Testing Weak Exogeneity in Multiplicative Error Models[☆]

Kul B Luintel[†] and Yongdeng Xu[‡]

Abstract

Empirical market microstructure literature widely employs the non-linear and non-Gaussian Multiplicative Error Class of Models (MEMs) in modelling the dynamics of trading duration and financial marks. It routinely maintains the weak exogeneity of duration vis-à-vis marks in estimations. However, microstructure theory states that trade duration, volume and transaction prices are simultaneously determined. We propose Lagrange-multiplier (LM) tests for weak exogeneity for the MEMs. Our LM tests are extensions of the weak exogeneity tests applicable to VAR or VECM models with Gaussian distribution. Empirical assessments show that (i) weak exogeneity is widely rejected by the data in the MEMs and (ii) the failure of weak exogeneity seriously biases parameter estimates. We hope our tests will be of interest in future empirical applications.

JEL Classification: C32, C12, G10

Key words: Weak exogeneity, Multiplicative Error Model, LM test, Market microstructure

---

[☆] An earlier version of this paper was circulated under the title "Weak exogeneity in the financial point processes" in Cardiff Economics Working Paper No. E2013/6. We would like to thank Luc Bauwens for his valuable comments and Simone Manganelli for providing us with the data used in this paper. We would also like to thank the participants in the 10th BMRC - DEMS Conference on Macro and Financial Economics and Econometrics in May 2014. Kul Luintel would like to thank CESifo Group Munich for their excellent hospitality during his research visit in July 2016 when a major part of the revision of this paper is completed.

[†] Economics Section, Cardiff Business School, Cardiff University, UK. Email: LuintelK@cardiff.ac.uk

[‡] Corresponding author: Economics Section, Cardiff Business School, Cardiff University, UK. Email: xuy16@cardiff.ac.uk

1 Introduction

The theory of market microstructure explains trading activity in financial markets as information and/or liquidity based. It maintains that trade duration and marks[1] convey important information about fundamental asset prices and market participants' behaviour. The empirical literature on market microstructure analyses the dynamics of trading duration, volume and price volatility. A large body of this literature (Dufour and Engle, 2000; Engle, 2000; Grammig and Wellner, 2002; Manganelli, 2005; Engle and Sun, 2007; Hautsch, 2008; Bowe et al., 2009; to name but a few) employs a vector Multiplicative Error Model (MEM)[2], proposed by Engle (2002) and Cipollini et al. (2007), to analyse these relationships. The basic idea of the MEM is to model the non-negative valued financial time series as the product of an autoregressive scale factor and an innovation process with non-negative support. The empirical evidence on the relationships between duration and marks is mixed, however.[3]

The vector MEM is usually estimated equation-by-equation, by assuming weak exogeneity of duration. This equation-by-equation approach reduces the multivariate setting into a series of univariate problems which makes estimation much simpler. However, this simplicity could prove costly unless the weak exogeneity of durations is sustained statistically. Conceptually, the exogeneity of duration is rather shaky because trade durations and marks tend to be highly correlated and there is no clarity on the flows of causality between them. Microstructure theory (Admati and Pfleiderer, 1988; Easley and O'Hara, 1992; Foster and Viswanathan, 1996) postulates that duration, volume and transaction prices tend to be simultaneously determined which raises concerns about the weak exogeneity of duration vis-à-vis marks. The failure of this assumption may result in biased and/or inconsistent parameter estimates and render inferences invalid (White, 1981, 1982). It is therefore important to assess this

[1]Durations are typically the time elapsing between trades of financial assets whereas market marks, commonly of most interest, are the trading volume, bid-ask spread, and the return volatility.

[2]Engle (2002) and Cipollini et al. (2007) propose a Multiplicative Error Model (MEM) for the dynamics of non-negative value processes. The MEMs incorporate the ACD model (Engle and Russell, 1998) for financial duration and the GARCH model for return volatility as a special case.

[3]Engle (2000) analyses the dynamics of duration and volatility recursively and finds that the longer duration leads to lower volatilities, which is also confirmed by Manganelli (2005) who analyses duration, volume and volatility jointly. In contrast, Grammig and Wellner (2002) formulate interdependent intraday duration and volatility models and report that lagged volatility significantly reduces transaction intensity. Dufour and Engle (2000), under the recursive VAR model, show that prices, bid-ask spreads and price volatility all increase when traders observe short duration. Bowe et al. (2009) analysing a trivariate VAR, find that duration is affected positively by volatility, which is opposite to the findings of Engle (2000).

issue and derive a formal test of weak exogeneity of duration applicable to these MEMs of market microstructure.

The issue of weak exogeneity is by no means new, yet the existing literature on this topic is mainly based on linear Gaussian models. For example, Dolado (1992), Boswijk and Urbain (1997) and Engle and Hendry (1993), among others, have derived weak exogeneity tests for VAR or VECM models with Gaussian distribution. Unfortunately, these tests are not directly applicable to MEMs as they are non-linear and non-Gaussian. This paper aims to bridge this gap and contributes to the literature in the following ways. First, we extend weak exogeneity tests to a non-linear and non-Gaussian framework which is applicable to the MEMs. In so doing, we formally illustrate the conditions under which weak exogeneity of duration vis-à-vis marks could be maintained in these models and propose LM tests of weak exogeneity. Our proposed tests show good power properties. Second, we examine the consequences of the failure of weak exogeneity in the MEMs through Monte Carlo simulations. Consistent with the findings vis-à-vis linear Gaussian VAR and VECM models, we also find that the failure of weak exogeneity results in biased parameter estimates for MEMs. This bias is accentuated with the increasing error correlation of duration and marks. Finally, we employ our proposed LM tests to investigate if the hypotheses of weak exogeneity, maintained by Manganelli (2005), could be sustained in his dataset. We find that the assumption of the weak exogeneity of duration is often rejected by frequently traded stocks, but less often by infrequently traded stocks. These findings largely go against the generally maintained hypothesis by the vast body of literature cited above and hence the general expectation that duration is weakly exogenous vis-à-vis marks in the Multiplicative Error Models.

The remainder of the paper is organised as follows. The following section sets out the model and illustrates the notion of weak exogeneity. Section 3 derives an LM test for weak exogeneity and discusses its power properties. Section 4 contains empirical applications and Section 5 concludes the paper.

2 Weak exogeneity in MEMs

The MEM, proposed by Engle (2002) and Cipollini et al. (2007), is widely used to model the dynamics of non-negative valued and highly persistent financial time series. This class of models has a non-linear mean and non-Gaussian distribution; it includes the Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998) for financial duration, and the GARCH model for return volatility. Since both duration and marks are non-negative valued and persistently clustered over time, it is commonplace to analyse their relationship using the MEMs. Let the joint density of duration (x_t) and mark (y_t) be $f(x_t, y_t | \Omega_t; \theta)$, where Ω_t denotes the information available up to period t and θ is a vector incorporating the parameters. Denote $z_t = \{x_t, y_t\}'$; its conditional expectations $E(z_t | \Omega_t) = \mu_t$; $\mu_t = (\mu_t^x, \mu_t^y)'$ and the error terms, η_t. The first order vector MEM is given by:

$$\begin{aligned} z_t &= \mu_t \odot \eta_t \\ \mu_t &= \omega + A z_{t-1} + B \mu_{t-1} + \Gamma z_t \end{aligned} \tag{1}$$

where $\odot$ denotes the Hadamard (element by element) product; η_t is the error term which has a mean vector I with all elements of unity and a general variance-covariance matrix Σ, i.e., $\eta_t | \Omega_t \sim D(I, \Sigma)$. The matrix Γ captures the contemporaneous dependence of duration and marks and is often specified as a matrix where only the lower triangular elements are non-zero. This specification is general and incorporates most of the MEMs, for example Manganelli (2005), Engle and Gallo (2006), Hautsch (2008), Engle et al. (2012).

A completely parametric formulation of the vector MEM requires a full specification of the conditional distribution of multivariate non-negative valued random variables η_t, which is not often available.[4] A common strategy is to reduce the multivariate setting into a series of univariate problems by imposing the assumption of weak exogeneity, and estimating each process separately (equation-by-equation). This is the strategy adopted by Engle (2000), Manganelli (2005), Engle and Gallo (2006), Hautsch (2008), Engle et al. (2012). Let us re-write (1), equation-by-equation, as a bivariate vector MEM:

$$\begin{aligned} & \left.\begin{aligned} x_t &= \mu_t^x(\theta_x; \Omega_t)\varepsilon_{1t} \\ y_t &= \mu_t^y(\theta_y; \Omega_t)\varepsilon_{2t} \end{aligned}\right., \qquad \operatorname{cov}(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{12} \\ & \mu_t^x = w_1 + a_{11} x_{t-1} + a_{12} y_{t-1} + b_{11} \mu_{t-1}^x + b_{12} \mu_{t-1}^y \\ & \mu_t^y = w_2 + a_{21} x_{t-1} + a_{22} y_{t-1} + b_{21} \mu_{t-1}^x + b_{22} \mu_{t-1}^y + \tau_0 x_t. \end{aligned} \tag{2}$$

[4] To our knowledge, the only available distribution for these processes is the Multivariate Gamma Distribution; however, it only admits positive error correlation which is too restrictive, as shown by Cipollini et al. (2007).

The focus of this paper is to investigate the conditions under which the weak exogeneity assumption in (2) is satisfied and derive formal tests of weak exogeneity.

Different definitions of exogeneity – viz., weak-, strong- and super-exogeneity – are clarified by Engle et al. (1983). Given a bivariate stochastic process $\{x_t, y_t\}$ and their joint density $f(x_t, y_t | \psi_t; \theta)$ where the joint density can be factorized into the product of the marginal density x_t and conditional density of y_t given x_t, such that $f(x_t, y_t | \psi_t; \theta) = f_x(x_t | \psi_t; \theta_x) f_{y|x}(y_t | x_t, \psi_t; \theta_y)$; Engle et al. (1983) formally define x_t as weakly exogenous for a set of parameters of interest ϑ if: (i) $\vartheta = f(\theta_y)$, ϑ is a function of parameters θ_y alone, and (ii) θ_x and θ_y are variation free, i.e. $(\theta_x, \theta_y) \in \Theta^x \times \Theta^y$. Weak exogeneity sets conditions under which the parameters of conditional density could be estimated without loss of information even if the marginal process is ignored.

In terms of the vector MEM model (2), the parameter of interest is $\vartheta = f(\theta_y)$; where, $\theta_y = \{w_2, a_{21}, a_{22}, b_{21}, b_{22}, \tau_0\}'$; and x_t is weakly exogenous of ϑ if

(i) $\sigma_{12} = 0$ (Orthogonality);

(ii) $b_{12} = b_{21} = 0$ (No lagged cross-dependence);

(iii) $\vartheta = f(\theta_y)$, ϑ is a function of parameters θ_y alone.

