



Taylor, N., & Xu, Y. (2016). The logarithmic vector multiplicative error model: an application to high frequency NYSE stock data. *Quantitative Finance*. Advance online publication.
<https://doi.org/10.1080/14697688.2016.1260756>

Peer reviewed version

Link to published version (if available):
[10.1080/14697688.2016.1260756](https://doi.org/10.1080/14697688.2016.1260756)

[Link to publication record on the Bristol Research Portal](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Taylor & Francis at <http://www.tandfonline.com/doi/full/10.1080/14697688.2016.1260756>. Please refer to any applicable terms of use of the publisher.

University of Bristol – Bristol Research Portal

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–28

The logarithmic vector multiplicative error model: An application to high frequency NYSE stock data

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). The log-vMEM improves on existing models in two ways. First, it is a more general form model as it allows the error terms to be cross-dependent and relaxes weak exogeneity restrictions. Second, the log-vMEM specification guarantees that the conditional means are non-negative without any restrictions imposed on the parameters. 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 of trading duration, volume and volatility, with the log-vMEM providing a better fit to the data than a competing model. 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

There is considerable interest in econometric models of irregularly spaced financial data. A seminal early example is the autoregressive conditional duration (ACD) model introduced by Engle and Russell (1998). In this univariate model rich dynamics are permitted by allowing conditional expectations of trading duration (that is, the time between consecutive transactions) to depend on past realised durations and past conditional expectations.¹ This model has subsequently been extended into a multivariate setting. For instance, Engle (2000) proposes a recursive model to represent the dynamics of duration and volatility.² 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 is proposed by Manganelli (2005), whereby volume is incorporated into the Engle (2000) model such that duration, volume and volatility are jointly considered.³ 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

*Corresponding author. Email: nick.taylor@bristol.ac.uk

¹Trading 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.

²Microstructure theory indicates that trading duration and volume convey information content with respect to fundamental asset prices; see, e.g., Easley and O'Hara (1992).

³The theoretical relationship between duration, volume and volatility has been the subject of considerable debate; see Diamond and Verrecchia (1987) and Easley and O'Hara (1992) for contrasting predictions.

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. We add to this body of work by specifying a flexible econometric model that avoids a number of restrictions imposed in existing models.

The virtue of the recursive framework is that it reduces model complexity, since each process is estimated separately. However, recursive-type models have limitations. First, they assume that the error terms are independent. To incorporate contemporaneous information, ad hoc causality assumptions are imposed. For instance, Manganelli (2005) specifies causality from duration to volume and from duration and volume to volatility (cf. Engle and Sun, 2007, and Hautsch, 2008). Second, they assume that the conditional expectation of one variable is a function only of its own past conditional expectation (see, e.g., Engle, 2000, Dufour and Engle, 2000, Manganelli, 2005, and Engle and Sun, 2007). However, for many applications, more general specifications of the conditional expectation equations should be considered.¹

As an extension of the recursive model, Cipollini et al. (2007) propose a vector multiplicative error model (vMEM), in which the variables of interest are assumed to be interdependent processes that evolve simultaneously. A complete parametric specification of this model requires full formulation of the conditional distribution of the multivariate non-negative random process. However, the specification of the conditional distribution of the 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 maximum likelihood (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. This is a general form vMEM since the coefficient matrix can be full.

However, both the recursive model and the general form vMEM face non-negativity challenges. Specifically, the conditional means of the variables under study almost always have to be non-negative by construction (e.g., duration, volatility and volume). A sufficient condition for non-negativity of the conditional means is that all parameters are non-negative (He and Teräsvirta, 2004). However, this is a potentially strong restriction. Indeed, previous studies such as Manganelli (2005), Engle and Gallo (2006), and Cipollini et al. (2007, 2013) choose not to impose this restriction. A consequence of this choice is that their results show that some of their estimated parameters are negative – thus violating the non-negativity condition and implying that predicted values could be negative.

Within the context of the above challenges we consider an alternative general form vMEM. In doing this we contribute to the existing literature in two ways. First, 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 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 methodology due to misspecification of the error distribution is trivial. The second contribution relates to the non-negativity of the conditional means of the variables under study. Specifically, we consider a logarithmic version of the vMEM, such that the conditional means are guaranteed to be positive without any restrictions placed on the parameters; see Bauwens and Giot (2000) and Bauwens et al. (2008) for use of the log specification within the context of ACD models. By using this model we build an unrestricted system that incorporates various causal and feedback effects among the 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 em-

¹For instance, Engle (2000) and Grammig et al. (2007) argue that the unexpected components of the trading process 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.

pirically 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 and volume rather than observed duration and volume that contain 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. The vMEM and log-vMEM specification

Define \mathbf{x}_t , $t = 1, \dots, T$, as the K -dimension vector of positively valued variables (e.g., measures of trading activity), with conditional mean vector $\boldsymbol{\mu}_t$ and error vector $\boldsymbol{\epsilon}_t$. Cipollini et al. (2007) propose the following vMEM for \mathbf{x}_t , $\boldsymbol{\mu}_t$, and $\boldsymbol{\epsilon}_t$:

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

where

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

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 $E(\mathbf{x}_t | \Omega_t) = \boldsymbol{\mu}_t$ and $\text{var}(\mathbf{x}_t | \Omega_t) = \boldsymbol{\mu}_t \boldsymbol{\mu}_t' \odot \boldsymbol{\Sigma}$, with the latter a positive definite matrix by construction.

The above vMEM specification has to be parameterized in a way that guarantees the non-negativity of the conditional mean $\boldsymbol{\mu}_t$ at all points in time. Since the vMEM has the same structure as the extended constant conditional correlation GARCH (ECCC-GARCH) model, theoretical results on the non-negativity of this model can be applied here. In general, a sufficient condition to guarantee non-negativity of $\boldsymbol{\mu}_t$ is $\boldsymbol{\omega} \geq 0$, $\mathbf{A}_i \geq 0$, $\mathbf{B}_i \geq 0$ (He and Teräsvirta, 2004). Conrad and Karanasos (2010) relax this restriction by allowing only one element of \mathbf{B}_i 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 non-negativity conditions. However, a more convincing way is to specify a model in which predicted values are guaranteed to be positive.

Motivated by the log-ACD model of Bauwens and Giot (2000), and Bauwens et al. (2008), we consider a logarithmic version of the vMEM so that the conditional mean is guaranteed to be positive without any restrictions placed on the parameters. Specifically,

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim D(\mathbf{1}, \boldsymbol{\Sigma}), \quad (3)$$

where

$$\ln \boldsymbol{\mu}_t = \tilde{\boldsymbol{\omega}} + \sum_{i=1}^p \tilde{\mathbf{A}}_i \ln \mathbf{x}_{t-i} + \sum_{i=1}^q \tilde{\mathbf{B}}_i \ln \boldsymbol{\mu}_{t-i}. \quad (4)$$

This is referred to as the log-vMEM.¹ Here tildes are applied to the coefficients to highlight the fact that the coefficients in the vMEM and log-vMEM are different from each other. 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.

