

Semiparametric vector MEM^{*}

Fabrizio Cipollini[†] Robert F. Engle[‡] Giampiero M. Gallo[§]

This version: September 2008

Abstract

In financial time series analysis we encounter several instances of non-negative valued processes (volumes, trades, durations, realized volatility, daily range, and so on) which exhibit clustering and can be modeled as the product of a vector of conditionally autoregressive scale factors and a multivariate iid innovation process (vector Multiplicative Error Model).

Two novel points are introduced in this paper relative to previous suggestions: a more general specification which sets this vector MEM apart from an equation by equation specification; and the adoption of a GMM-based approach which bypasses the complicated issue of specifying a general multivariate non-negative valued innovation process.

A vMEM for volumes, number of trades and realized volatility reveals empirical support for a dynamically interdependent pattern of relationships among the variables on a number of NYSE stocks.

^{*}This paper develops some ideas introduced in Cipollini, Engle and Gallo (2006) where estimation was based in the framework of estimating functions. Without implicating, we acknowledge comments by Nour Meddahi and Kevin Sheppard which led us to present the Estimating Functions approach in a more familiar GMM notation. We acknowledge financial support from the Italian MIUR under grant PRIN 2006131140.004.

[†]Dipartimento di Statistica “G. Parenti”, Università di Firenze, Italy. e-mail: cipollin@ds.unifi.it

[‡]Department of Finance, Stern School of Business, New York University; email rengle@stern.nyu.edu

[§]Corresponding author: Dipartimento di Statistica “G. Parenti”, Università di Firenze, Viale G.B. Morgagni, 59 - 50134 Firenze – Italy. e-mail: gallog@ds.unifi.it

1 Introduction

In financial time series analysis we encounter several instances of non-negative valued processes (volumes, trades, durations, realized volatility, daily range, and so on). The Multiplicative Error Model, as introduced by Engle (2002), exploits the stylized fact that these time series share similar persistence and clustering features as the absolute or squared returns. As in the GARCH approach (Bollerslev (1986)), the conditional expectation of the variable of interest can be modeled as a time-varying function of the information set (past value of the variable and past value of the conditional expectation) and an iid (unit mean) innovation process.

The MEM is a generalization of the ACD model (Engle and Russell (1998)); other examples of MEMs are the CARR model by Chou (2005) or the model by Manganelli (2005). Engle and Gallo (2006) have specified a simultaneous MEM where three different measures of volatility (absolute returns, daily range and realized volatility) are jointly modelled introducing lagged values of all variables in each equation for the conditional expectations. Consistent parameter estimation equation by equation of such a model is very simple and is based on the assumption that the covariance matrix of the innovation term is diagonal. The empirical results show a dynamic interdependence of the indicators, suggesting that a vector approach is advisable to gain further efficiency and have more reliable interpretation of the significance of the links.

A first suggestion in order to obtain estimation in a vector MEM context is provided in Cipollini *et al.* (2008). The equation by equation approach is used as a first stage (only observable variables are involved) to reconstruct the marginals and then a copula function is used to reconstruct a plausible joint pdf for the variables of interest.

In this paper we extend the results on VMEM in two directions. The first is the specification of the vMEM in a more general form, namely one in which the conditional expectation of one variable is a function not only of its own past conditional expectation, but also of past conditional expectations of other variables. This bars the equation by equation *cum* copula approach from being used and suggests the adoption of a novel GMM approach which allows us to bypass the issue of reconstructing a multivariate non-negative valued pdf for the vector of innovations.

The structure of the paper is as follows. In Section 2 we present the Multiplicative Error Model summarizing its main features; Section 3 provides the main details on the GMM derivation of the parameter estimates and on the issue of the nuisance parameter estimation (the innovation covariance matrix). We introduce an illustration of the capability of the model in reference with three variables, volumes, trades and realized volatility.