If conditions (i) through (iii) are satisfied then x_t is weakly exogenous for ϑ and there is no loss of information in estimating ϑ even if the process of x_t is ignored. If the process of x_t is also of interest, then the weak exogeneity of x_t implies that the estimations of θ_x (where, $\theta_x = \{w_1, a_{11}, a_{12}, b_{11}, b_{12}\}'$) and θ_y can be undertaken separately without loss of information. It should be noted that in the vector MEM given by (2), there are two sources of the breakdown of weak exogeneity: (i) the contemporaneous error correlation between the two processes and (ii) the existence of lagged cross-dependence between the two processes. The first cross restriction is precisely given by $\sigma_{12} \neq 0$, which implies θ_x and θ_y are not variation free. The second cross restriction is given by $b_{12} \neq b_{21} \neq 0$, which implies that $\mu_t^x(\theta_x)$ is a function of $\mu_{t-1}^y(\theta_y)$ and μ_t^y is a function of $\mu_{t-1}^x(\theta_x)$; hence, θ_x and θ_y are cross-restricted. If conditions (i) and (ii) hold then they imply that there is no restriction across duration and marks processes. The non-rejection of conditions (i) and (ii), by definition, implies (iii); i.e., the duration model parameters θ_x and the marks model parameters θ_y are independent and each process can be estimated separately without the loss of any information. Therefore, these conditions require tests of $b_{12} = b_{21} = 0$ and $\sigma_{12} = 0$ as the tests of weak exogeneity of x_t in (2). [5]

It is notable that if the further restrictions of $a_{12} = a_{21} = 0$ are imposed, this implies Granger no-causality between duration and marks processes. The weak exogeneity of x_t coupled with its Granger non-causality to y_t imply that x_t is strongly exogenous vis-à-vis y_t. Unfortunately, condition (i) is not directly testable empirically. The test of $\sigma_{12} = 0$ is discussed in the VAR or VECH model by Dolado (1992) and Boswijk and Urbain (1997). Following their approach, we derive the following Lemma:

Lemma 1: In the vector MEM given by (2), where $\vartheta = f(\theta_y)$ is the parameter of interest, it is sufficient to test $\tau_1 = 0$ defined below in (3) in order to test $\sigma_{12} = 0$ defined by (2)[6]

$$\begin{aligned}
x_t &= \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t},\ \varepsilon_{1t} \sim i.i.d(1,\sigma_{\varepsilon_1}) \\
\mu_t^x &= w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x + b_{12}\mu_{t-1}^y. \\
y_t &= \mu_t^y(\theta_y;\Omega_t)\varepsilon_{12,t},\ \varepsilon_{12,t} \sim i.i.d(1,\sigma_{\varepsilon_2}) \\
\mu_t^y &= w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{21}\mu_{t-1}^x + b_{22}\mu_{t-1}^y + \tau_0' x_t + \tau_1 \mu_t^x.
\end{aligned} \tag{3}$$

Proofs of Lemma 1 are provided in Appendix1.

It is worth noting that the orthogonality condition is rather strong in practice. However, its requirement depends on the model or the parameters of interest as illustrated by Engle et al. (1983). Clearly, in (2) the focus is on the joint density of x_t and y_t – i.e., the parameters of both processes - which requires orthogonality condition to be satisfied for the weak exogeneity of x_t, as outlined. If one is interested in conditional process alone, then the orthogonality condition may not be

[5] Although the specification (2) is of first order, this test principle can easily accommodate higher order autoregressive terms.

[6] Note this test principle is similar to Hausman's two-stage testing for weak exogeneity.

necessary. The latter may require additional parameter restrictions, which we illustrate in Appendix 2.[7]

3 LM test of weak exogeneity

In this section, we propose a Lagrange-multiplier (LM) or efficient score test of weak exogeneity under the conditions shown above. As is well known, the LM score test only requires the estimation of a restricted model. Engle (1982) establishes the optimality of the principle of the LM test. Following this, we derive an LM test principle for weak exogeneity for the MEMs. As shown above, we begin with bivariate vector MEM as in (2).

The null hypotheses are:

$H_0^1 : \sigma_{12} = 0$ (Orthogonality)

$H_0^2 : b_{12} = 0, b_{21} = 0$ (No lagged cross-dependence)

$H_0 = H_0^1 \cap H_0^2 : \sigma_{12} = 0, b_{12} = 0, b_{21} = 0.$

Let H_0^1 and H_0^2 denote the partial null hypotheses, and $H_0 = H_0^1 \cap H_0^2$ denote the joint null hypothesis. Theoretically, we have to test the joint null hypothesis in order to test the weak exogeneity from both sources. However, these two sources of potential breakdown of weak exogeneity are rarely discussed or allowed jointly in empirical work. For example, in the ACD-GARCH models of Engle (2000) only expected durations enter the volatility equation as covariate, which only account for the potential violation of orthogonality conditions. Likewise, Manganelli (2005) and Hautsch (2008) also assume away the lagged cross-dependence. In contrast, Grammig and Wellner (2002) consider the effect of lagged conditional volatility on duration equations but they assume away the potential contemporaneous error correlations. Given that only one of the two sources of potential breakdown of weak exogeneity (i.e., the partial hypothesis) is engaged in the empirical literature, we proceed to derive separate LM test statistics for H_0 against $H_1^1 \cap H_0^2$ (testing orthogonality), and H_0 against $H_0^1 \cap H_1^2$ (testing lagged cross-dependence). Interestingly, the joint test statistic turns out to be the sum of two partial test statistics which we elaborate below (immediately following the *Type II* LM test).

<u>Type I</u>: Testing H_0 against $H_1^1 \cap H_0^2 : \sigma_{12} = 0$ (Orthogonality)

[7] We would like to thank an anonymous referee for encouraging us to illustrate this point.

As shown in Section 2 (Lemma 1), it suffices to test $H_0 : \tau_1 = 0$ in (3) in order to test for the $\sigma_{12} = 0$ in (2). Under H_0:

$$
\begin{aligned}
& x_t = \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t},\ \varepsilon_{1t} \sim i.i.d(1,\sigma_{\varepsilon_1}) \\
& \mu_t^x = w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x \\
& y_t = \mu_t^y(\theta_y;\Omega_t)\varepsilon_{12,t}, \varepsilon_{12,t} \sim i.i.d(1,\sigma_{\varepsilon_{12}}) \\
& \mu_t^y = w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{22}\mu_{t-1}^y + \tau_0 x_t .
\end{aligned}
\tag{4}
$$

Under H_1

$$
\begin{aligned}
& x_t = \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t},\ \varepsilon_{1t} \sim i.i.d(1,\sigma_{\varepsilon_1}) \\
& \mu_t^x = w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x \\
& y_t = \mu_t^y(\theta_y;\Omega_t)\varepsilon_{12,t},\ \varepsilon_{12,t} \sim i.i.d(1,\sigma_{\varepsilon_2}) \\
& \mu_t^y = w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{22}\mu_{t-1}^y + \tau_0' x_t + \tau_1 \mu_t^x .
\end{aligned}
\tag{5}
$$

The error terms ε_{1t} and $\varepsilon_{12,t}$ can follow any distribution that has non-negative support, e.g. exponential, Weibull, lognormal or Gamma distribution. For simplicity, we assume ε_{1t} and $\varepsilon_{12,t}$ both have exponential density and hence their variance σ_{ε_1} and $\sigma_{\varepsilon_{12}}$ is equal to one. The joint log-likelihood function is then:

$$
\begin{aligned}
L(\theta) &= l_y(y \mid \theta_y, \mu^x(\theta_x)) + l_x(x \mid \theta_x) \\
&= -\sum_{t=1}^{T}(\log \mu_t^y + y_t / \mu_t^y) - \sum_{t=1}^{T}(\log \mu_t^x + x_t / \mu_t^x).
\end{aligned}
\tag{6}
$$

where l_y and l_x are the log-likelihood functions for marks and the duration processes. Moreover, under H_0 of weak exogeneity, the two processes can be estimated separately.

The general theory of ML leads to a simple score test for $\tau_1 = 0$ in (5). Then the score LM test (S_1) has the familiar form:

$$
S_1 = -\hat{i}^{\tau_1}(\hat{\theta}_c)' \hat{I}^{\tau_1}(\hat{\theta}_c)^{-1} \hat{i}^{\tau_1}(\hat{\theta}_c) \tag{7}
$$

where $\hat{i}^{\tau_1}(\hat{\theta}_c) = \dfrac{\partial L}{\partial \theta}$ and $\hat{I}^{\tau_1}(\hat{\theta}_c) = \dfrac{\partial L}{\partial \theta \partial \theta'}$ are the components corresponding to τ_1 in the empirical score and the Hessian from the unconstrained model. Under mild regularity conditions, it is well known that the score test has an asymptotically $\chi^2(1)$ distribution under H_0. It should be noted that even if error processes follow a different distribution (e.g. Weibull, lognormal or Gamma distribution), that only changes the likelihood function in (6). The test statistics in (7) and the rest of the derivations of this section remain exactly the same.

The Hessian matrix can be consistently estimated by (6). Due to the fact that the score under the null (4) has zero elements, we only require the relevant part of the inverse of the Hessian matrix to derive the LM statistic. The score matrix is partitioned as:

$$\hat{i}(\hat{\theta}_c) = \begin{pmatrix} \hat{i}_y(\hat{\theta}_c^y) \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{i}_y^{\tau_1}(\hat{\theta}_c^y) \\ \hat{i}_y^{\theta_{y\dagger}}(\hat{\theta}_c^y) \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{i}_y^{\tau_1}(\hat{\theta}_c^y) \\ 0 \\ 0 \end{pmatrix} \tag{8}$$

where $\theta_{y\dagger} = \theta_y \setminus \tau_1$; i.e., all other parameters in θ_y, except the one pertaining to τ_1.