A complete parametric formulation of the vMEM or log-vMEM requires full specification of the conditional distribution of the non-negative random processes in $\boldsymbol{\epsilon}_t$. In the vMEM literature, Cipollini et al. (2007) adopt a copula-based approach. However, to enable use of ML estimation, the \mathbf{B}_i 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. This is considered a general form vMEM in the sense that the \mathbf{B}_i matrix is full. However, use of GMM estimation in the context of the log-vMEM is complicated as there are no analytical solutions for the conditional first and second moments of \mathbf{x}_t ; see Bauwens and Giot (2000) for a discussion in the univariate case. For this reason, we propose a multivariate lognormal distribution for the errors, so that ML estimation can be employed. Our linear vMEM is close to Cipollini et al.'s (2013) in the sense that both have a general form. However, we propose to use a multivariate lognormal distribution for the errors, so that we have a full parametric (log-)vMEM.

In a 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 realised 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 non-negative distributions, and is also sufficiently accurate so as not to induce unnecessary numerical difficulties. A possible limitation of this distribution 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 (e.g., duration, number of trades, volume, bid-ask spread, realised volatility, and daily high-low range) are actually positively valued.

2.2. The ML estimator and its asymptotic properties

The following ML estimator is proposed:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} -l(\boldsymbol{\theta}), \quad (5)$$

where $\boldsymbol{\theta}$ is a vector incorporating the parameters of interest, and $l(\boldsymbol{\theta})$ 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 $\boldsymbol{\epsilon}_t \sim \ln N(\mathbf{M}, \mathbf{V})$, with its density function given

¹This particular specification can be augmented in a number of ways. For instance, asymmetric effects could be incorporated by adding interaction dummy variables that condition on the values of \mathbf{x}_{t-i} (or similarly defined variables); see Cipollini et al. (2013) for such an augmentation within the context of the vMEM.

by

$$f(\boldsymbol{\epsilon}_t) = (2\pi)^{-K/2} |\mathbf{V}|^{-1/2} \prod_{i=1}^K \epsilon_{i,t}^{-1} \exp(-(\ln \boldsymbol{\epsilon}_t - \mathbf{M})' \mathbf{V}^{-1} (\ln \boldsymbol{\epsilon}_t - \mathbf{M})/2). \quad (6)$$

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

$$f(\mathbf{x}_t | \boldsymbol{\theta}) = (2\pi)^{-K/2} |\mathbf{V}|^{-1/2} \prod_{i=1}^K x_{i,t}^{-1} \exp(-(\ln \mathbf{x}_t - \ln \boldsymbol{\mu}_t - \mathbf{M})' \mathbf{V}^{-1} (\ln \mathbf{x}_t - \ln \boldsymbol{\mu}_t - \mathbf{M})/2). \quad (7)$$

The log likelihood of the model is then

$$l(\boldsymbol{\theta}) = \sum_{t=1}^T l_t(\boldsymbol{\theta}) = \sum_{t=1}^T \ln f(\mathbf{x}_t | \boldsymbol{\theta}), \quad (8)$$

where

$$\ln f(\mathbf{x}_t | \boldsymbol{\theta}) = -\frac{K}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{V}| - \sum_{i=1}^K \ln x_{i,t} - \frac{1}{2} (\ln \mathbf{x}_t - \ln \boldsymbol{\mu}_t - \mathbf{M})' \mathbf{V}^{-1} (\ln \mathbf{x}_t - \ln \boldsymbol{\mu}_t - \mathbf{M}). \quad (9)$$

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

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 $\boldsymbol{\rho} = \text{vech}(\mathbf{V})$, 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

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

Proof. Standard vector/matrix differentiation of the log likelihood function in (8) eventually leads to the above expression. \square

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

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

Proof. Standard vector/matrix differentiation of the log likelihood function in (8) eventually leads to the above expression. \square

The consistency and asymptotic normality of the ML estimator $\widehat{\boldsymbol{\theta}}$ follows from a more general ML 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 by¹

$$\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, J^{-1}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)J^{-1}(\boldsymbol{\theta}_0)), \quad (10)$$

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,

$$I(\boldsymbol{\theta}_0) = \frac{1}{T} E(S(\boldsymbol{\theta}_0)S(\boldsymbol{\theta}_0)') = E(S_t(\boldsymbol{\theta}_0)S_t(\boldsymbol{\theta}_0)'), \quad (11)$$

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

$$J(\boldsymbol{\theta}_0) = -\frac{1}{T} E(H(\boldsymbol{\theta}_0)) = -E(H_t(\boldsymbol{\theta}_0)). \quad (12)$$

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

2.3. 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

¹The 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.

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 ϵ_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 standard deviation and the two correlation scenarios. We assume that $M_i = -V_{ii}/2 = -0.5$ to ensure that $E(\epsilon_t) = \mathbf{1}$ in the lognormal distribution. Estimated vMEM parameter means and root mean square error (RMSE) values using the GMM and ML estimation methodologies are reported 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 the cases, with the efficiency loss relative to the 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 outperforms 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 only 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 to construct measures of duration (defined as the time elapsing between consecutive trades), volume (the trade size associated with each transaction), volatility (measured by the absolute return), and returns (calculated using the mid-quote price change).² The first dataset is obtained from the NYSE-based Trades and Quotes (TAQ) dataset,

¹Within 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.

²To eliminate exact zero problems associated with absolute returns, we add a small constant (10 percent of the mean) to the absolute return series. This simple approach alleviates exact zero problems, but does not adversely affect the dynamics of absolute returns. 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.

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, 1999.³ 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.⁴

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.¹ This process of stock selection and preparation is also used in the construction of the 2012 dataset.² The tickers of the ten stocks in the 1998 and 2012 datasets are reported in Table 2.

Insert Table 2 here

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 1.³ 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 lends support to the use of the lognormal (log)-vMEM.

Insert Figure 1 here

³These data were kindly supplied by Simone Manganelli.

⁴Since 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.

¹Note that trade durations are positive by construction. The minimum permitted time between trades is one second. Transactions that occur within one second are treated as one transaction and the volumes are aggregated; see Hautsch et al. (2014) for an alternative approach.

²The 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.

³The 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.

3.2. Estimated models

Define $\mathbf{x}_t = (d_t, v_t, |r_t|)'$, $t = 1, \dots, T$, as the three-dimension time series associated with the duration, volume and volatility processes, respectively. We apply two first-order models to these data: the restricted vMEM and unrestricted log-vMEM. Both models are estimated using the proposed ML estimation methodology. The vMEM is a restricted model in the sense that we have to impose restrictions to ensure that the conditional means are non-negative. If the estimated \mathbf{A} and \mathbf{B} matrices are all non-negative, then the vMEM would be unrestricted. However, if these matrices have negative elements, then we are required to restrict them to be non-negative to ensure that predicted values are non-negative.⁴ 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 non-negative without any restrictions.