2 Multiplicative Error Models

The Multiplicative Error Model (MEM) extends the GARCH approach to processes x_t with non-negative support (Engle (2002), Engle and Gallo (2006)). In the univariate case,

the model is specified as

$$x_t = \mu_t \varepsilon_t. \quad (1)$$

μ_t is a scale factor assumed to evolve deterministically conditionally upon \mathcal{F}_{t-1} : using a general but rather informal notation we can denote this as $\mu_t = \mu(\mathcal{F}_{t-1}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters ruling the dynamics of μ_t . ε_t is an innovation term assumed to have non-negative support, conditional mean 1 and unknown variance σ^2 : $\varepsilon_t | \mathcal{F}_{t-1} \sim D(1, \sigma^2)$.

The previous assumptions on μ_t and ε_t give

$$E(x_t | \mathcal{F}_{t-1}) = \mu_t \quad (2)$$

$$V(x_t | \mathcal{F}_{t-1}) = \sigma^2 \mu_t^2. \quad (3)$$

The Autoregressive Conditional Duration model by Engle and Russell (1998) is a special case of MEM, but absolute returns, high-low range, number of trades in a certain interval, volume, various versions of ultra-high frequency based measures of volatility can be modeled with MEMs. One of the advantages of such a model is to avoid the need to resort to logs (not possible when zeros are present in the data) and to provide conditional expectations of the variables of interest directly (rather than expectations of the logs). Empirical results show a good performance of these types of models in capturing the stylized facts of the observed series (e.g. for daily range, Chou (2005); for volume, Manganello (2005)).

However, one can be interested to model jointly the dynamics of two or more non-negative time series. Examples are: volatility forecasting using different measures; volatility spillovers for studying contagion among markets; order execution dynamics in order-driven markets; joint dynamics of duration, volume and volatility for the same asset (see Cipollini *et al.* (2007) and references therein for details about these examples). This motivates the following multivariate extension of the MEM.

Let \mathbf{x}_t be a K -dimensional process with non-negative components¹. A vector MEM (or vMEM for shortly) for \mathbf{x}_t is defined as

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t = \text{diag}(\boldsymbol{\mu}_t) \boldsymbol{\varepsilon}_t, \quad (4)$$

where \odot indicates the Hadamard (element-by-element) product and $\text{diag}(\cdot)$ denotes a diagonal matrix with the vector in the argument as main diagonal. Conditionally on the information set \mathcal{F}_{t-1} , $\boldsymbol{\mu}_t$ can be defined as above, that is

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}(\mathcal{F}_{t-1}; \boldsymbol{\theta}), \quad (5)$$

except that now we are dealing with a K -dimensional vector depending on a (larger) vector of parameters $\boldsymbol{\theta}$. The innovation vector $\boldsymbol{\varepsilon}_t$ is a K -dimensional random variable defined over a $[0, +\infty)^K$ support, with unit vector $\mathbf{1}$ as expectation and a general variance-

¹In what follows we will adopt the following conventions: if \mathbf{x} is a vector or a matrix and a is a scalar, then the expressions $\mathbf{x} \geq \mathbf{0}$ and \mathbf{x}^a are meant element by element; if $\mathbf{x}_1, \dots, \mathbf{x}_K$ are (m, n) matrices then $(\mathbf{x}_1; \dots; \mathbf{x}_K)$ means the (mK, n) matrix obtained stacking the matrices \mathbf{x}_i 's columnwise.

covariance matrix Σ ,

$$\varepsilon_t | \mathcal{F}_{t-1} \sim D(\mathbf{1}, \Sigma). \quad (6)$$

From the previous conditions follow that

$$E(\mathbf{x}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \quad (7)$$

$$V(\mathbf{x}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \boldsymbol{\mu}_t' \odot \Sigma = \text{diag}(\boldsymbol{\mu}_t) \Sigma \text{diag}(\boldsymbol{\mu}_t), \quad (8)$$

where the latter is a positive definite matrix by construction.