Equivalently, the Hessian matrix can be partitioned as:

$$\hat{I}(\hat{\theta}_c) = \begin{pmatrix} \hat{I}_y(\hat{\theta}_c^y) & \hat{I}_{y,x}(\hat{\theta}_c^{y,x}) \\ \hat{I}_{x,y}(\hat{\theta}_c^{x,y}) & \hat{I}_x(\hat{\theta}_c^x) \end{pmatrix} = \begin{pmatrix} \hat{I}_y(\hat{\theta}_c^y) & 0 \\ 0 & \hat{I}_x(\hat{\theta}_c^x) \end{pmatrix}. \tag{9}$$

And the submatrix $\hat{I}_y(\hat{\theta}_c^y)$ can be partitioned as

$$\hat{I}_y(\hat{\theta}_c^y) = \begin{pmatrix} \hat{I}_y^{\tau_1}(\hat{\theta}_c^y) & \hat{I}_y^{\tau_1,\theta_{y\dagger}}(\hat{\theta}_c^y) \\ \hat{I}_y^{\theta_{y\dagger},\tau_1}(\hat{\theta}_c^y) & \hat{I}_y^{\theta_{y\dagger}}(\hat{\theta}_c^y) \end{pmatrix}. \tag{10}$$

Only the inverse of $\hat{I}_y(\hat{\theta}_c^y)$ is needed to derive the LM test. Applying the formula for the inverse of a partitioned matrix to this, we obtain $\hat{I}_y^{\tau_1}(\hat{\theta}_c^y)^{-1} = \left[\hat{I}_y^{\tau_1}(\hat{\theta}_c^y) - \hat{I}_y^{\tau_1,\theta_{y\dagger}}(\hat{\theta}_c^y)\hat{I}_y^{\theta_{y\dagger}}(\hat{\theta}_c^y)^{-1}\hat{I}_y^{\theta_{y\dagger},\tau_1}(\hat{\theta}_c^y)\right]^{-1}$. Hence, the Type I LM test has the following form:

$$S_1 = -\hat{i}_y^{\tau_1}(\hat{\theta}_c^y)'\hat{I}_y^{\tau_1}(\hat{\theta}_c^y)^{-1}\hat{i}_y^{\tau_1}(\hat{\theta}_c^y). \tag{11}$$

<u>Type II</u>: Testing H_0 against $H_0^1 \cap H_1^2$: $b_{12} = b_{21} = 0$ (No lagged cross-dependence)

Under H_0:

$$\begin{aligned} x_t &= \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t} \\ \mu_t^x &= w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x \\ y_t &= \mu_t^y(\theta_y;\Omega_t)\varepsilon_{12,t} \\ \mu_t^y &= w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{22}\mu_{t-1}^y + \tau_0 x_t. \end{aligned} \tag{12}$$

Under H_1:

$$
\begin{aligned}
&x_t = \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t}\\
&\mu_t^x = w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x + b_{12}\mu_{t-1}^y\\
&y_t = \mu_t^y(\theta_y;\Omega_t)\varepsilon_{12,t}\\
&\mu_t^y = w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{21}\mu_{t-1}^x + b_{22}\mu_{t-1}^y + \tau_0 x_t.
\end{aligned}
\tag{13}
$$

Assume ε_{1t} and $\varepsilon_{12,t}$ both have exponential density, then the associated log-likelihood function is:

$$
\begin{aligned}
L(\theta) &= l_y(y\,|\,\theta_y, \mu^x(\theta_x)) + l_x(x\,|\,\theta_x, \mu^y(\theta_y))\\
&= -\sum_{t=1}^{T}(\log \mu_t^y + y_t / \mu_t^y) - \sum_{t=1}^{T}(\log \mu_t^x + x_t / \mu_t^x).
\end{aligned}
\tag{14}
$$

Moreover, under H_0 of weak exogeneity, the marginal and conditional models can be estimated separately. Then the score LM test (S_2) has the familiar form:

$$
S_2 = -\hat{i}^{b_{ij,i\neq j}}(\hat{\theta}_c)'\hat{I}^{b_{ij,i\neq j}}(\hat{\theta}_c)^{-1}\hat{i}^{b_{ij,i\neq j}}(\hat{\theta}_c)
\tag{15}
$$

where $\hat{i}^{b_{ij,i\neq j}}(\hat{\theta}_c) = \dfrac{\partial L}{\partial \theta}$ and $\hat{I}^{b_{ij,i\neq j}}(\hat{\theta}_c) = \dfrac{\partial L}{\partial\theta\partial\theta'}$ are the components corresponding to $b_{ij,i\neq j}$ in the empirical score and the Hessian from the unconstrained model. Under mild regularity conditions, it is well known that the score test has an asymptotically $\chi^2(2)$ distribution under H_0.

We only require the relevant part of the inverse of the Hessian matrix to derive the LM statistic. The score matrix is partitioned as:

$$
\hat{i}(\hat{\theta}_c) = \begin{pmatrix} \hat{i}_y(\hat{\theta}_c^y) \\ \hat{i}_x(\hat{\theta}_c^x) \end{pmatrix} = \begin{pmatrix} \hat{i}_y^{b_{21}}(\hat{\theta}_c^y) \\ \hat{i}_y^{\theta_{y\dagger}}(\hat{\theta}_c^y) \\ \hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \\ \hat{i}_x^{\theta_{x\dagger}}(\hat{\theta}_c^x) \end{pmatrix} = \begin{pmatrix} \hat{i}_y^{b_{21}}(\hat{\theta}_c^y) \\ 0 \\ \hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \\ 0 \end{pmatrix}
\tag{16}
$$

where $\theta_{y\dagger} = \theta_y \setminus b_{21}$; i.e., all other parameters in θ_y, except the one pertaining to b_{21} and equivalently for $\theta_{x\dagger} = \theta_x \setminus b_{12}$. The Hessian matrix can be partitioned as:

$$
\hat{I}(\hat{\theta}_c) = \begin{pmatrix} \hat{I}_y(\hat{\theta}_c^y) & \hat{I}_{y,x}(\hat{\theta}_c^{y,x}) \\ \hat{I}_{x,y}(\hat{\theta}_c^{x,y}) & \hat{I}_x(\hat{\theta}_c^x) \end{pmatrix} = \begin{pmatrix} \hat{I}_y(\hat{\theta}_c^y) & 0 \\ 0 & \hat{I}_x(\hat{\theta}_c^x) \end{pmatrix}.
\tag{17}
$$

Equivalently, the submatrix $\hat{I}_y(\hat{\theta}_c^y)$ and $\hat{I}_x(\hat{\theta}_c^x)$ can be partitioned:

$$
\hat{I}_y(\hat{\theta}_c^y) = \begin{pmatrix} \hat{I}_y^{b_{21}}(\hat{\theta}_c^y) & \hat{I}_y^{b_{21},\theta_{y\dagger}}(\hat{\theta}_c^y) \\ \hat{I}_y^{\theta_{y\dagger},b_{21}}(\hat{\theta}_c^y) & \hat{I}_y^{\theta_{y\dagger}}(\hat{\theta}_c^y) \end{pmatrix}, \quad \hat{I}_x(\hat{\theta}_c^x) = \begin{pmatrix} \hat{I}_x^{b_{12}}(\hat{\theta}_c^x) & \hat{I}_x^{b_{21},\theta_{x\dagger}}(\hat{\theta}_c^x) \\ \hat{I}_x^{\theta_{x\dagger},b_{21}}(\hat{\theta}_c^x) & \hat{I}_x^{\theta_{x\dagger}}(\hat{\theta}_c^x) \end{pmatrix}.
\tag{18}
$$

Applying the formula for the inverse of a partitioned matrix, we get

$\widehat{I}_y^{b_{21}}(\widehat{\theta}_c^y)^{-1} = \left[\widehat{I}_y^{b_{21}}(\widehat{\theta}_c^y) - \widehat{I}_y^{b_{21},\theta_{y\dagger}}(\widehat{\theta}_c^y)\widehat{I}_y^{\theta_{y\dagger}}(\widehat{\theta}_c^y)^{-1}\widehat{I}_y^{\theta_{y\dagger},b_{21}}(\widehat{\theta}_c^y)\right]^{-1}$ and

$\widehat{I}_x^{b_{12}}(\widehat{\theta}_c^x)^{-1} = \left[\widehat{I}_x^{b_{12}}(\widehat{\theta}_c^x) - \widehat{I}_x^{b_{12},\theta_{x\dagger}}(\widehat{\theta}_c^x)\widehat{I}_x^{\theta_{x\dagger}}(\widehat{\theta}_c^x)^{-1}\widehat{I}_x^{\theta_{x\dagger},b_{12}}(\widehat{\theta}_c^x)\right]^{-1}$.

Hence, the Type II LM test has the form:

$$\begin{aligned} S_2 &= -\begin{pmatrix} \hat{i}_y^{b_{21}}(\hat{\theta}_c^y) & \hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \end{pmatrix} \begin{pmatrix} \hat{I}_y^{b_{21}}(\hat{\theta}_c^y)^{-1} & 0 \\ 0 & \hat{I}_x^{b_{12}}(\hat{\theta}_c^x)^{-1} \end{pmatrix} \begin{pmatrix} \hat{i}_y^{b_{21}}(\hat{\theta}_c^y) \\ \hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \end{pmatrix} \\ &= -\hat{i}_y^{b_{21}}(\hat{\theta}_c^y)\hat{I}_y^{b_{21}}(\hat{\theta}_c^y)^{-1}\hat{i}_y^{b_{21}}(\hat{\theta}_c^y) - \hat{i}_x^{b_{12}}(\hat{\theta}_c^x)\hat{I}_x^{b_{12}}(\hat{\theta}_c^x)^{-1}\hat{i}_x^{b_{12}}(\hat{\theta}_c^x). \end{aligned} \tag{19}$$

The tests for the joint hypothesis can be derived similarly. Given that the score has zero elements under the null, the joint test statistic turns out to be the sum of two partial test statistics. Under some mild regularity conditions, the joint test has asymptotically $\chi^2(3)$ distribution under H_0; this is shown in Appendix 3.

Power of the test

We first study the power of the Type I LM test as set out in (7). We directly test $\tau_1 = 0$ against $\tau_1 \neq 0$, since it is hard to generate data from a flexible, bivariate, non-negative random distribution. Hence,

H_0: $\tau_1 = 0$

H_1: $\tau_1 \neq 0$

We generate data under the alternative hypothesis and estimate the model under the null hypothesis.[8] These hypotheses are set out in (5) and (4), respectively. In the data generation process, the parameter values are $w_1 = w_2 = 0.1$, $a_{11} = a_{22} = a_{12} = a_{21} = 0.05$, $b_{22} = 0.80$ and $\tau_0 = 0.1$. The signs and magnitudes of these parameters are fairly standard (e.g. Manganelli, 2005; Taylor and Xu, 2016). The parameter τ_1 varies between -0.2 and 0.2 with step 0.025. The empirical size of our tests, shown in Table 1(a), reveals that for the 5% significance level, the empirical size of the test is 8.9% when sample size n=2000; 7.2% when sample size n=5000; and 6.7% when sample size n=10000. The empirical size is close to the theoretical size especially when the sample size is large. This implies that the proposed LM test

[8] To avoid the negative value of volume, we use a logarithmic version of the ACD model (see e.g. Bauwens and Giot, 2000) for DGP process and estimation.

has a good size property. We use the empirical size of 5% to explore the power of the LM test. The results are reported in Table 1(b). The power of the LM test increases with the increases in sample size and the power grows quickly to 1 as τ_1 moves away from zero. The powers are appropriately symmetrical when the sample is large. Overall, our proposed Type I LM test has good size and power properties.

The power of Type II LM test H_0 against $H_1^1 \cap H_0^2$ ($b_{12} = b_{21} = 0$), can be conducted in a similar way. However, our Type II LM test turns out to be similar to the LM test of Nakatani and Teräsvirta (2009), derived separately for the volatility interactions in a CCC-GARCH model. Nakatani and Teräsvirta (2009) report good power properties of their test; although they are concerned with reduced form systems nonetheless their results could be generalised vis-à-vis our test.

4 An empirical application

We address two issues in this section. First, we numerically assess the consequences of the breakdown of weak exogeneity. Second, we employ our proposed LM tests to investigate if the hypotheses of weak exogeneity, maintained by Manganelli (2005), could be sustained in his dataset.