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

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + (\mathbf{A} + \mathbf{B})\mathbf{x}_{t-1} - \mathbf{B}\mathbf{e}_{t-1}, \quad (13)$$

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

$$\ln \boldsymbol{\mu}_t = \tilde{\boldsymbol{\omega}} + (\tilde{\mathbf{A}} + \tilde{\mathbf{B}}) \ln \mathbf{x}_{t-1} - \tilde{\mathbf{B}} \ln \boldsymbol{\epsilon}_{t-1}. \quad (14)$$

This model specification is henceforth denoted M2.

Various causal and feedback effects among the three variables can be examined by using the above specifications. For example in M1, $a_{31} + b_{31}$ ($a_{32} + b_{32}$) 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 a superior fit to the data. Moreover, many off-diagonal elements in the \mathbf{A} and \mathbf{B} matrices associated with M1 are zero. This indicates that the non-negativity constraints have been hit with the corresponding parameters forced to equal zero. Given the inferior fit of M1 this restriction suggests a loss of useful information.

Insert Table 3 here

The proposed log-vMEM allows the coefficient matrix $\tilde{\mathbf{B}}$ and the covariance matrix \mathbf{V} to have non-zero off-diagonal elements. Consequently, the virtue of our model can be examined by conducting Wald tests in which $\tilde{\mathbf{B}}$ and \mathbf{V} have zero off-diagonal restrictions imposed. The results in Table 3 indicate that the tests applied to all off-diagonal elements of $\tilde{\mathbf{B}}$ are almost always significant for all datasets. Moreover, the tests applied to the off-diagonal elements of \mathbf{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.

⁴To 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 non-negative. 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, the non-negativity condition of Conrad and Karanasos (2012) is violated (that is, $\mathbf{B}\mathbf{A} < \mathbf{0}$ or $\mathbf{B}^2\mathbf{A} < \mathbf{0}$), then we restrict all elements to be non-negative and re-estimate the model.

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 hypotheses.

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 realised values and shocks (and their impact). If, however, duration (volume) is predictable then realised 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 ($\tilde{a}_{31} + \tilde{b}_{31}$) and duration shock coefficient ($-\tilde{b}_{31}$) 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, $\tilde{a}_{31} + \tilde{b}_{31}$ is positive and $-\tilde{b}_{31}$ 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 ($\tilde{a}_{32} + b_{32}$) and volume shock coefficient ($-\tilde{b}_{32}$) 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, since many of the coefficients (\tilde{a}_{31} , \tilde{b}_{31} , \tilde{a}_{32} , \tilde{b}_{32}) are insignificant. As transactions in these illiquid stocks occur only every 28 to 74 minutes (see Table 2), it follows that these transactions may contain far less information compared with the frequently observed transactions. 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 ($\tilde{a}_{13} + \tilde{b}_{13}$) is positive and significant, while the volatility shock coefficient ($-\tilde{b}_{13}$) is negative and significant for both the 1998 and 2012 datasets. 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 $\mathbf{A} + \mathbf{B}$ (M1) or $\tilde{\mathbf{A}} + \tilde{\mathbf{B}}$ (M2) 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 absorption 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 vMEM and log-vMEM and show that the vMEM tends to be more restrictive in terms of permitted parameter values than the log-vMEM. This manifests itself in the log-vMEM having a superior fit to the data.
- (ii) 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 non-negative 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 (Engle and Gallo, 2006). One possibility is to consider absolute daily returns, daily high-low range and daily realised 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. *Quantitative and Qualitative Analysis in Social Sciences* 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, 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.

- Hautsch N, Malec P, Schienle M. 2014. Capturing the zero: A new class of zero-augmented distributions and multiplicative error processes. *Journal of Financial Econometrics* 12: 89–121.
- He C, Terasvirta 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. The vMEM derivation

Consider the following vMEM:

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

where

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

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

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

It follows that

$$\boldsymbol{\mu}_t = \mathbf{x}_t - \mathbf{e}_t, \quad (\text{A.4a})$$

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

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

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

where $\mathbf{C}_i = \mathbf{A}_i + \mathbf{B}_i$, 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 $|\mathbf{I} - \mathbf{C}_1 z - \mathbf{C}_2 z^2 \dots \mathbf{C}_r z^r| = 0$ are all greater than one, and invertible if the modulus of the roots of $|\mathbf{I} - \mathbf{B}_1 z - \mathbf{B}_2 z^2 \dots \mathbf{B}_q z^q| = 0$ are all greater than one.

A.2. The log-vMEM derivation

Consider the following log-vMEM:

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

where

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

Taking logs of (A.6) we obtain

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

It follows that

$$\ln \boldsymbol{\mu}_t = \ln \mathbf{x}_t - \mathbf{c} - \mathbf{e}_t, \quad (\text{A.9a})$$

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

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

$$\begin{aligned} \ln \mathbf{x}_t &= \bar{\mathbf{c}} + \sum_{i=1}^p \tilde{\mathbf{A}}_i \ln \mathbf{x}_{t-i} + \sum_{i=1}^q \tilde{\mathbf{B}}_i \ln \mathbf{x}_{t-i} + \mathbf{e}_t - \sum_{i=1}^q \tilde{\mathbf{B}}_i \mathbf{e}_{t-i}, \\ &= \bar{\mathbf{c}} + \sum_{i=1}^r \tilde{\mathbf{C}}_i \ln \mathbf{x}_{t-i} + \mathbf{e}_t - \sum_{i=1}^q \tilde{\mathbf{B}}_i \mathbf{e}_{t-i}, \end{aligned} \quad (\text{A.10})$$

where $\bar{\mathbf{c}} = \mathbf{c} + \boldsymbol{\omega} - \sum_{i=1}^q \tilde{\mathbf{B}}_i \mathbf{c}$, $\tilde{\mathbf{C}}_i = \tilde{\mathbf{A}}_i + \tilde{\mathbf{B}}_i$, 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 $|\mathbf{I} - \tilde{\mathbf{C}}_1 z - \tilde{\mathbf{C}}_2 z^2 - \dots - \tilde{\mathbf{C}}_r z^r| = 0$ are all greater than one, and invertible if the modulus of the roots of $|\mathbf{I} - \tilde{\mathbf{B}}_1 z - \tilde{\mathbf{B}}_2 z^2 - \dots - \tilde{\mathbf{B}}_q z^q| = 0$ are all greater than one.

Appendix B: Impulse response functions

Under stationary conditions, the impulse response functions associated with the first-order vMEM and log-vMEM are derived. We use the methodology used in Engle et al. (2012) to derive these functions.