Both in the univariate and in the vector case, empirical applications require to complete the specification of the model. We discuss this in the vector case.

About the scale term $\boldsymbol{\mu}_t$, a sufficiently general specification can be

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \sum_{l=1}^L \left[\boldsymbol{\alpha}_l \mathbf{x}_{t-l} + \boldsymbol{\gamma}_l \mathbf{x}_{t-l}^{(-)} + \boldsymbol{\beta}_l \boldsymbol{\mu}_{t-l} \right]. \quad (9)$$

Among the parameters (whose nonzero elements are arranged in the vector $\boldsymbol{\theta}$) $\boldsymbol{\omega}$ has dimension $(K, 1)$, whereas $\boldsymbol{\alpha}_l$, $\boldsymbol{\gamma}_l$ and $\boldsymbol{\beta}_l$ have dimension (K, K) . Some of the elements of this second set of coefficients can be constrained to zero if the corresponding component is not interesting for structuring $\boldsymbol{\mu}_t$. In this sense, L denotes the maximal lag that, for at least one of these coefficients, includes non zero elements. The terms $\boldsymbol{\gamma}_l \mathbf{x}_{t-l}^{(-)}$ aim to capture asymmetric effects associated with the sign of an observed variable: the vector $\mathbf{x}_t^{(-)}$ contain $x_{t,i}$'s multiplied by a function related to a signed variable, be it a return (0, 1 values) or a signed trade (buy or sell 1, -1 values). For example, when different volatility indicators of the same asset are considered, such an indicator assumes value one when its previous day's return r_{t-1} is negative. In a market volatility spillover study, each market would have its own indicator function built from the sign its own returns $r_{t-1,i}$. Finally, in a microstructure context, we can think of assigning positive or negative values to volumes according to whether the trade was a buy or a sell.

As far as the error term ε_t , a completely parametric formulation of the vMEM requires full specification of its conditional distribution. Some proposals about this are provided in Cipollini *et al.* (2007), in which marginals for the components $\varepsilon_{t,i}$ satisfying above requirements are linked together using copulas. However, full parametric formulation of the distribution of the error component can be quite tiring for a number of reasons.

1. Multivariate distributions defined on the non-negative orthant are often not sufficiently flexible. Furthermore, sometimes they are defined via the characteristics function without an explicit pdf, thus complicating considerably parameter estimation (see, as an example, the discussion on the multivariate gamma in Cipollini *et al.* (2007) and the work of Ahoniemi and Lanne (2007)). These are the main reasons that motivate recurring to copulas.
2. For what concerns the copula approach, although there is a one-to-one correspondence between the cdf of an absolutely continuous rv and its copula representation,

practical choices are often driven by convenience reasons. For instance, Elliptical copulas are appealing because they can be employed also for moderately large K and can accommodate tail dependency. However they have an elliptically symmetric behavior that can represent a limit in some contexts. On the contrary, some Archimedean copulas can accommodate asymmetric dependency profiles but appear less usable when the dimension K tends to increase (McNeil *et al.* (2005)). As a consequence, even when combined with correct formulations of the marginals, copulas are not always able to model adequately the association among components of the error term.

3. As far as the marginals, in principle each one of them can follow a different distribution. This could require an expensive tuning of the model, in which different choices for the marginals are compared.
4. As a final point, full specification of the distribution of ε_t may be not interesting if the main focus of the analysis is on the dynamics of the μ_t component.

Hence, we assume in this work that the error component has an unspecified distribution with conditional moments as in (6), thus leading to a semiparametric specification of the vMEM.