4.1 Consequences of the failure of weak exogeneity assumption

Hendry (1995) vividly shows the impact of the failure of weak exogeneity in linear Gaussian models. We set up our Monte Carlo simulation, by extending his generic approach, and examine the consequences of the breakdown of weak exogeneity (assumption) in vector MEMs. We consider two cases of potential breakdown of weak exogeneity in these models and analyse them through three different data generating processes (DGPs) as summarized in Table 2. DGPs I and II allow us to explore the consequences of ignoring the contemporaneous error correlation between duration and marks processes (or the violation of Type I weak exogeneity). We generate data by choosing $\sigma_{12} = 0.1$ in DGP I and $\sigma_{12} = 0.5$ in DGP II. DGP III permits us to investigate the consequences of ignoring the lagged cross-dependence between duration and marks (or the violation of Type II weak exogeneity). We generate data by choosing $b_{12} = b_{21} = -0.05$ for DGP III. Other parameter values in DGPs, as shown in Table 2, are chosen from the empirical results of Manganelli (2005) and Taylor and Xu (2016). Simulation results are based on a sample size of 5,000 observations and 2,000 replications for each DGP. We use exponential distribution with a mean value of unity to generate the random disturbances. We estimate the vector MEM, equation-by-equation, maintaining that exogeneity conditions hold. In other words, we estimate duration and marks equations separately.

While assessing the implications of the breakdown of weak exogeneity, we focus on two issues. First, we assess whether the equation-by-equation estimates of a_{21} and τ_0 are affected by the misspecification of weak exogeneity. These parameters measure the impact of duration on marks, hence are of interest in the market microstructure. Second, we examine if the predicted conditional expectations, μ_t^x and μ_t^y, are also affected by the misspecification of weak exogeneity. This is important because a class of ACD-GARCH models (Engle, 2000) uses predicted conditional durations as regressors in volatility equations. The predictions of μ_t^x and μ_t^y are also important for vector MEM models of volatility forecasting (e.g. Engle and Gallo, 2006).

Simulation results – mean, standard deviation and Root Mean Square Error (RMSE) for the estimated parameters – are reported in Table 3. Taking the results of DGPs I and II first, it is evident that the parameter estimates of the duration process are unbiased while those of the marks process are biased, in general. In particular, the equation-by-equation estimates of a_{21} and τ_0 are seriously biased when Type I weak exogeneity is violated. The equation-by-equation estimates of a_{21} appear negative and this negativity increases dramatically with the magnitude of error correlation between the two processes (DGP II). Give that the true parameter values are positive, the negative point estimates imply that the bias due to the failure of Type I weak exogeneity could be severe, leading to an altogether different interpretation of the relationship between duration and marks. The estimate of τ_0 is biased upward by a factor of 2.8 and this bias also increases with increasing error correlation between these two processes. The RMSEs of these two parameters are also very large. These findings are also confirmed in the out of sample forecasts which are as shown in

Figures 1 and 2.[9] The out of sample forecasts of duration (μ_t^x) are very close to their population value while the same forecasts of marks (μ_t^y) show divergences from their population values and these divergence become more pronounced in DGP II (Figure 2). This suggests that the forecasts of conditional volatility, under the assumption of weak exogeneity contained in Engle and Gallo (2006), may not be without concern. Overall, our simulation results show that the violation of Type I weak exogeneity has a serious effect on the estimates of the marks process.

Considering the results of DGP III, it is interesting that the equation-by-equation estimates of a_{21} and τ_0 are generally unbiased in the marks process, implying that the misspecification of Type II weak exogeneity does not affect the relationship between duration and marks. However, it shows small downward biases in the estimates of persistent parameter, both in the duration process (a_{11}, b_{11}) and marks process (a_{22}, b_{22}), suggesting smaller persistence estimates of both duration and marks. The out of sample forecasts of conditional expectation μ_t^x and μ_t^y, plotted in Figure 3, are also close to the population values, although the former shows some deviations. Overall, the violation of Type II weak exogeneity appears to have little consequence.

4.2 Test of the weak exogeneity of duration for return volatility

We employ the LM tests proposed in Section 3 to test the weak exogeneity of duration vis-à-vis return volatility. We use the Trades and Quotes (TAQ) dataset of the NYSE analysed by Engle and Patton (2004) and Manganelli (2005). Engle and Patton (2004) construct 10 deciles of stocks covering a period from 1 January, 1998 to 30 June, 1999 using the total number of trades of all stocks quoted on the NYSE in 1997. Out of this sample, Manganelli (2005) randomly selected five stocks from the eighth decile (frequently traded stocks) and another five stocks from the second decile (infrequently traded stocks) and analysed them in his study. We use the raw dataset of Manganelli and follow the approach outlined therein to compute duration, volume and return volatility (including diurnal adjustment for intraday patterns).[10] Specifically, we compute return sequences that are free from bid-ask bounces by following Ghysels

[9] We generate 5200 data and use the first 5000 data for the estimation and the remaining 200 for out of sample forecasting.

[10] See subsection 4.1 in Manganelli (2005) for a concise description of data prepared; we follow the same approach.

et al. (2004) and use the residuals of the ARMA (1,1) model on the return data. We also adjust the time-of-the-day effect following the method of Engle (2000). We regress durations, volumes and squared returns on a piecewise cubic spline with knots at 9:30, 10:00, 11:00, 12:00, 13:00, 14:00, 15:00, 15:30 and 16:00. The original series is then divided by the spline forecast to obtain the adjusted series for econometric analyses. The tickers' names and summary statistics of the ten stocks are reported in Table 4.

The number of observations range from 46,827 to 88,918 for frequently traded stocks in our sample and the average trading duration ranges from 99 seconds to 187 seconds. For the infrequently traded stocks, the number of observation ranges from 1,969 to 5,155 with an average trading duration of 1,693 seconds to 4,441 seconds. It is evident that there is no clear pattern in trading volumes between frequently and infrequently traded stocks in the data. The Ljung-Box (LB) statistics show duration, volume and volatility to be significantly serially autocorrelated; the precision of autocorrelation is particularly high for frequently traded stocks. As expected, these summary statistics are very similar to those reported by Manganelli.

The maximum likelihood estimation results of the MEMs under the null of weak exogeneity are presented in Tables 5 and 6. Table 5 contains the results of the frequently traded stocks and Table 6 reports those of the infrequently traded stocks.

We find large persistent parameters (b_{11}, b_{22}) which are consistent with the literature. The contemporaneous duration effect (τ_0) and the lagged duration effect (a_{21}) on return volatility are statistically significant. The overall effect $(a_{21} + \tau_0 b_{22})$[11] is negative which is consistent with the market microstructure predictions (Easley and O'Hara, 1992). This reinforces the findings elsewhere (e.g. Engle, 2000) that the time of great activity coincides with the probability of informed traders in the market, thereby increasing return volatility. We also find that the feedback effect from return volatility to duration (a_{12}) is negative. Our parameter estimates for duration and return volatility are in line with those of Manganelli (2005) despite the fact that he estimates three processes (viz., duration, volume and return volatility) whereas we model only duration and return volatility. Given our focus on weak exogeneity, modelling duration and return volatility is sufficient for the purpose at hand.

[11] The formula is from Manganelli (2005) Equation (12).

On weak exogeneity, the Type I LM tests reject the null of weak exogeneity (i.e., orthogonality) of duration vis-à-vis return volatility in four of the five cases for frequently traded stocks at the 5% significance level or better. The Type II LM tests reject the weak exogeneity of duration (i.e., no lagged cross-dependence) in all cases for frequently traded stocks. These are interesting findings which go against the assumption of weak exogeneity of duration maintained by empirical literature while analysing frequently traded stocks in the MEM framework. However, the picture changes somewhat vis-à-vis the infrequently traded stocks (Table 6). The null of weak exogeneity is only rejected in two of the five stocks under analyses by both Type I and Type II LM tests.

The correct specification of the conditional mean of the MEMs is fundamental for the validity of weak exogeneity tests. This is because the rejection of the null hypothesis could be due to either the rejection of weak exogeneity or the misspecification of the conditional mean. To ensure that our results do not suffer from the latter, we re-specify the duration as the Augmented ACD (AACD) model (Fernandes and Grammig, 2006) and return volatility as the APGARCH process (Ding et al., 1993), and re-compute our proposed LM tests of weak exogeneity. Given that the AACD model is flexible enough to encompass most of the extant ACD models, the rejection of the null is less likely to be due to misspecification of the conditional mean. These results are reported in the last two rows of Tables 5 and 6.[12] These LM tests confirm the robustness of our results. Overall, our results suggest that the assumption of weak exogeneity is not data sustainable for frequently traded stocks, whereas there could be some exceptions vis-à-vis infrequently traded stocks. This suggests that duration and marks should be estimated jointly.

5 Conclusion

A common strategy in modelling financial duration and marks under MEMs has been to assume weak exogeneity of duration vis-à-vis marks and estimate each process separately. However, microstructure theory suggests that duration, volume and transaction prices are jointly determined which raises doubt on the assumption of

[12] The details of the AACD/APGARCH model and estimated results are in Appendix 4. The estimation results shows that the (asymmetric) log-ACD models are appropriate for the duration process, while the (asymmetric) linear GARCH models are appropriate for the volatility process in general.

weak exogeneity maintained by voluminous empirical work. While the issue of weak exogeneity is not new in the literature, the relevant test statistics are mainly derived for linear Gaussian VAR or VECM models, which are not directly applicable to these non-linear and non-Gaussian MEMs. In this paper, we propose two types of weak exogeneity tests applicable to this class of models. The first test statistic, Type I LM test, tests the null of no contemporaneous error correlations between the innovations of duration and marks processes. The second test statistic, Type II LM test, tests the null of no lagged cross-dependence between these processes. Our simulation exercises reveal that the proposed tests have good power properties.

Through Monte Carlo simulations, we show that the failure of weak exogeneity results in biased parameter estimates in the MEMs, a finding which is also fairly standard vis-à-vis linear VAR or VECM models. This bias becomes accentuated with increasing error correlation between the duration and marks processes. This shows that the issue of weak exogeneity is an important one in estimating these MEMs requiring a formal test rather than simply assuming it away. As an empirical illustration, we assess the weak exogeneity assumption maintained by Manganelli (2005) through our LM tests in his dataset and find that the assumption is widely rejected – four of the five highly frequently traded stocks and two of the five infrequently traded stocks reject the null of weak exogeneity. This implies that the routine assumption of weak exogeneity maintained by the MEMs of market microstructure may not be data congruent, where our proposed tests might come in handy in establishing their weak exogeneity or otherwise.

References

Admati, A. R. and P. Pfleiderer (1988). "A theory of intraday patterns: Volume and price variability." Review of Financial Studies **1**(1): 3-40.

Allen, D., et al. (2008). "Finite sample properties of the QMLE for the Log-ACD model: Application to Australian stocks." Journal of Econometrics **147**(1): 163-185.

Bauwens, L. and P. Giot (2000). "The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks." Annales d'Economie et de Statistique **60**: 117-149.

Boswijk, H. P. and J.-P. Urbain (1997). "Lagrange-multiplier tests for weak exogeneity: a synthesis." Econometric Reviews **16**(1): 21-38.

Bowe, M., S. Hyde and L. McFarlane (2009). "The interrelationship between trading volume, duration and futures contract price volatility." Manchester Business School working paper.

Cipollini, F., et al. (2007). "A model for multivariate non-negative valued processes in financial econometrics." Unpublished paper: Stern School of Business, New York University.