B.1. The vMEM derivation

Consider the following first-order vMEM:

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

where

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

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

$$\mathbf{x}_t = \boldsymbol{\mu}_t, \quad (\text{B.3})$$

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

with values of $\boldsymbol{\mu}_t$ conditional on shocks occurring at $t = 0$ given by

$$\boldsymbol{\mu}_{t|0} = \bar{\boldsymbol{\omega}} + (\mathbf{A} + \mathbf{B})^t \mathbf{x}_0 \quad \forall t > 0, \quad (\text{B.5})$$

where $\bar{\boldsymbol{\omega}} = \sum_{i=1}^t (\mathbf{A} + \mathbf{B})^{i-1} \boldsymbol{\omega}$. If a shock occurs to \mathbf{x}_t at time $t = 0$ then the impulse response function for $t > 0$ is given by

$$\frac{\partial \text{E}(\boldsymbol{\mu}_t | \Omega_0)}{\partial \mathbf{x}'_0} = \boldsymbol{\Lambda}^t, \quad (\text{B.6})$$

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 i th element of $\boldsymbol{\epsilon}_0$ (denoted ϵ_{i0}). The size of the shock is assumed to equal the unconditional standard deviation of the i th element of $\boldsymbol{\epsilon}_t$. The effect of this shock on \mathbf{x}_0 is given by

$$\frac{\partial \mathbf{x}_0}{\partial \epsilon_{i0}} = \boldsymbol{\mu}_0 \odot \mathbf{s}_i, \quad (\text{B.7})$$

where

$$\mathbf{s}_i = \left(\sigma_i \frac{\sigma_{i1}}{\sigma_i^2}, \sigma_i \frac{\sigma_{i2}}{\sigma_i^2}, \sigma_i \frac{\sigma_{i3}}{\sigma_i^2} \right)'. \quad (\text{B.8})$$

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 $\text{E}(\epsilon_{jt} | \epsilon_{it} = 1 + \sigma_i) = 1 + \sigma_i \sigma_{ij} / \sigma_i^2$.

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

$$\frac{\partial \text{E}(\boldsymbol{\mu}_t | \Omega_0)}{\partial \boldsymbol{\epsilon}'_0} = \boldsymbol{\Lambda}^t \text{dg}(\boldsymbol{\mu}_0) \boldsymbol{\Xi}, \quad (\text{B.9})$$

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

$$\boldsymbol{\Xi} = \begin{bmatrix} \sigma_1 & \frac{\sigma_{21}}{\sigma_2} & \frac{\sigma_{31}}{\sigma_3} \\ \frac{\sigma_{12}}{\sigma_1} & \sigma_2 & \frac{\sigma_{32}}{\sigma_3} \\ \frac{\sigma_{13}}{\sigma_1} & \frac{\sigma_{23}}{\sigma_2} & \sigma_3 \end{bmatrix}. \quad (\text{B.10})$$

B.2. The log-vMEM derivation

Consider the following first-order log-vMEM:

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \ln \text{N}(\mathbf{M}, \mathbf{V}), \quad (\text{B.11})$$

where

$$\ln \boldsymbol{\mu}_t = \tilde{\boldsymbol{\omega}} + \tilde{\mathbf{A}} \ln \mathbf{x}_{t-1} + \tilde{\mathbf{B}} \ln \boldsymbol{\mu}_{t-1}. \quad (\text{B.12})$$

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

$$\ln \mathbf{x}_t = \ln \boldsymbol{\mu}_t, \quad (\text{B.13})$$

$$\ln \boldsymbol{\mu}_t = (\mathbf{I} - \tilde{\mathbf{A}} - \tilde{\mathbf{B}})^{-1} \tilde{\boldsymbol{\omega}} \quad \forall t < 0, \quad (\text{B.14})$$

with values of $\ln \boldsymbol{\mu}_t$ conditional on shocks occurring at $t = 0$ given by

$$\ln \boldsymbol{\mu}_{t|0} = \tilde{\boldsymbol{\omega}} + (\tilde{\mathbf{A}} + \tilde{\mathbf{B}})^t \ln \mathbf{x}_0 \quad \forall t > 0, \quad (\text{B.15})$$

where $\tilde{\boldsymbol{\omega}} = \sum_{i=1}^t (\tilde{\mathbf{A}} + \tilde{\mathbf{B}})^{i-1} \tilde{\boldsymbol{\omega}}$. If a shock occurs to \mathbf{x}_t at time $t = 0$ then the impulse response function for $t > 0$ is given by

$$\frac{\partial \text{E}(\ln \boldsymbol{\mu}_t | \Omega_0)}{\partial \ln \mathbf{x}'_0} = \tilde{\boldsymbol{\Lambda}}^t, \quad (\text{B.16})$$

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

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 i th element of $\boldsymbol{\epsilon}_0$ (denoted ϵ_{i0}). The size of the shock is assumed to equal the unconditional standard deviation of the i th element of $\boldsymbol{\epsilon}_t$. The effect of this shock on $\ln \mathbf{x}_0$ is given by

$$\frac{\partial \ln \mathbf{x}_0}{\partial \epsilon_{i0}} = \boldsymbol{\mu}_0 \odot \frac{1}{\mathbf{x}_0} \odot \mathbf{s}_i, \quad (\text{B.17})$$

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

$$\frac{\partial \text{E}(\ln \boldsymbol{\mu}_t | \Omega_0)}{\partial \epsilon'_{i0}} = \tilde{\boldsymbol{\Lambda}}^t \boldsymbol{\Xi}, \quad (\text{B.18})$$

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

$$\frac{\partial \text{E}(\boldsymbol{\mu}_t | \Omega_0)}{\partial \epsilon'_{i0}} = \text{dg}(1/\boldsymbol{\mu}_t) \tilde{\boldsymbol{\Lambda}}^t \boldsymbol{\Xi}, \quad (\text{B.19})$$

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

$$\boldsymbol{\mu}_{t|0} = \exp(\tilde{\boldsymbol{\omega}} + \tilde{\mathbf{A}} \ln \boldsymbol{\mu}_{t-1|0} + \tilde{\mathbf{B}} \ln \boldsymbol{\mu}_{t-1|0}) \quad \forall t > 0, \quad (\text{B.20})$$