3 Model Inference

In this section, we illustrate how to obtain inferences on the semiparametric specification of the vector MEM proposed in section 2. The parameter of main interest is that ruling the dynamics of μ_t , i.e. θ , whose dimension is denoted by p ; relative to it, the variance matrix of the error term, i.e. Σ , represents a nuisance parameter of dimension $K(K + 1)/2$.²

The key ingredient for the inference is represented by the two conditional moments of the observed variable x_t , (7) and (8), that follow from model assumptions. Even if different estimation methods, relying essentially on first and second moments, can be taken into account, in order to simplify the exposition we focus attention on Generalized Method of Moments (GMM). Connections with other estimation methods leading to comparable inferences are illustrated in appendix A.

3.1 Efficient GMM inference

Let us define

$$\mathbf{u}_t = \mathbf{x}_t \otimes \mu_t - \mathbb{1}, \quad (10)$$

²Because of the symmetry, the nuisance parameter could be better denoted with $\text{vech}(\Sigma)$, the vector obtained stacking the portion of each column up to the main diagonal included. However, in the following we indicate the nuisance parameter as Σ order to simplify the exposition.

where \oslash indicates the element-by-element division and we suppressed dependency of \mathbf{u}_t on the parameter $\boldsymbol{\theta}$, on the information \mathcal{F}_{t-1} and on the current value of the dependent variable \mathbf{x}_t .

From (7) and (8) one obtains immediately that \mathbf{u}_t is a conditionally homoskedastic martingale difference, that is

$$E(\mathbf{u}_t | \mathcal{F}_{t-1}) = \mathbf{0} \quad (11)$$

$$V(\mathbf{u}_t | \mathcal{F}_{t-1}) = \boldsymbol{\Sigma}. \quad (12)$$

Following Wooldridge (1994, sect. 7) a conditional moment restriction like (11) can be used as a key ingredient for estimation. By means of the law of iterated expectations, this equation gives rise to K moment conditions of the form $E(\mathbf{u}_t) = \mathbf{0}$, usually not enough for estimating the p -dimensional parameter $\boldsymbol{\theta}$. However, any (M, K) matrix \mathbf{G}_t depending deterministically on the information \mathcal{F}_{t-1} gives

$$E(\mathbf{G}_t \mathbf{u}_t | \mathcal{F}_{t-1}) = \mathbf{0}, \quad (13)$$

and then, by the law of iterated expectations,

$$E(\mathbf{G}_t \mathbf{u}_t) = \mathbf{0}, \quad (14)$$

so that \mathbf{G}_t is uncorrelated with \mathbf{u}_t .³ A matrix with the characteristics of \mathbf{G}_t (namely dependent on \mathcal{F}_{t-1} and uncorrelated with \mathbf{u}_t) is usually named *instrument* or *instrumental variable* and may depend on one vector of nuisance parameters $\boldsymbol{\psi}$ (assumed for the moment a known constant).

Equation (14) provides M moment conditions. If $M = p$, we have as many equations as the dimension of $\boldsymbol{\theta}$, thus leading to the MM criterion

$$\bar{\mathbf{g}} = \frac{1}{T} \sum_{t=1}^T \mathbf{g}_t = \mathbf{0} \quad (15)$$

where $\mathbf{g}_t = \mathbf{G}_t \mathbf{u}_t$.

Under correct specification of the equation arising (14) (the $\boldsymbol{\mu}_t$ equation in our case) and some regularity conditions, the GMM estimator $\widehat{\boldsymbol{\theta}}_T$, obtained solving (15) for $\boldsymbol{\theta}$, is consistent (Wooldridge (1994, th. 7.1)). Furthermore, under some additional regularity conditions we have asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$, with asymptotic variance matrix

$$\text{Avar}(\widehat{\boldsymbol{\theta}}_T) = \frac{1}{T} (\mathbf{S}' \mathbf{V}^{-1} \mathbf{S})^{-1}. \quad (16)$$

³As remarked by Wooldridge (1994, p. 2693), equation (14) requires that the absolute value of each component of \mathbf{u}_t and $\mathbf{G}_t \mathbf{u}_t$ has finite expectation(s?).

