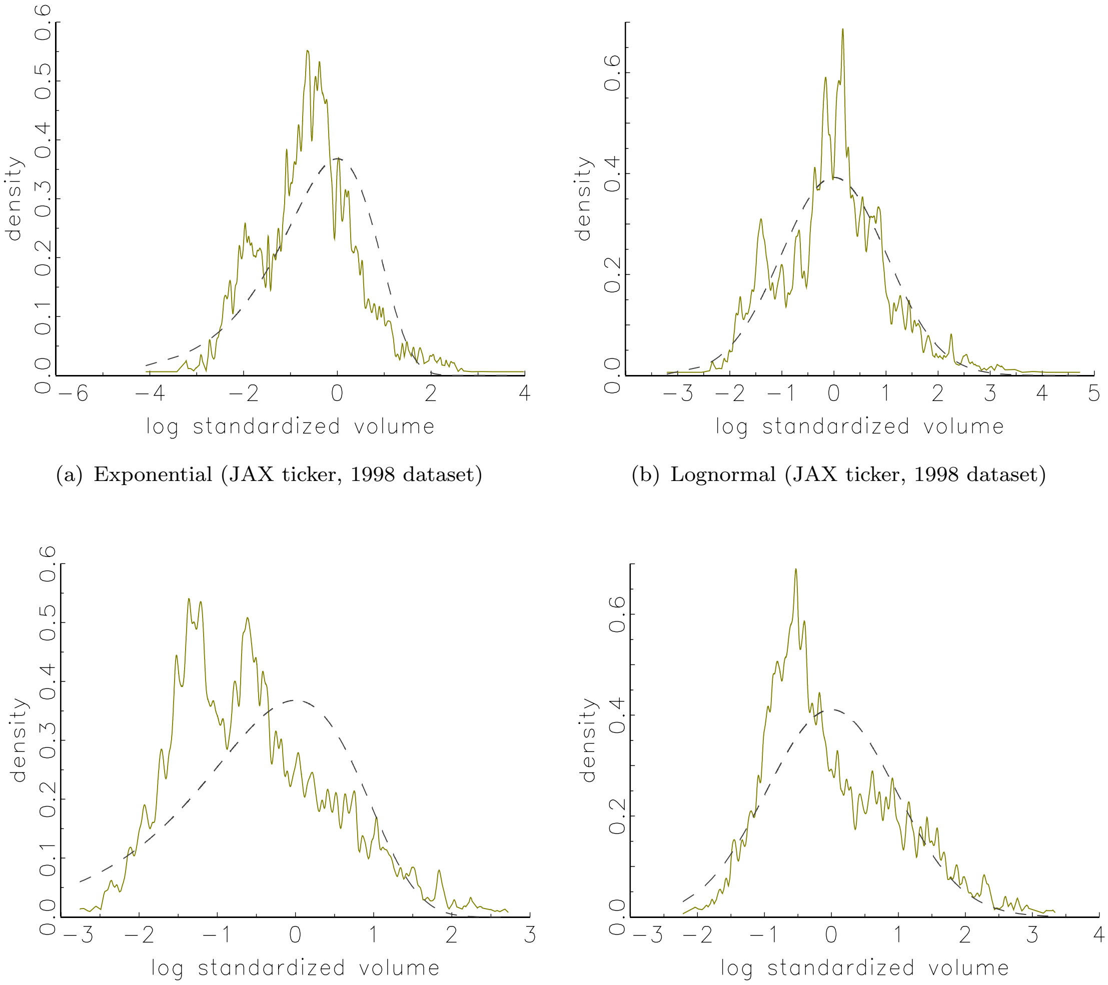
	S1		S2		S3		S4		S5	
Parameter	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Panel D: 2012 dataset (frequently traded stocks)										
$a_{11}+b_{11}$	0.94**	0.88**	0.89**	0.89**	0.96**	0.90**	0.95**	0.98**	0.95**	0.93**
$a_{12}+b_{12}$	0.00	0.15**	0.00	0.15**	0.00	0.13**	0.00	0.18**	0.00	0.12**
$a_{13}+b_{13}$	0.01	0.14**	0.01	0.10**	0.01	0.09**	0.00	0.11**	0.01	0.07**
$a_{21}+b_{21}$	0.00	0.02**	0.00	0.02**	0.00	0.03**	0.00	0.00	0.00	0.03**
$a_{22}+b_{22}$	0.69**	0.96**	0.65**	0.96**	0.74**	0.94**	0.89**	0.93**	0.67**	0.94**
$a_{23}+b_{23}$	0.00	−0.02**	0.00	−0.02**	0.00	−0.03**	0.00	−0.04**	0.02**	−0.03**
$a_{31}+b_{31}$	0.02	0.12**	0.00	0.10**	0.03	0.11**	0.01	0.03	0.01	0.08**
$a_{32}+b_{32}$	−0.76**	−0.15*	−0.78**	−0.13**	−0.28**	−0.13**	−0.11**	−0.22**	−0.47**	−0.09**
$a_{33}+b_{33}$	0.96**	0.81**	0.97**	0.88**	0.93**	0.86**	0.96**	0.84**	0.97**	0.89**
$-b_{11}$	−0.79**	−0.75**	−0.73**	−0.76**	−0.83**	−0.77**	−0.83**	−0.87**	−0.79**	−0.80**
$-b_{12}$	0.00	−0.30**	0.00	−0.29**	0.00	−0.28**	0.00	−0.27**	0.00	−0.27**
$-b_{13}$	0.00	−0.15**	0.00	−0.10**	0.00	−0.08**	0.00	−0.11**	0.00	−0.07**
$-b_{21}$	0.00	−0.03**	0.00	−0.04**	0.00	−0.06**	0.00	−0.03**	0.00	−0.07**