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

Table 1. Simulation results

Parameter	True	Weak Correlation ($\rho = 0.4$)				Strong Correlation ($\rho = 0.8$)			
		GMM		ML		GMM		ML	
		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 vMEM parameter mean and root mean square error (RMSE) values associated with the GMM and ML estimation methodologies. 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 is 25.00 and the corresponding value associated with the MLB test statistic is 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		S2		S3		S4		S5	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Panel A: 1998 dataset (infrequently traded stocks)										
$a_{11} + b_{11}$ (M1) or $\tilde{a}_{11} + \tilde{b}_{11}$ (M2)	0.99**	0.65**	0.92**	0.84**	0.99**	0.96**	0.99**	0.33**	0.99**	0.33
$a_{12} + b_{12}$ (M1) or $\tilde{a}_{12} + \tilde{b}_{12}$ (M2)	0.00	-0.37	0.07	0.07	0.00	0.14**	0.00	0.63	0.00	0.67
$a_{13} + b_{13}$ (M1) or $\tilde{a}_{13} + \tilde{b}_{13}$ (M2)	0.00	0.24	0.00	0.02	0.00	0.11**	0.00	-0.57	0.00	-0.87
$a_{21} + b_{21}$ (M1) or $\tilde{a}_{21} + \tilde{b}_{21}$ (M2)	0.00	-0.27*	0.01	0.00	0.00	0.00	0.00	0.17**	0.00	0.12
$a_{22} + b_{22}$ (M1) or $\tilde{a}_{22} + \tilde{b}_{22}$ (M2)	0.20	0.19	0.67**	0.99**	0.55**	0.72**	0.96**	0.78**	0.61**	0.66*
$a_{23} + b_{23}$ (M1) or $\tilde{a}_{23} + \tilde{b}_{23}$ (M2)	0.08	0.11	0.00	-0.11**	0.00	-0.12**	0.01	0.18	0.14	0.23
$a_{31} + b_{31}$ (M1) or $\tilde{a}_{31} + \tilde{b}_{31}$ (M2)	0.02	-0.25*	0.00	-0.01	0.01**	0.01	0.00	0.18*	0.00	0.07
$a_{32} + b_{32}$ (M1) or $\tilde{a}_{32} + \tilde{b}_{32}$ (M2)	-0.39	0.63**	-0.32	0.04	-0.18**	-0.19**	0.00	-0.14	0.00	-0.13
$a_{33} + b_{33}$ (M1) or $\tilde{a}_{33} + \tilde{b}_{33}$ (M2)	0.93**	1.10**	0.97**	0.56**	0.98**	0.88**	0.98**	1.09**	0.96**	1.03**
$-b_{11}$ (M1) or $-\tilde{b}_{11}$ (M2)	-0.83**	-0.58**	-0.77**	-0.81**	-0.84**	-0.90**	-0.74**	-0.27*	-0.48**	-0.25
$-b_{12}$ (M1) or $-\tilde{b}_{12}$ (M2)	0.00	0.27	-0.07	-0.14	0.00	-0.24**	0.00	-0.85*	0.00	-0.98*
$-b_{13}$ (M1) or $-\tilde{b}_{13}$ (M2)	0.00	-0.29	0.00	-0.03	0.00	-0.14**	0.00	0.53	0.00	0.88
$-b_{21}$ (M1) or $-\tilde{b}_{21}$ (M2)	0.00	0.22*	0.00	-0.01	0.00	-0.02	0.00	-0.18**	0.00	-0.13
$-b_{22}$ (M1) or $-\tilde{b}_{22}$ (M2)	-0.07	-0.03	-0.51**	-0.90**	-0.45**	-0.56**	-0.94**	-0.68**	-0.51**	-0.48
$-b_{23}$ (M1) or $-\tilde{b}_{23}$ (M2)	-0.07	-0.08	0.00	0.12**	0.00	0.14**	-0.01	-0.17	-0.14	-0.24
$-b_{31}$ (M1) or $-\tilde{b}_{31}$ (M2)	-0.02	0.22*	0.00	-0.04	-0.01	-0.04	0.00	-0.21**	0.00	-0.10
$-b_{32}$ (M1) or $-\tilde{b}_{32}$ (M2)	0.47	0.73**	0.49	0.27	0.22**	0.30**	0.00	0.21	0.01	0.23
$-b_{33}$ (M1) or $-\tilde{b}_{33}$ (M2)	-0.85**	-1.06**	-0.93**	-0.49**	-0.92**	-0.82**	-0.92**	-1.05**	-0.91**	-0.99**
Wald ($\tilde{b}_{ij, i \neq j} = 0$)	R	R	R	R	R	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	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 M1 (vMEM) and M2 (log-vMEM) specifications. Γ denotes the largest eigenvalue of $\mathbf{A} + \mathbf{B}$ (M1) or $\tilde{\mathbf{A}} + \tilde{\mathbf{B}}$ (M2), LL is the system log-likelihood, AIC is the system Akaike information criterion, and BIC is the system Bayesian information criterion. Wald ($\tilde{b}_{ij, i \neq j} = 0$) 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.)