Ding, Z., et al. (1993). "A long memory property of stock market returns and a new model." Journal of Empirical Finance **1**(1): 83-106.

Dolado, J. J. (1992). "A note on weak exogeneity in VAR cointegrated models." Economics Letters **38**(2): 139-143.

Dufour, A. and R. F. Engle (2000). "Time and the price impact of a trade." Journal of Finance **55**(6): 2467-2498.

Easley, D. and M. O'Hara (1992). "Time and the process of security price adjustment." Journal of Finance **47**(2): 577-605.

Engle, R. F. (1982). "A general approach to Lagrange multiplier model diagnostics." Journal of Econometrics **20**(1): 83-104.

Engle, R. F. (2000). "The Econometrics of Ultra-high-frequency Data." Econometrica **68**(1): 1-22.

Engle, R. F. (2002). "New frontiers for ARCH models." Journal of Applied Econometrics **17**(5): 425-446.

Engle, R. F. and G. M. Gallo (2006). "A multiple indicators model for volatility using intra-daily data." Journal of Econometrics **131**(1): 3-27.

Engle, R. F. and D. F. Hendry (1993). "Testing Superexogeneity and Invariance in Regression Models." Journal of Econometrics **56**(1–2): 119–139.

Engle, R. F., et al. (2012). "Liquidity, Volatility, and Flights to Safety in the U.S. Treasury Market: Evidence from a New Class of Dynamic Order Book Models." Federal Reserve Bank of New York Staff Report.

Engle, R. F., et al.(1983). "Exogeneity." Econometrica**51**(2): 277-304.

Engle, R. F. and A. J. Patton (2004). "Impacts of trades in an error-correction model of quote prices." Journal of Financial Markets**7**(1): 1-25.

Engle, R. F. and J. R. Russell (1998). "Autoregressive conditional duration: A new model for irregularly spaced transaction data." Econometrica**66**(5): 1127-1162.

Engle, R. F. and Z. Sun (2007). "When is noise not noise–a microstructure estimate of realized volatility." Manuscript, Stern School, New York University.

Fernandes, M. and J. Grammig (2006). "A family of autoregressive conditional duration models." Journal of Econometrics**130**(1): 1-23.

Foster, F. D. and S. Viswanathan (1996). "Strategic trading when agents forecast the forecasts of others." Journal of Finance**51**(4): 1437-1478.

Ghysels E, Gourieroux C, Jasiak J. (2004). "Stochastic volatility duration models. " Journal of Econometrics **119**: 413-433.

Grammig, J. and M. Wellner (2002). "Modeling the interdependence of volatility and inter-transaction duration processes." Journal of Econometrics**106**(2): 369-400.

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**(12): 3978-4015.

Hendry, D. F. (1995). "On the interactions of unit roots and exogeneity." Econometric Reviews**14**(4): 383-419

Manganelli, S. (2005). "Duration, volume and volatility impact of trades." Journal of Financial Markets**8**(4): 377-399.

Nakatani, T. and T. Teräsvirta (2009). "Testing for volatility interactions in the Constant Conditional Correlation GARCH model." The Econometrics Journal**12**(1): 147-163.

White, H. (1981). "Consequences and detection of misspecified nonlinear regression models." Journal of the American Statistical Association**76**(374): 419-433.

White, H. (1982). "Maximum likelihood estimation of misspecified models." Econometrica**50**(1): 1-25.

Taylor, N. and Xu, Y. (2016). "The logarithmic vector multiplicative error model: An application to high frequency NYSE stock data ", Quantitative Finance, forthcoming.

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**(1).

Table 1: Percentage rejections of the LM tests for testing $\tau_1 = 0$ against $\tau_1 \neq 0$

(a)	Empirical size of the test		
τ_1	n=2000	n=5000	n=10000
0.000	0.089	0.072	0.067

(b)	Power of the test		
τ_1	n=2000	n=5000	n=10000
-0.200	0.825	0.989	1.000
-0.175	0.736	0.977	1.000
-0.150	0.614	0.933	0.999
-0.125	0.457	0.803	0.972
-0.100	0.354	0.610	0.878
-0.075	0.232	0.403	0.600
-0.050	0.149	0.210	0.347
-0.025	0.084	0.093	0.141
0.000	0.052	0.051	0.051
0.025	0.041	0.057	0.098
0.050	0.054	0.120	0.236
0.075	0.105	0.277	0.502
0.100	0.164	0.458	0.781
0.125	0.260	0.684	0.938
0.150	0.414	0.848	0.987
0.175	0.547	0.924	0.997
0.200	0.642	0.970	1.000

Note: The empirical size of the test is the percentage rejection of LM test at 5% theoretical significance level for testing $\tau_1 = 0$ against $\tau_1 = 0$. The power of the test is the percentage rejections of the LM tests at empirical size for testing $\tau_1 = 0$ against $\tau_1 \neq 0$.

Table 2: Data generating Processes

	ω	A	B	τ_0	σ_{12}
DGP I	$\begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}$	$\begin{pmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{pmatrix}$	$\begin{pmatrix} 0.85 & - \\ - & 0.80 \end{pmatrix}$	0.1	0.1
DGP II	$\begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}$	$\begin{pmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{pmatrix}$	$\begin{pmatrix} 0.85 & - \\ - & 0.80 \end{pmatrix}$	0.1	0.5
DGP III	$\begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix}$	$\begin{pmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{pmatrix}$	$\begin{pmatrix} 0.85 & -0.05 \\ -0.05 & 0.80 \end{pmatrix}$	0.1	–

Model:

$$
\begin{aligned}
x_t &= \mu_t^x(\theta_x;\Omega_{t-1})\varepsilon_{1t}, \\
y_t &= \mu_t^y(\theta_y;\Omega_t)\varepsilon_{2t}, \quad \operatorname{cov}(\varepsilon_{1t},\varepsilon_{2t}) = \sigma_{12}; \\
\mu_t^x &= w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x + b_{12}\mu_{t-1}^y, \\
\mu_t^y &= w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{21}\mu_{t-1}^x + b_{22}\mu_{t-1}^y + \tau_0 x_t.
\end{aligned}
$$

The population parameter values are

$w_1 = w_2 = 0.1$, $a_{11} = a_{22} = a_{12} = a_{21} = 0.05$,

$b_{11} = 0.85$, $b_{22} = 0.80$, $b_{12} = b_{21} = -0.05$, $\tau_0 = 0.1$.

The population parameter values (particularly the sign) are from the empirical results of Manganelli (2005) and Taylor and Xu (2016).

Table 3: Simulation results (consequences of the failure of weak exogeneity).

	DGP I ($\rho = 0.1$)		DGP II ($\rho = 0.5$)		DGP III ($b_{12} = b_{21} = -0.05$)	
	Mean	SD	Mean	SD	Mean	SD
Duration						
w_1=0.10	0.102	0.014	0.102	0.012	0.124	0.018
a_{11}=0.05	0.050	0.006	0.050	0.007	0.039	0.007
a_{12}=0.05	0.050	0.005	0.050	0.005	0.053	0.006
b_{11}=0.85	0.849	0.011	0.849	0.009	0.767	0.026
Marks						
w_2=0.10	0.084	0.018	0.036	0.018	0.083	0.014
a_{21}=0.05	**-0.102**	0.019	**-0.534**	0.036	0.055	0.017
a_{22}=0.05	0.041	0.007	0.019	0.008	0.047	0.007
b_{22}=0.80	0.793	0.015	0.772	0.028	0.775	0.019
τ_0=0.10	**0.280**	0.017	**0.794**	0.022	0.100	0.015
RMSE(a_{12})	0.005		0.005		0.006	
RMSE(a_{21})	0.153		0.586		0.018	
RMSE(τ_0)	0.181		0.695		0.015	

Notes: Mean denotes the average value; SD denotes the standard deviation. The true parameter values used in DGP are reported in the first column.

Estimated Model under null hypothesis of weak exogeneity:
Duration equation

$$x_t = \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t},\ \varepsilon_{1t} \sim i.i.d(1,\sigma_{\varepsilon_1})$$

$$\mu_t^x = w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x$$

Marks equation

$$y_t = \mu_t^y(\theta_y;\Omega_t)\varepsilon_{12,t}, \varepsilon_{12,t} \sim i.i.d(1,\sigma_{\varepsilon_{12}})$$

$$\mu_t^y = w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{22}\mu_{t-1}^y + \tau_0 x_t.$$

Under null hypothesis, duration and marks equation are estimated separately.

Table 4: Summary statistics of the ten sample stocks.

Ticker	Obs	Mean			LB(15)		
		Duration	Volume	Volatility	Duration	Volume	Volatility
DLP	65305	134.15	1486.35	7.42	19672.60	1775.60	2790.50
GAP	46827	187.30	824.63	5.57	4962.56	174.97	2105.26
CP	71673	122.42	2892.23	5.44	3713.70	769.86	4315.86
COX	88918	98.60	2678.86	5.93	29905.30	364.91	4109.56
AVT	58390	150.02	1070.01	4.57	6772.70	413.04	1751.77
JAX	2766	3164.67	1000.04	2.63	109.03	104.80	259.14
GSE	1969	4441.14	1523.77	1.69	99.65	117.59	188.21
GBX	5155	1693.45	1434.04	1.82	1047.46	170.84	683.37
FTD	3625	2417.11	736.25	0.53	19.29	69.24	289.32
DTC	4162	2093.91	2136.83	2.25	155.35	142.41	239.07

Notes: The first 5 stocks are frequently traded stocks and the last 5 stocks are infrequently traded stocks. The mean of volatility is given in per second and is obtained by dividing the mean of Abs(Return) by the mean duration and multiplying by 10^6. The Ljung-Box (LB) test statistic is based on 15 lags of duration, volume, or volatility (given by absolute return). The 95% critical value associated with the LB test statistic equals 25.00. Mean statistics pertain to the series before diurnal adjustment, while the LB statistics pertain to the series after diurnal adjustment.

Table 5: Results of MEM models under the null hypothesis – frequently traded stocks.

	DLP	GAP	CP	COX	AVT
Duration					
a_{11}	0.068**	0.044**	0.029**	0.062**	0.089**
a_{12}	-0.001**	0.001	-0.002**	0.000	-0.007**
b_{11}	0.932**	0.952**	0.962**	0.936**	0.891**
Volatility					
a_{21}	-0.252**	-0.245**	-0.202**	-0.211**	-0.207**
a_{22}	0.026**	0.034**	0.057**	0.043**	0.031**
b_{22}	0.962**	0.947**	0.885**	0.934**	0.941**
τ_0	0.260**	0.255**	0.208**	0.221**	0.212**
$a_{21} + \tau_0 b_{22}$	-0.002	-0.004	-0.018	-0.005	-0.008
LB-duration	113.18	57.59	117.84	99.89	76.61
LB-AR	59.55	109.76	159.38	111.67	26.88
Type I LM	1.20	24.52**	43.33**	716.74**	141.78**
Type II LM	11.19**	25.23**	237.71**	1141.79**	145.06**
AACD Model:					
Type I LM	20.53**	4.28*	36.33**	14.58**	2.24
Type II LM	109.89**	39.19**	32.56**	256.10**	14.40**

Note: ** denotes significant at 1% level. * denotes significant at 5% level. AR denotes absolute return. Type I LM test H_0: $\sigma_{12} = 0$ (Orthogonality). Type II LM test H_0: $b_{12} = b_{21} = 0$ (No lagged cross-dependence). Critical values $\chi^2(1)_{0.05}$ =3.84, $\chi^2(1)_{0.01}$ =6.64; $\chi^2(2)_{0.05}$ =5.99, $\chi^2(2)_{0.01}$ =9.12. AACD denotes Augmented ACD model as discussed in Section 4 of the text.