$-b_{22}$	−0.66**	−0.93**	−0.62**	−0.92**	−0.68**	−0.86**	−0.86**	−0.89**	−0.61**	−0.85**
$-b_{23}$	0.00	0.02**	0.00	0.01**	0.00	0.03**	0.00	0.03**	−0.02**	0.03**
$-b_{31}$	−0.02**	−0.25**	0.00	−0.24**	−0.03**	−0.24**	−0.01**	−0.16**	−0.01**	−0.22**
$-b_{32}$	0.87**	0.30**	0.87**	0.26**	0.37**	0.31**	0.15**	0.32**	0.59**	0.28**
$-b_{33}$	−0.91**	−0.75**	−0.91**	−0.81**	−0.85**	−0.78**	−0.92**	−0.78**	−0.92**	−0.82**
Wald ($b_{ij,i\neq j}=0$)		R		R		R		R		R
σ_{11}	1.58**	1.51**	1.30**	1.23**	1.59**	1.50**	1.08**	1.03**	1.31**	1.24**
σ_{21}	0.04**	0.04**	0.02**	0.02**	0.01	0.00	0.02**	0.02**	−0.01**	−0.02**
σ_{22}	0.79**	0.78**	0.79**	0.78**	0.92**	0.91**	0.78**	0.77**	0.97**	0.96**
σ_{31}	0.82**	0.81**	0.69**	0.68**	0.72**	0.70**	0.69**	0.68**	0.59**	0.58**
σ_{32}	0.01**	0.01**	−0.01	−0.01*	0.00	−0.01	0.00	−0.01*	0.00	−0.01**
σ_{33}	2.06**	2.00**	2.05**	1.99**	2.24**	2.13**	1.92**	1.88**	2.23**	2.13**
Wald ($\sigma_{ij,i\neq J}=0$)		R		R		R		R		R
Γ	0.97	0.99	0.97	0.99	0.97	0.98	0.97	1.00	0.95	0.99
−LL	47.71	47.25	66.43	65.69	48.03	47.38	93.36	92.50	63.54	62.74
AIC	95.43	94.51	132.86	131.38	96.07	94.77	186.73	185.00	127.08	125.49
BIC	95.45	94.53	132.89	131.40	96.09	94.79	186.76	185.03	127.11	125.51