where

$$\mathbf{S} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(\nabla_{\theta'} \mathbf{g}_t) \quad (17)$$

$$\mathbf{V} = \lim_{T \rightarrow \infty} \frac{1}{T} V \left(\sum_{t=1}^T \mathbf{g}_t \right) \quad (18)$$

(Wooldridge (1994, th. 7.2)). The structure of the vector MEM permits some considerable simplifications, stemming in particular from \mathbf{u}_t being a martingale difference (see Wooldridge (1994)). In fact, such characteristic implies that $\mathbf{g}_t = \mathbf{G}_t \mathbf{u}_t$ also is a martingale difference (equation (14)): this leads to simplifications in the assumptions needed for the asymptotic normality, by virtue of the martingale CLT, and is a sufficient condition for making terms \mathbf{g}_t 's into (18) serially uncorrelated, thus leading to

$$\mathbf{V} = \lim_{T \rightarrow \infty} \left[\frac{1}{T} \sum_{t=1}^T E(\mathbf{g}_t \mathbf{g}_t') \right]. \quad (19)$$

The martingale difference structure of \mathbf{u}_t gives also a simple formulation for the *efficient* choice of the instrument \mathbf{G}_t , where efficient is meant producing the 'smallest' asymptotic variance among the GMM estimators arisen by $\bar{\mathbf{g}}$ functions structured as in (15), with $\mathbf{g}_t = \mathbf{G}_t \mathbf{u}_t$ and \mathbf{G}_t being an instrument. Such efficient choice is

$$\mathbf{G}_t^* = -E(\nabla_{\theta} \mathbf{u}_t' | \mathcal{F}_{t-1}) V(\mathbf{u}_t | \mathcal{F}_{t-1})^{-1}. \quad (20)$$

By computing $E(\mathbf{g}_t \mathbf{g}_t')$ into (19) and $E(\nabla_{\theta'} \mathbf{g}_t)$ into (17) we obtain $E(\mathbf{g}_t \mathbf{g}_t') = -E(\nabla_{\theta'} \mathbf{g}_t) = E(\mathbf{G}_t^* \Sigma \mathbf{G}_t^{*'})$, so that

$$\mathbf{V} = -\mathbf{S} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(\mathbf{G}_t^* \Sigma \mathbf{G}_t^{*'})$$

and (16) specializes as

$$\text{Avar}(\hat{\boldsymbol{\theta}}_T) = -\frac{1}{T} \mathbf{S}^{-1} = \frac{1}{T} \mathbf{V}^{-1}. \quad (21)$$

Considering the analytical structure of \mathbf{u}_t in the vector MEM (equation (10)), we have

$$\nabla_{\theta} \mathbf{u}_t' = -\nabla_{\theta} \boldsymbol{\mu}_t' \text{diag}(\boldsymbol{\mu}_t)^{-1} \text{diag}(\mathbf{u}_t + \mathbf{1}),$$

so that (20) becomes

$$\mathbf{G}_t^* = \nabla_{\theta} \boldsymbol{\mu}_t' \text{diag}(\boldsymbol{\mu}_t)^{-1} \Sigma^{-1}.$$

Replacing it into $\mathbf{g}_t = \mathbf{G}_t \mathbf{u}_t$ and this, in turn, into (15), we obtain that the GMM estimator of $\boldsymbol{\theta}$ in the vector MEM solves the MM equation

$$\frac{1}{T} \sum_{t=1}^T \nabla_{\theta} \boldsymbol{\mu}_t' [\text{diag}(\boldsymbol{\mu}_t) \Sigma \text{diag}(\boldsymbol{\mu}_t)]^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_t) = \mathbf{0} \quad (22)$$

and has asymptotic variance matrix

$$\text{Avar}(\widehat{\boldsymbol{\theta}}_T) = \frac{1}{T} \left[\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}'_t [\text{diag}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma} \text{diag}(\boldsymbol{\mu}_t)]^{-1} \nabla_{\boldsymbol{\theta}'} \boldsymbol{\mu}_t \right] \right]^{-1}. \quad (23)$$