Parameter	S1		S2		S3		S4		S5	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Panel B: 1998 dataset (frequently traded stocks)										
$a_{11} + b_{11}$ (M1) or $\tilde{a}_{11} + \tilde{b}_{11}$ (M2)	0.99**	0.97**	0.99**	0.93**	0.99**	1.04**	0.99**	0.96**	0.99**	0.97**
$a_{12} + b_{12}$ (M1) or $\tilde{a}_{12} + \tilde{b}_{12}$ (M2)	0.00	0.09**	0.00	0.05**	0.00	0.02	0.00	0.03**	0.00	0.07**
$a_{13} + b_{13}$ (M1) or $\tilde{a}_{13} + \tilde{b}_{13}$ (M2)	0.00	0.29**	0.00	0.22**	0.00	0.24**	0.00	0.13**	0.00	0.09**
$a_{21} + b_{21}$ (M1) or $\tilde{a}_{21} + \tilde{b}_{21}$ (M2)	0.00	-0.01	0.00	0.04**	0.00	-0.06*	0.00	0.03**	0.00	0.00
$a_{22} + b_{22}$ (M1) or $\tilde{a}_{22} + \tilde{b}_{22}$ (M2)	0.88**	0.88**	0.99**	0.93**	0.53**	0.96**	0.93**	0.92**	0.88**	0.77**
$a_{23} + b_{23}$ (M1) or $\tilde{a}_{23} + \tilde{b}_{23}$ (M2)	0.00	-0.19**	0.00**	-0.10**	0.01	-0.24**	0.00	-0.14**	0.00	-0.16**
$a_{31} + b_{31}$ (M1) or $\tilde{a}_{31} + \tilde{b}_{31}$ (M2)	0.01	0.02	0.01	0.08**	0.00	-0.13*	0.01**	0.06**	0.01	0.02
$a_{32} + b_{32}$ (M1) or $\tilde{a}_{32} + \tilde{b}_{32}$ (M2)	-0.06	-0.19**	0.00	-0.06**	-0.37	-0.06	-0.02**	-0.09**	-0.06	-0.24**
$a_{33} + b_{33}$ (M1) or $\tilde{a}_{33} + \tilde{b}_{33}$ (M2)	0.93**	0.53**	0.96**	0.75**	0.93**	0.44**	0.97**	0.73**	0.97**	0.76**
$-b_{11}$ (M1) or $-\tilde{b}_{11}$ (M2)	-0.85**	-0.87**	-0.86**	-0.82**	-0.94**	-0.98**	-0.91**	-0.89**	-0.88**	-0.91**
$-b_{12}$ (M1) or $-\tilde{b}_{12}$ (M2)	0.00	-0.20**	0.00	-0.16**	0.00	-0.11**	0.00	-0.09**	0.00	-0.13**
$-b_{13}$ (M1) or $-\tilde{b}_{13}$ (M2)	0.00	-0.33**	0.00	-0.26**	0.00	-0.27**	0.00	-0.15**	0.00	-0.11**
$-b_{21}$ (M1) or $-\tilde{b}_{21}$ (M2)	0.00	0.00	0.00	-0.06**	0.00	0.03	0.00	-0.05**	0.00	-0.02*
$-b_{22}$ (M1) or $-\tilde{b}_{22}$ (M2)	-0.85**	-0.75**	-0.99**	-0.83**	-0.41**	-0.85**	-0.88**	-0.79**	-0.85**	-0.63**
$-b_{23}$ (M1) or $-\tilde{b}_{23}$ (M2)	0.00	0.19**	0.00	0.10**	0.00	0.25**	0.00	0.15**	0.00	0.17**
$-b_{31}$ (M1) or $-\tilde{b}_{31}$ (M2)	-0.01**	-0.10**	-0.01**	-0.18**	0.00	0.05	0.00	-0.12**	-0.01	-0.07**
$-b_{32}$ (M1) or $-\tilde{b}_{32}$ (M2)	0.08*	0.36**	0.00	0.20**	0.49	0.27**	0.04**	0.22**	0.08	0.37**
$-b_{33}$ (M1) or $-\tilde{b}_{33}$ (M2)	-0.87**	-0.48**	-0.92**	-0.69**	-0.87**	-0.38**	-0.94**	-0.68**	-0.92**	-0.69**
Wald ($\tilde{b}_{ij, i \neq j} = 0$)	R	R	R	R	R	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 ($\sigma_{ij, i \neq j} = 0$)	R	R	R	R	R	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 M1 (vMEM) and M2 (log-vMEM) specifications. Γ denotes the largest eigenvalue of $\mathbf{A} + \mathbf{B}$ (M1) or $\mathbf{A} + \mathbf{B}$ (M2), LL is the system log-likelihood, AIC is the system Akaike information criterion, and BIC is the system Bayesian information criterion. Wald ($\tilde{b}_{ij, i \neq j} = 0$) 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.)

Parameter	S1		S2		S3		S4		S5	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Panel C: 2012 dataset (infrequently traded stocks)										
$a_{11} + b_{11}$ (M1) or $\tilde{a}_{11} + \tilde{b}_{11}$ (M2)	0.99**	0.79**	0.97**	0.70**	0.99**	0.48**	0.93**	0.74**	0.99**	0.78**
$a_{12} + b_{12}$ (M1) or $\tilde{a}_{12} + \tilde{b}_{12}$ (M2)	0.00	0.06	0.00	-0.10	0.00	0.15	0.00	0.16	0.00	0.47**
$a_{13} + b_{13}$ (M1) or $\tilde{a}_{13} + \tilde{b}_{13}$ (M2)	0.00	0.06	0.02	0.29**	0.00	0.00	0.06	0.02	0.00	-0.14
$a_{21} + b_{21}$ (M1) or $\tilde{a}_{21} + \tilde{b}_{21}$ (M2)	0.02**	0.01	0.04	-0.02	0.04	0.10**	0.05**	0.07**	0.04	0.02
$a_{22} + b_{22}$ (M1) or $\tilde{a}_{22} + \tilde{b}_{22}$ (M2)	0.92**	0.94**	0.86**	0.90**	0.80**	0.77**	0.67**	0.83**	0.77	0.75**
$a_{23} + b_{23}$ (M1) or $\tilde{a}_{23} + \tilde{b}_{23}$ (M2)	0.00	-0.09**	0.00	0.02	0.00	0.05	0.00	-0.01	0.02	0.13*
$a_{31} + b_{31}$ (M1) or $\tilde{a}_{31} + \tilde{b}_{31}$ (M2)	0.00	0.01	0.06**	0.10**	0.03	0.12**	0.00	0.09**	0.03	0.06**
$a_{32} + b_{32}$ (M1) or $\tilde{a}_{32} + \tilde{b}_{32}$ (M2)	0.04	-0.06*	0.10	0.10*	0.21	0.08	0.29	0.19	0.11	0.36**
$a_{33} + b_{33}$ (M1) or $\tilde{a}_{33} + \tilde{b}_{33}$ (M2)	0.69**	0.72**	0.70**	0.69**	0.57	0.76**	0.45*	0.58**	0.77	0.56**
$-b_{11}$ (M1) or $-\tilde{b}_{11}$ (M2)	-0.43**	-0.55**	-0.34**	-0.46**	-0.34**	-0.33**	-0.17**	-0.58**	-0.52**	-0.59**
$-b_{12}$ (M1) or $-\tilde{b}_{12}$ (M2)	0.00	-0.14**	0.00	0.00	0.00	-0.34**	0.00	-0.37*	0.00	-0.71**
$-b_{13}$ (M1) or $-\tilde{b}_{13}$ (M2)	0.00	-0.03	0.00	-0.29*	0.00	-0.01	0.00	0.01	0.00	0.12
$-b_{21}$ (M1) or $-\tilde{b}_{21}$ (M2)	-0.01	-0.02	0.00	0.06*	-0.02	-0.10*	-0.04**	-0.09**	-0.01	-0.01
$-b_{22}$ (M1) or $-\tilde{b}_{22}$ (M2)	-0.85**	-0.79**	-0.74**	-0.76**	-0.71**	-0.57**	-0.54**	-0.65**	-0.69	-0.60**
$-b_{23}$ (M1) or $-\tilde{b}_{23}$ (M2)	0.00	0.10**	0.00	-0.01	0.00	-0.06	0.00	0.01	-0.02	-0.14*
$-b_{31}$ (M1) or $-\tilde{b}_{31}$ (M2)	0.00	-0.10**	-0.06**	-0.15**	-0.03	-0.19**	0.00	-0.19**	-0.03	-0.10**
$-b_{32}$ (M1) or $-\tilde{b}_{32}$ (M2)	0.11**	0.23**	-0.09	-0.04	-0.07	0.05	-0.12	-0.07	-0.04	-0.35**
$-b_{33}$ (M1) or $-\tilde{b}_{33}$ (M2)	-0.56**	-0.59**	-0.59**	-0.53**	-0.43	-0.64**	-0.33	-0.44**	-0.63	-0.38**
Wald ($\tilde{b}_{i,j,i \neq j} = 0$)	R	R	R	R	R	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 ($\sigma_{i,j,i \neq j} = 0$)	R	R	R	R	R	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 M1 (vMEM) and M2 (log-vMEM) specifications. Γ denotes the largest eigenvalue of $\mathbf{A} + \mathbf{B}$ (M1) or $\mathbf{A} + \mathbf{B}$ (M2), LL is the system log-likelihood, AIC is the system Akaike information criterion, and BIC is the system Bayesian information criterion. Wald ($\tilde{b}_{i,j,i \neq j} = 0$) and Wald ($\sigma_{i,j,i \neq j} = 0$) 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.)