Table 6: Results of MEM model under null hypothesis – infrequently traded stocks.

	DTC	FTD	GBX	GSE	JAX
Duration					
a_{11}	0.075**	0.026**	0.065**	0.150**	0.042**
a_{12}	0.006	-0.006	0.004*	-0.056**	-0.002
b_{11}	0.890**	0.864**	0.931**	0.679**	0.949**
Volatility					
a_{21}	-0.139**	-0.153**	-0.139**	-0.158**	-0.196**
a_{22}	0.095**	0.052**	0.050**	0.044**	0.054**
b_{22}	0.800**	0.919**	0.938**	0.935**	0.894**
τ_0	0.174**	0.139**	0.147**	0.159**	0.204**
$a_{21} + \tau_0 b_{22}$	0.0002	-0.025	-0.001	-0.009	-0.014
LB-duration	12.07	9.61	15.55	11.19	9.57
LB-AR	20.85	30.67	28.56	9.32	17.13
Type I LM	36.53**	1.26	0.12	2.06	11.48**
Type II LM	67.39**	1.20	2.32	3.28	11.48**
AACD Model:					
Type I LM	23.88**	0.00	0.37	3.69	9.64**
Type II LM	30.68**	1.05	0.31	60.56**	24.27**

Note: ** denotes significant at 1% level. * denotes significant at 5% level. AR denotes absolute return. Type I LM test H_0: $\sigma_{12} = 0$ (Orthogonality). Type II LM test H_0: $b_{12} = b_{21} = 0$ (No lagged cross-dependence). Critical values $\chi^2(1)_{0.05}$ =3.84, $\chi^2(1)_{0.01}$ =6.64; $\chi^2(2)_{0.05}$ =5.99, $\chi^2(2)_{0.01}$ =9.12. AACD denotes Augmented ACD model as discussed in Section 4 of the text.

Figure 1: Out of sample forecasting - DGP I

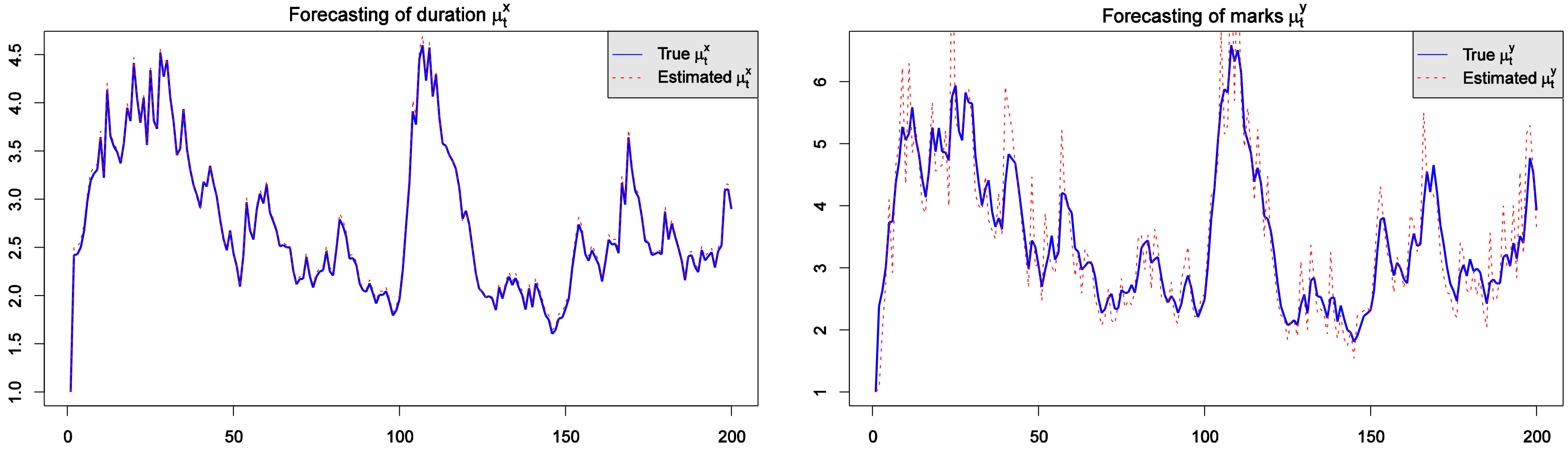


Figure 2: Out of sample forecasting - DGP II

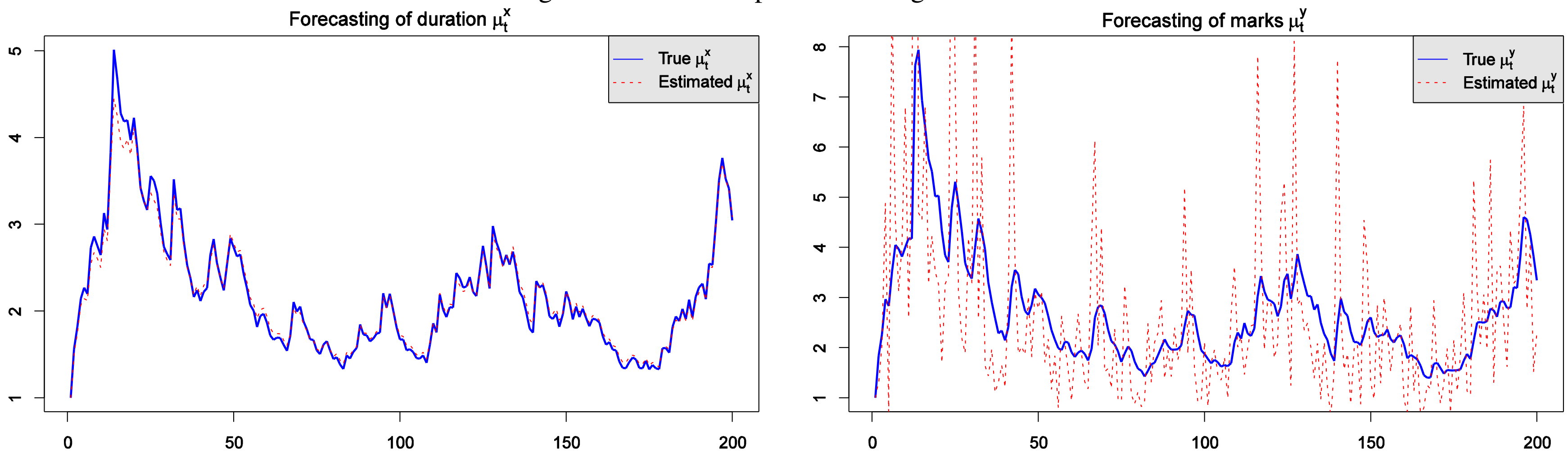
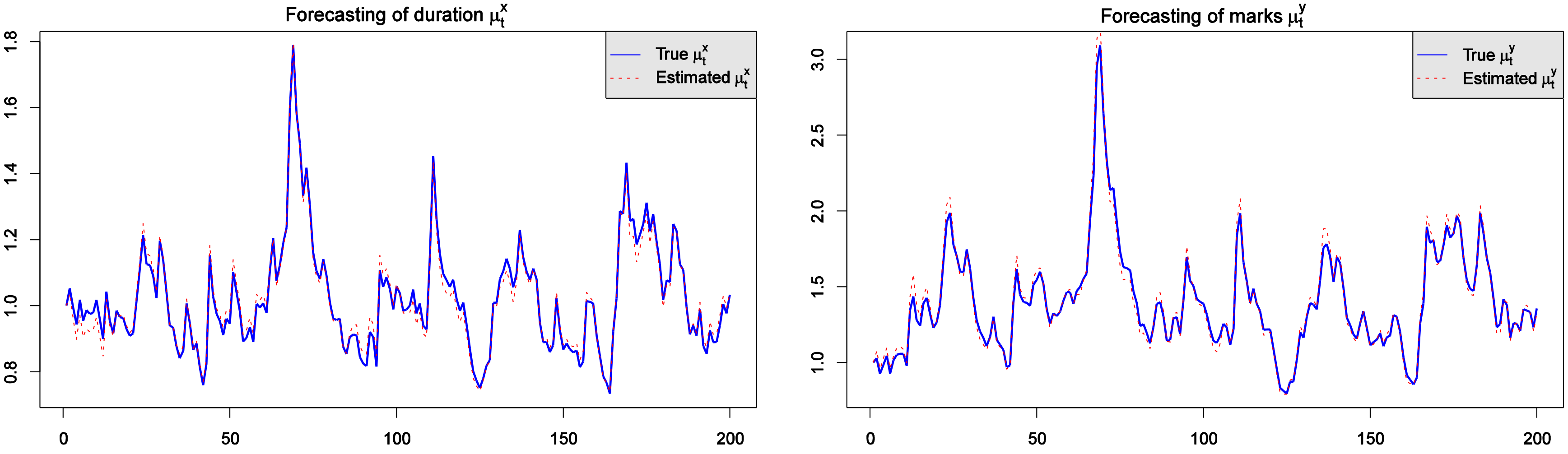


Figure 3: Out of sample forecasting - DGP III



Note: True μ_t^x denotes the out of sample forecast of duration based on population paramter values. Estimated μ_t^x denotes the out of sample forecast of duration based on estimated parameter values. This is equivalent to μ_t^y .

Appendix

Appendix 1: Proof of Lemma 1

Xu (2013) and Allen et al. (2008) show the equivalence of the ACD model to the ARMA model. Taylor and Xu (2016) shows that Vector MEM has a Vector ARMA specification. On similar lines, we transform the MEM model (2) into a VAR model and apply the results of Dolado (1992) and Boswijk and Urbain (1997) to derive Lemma 1.

First, writing (2) in matrix form:

$$\begin{aligned} z_t &= \mu_t \odot \varepsilon_t \\ \mu_t &= \omega + A z_{t-1} + B\mu_{t-1} + \Gamma z_t \end{aligned} \tag{20}$$

where $\varepsilon_t \,|\, \Omega_t \sim IID(I,\Sigma)$, $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ and metric Γ is restricted to $\begin{pmatrix} 0 & 0 \\ \tau_0 & 0 \end{pmatrix}$.