It is interesting to remark that (22) specializes in the univariate MEM as

$$\sigma^{-2} \frac{1}{T} \sum_{t=1}^T \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}'_t \frac{x_t - \mu_t}{\mu_t^2} = \mathbf{0}, \quad (24)$$

that represents just the 1-order condition of the univariate MEM under Gamma assumption of the multiplicative error term (Engle and Gallo (2006)). The only substantial difference is the impossibility, in the vector version, of removing the variance of the error term, $\boldsymbol{\Sigma}$, from the equation. This motivates the importance of investigating the role of this nuisance parameter in making inference about $\boldsymbol{\theta}$ (section 3.2). In the appendix A we discuss some interesting connections with other estimation methods, like Estimating Functions (EFs) and Weighted Nonlinear Least Squares (WNLS).

3.2 Inference on $\boldsymbol{\Sigma}$

In section 3.1, in discussing GMM inference on $\boldsymbol{\theta}$ we admitted the possibility that the instrument \mathbf{G}_t may depend on a vector of nuisance parameters but we assumed to know it, in order to simplify matters. Now, it's time to remove this restriction and to investigate how inferences on the nuisance parameter affect those on the main parameter. This is especially important in vector MEM because the proposed efficient MM equation (22) for estimating $\boldsymbol{\theta}$ depends on the nuisance parameter $\boldsymbol{\Sigma}$.

A clear handling of the role of nuisance parameters in the context of GMM inference is provided in Newey and McFadden (1994, sect. 6). The theory of such paper is developed by using MM conditions for estimating both the main and the nuisance parameters. Considering the remarked connections between MM and EFs (appendix A), other useful references can then be found in Liang and Zeger (1995), Jørgensen and Knudsen (2004). Whatever not being referenced to dependent data, many arguments of these paper can be trivially adapted to our framework simply adjusting notation.

For investigating the role of nuisance parameters $\boldsymbol{\psi}$ in inferencing the main parameters $\boldsymbol{\theta}$, the basic idea is to stack the moment functions of both in a joint GMM estimator. Partitioning correspondingly the expectation of the derivative w.r. to parameters, \mathbf{S} , and the variance matrix, \mathbf{V} , of such stacked moment function, we can study how $\text{Avar}(\widehat{\boldsymbol{\theta}}_T)$ depends on $\widehat{\boldsymbol{\psi}}_T$

In the following, we denote as $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1; \boldsymbol{\lambda}_2) = (\boldsymbol{\theta}; \boldsymbol{\psi})$ the stacked vector of parameters, as

$$\bar{\mathbf{g}}_i = \frac{1}{T} \sum_{t=1}^T \mathbf{g}_{t,i} \quad (25)$$

the corresponding stacked moment functions and as

$$\mathbf{S}_{ij} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(\nabla_{\mathbf{x}_j} \mathbf{g}_{t,i}) \quad (26)$$

$$\mathbf{V}_{ij} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(\mathbf{g}_{t,i} \mathbf{g}'_{t,j}) \quad (27)$$

($i, j = 1, 2$) the corresponding portion of the \mathbf{S} and \mathbf{V} matrices relative to $\bar{\mathbf{g}}$.

Using this representation, Newey and McFadden (1994) and Jørgensen and Knudsen (2004) develop concepts and theories sharing many common points. One of the main differences between the two works can be explained as follows. Newey and McFadden (1994) emphasize two-step estimators: a first-step estimate of $\boldsymbol{\psi}$ is obtained by solving $\bar{\mathbf{g}}_2 = \mathbf{0}$, where $\bar{\mathbf{g}}_2$ is made to depend on $\boldsymbol{\psi}$ but not on $\boldsymbol{\theta}$; a second step estimate of $\boldsymbol{\theta}$ is then obtained by plugging the first step estimate $\hat{\boldsymbol{\psi}}_T$ into $\bar{\mathbf{g}}_1$ and then solving $\bar{\mathbf{g}}_1 = \mathbf{0}$ for $\boldsymbol{\theta}$. Since $\bar{\mathbf{g}}_2$ does not depend on $\boldsymbol{\theta}$, this framework implies $\mathbf{S}_{21} = \mathbf{0}$, an assumption not made by Jørgensen and Knudsen (2004). Of course, by means of iteration the two-step approach in Newey and McFadden (1994) can be extended to a more general iterative approach.