Parameter	S1		S2		S3		S4		S5	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
Panel D: 2012 dataset (frequently traded stocks)										
$a_{11} + b_{11}$ (M1) or $\tilde{a}_{11} + \tilde{b}_{11}$ (M2)	0.94**	0.88**	0.89**	0.89**	0.96**	0.90**	0.95**	0.98**	0.95**	0.93**
$a_{12} + b_{12}$ (M1) or $\tilde{a}_{12} + \tilde{b}_{12}$ (M2)	0.00	0.15**	0.00	0.15**	0.00	0.13**	0.00	0.18**	0.00	0.12**
$a_{13} + b_{13}$ (M1) or $\tilde{a}_{13} + \tilde{b}_{13}$ (M2)	0.01	0.14**	0.01	0.10**	0.01	0.09**	0.00	0.11**	0.01	0.07**
$a_{21} + b_{21}$ (M1) or $\tilde{a}_{21} + \tilde{b}_{21}$ (M2)	0.00	0.02**	0.00	0.02**	0.00	0.03**	0.00	0.00	0.00	0.03**
$a_{22} + b_{22}$ (M1) or $\tilde{a}_{22} + \tilde{b}_{22}$ (M2)	0.69**	0.96**	0.65**	0.96**	0.74**	0.94**	0.89**	0.93**	0.67**	0.94**
$a_{23} + b_{23}$ (M1) or $\tilde{a}_{23} + \tilde{b}_{23}$ (M2)	0.00	-0.02**	0.00	-0.02**	0.00	-0.03**	0.00	-0.04**	0.02**	-0.03**
$a_{31} + b_{31}$ (M1) or $\tilde{a}_{31} + \tilde{b}_{31}$ (M2)	0.02	0.12**	0.00	0.10**	0.03	0.11**	0.01	0.03	0.01	0.08**
$a_{32} + b_{32}$ (M1) or $\tilde{a}_{32} + \tilde{b}_{32}$ (M2)	-0.76**	-0.15**	-0.78**	-0.13**	-0.28**	-0.13**	-0.11**	-0.22**	-0.47**	-0.09**
$a_{33} + b_{33}$ (M1) or $\tilde{a}_{33} + \tilde{b}_{33}$ (M2)	0.96**	0.81**	0.97**	0.88**	0.93**	0.86**	0.96**	0.84**	0.97**	0.89**
$-b_{11}$ (M1) or $-\tilde{b}_{11}$ (M2)	-0.79**	-0.75**	-0.73**	-0.76**	-0.83**	-0.77**	-0.83**	-0.87**	-0.79**	-0.80**
$-b_{12}$ (M1) or $-\tilde{b}_{12}$ (M2)	0.00	-0.30**	0.00	-0.29**	0.00	-0.28**	0.00	-0.27**	0.00	-0.27**
$-b_{13}$ (M1) or $-\tilde{b}_{13}$ (M2)	0.00	-0.15**	0.00	-0.10**	0.00	-0.08**	0.00	-0.11**	0.00	-0.07**
$-b_{21}$ (M1) or $-\tilde{b}_{21}$ (M2)	0.00	-0.03**	0.00	-0.04**	0.00	-0.06**	0.00	-0.03**	0.00	-0.07**
$-b_{22}$ (M1) or $-\tilde{b}_{22}$ (M2)	-0.66**	-0.93**	-0.62**	-0.92**	-0.68**	-0.86**	-0.86**	-0.89**	-0.61**	-0.85**
$-b_{23}$ (M1) or $-\tilde{b}_{23}$ (M2)	0.00	0.02**	0.00	0.01**	0.00	0.03**	0.00	0.03**	-0.02**	0.03**
$-b_{31}$ (M1) or $-\tilde{b}_{31}$ (M2)	-0.02**	-0.25**	0.00	-0.24**	-0.03**	-0.24**	-0.01**	-0.16**	-0.01**	-0.22**
$-b_{32}$ (M1) or $-\tilde{b}_{32}$ (M2)	0.87**	0.30**	0.87**	0.26**	0.37**	0.31**	0.15**	0.32**	0.59**	0.28**
$-b_{33}$ (M1) or $-\tilde{b}_{33}$ (M2)	-0.91**	-0.75**	-0.91**	-0.81**	-0.85**	-0.78**	-0.92**	-0.78**	-0.92**	-0.82**
Wald ($\tilde{b}_{ij, i \neq j} = 0$)	R	R	R	R	R	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	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 M1 (vMEM) and M2 (log-vMEM) specifications. Γ denotes the largest eigenvalue of $\mathbf{A} + \mathbf{B}$ (M1) or $\mathbf{A} + \mathbf{B}$ (M2), LL is the system log-likelihood, AIC is the system Akaike information criterion, and BIC is the system Bayesian information criterion. Wald ($\tilde{b}_{ij, i \neq j} = 0$) 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 $\tilde{\mathbf{B}}$ and \mathbf{V} , respectively. Significance at the 1% level is indicated by **, and at the 5% level by *.

Table 4. Impulse response function absorption times

Shock	S1		S2		S3		S4		S5	
	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}). The calendar time the system takes to return to its long-run equilibrium is approximated by multiplying the number of transactions by their average duration.