As $E(z_t \,|\, \Omega_t) = \mu_t$, we define the martingale difference as $e_t = z_t - \mu_t$. Assuming e_t as homoscedastic, the Gaussian mean innovation process with a constant covariance matrix D (see Engle et al., 1983), i.e., $e_t \,|\, \Omega_t \sim IID(0,D)$, $D = \begin{bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{bmatrix}$, the second moment conditions of z_t from (20) and from the martingale difference are given by:

$$Var(z_t \,|\, \Omega_t) = \mu_t \mu_t' \odot \Sigma = diag(\mu_t)\Sigma\, diag(\mu_t)$$

$$Var(z_t \,|\, \Omega_t) = Var(e_t \,|\, \Omega_t) = D\,.$$

Therefore, testing $\sigma_{12} = 0$ is equivalent to testing $d_{12} = 0$. If $\sigma_{12} = 0$, then $d_{12} = 0$ and vice versa. We then transform the vector MEM in (20) into a VAR form model as follows:

$$\begin{aligned} z_t &= \omega + A z_{t-1} + B\mu_{t-1} + \Gamma z_t + z_t - \mu_t \\ &= \omega + A z_{t-1} + B\mu_{t-1} + \Gamma z_t + e_t \\ &= \lambda' \chi_{t-1} + \Gamma z_t + e_t \end{aligned} \tag{21}$$

where $\chi_{t-1} = (1, x_{t-1}, y_{t-1}, \mu^x_{t-1}, \mu^y_{t-1})$ and $\lambda = \{w, A, B\}'$; $e_t \,|\, \Omega_t \sim IID(0,D)$, $D = \begin{bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{bmatrix}$. And re-writing (21) equation-by-equation as

$$x_t = a'\chi_{t-1} + e_{1t}$$
$$y_t = \beta'\chi_{t-1} + \tau_o x_t + e_{2t} \tag{22}$$

where, $a = \{w_1, a_{11}, a_{12}, b_{11}, b_{12}\}'$ and $\beta = \{w_2, a_{21}, a_{22}, b_{21}, b_{22}\}'$. Since e_t is a homoscedastic Gaussian mean innovation process relative to χ_{t-1} (Engle et al., 1983) with a constant covariance matrix D, we follow Engle and Hendry (1993) and Boswijk and Urbain (1997) and partition (22) as:

$$x_t = a'\chi_{t-1} + e_{1t}$$
$$y_t = \beta'\chi_{t-1} + \tau_0' x_t + \tau_1 \hat{x}_t + e_{12,t} \tag{23}$$

where $\tau_1 = -d_{12} / d_{22}$, $\tau_0' = \tau_0' - \tau_1$ and $\hat{x}_t$ is the predicted value of x_t. Now $e_{12,t}$ is independent of e_{1t} by construction. Therefore, it is sufficient to test τ_1=0 in (23) in order to test $d_{12} = 0$ in (21). Further, this also implies that it is sufficient to test τ_1=0 in (23) in order to test $\sigma_{12} = 0$ in (20). Transforming the VAR model (23) into the vector MEM model, we obtain:

$$x_t = \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t}$$
$$\mu_t^x = w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x$$
$$y_t = \mu_t^y(\theta_y;\Omega_t)\varepsilon_{12,t} \tag{24}$$
$$\mu_t^y = w_2 + a_{21}x_{t-1} + a_{22}v_{t-1} + b_{22}\mu_{t-1}^y + \tau_0' x_t + \tau_1 \mu_t^x.$$

It is plausible to use the expected value rather than the predicted value in the MEM models, hence the last term of the μ_t^y equation in (24) contains μ_t^x rather than $\hat{x}_t$. Because $e_{12,t}$ is independent of e_{1t}, $\varepsilon_{12,t}$ is independent of ε_{1t}. This implies that it is sufficient to test τ_1=0 in (24) in order to test $\sigma_{12} = 0$ in (20). This proves Lemma 1.

Appendix 2: An example to show where the orthogonality is not a necessary condition for the weak exogeneity of x_t

Assuming no lagged cross dependence i.e., $b_{12} = b_{21} = 0$, the joint density of x_t and y_t follows a vector MEM:

$$
\begin{aligned}
& \left.\begin{aligned} x_t &= \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t} \\ y_t &= \mu_t^y(\theta_y;\Omega_t)\varepsilon_{2t} \end{aligned}\right., \qquad \operatorname{cov}(\varepsilon_{1t},\varepsilon_{2t}) = \sigma_{12} \\
& \mu_t^x = w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x \\
& \mu_t^y = w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{22}\mu_{t-1}^y + \tau_0 x_t .
\end{aligned} \tag{25}
$$

If the parameters of interest are $\theta_y = \{w_2, a_{21}, a_{21}, b_{22}, \tau_0\}'$, the orthogonality is a requirement for the weak exogeneity of x_t vis-à-vis θ_y.

However, if one is only interested in the conditional density of y_t given x_t then the orthogonality condition might be not necessary but it requires additional restrictions. We illustrate this by making use of Example 3.2 of Engle et al. (1983). Let's transform the vector MEM into a VAR form model similar to Appendix 1 (equation 21):

$$
\begin{aligned}
x_t &= w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x + e_{1t} \\
y_t &= w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{22}\mu_{t-1}^y + \tau_0 x_t + e_{2t}
\end{aligned} \tag{26}
$$

where $e_t = \{e_{1t}, e_{2t}\}'$ and $e_t \mid \Omega_t \sim IID(0, D)$, $D = \begin{bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{bmatrix}$. Following Engle et al. (1983; Example 3.2), we can derive the conditional density of y_t given x_t for (26) as:

$$
\begin{aligned}
& y_t = bx_t + (c_1 x_{t-1} + c_2 y_{t-1}) + (w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{22}\mu_{t-1}^y) + e_{2t}^*, \\
& e_{2t}^* \sim IID(0, d^2)
\end{aligned} \tag{27}
$$

where $b = \tau_o + \dfrac{d_{12}}{d_{11}}, c_1 = a_{11}\dfrac{d_{12}}{d_{11}}, c_2 = a_{12}\dfrac{d_{12}}{d_{11}}, d^2 = d_{11} - d_{12}^2 / d_{11}$. Rewriting it as:

$$
y_t = w_2 + a_{21}^* x_{t-1} + a_{22}^* y_{t-1} + b_{22}\mu_{t-1}^y + bx_t + e_{2t}^*, \tag{28}
$$

where $a_{21}^* = a_{21} + c_1$ and $a_{22}^* = a_{22} + c_2$.

Now transforming (28) into the MEM model similar to Appendix 1 (equation 24), we obtain the conditional density of y_t given x_t for the vector MEM of (25):

$$y_t = \mu_t^y(\theta_t^*;\Omega_t)\varepsilon_{2t}^*, \varepsilon_{2t}^* \sim IID(1,\sigma_2^*)$$
$$\mu_t^y = w_2 + a_{21}^* x_{t-1} + a_{22}^* y_{t-1} + b_{22}\mu_{t-1}^y + bx_t. \tag{29}$$

The marginal density of x_t remains unchanged. Now the conditional density and the marginal density are independent by construction. Under this formulation, if the parameters of interest are $\theta_y^* = \{w_2, a_{21}^*, a_{22}^*, b_{22}, b_0\}'$, then the orthogonality is not a necessary condition for the weak exogeneity of x_t. However, it should be noted that x_t is still not weakly exogenous for θ_y^* not because of the breakdown of orthogonality but due to the emergence of new cross-restriction between θ_x and θ_y^*. This cross-restriction is: $c_1 a_{12} = c_2 a_{11}$, which is also shown in Engle et al. (1983; Example 3.2). Obviously, further restrictions are required to operate a sequence cut on this cross-restriction which ensures the weak exogeneity of x_t. Following Engle et al. (1983), it suffices to impose $a_{12} = 0$, which effectively implies no feedback effect from marks to duration in our model.

Appendix 3: Testing the joint hypothesis

Testing joint hypothesis $H_0: \ \tau_1 = b_{12} = b_{21} = 0$

Under H_0:

$$\begin{aligned} x_t &= \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t} \\ \mu_t^x &= w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x \\ y_t &= \mu_t^y(\theta_y;\Omega_t)\varepsilon_{2t} \\ \mu_t^y &= w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{22}\mu_{t-1}^y + \tau_0 x_t. \end{aligned} \tag{30}$$

Under H_1:

$$\begin{aligned} x_t &= \mu_t^x(\theta_x;\Omega_t)\varepsilon_{1t} \\ \mu_t^x &= w_1 + a_{11}x_{t-1} + a_{12}y_{t-1} + b_{11}\mu_{t-1}^x + b_{12}\mu_{t-1}^y \\ y_t &= \mu_t^y(\theta_y;\Omega_t)\varepsilon_{12,t} \\ \mu_t^y &= w_2 + a_{21}x_{t-1} + a_{22}y_{t-1} + b_{21}\mu_{t-1}^x + b_{22}\mu_{t-1}^y + \tau_0 x_t + \tau_1 \mu_t. \end{aligned} \tag{31}$$

Assume ε_{1t} and $\varepsilon_{12,t}$ both have exponential density, the associated log-likelihood function is:

$$\begin{aligned} L(\theta) &= l_y(y \mid \theta_y, \mu^x(\theta_x)) + l_x(x \mid \theta_x, \mu^y(\theta_y)) \\ &= -\sum_{t=1}^{T}(\log \mu_t^y + y_t / \mu_t^y) - \sum_{t=1}^{T}(\log \mu_t^x + x_t / \mu_t^x). \end{aligned} \tag{32}$$

Moreover, under H_0 of weak exogeneity, the marginal and conditional models can be estimated separately. Then the joint score LM test has the familiar form

$$S = -\hat{i}^{\{b_{ij,i\neq j},\tau_1\}}(\hat{\theta}_c)'\hat{I}^{\{b_{ij,i\neq j},\tau_1\}}(\hat{\theta}_c)^{-1}\hat{i}^{\{b_{ij,i\neq j},\tau_1\}}(\hat{\theta}_c) \tag{33}$$

where $\hat{i}^{\{b_{ij,i\neq j},\tau_1\}}(\hat{\theta}_c) = \dfrac{\partial L}{\partial \theta}$ and $\hat{I}^{\{b_{ij,i\neq j},\tau_1\}}(\hat{\theta}_c) = \dfrac{\partial L}{\partial \theta \partial \theta'}$ are the components corresponding to $\{b_{ij,i\neq j},\tau_1\}$ in the empirical score and the Hessian from the unconstrained model. Under mild regularity conditions, it is well known that the score test has an asymptotically $\chi^2(3)$ distribution under H_0.

We only require the relevant part of the inverse of the Hessian matrix to derive the LM statistic. The score matrix is partitioned as:

$$\hat{i}(\hat{\theta}_c) = \begin{pmatrix} \hat{i}_y(\hat{\theta}_c^y) \\ \hat{i}_x(\hat{\theta}_c^x) \end{pmatrix} = \begin{pmatrix} \hat{i}_y^{\{b_{21},\tau_1\}}(\hat{\theta}_c^y) \\ \hat{i}_y^{\theta_y\dagger}(\hat{\theta}_c^y) \\ \hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \\ \hat{i}_x^{\theta_x\dagger}(\hat{\theta}_c^x) \end{pmatrix} = \begin{pmatrix} \hat{i}_y^{\{b_{21},\tau\}}(\hat{\theta}_c^y) \\ 0 \\ \hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \\ 0 \end{pmatrix} \tag{34}$$

where $\theta_{y\dagger} = \theta_y \setminus \{b_{21}, \tau_1\}$ i.e. all other the parameters in θ_y, except the one pertaining to $\{b_{21}, \tau_1\}$ and equivalently for $\theta_{x\dagger} = \theta_x \setminus b_{12}$.