In general, if $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\psi}}_T$ are consistent estimators of the corresponding parameters and the regularity conditions for asymptotic normality are satisfied, then

$$\text{Avar}(\hat{\boldsymbol{\theta}}_T) = \frac{1}{T} [\mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}]_{11} \quad (28)$$

(see (16)). Using some block matrix algebra we can show that

$$\begin{aligned} [\mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}]_{11} &= [(\mathbf{S}^{-1})_{11} \mathbf{V}_{11} + (\mathbf{S}^{-1})_{12} \mathbf{V}_{21}] (\mathbf{S}^{-1})'_{11} \\ &\quad + [(\mathbf{S}^{-1})_{11} \mathbf{V}_{12} + (\mathbf{S}^{-1})_{12} \mathbf{V}_{22}] (\mathbf{S}^{-1})'_{12} \\ &= (\mathbf{S}^{-1})_{11} \left[\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T V [\mathbf{g}_{t,1} + (\mathbf{S}^{-1})_{11}^{-1} (\mathbf{S}^{-1})_{12} \mathbf{g}_{t,2}] \right] (\mathbf{S}^{-1})'_{11}, \end{aligned} \quad (29)$$

a result slightly different from Newey and McFadden (1994, th. 6.1) in order to encompass also the case $\mathbf{S}_{21} \neq \mathbf{0}$ considered by Jørgensen and Knudsen (2004).

Equation (29) shows that, in general, the moment function of the nuisance parameter, $\bar{\mathbf{g}}_2$, impacts on $\text{Avar}(\hat{\boldsymbol{\theta}}_T)$: ignoring it can lead to inconsistent standard errors for the components of $\hat{\boldsymbol{\theta}}_T$. We can easily check that a sufficient condition for ignoring the effect of $\bar{\mathbf{g}}_2$ is

$$\mathbf{S}_{12} = \mathbf{0}, \quad (30)$$

a condition named *Nuisance Parameter Insensitivity* (NPI) by Jørgensen and Knudsen (2004). In such case

$$[\mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}]_{11} = \mathbf{S}_{11}^{-1} \mathbf{V}_{11} \mathbf{S}_{11}^{-1}. \quad (31)$$

An interesting interpretation of NPI provided by Newey and McFadden (1994, th. 6.2) (but contained also in Jørgensen and Knudsen (2004, sect. 3.2), whatever in a slightly different form) is that, under certain regularity conditions, we have $\mathbf{S}_{12} \neq \mathbf{0}$ if and only if inconsistency in estimating $\boldsymbol{\psi}$ in the first step leads to inconsistency in estimating $\boldsymbol{\theta}$ in the second step. This means that $\bar{\mathbf{g}}_1$ is $\boldsymbol{\psi}$ -insensitive, if and only if an inconsistent estimate of $\boldsymbol{\psi}$ does not affect consistency of $\hat{\boldsymbol{\theta}}_T$.