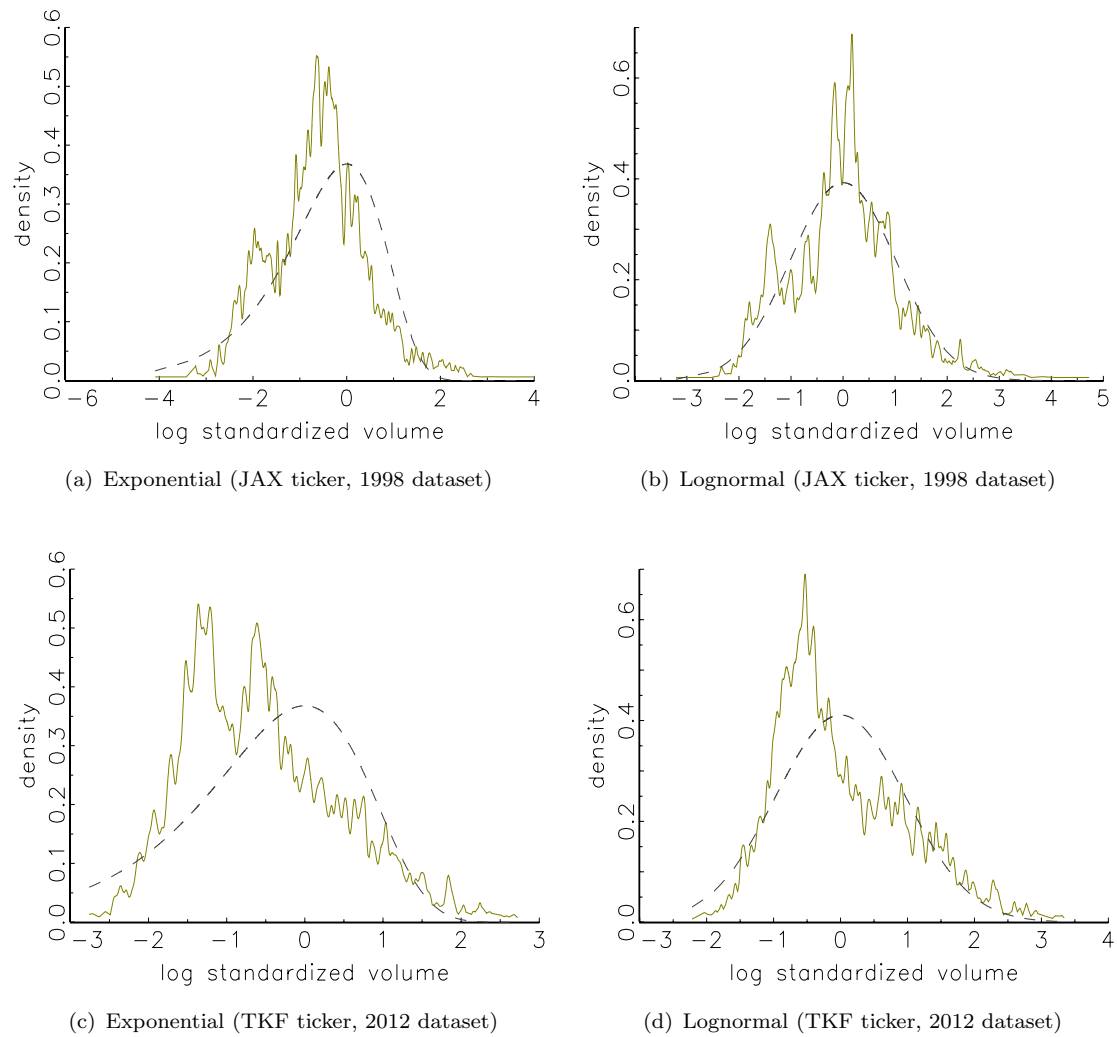


Figure 1. A comparison of parametric and non-parametric densities

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

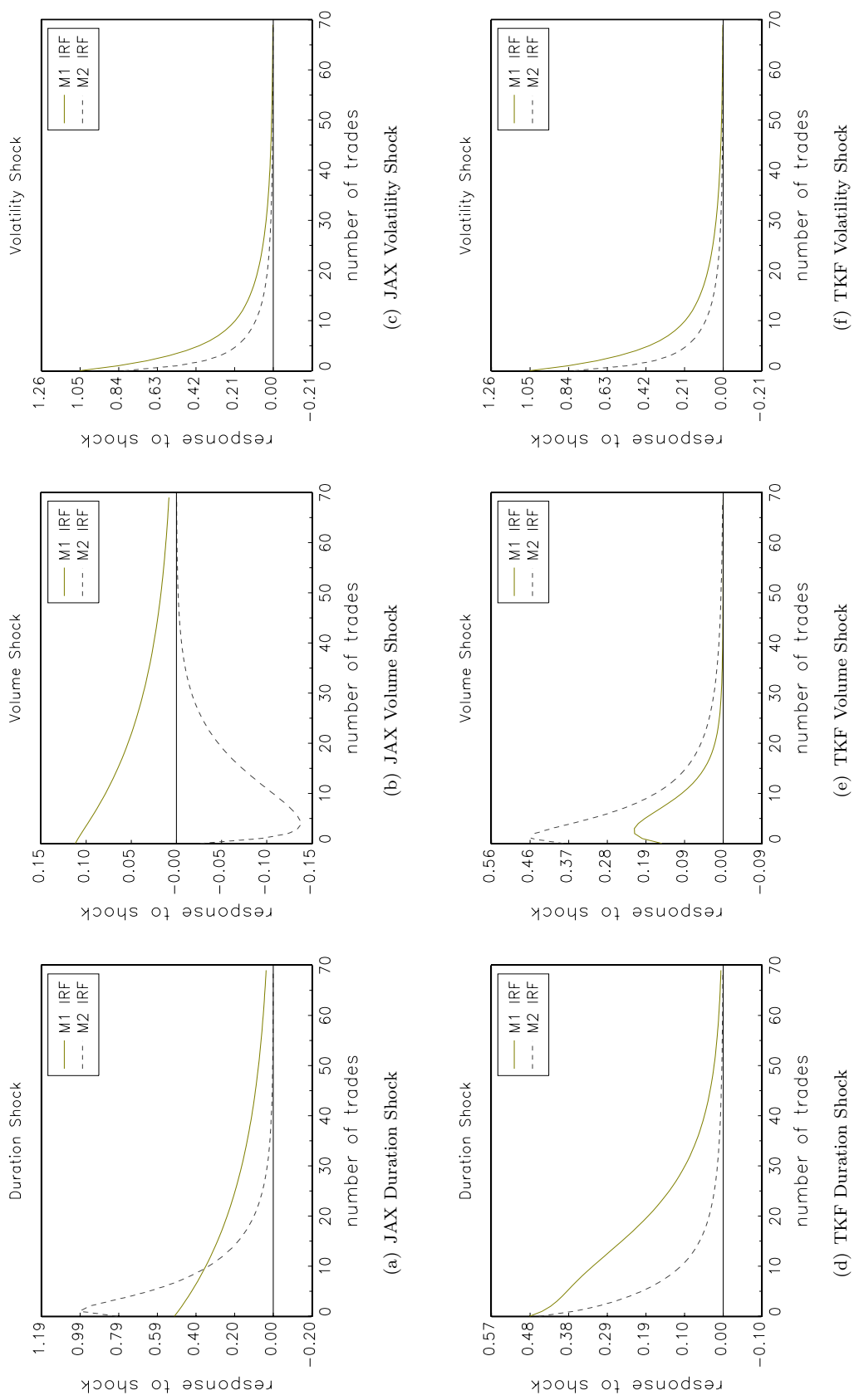


Figure 2. Impulse response functions associated with the volatility reaction to various shocks using parameters associated with the M1 (vMEM) and M2 (log-vMEM) specifications.