Notes: This table contains estimated parameters associated with the VMEM (M1) and log-VMEM (M2) specifications. $\mathbf{\Gamma}$ denotes the largest eigenvalue of $\mathbf{A}+\mathbf{B}$, LL is the system log-likelihood, AIC is the system Akaike information criterion, and BIC is the system Bayesian information criterion. Wald ($b_{ij,i\neq j}=0$) and Wald ($\sigma_{ij,i\neq j}=0$) provide an indication of Wald test rejection (R) or non-rejection (N) at the 5% level of the null of zero off-diagonal elements in $\mathbf{B}$ and $\mathbf{V}$, respectively. Significance at the 1% level is indicated by $**$, and at the 5% level by $*$.

Table 4. Impulse response function absorbtion times

	S1		S2		S3		S4		S5	
Shock	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Panel A: 1998 dataset (infrequently traded stocks)										
Duration	894.3	365.4	297.4	357.8	784.0	737.4	767.0	690.6	356.9	123.1
Volume	567.3	418.9	384.0	435.0	621.6	677.6	662.2	670.9	322.6	119.5
Volatility	830.9	429.4	391.3	400.7	725.6	749.6	823.8	689.4	380.6	129.2
Panel B: 1998 dataset (frequently traded stocks)										
Duration	59.3	71.5	40.2	133.0	44.2	60.4	56.6	89.7	81.0	57.5
Volume	46.8	67.0	32.2	98.0	31.8	60.2	40.2	77.7	64.2	52.0
Volatility	54.6	71.0	37.1	132.3	39.4	60.4	52.7	88.4	75.1	56.5
Panel C: 2012 dataset (infrequently traded stocks)										
Duration	69.9	19.8	100.9	19.7	14.1	8.1	14.9	8.8	219.5	30.8
Volume	56.2	22.7	85.4	20.8	94.6	8.1	10.9	9.0	166.8	31.7
Volatility	62.2	21.5	94.7	19.1	105.4	8.1	13.9	8.3	206.5	30.6
Panel D: 2012 dataset (frequently traded stocks)										
Duration	1.4	4.9	1.4	4.1	2.1	3.3	0.9	5.7	0.9	4.9
Volume	1.5	5.3	1.5	4.3	1.8	3.5	0.9	5.8	0.9	5.1
Volatility	1.4	4.8	1.4	4.0	2.0	3.0	0.9	5.5	0.9	4.5

Notes: This table contains the time in hours for a shock to be absorbed into the volatility equation (that is, when the variance of the volatility response is less than 10^{-7}). We use the Manganelli (2005) method to approximate the calendar time the system takes to return to its long-run equilibrium, by multiplying the number of transactions by their average duration.