Hence, the joint LM test has the form:

$$S = -\begin{pmatrix} \hat{i}_y^{\{b_{21},\tau_1\}}(\hat{\theta}_c^y) & \hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \end{pmatrix} \begin{pmatrix} \hat{I}_y^{\{b_{21},\tau_1\}}(\hat{\theta}_c^y)^{-1} & 0 \\ 0 & \hat{I}_x^{b_{12}}(\hat{\theta}_c^x)^{-1} \end{pmatrix} \begin{pmatrix} \hat{i}_y^{\{b_{21},\tau_1\}}(\hat{\theta}_c^y) \\ \hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \end{pmatrix}. \tag{35}$$

where $\widehat{I}_y^{\{b_{21},\tau_1\}}(\widehat{\theta}_c^y)^{-1} = \left[\widehat{I}_y^{\{b_{21},\tau_1\}}(\widehat{\theta}_c^y) - \widehat{I}_y^{\{b_{21},\tau_1\},\theta_{y\dagger}}(\widehat{\theta}_c^y)\widehat{I}_y^{\theta_{y\dagger}}(\widehat{\theta}_c^y)^{-1}\widehat{I}_y^{\theta_{y\dagger},\{b_{21},\tau_1\}}(\widehat{\theta}_c^y)\right]^{-1}$ and $\widehat{I}_x^{b_{12}}(\widehat{\theta}_c^x)^{-1} = \left[\widehat{I}_x^{b_{12}}(\widehat{\theta}_c^x) - \widehat{I}_x^{b_{12},\theta_{x\dagger}}(\widehat{\theta}_c^x)\widehat{I}_x^{\theta_{x\dagger}}(\widehat{\theta}_c^x)^{-1}\widehat{I}_x^{\theta_{x\dagger},b_{12}}(\widehat{\theta}_c^x)\right]^{-1}$.

Furthermore,

$$\begin{aligned} S &= -\hat{i}_y^{\{b_{21},\tau_1\}}(\hat{\theta}_c^y)\hat{I}_y^{\{b_{21},\tau_1\}}(\hat{\theta}_c^y)^{-1}\hat{i}_y^{\{b_{21},\tau_1\}}(\hat{\theta}_c^y) - \hat{i}_x^{b_{12}}(\hat{\theta}_c^x)\hat{I}_x^{b_{12}}(\hat{\theta}_c^x)^{-1}\hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \\ &= -\hat{i}_y^{\tau_1}(\hat{\theta}_c^y)\hat{I}_y^{\tau_1}(\hat{\theta}_c^y)^{-1}\hat{i}_y^{\tau_1}(\hat{\theta}_c^y) - \hat{i}_y^{b_{21}}(\hat{\theta}_c^y)\hat{I}_y^{b_{21}}(\hat{\theta}_c^y)^{-1}\hat{i}_y^{b_{21}}(\hat{\theta}_c^y) - \hat{i}_x^{b_{12}}(\hat{\theta}_c^x)\hat{I}_x^{b_{12}}(\hat{\theta}_c^x)^{-1}\hat{i}_x^{b_{12}}(\hat{\theta}_c^x) \\ &= S_1 + S_2. \end{aligned} \tag{36}$$

The joint LM test statistics tends to be the sum of the two partial LM test statistics.

Appendix 4: AACD Models and estimation results

The AACD model is given by:

$$d_t = \psi_t(\theta_d;\Omega_t)\varepsilon_t$$

$$\psi_t^{\lambda} = \omega + \alpha\phi_{t-1}^{\lambda}[|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^{v} + \beta\psi_{t-1}^{\lambda}$$

The AACD model is flexible enough to permit the conditional duration process { ψ_t } and respond distinctly to small and large shocks. The shock impact curve $g(\varepsilon_i) = [|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^{v}$ incorporates asymmetric responses through the shift and rotation parameters b and c, respectively. The shape parameter v plays a similar role to λ, which determines if the Box-Cox transformation is concave ($\lambda \le 1$) or convex ($\lambda \ge 1$). Below we provide a typology showing how ACD models are nested under the AACD model. Augmented ACD:

$$\psi_t^{\lambda} = \omega + \alpha\phi_{t-1}^{\lambda}[|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^{v} + \beta\psi_{t-1}^{\lambda}$$

Asymmetric power ACD ($\lambda = v$)

$$\psi_t^{\lambda} = \omega + \alpha\phi_{t-1}^{\lambda}[|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^{\lambda} + \beta\psi_{t-1}^{\lambda}$$

Asymmetric logarithmic ACD ($\lambda \to 0$ and $v = 1$)

$$\log\psi_t = \omega + \alpha[|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^{\lambda} + \beta\log\psi_{t-1}$$

Asymmetric ACD ($v = \lambda = 1$)

$$\psi_t = \omega + \alpha[|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)]^{\lambda} + \beta\psi_{t-1}$$

Power ACD ($\lambda = v$ and b=c=0)

$$\psi_t^{\lambda} = \omega + \alpha x_{t-1}^{\lambda} + \beta\psi_{t-1}^{\lambda}$$

Box-Cox ACD ($\lambda \to 0$ and b=c=0) Dufour and Engle (2000)

$$\log\psi_t = \omega + \alpha\varepsilon_{t-1}{}^{\lambda} + \beta\log\psi_{t-1}$$

Logarithmic ACD Type I ($\lambda, v \to 0$ and b=c=0) Bauwens and Giot (2000)

$$\log\psi_t = \omega + \alpha\log x_{t-1} + \beta\log\psi_{t-1}$$

Logarithmic ACD Type II ($\lambda \to 0, v = 1$ and b=c=0) Bauwens and Giot (2000)

$$\log\psi_t = \omega + \alpha\varepsilon_{t-1} + \beta\log\psi_{t-1}$$

Linear ACD ($\lambda = v = 1$ and b=c=0) Engle and Russell (1998)

$$\psi_t = \omega + \alpha x_{t-1} + \beta\psi_{t-1}$$

The estimation results of the AACD and APGARCH under null hypothesis of weak exogeneity are reported in Tables 7 and 8.

Table 7: Estimation results – AACD-APGARCH model

Frequently traded stocks	DLP	GAP	CP	COX	AVT
Duration					
α	0.190**	0.020**	0.027**	0.110**	0.091**
β	0.797**	0.973**	0.965**	0.962**	0.864**
λ	0.075**	0.096**	0.493**	0.393**	0.033**
v	0.072**	0.568**	0.623**	0.723**	0.102**
b	-1.171**	-1.114**	-0.040	0.029**	-1.823**
c	0.109**	0.597**	0.006	0.778**	0.047
a_{12}	0.000**	0.000	-0.001*	0.000	0.000**
Volatility					
α	0.025**	0.022**	0.053**	0.021**	0.052**
β	0.956**	0.948**	0.893**	0.959**	0.925**
λ	1.210**	1.130**	0.932**	0.474**	1.035**
v	1.087**	1.211**	1.189**	1.082**	0.778**
b	0.454**	0.169	0.381	0.018*	0.804**
c	-0.366**	-0.376*	0.041	0.120	-0.433**
α_{21}	-0.251**	-0.254**	-0.196**	-0.077**	-0.203**
τ	0.262**	0.266**	0.202**	0.079**	0.211**
$a_{21} + \tau_0 b_{22}$	-0.001	-0.002	-0.016	-0.001	-0.008
LB-duration	118.52	48.17	119.49	99.87	59.71
LB-AR	51.39	89.65	101.78	102.54	22.86
Type I LM	20.53**	4.28*	36.33**	14.58**	2.24
Type II LM	109.89**	39.19**	32.56**	256.10**	14.40**

Note: ** denotes significant at 1% level. * denotes significant at 5% level.

LM Critical values $\chi^2(1)_{0.05}$ =3.84, $\chi^2(1)_{0.01}$ =6.64; $\chi^2(2)_{0.05}$ =5.99, $\chi^2(2)_{0.01}$ =9.21

Table 8: Estimation results- AACD-APGARCH model

Infrequently traded stocks	DTC	FTD	GBX	GSE	JAX
Duration					
α	0.195**	0.101**	0.104**	0.077	0.041**
β	0.839**	0.763**	0.920**	0.799**	0.948**
λ	1.017**	1.025**	0.636**	0.430**	0.073**
v	0.509**	0.371**	0.557**	1.010**	0.062**
b	0.042**	0.176**	0.177**	-0.716	0.187**
c	0.292**	0.031	0.219**	0.444	-0.013**
a_{12}	0.008**	-0.007	0.003**	-0.010*	0.000**
Volatility					
α	0.168**	0.161**	0.087**	0.131	0.059**
β	0.813**	0.839**	0.927**	0.933**	0.889**
λ	1.039**	0.832**	1.348**	0.850**	1.152**
v	1.858**	0.592**	1.045**	1.748**	1.207**
b	-0.221	0.556**	0.019	-1.422**	0.365**
c	0.603**	-0.032**	0.275	0.791**	-0.047
α_{21}	-0.132**	-0.153**	-0.159**	-0.135*	-0.198**
τ	0.176**	0.144**	0.170**	0.134*	0.209**
$a_{21} + \tau_0 b_{22}$	0.011	-0.032	-0.001	-0.001	-0.012
LB_D	11.42	9.02	14.94	7.78	11.35
LB_AR	13.90	21.24	29.24	9.25	17.43
Type I LM	23.88**	0.00	0.37	3.69	9.64**
Type II LM	30.68**	1.05	0.31	60.56**	24.27**

Note: ** denotes significant at 1% level. * denotes significant at 5% level.
LM Critical values $\chi^2(1)_{0.05}$ =3.84, $\chi^2(1)_{0.01}$ =6.64; $\chi^2(2)_{0.05}$ =5.99, $\chi^2(2)_{0.01}$ =9.21

Results in Tables 7 and 8 reveal that, for the duration process, the Box-Cox parameters (λ), are close to zero for the three stocks: DLP, GAP and AVT; and range between 0.393 and 0.493 for the remaining two frequently traded stocks. The shape parameter (v), is close to zero for DLP and AVT and ranges between 0.568 and 0.723 for the remainder. The parameters incorporating asymmetric effect[13] (c) are all positive and most of them are significant. This suggests (asymmetric) Log-ACD models are appropriate specifications for the duration of these stocks. For the other six stocks, neither ACD nor Log-ACD models are appropriate specifications. For the volatility process, both the Box-Cox parameters (λ) and shape parameters (v) are

[13] The asymmetric effect of transaction is driven by a sell or buy process.

close to 1, while the asymmetric effect parameters (c) are significantly different from zero. These suggest an (asymmetric) linear GARCH model is an appropriately good choice for modelling return volatility. In general, if no more information is available, the (asymmetric) log-ACD models are appropriate for the duration process, while the (asymmetric) linear GARCH models are appropriate for the volatility process.