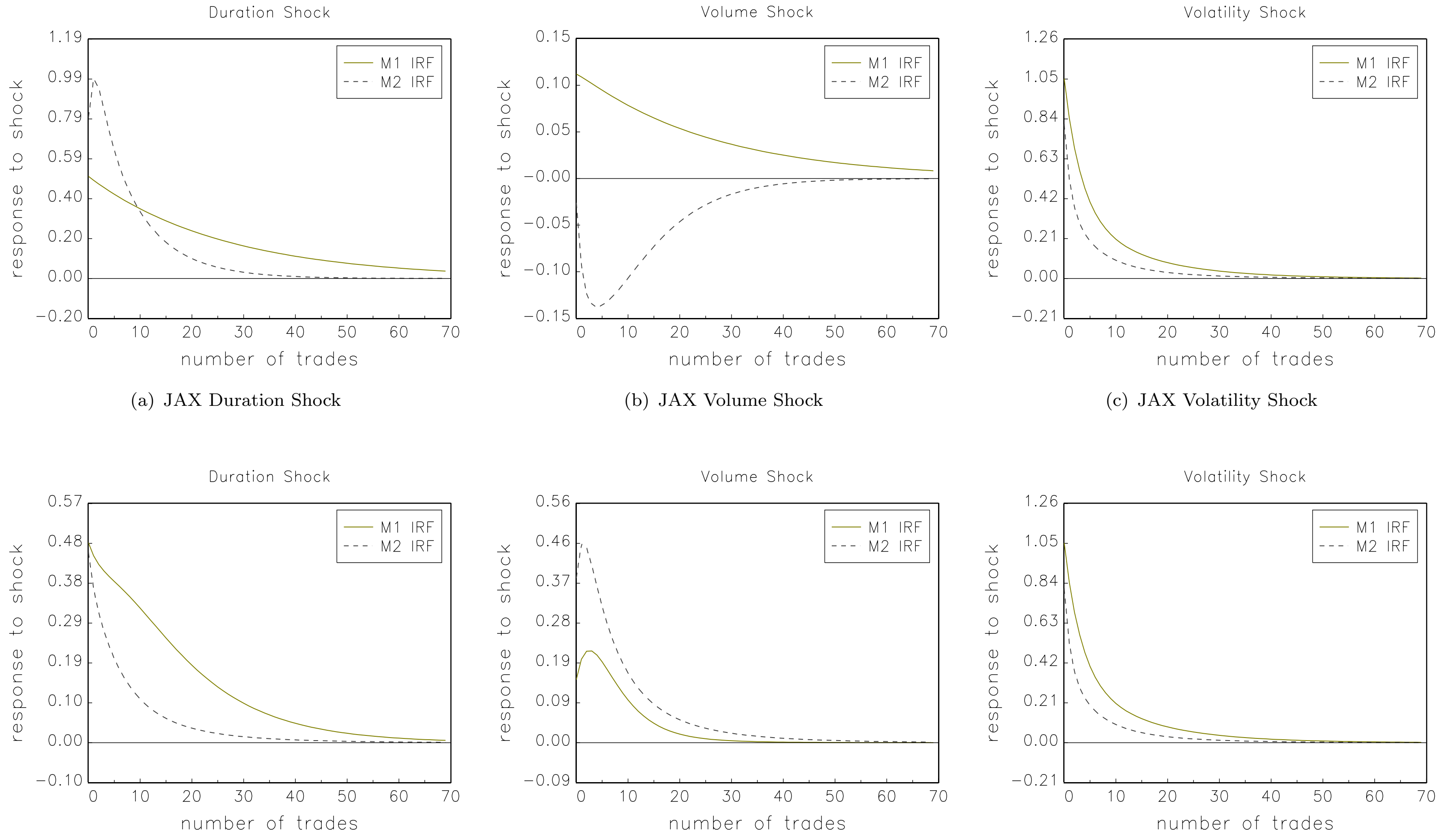
(a) Exponential (JAX ticker, 1998 dataset)

(b) Lognormal (JAX ticker, 1998 dataset)

(c) Exponential (TKF ticker, 2012 dataset)

(d) Lognormal (TKF ticker, 2012 dataset)

Figure 1. A comparison of parametric and non-parametric densities
This figure contains distribution plots of trading volume associated with parametric and non-parametric densities. The first column of panels contain plots of non-parametric and exponential densities; the second column of panels contain plots of non-parametric and lognormal densities. In both cases the non-parametric (parametric) densities are given by the solid (dashed) line.



(a) JAX Duration Shock

(b) JAX Volume Shock

(c) JAX Volatility Shock

(d) TKF Duration Shock

(e) TKF Volume Shock

(f) TKF Volatility Shock

Figure 2. Impulse response functions

This figure contains impulse response functions associated with the volatility reaction to various shocks using M1 and M2 parameter values.