Jørgensen and Knudsen (2004, th. 1) states that NPI is not only sufficient, but also necessary for (31). However, a careful inspection of assumptions for proving the necessary part (sect. 6.3 of the cited paper) reveals that this requires also $\mathbf{S}_{21} = -\mathbf{V}_{21}$. Thus, this result does not cover the case considered by Newey and McFadden (1994), in which $\mathbf{S}_{21} = \mathbf{0}$.⁴

Returning now to the vector MEM, we can check that the moment function for estimating $\boldsymbol{\theta}$ in equation (22), denoted here as $\bar{\mathbf{g}}_1$, is $\boldsymbol{\Sigma}$ -insensitive. In fact, expressing it as sum of martingale differences $\mathbf{g}_{t,1} = \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}'_t [\text{diag}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma} \text{diag}(\boldsymbol{\mu}_t)]^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_t)$, we have

$$\nabla_{\sigma_{ij}} \mathbf{g}_{t,1} = -\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}'_t \text{diag}(\boldsymbol{\mu}_t)^{-1} \boldsymbol{\Sigma}^{-1} \nabla_{\sigma_{ij}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \text{diag}(\boldsymbol{\mu}_t)^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_t),$$

whose conditional expectation is $\mathbf{0}$. By the law of iterated expectations, $\mathbf{S}_{12} = \mathbf{0}$ also. This implies that (23) is the asymptotic variance matrix of $\hat{\boldsymbol{\theta}}_T$ even when $\boldsymbol{\Sigma}$ is estimated, instead of being known as assumed in section 3.1, and can be consistently estimated by

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}_T) = \left[\sum_{t=1}^T \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}'_t [\text{diag}(\boldsymbol{\mu}_t) \hat{\boldsymbol{\Sigma}}_T \text{diag}(\boldsymbol{\mu}_t)]^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t \right]^{-1}, \quad (32)$$

where $\boldsymbol{\mu}_t$ is computed on the basis of $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\Sigma}}_T$ is a consistent estimator of $\boldsymbol{\Sigma}$. By invoking above interpretation of NPI, we note as the $\boldsymbol{\Sigma}$ -insensitivity of $\bar{\mathbf{g}}_1$ implies that by using the identity matrix in place of $\hat{\boldsymbol{\Sigma}}_T$ (equivalent to equation by equation estimation of the vector MEM) we likewise have consistency (even if not efficiency) of the corresponding estimator $\hat{\boldsymbol{\theta}}_T$.

Finally, equation (12) suggests that a natural estimator for the nuisance parameter $\boldsymbol{\Sigma}$ can be

$$\hat{\boldsymbol{\Sigma}}_T = \frac{1}{T} \sum_{t=1}^T \mathbf{u}_t \mathbf{u}'_t \quad (33)$$

where \mathbf{u}_t is the working residual (10) computed by using current values of $\hat{\boldsymbol{\theta}}_T$. An interesting characteristic of such estimator, is that it is not compromised by zeros in the

⁴In fact, using block matrix algebra we can check that (31) is equivalent to $[\mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}]_{12} = \mathbf{0}$. On the other side, this equation can be written as $(\mathbf{S}'_{11} - \mathbf{S}'_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21})(\mathbf{V}^{-1})_{11} \mathbf{S}_{12} + (\mathbf{S}'_{21} - \mathbf{S}'_{11} \mathbf{V}_{11}^{-1} \mathbf{V}_{12})(\mathbf{V}^{-1})_{22} \mathbf{S}_{22} = \mathbf{0}$, so that:

- when $\mathbf{S}_{21} = -\mathbf{V}_{21}$ and $\bar{\mathbf{g}}_1$ is in standardized form (i.e. $\mathbf{S}_{11} = -\mathbf{V}_{11}$), (31) implies $\mathbf{S}_{12} = \mathbf{0}$ (it is the Jørgensen and Knudsen (2004) case; we note that this proof is different from that of their lemma 2);
- when $\mathbf{S}_{21} = \mathbf{0}$, (31) implies $\mathbf{S}_{12} = (\mathbf{V}^{-1})_{11}^{-1} \mathbf{V}_{11}^{-1} \mathbf{V}_{12} (\mathbf{V}^{-1})_{22} \mathbf{S}_{22}$ (it is the Newey and McFadden (1994) case).

data.

4 Interdependence in Trading Activity

After the consolidated success of GARCH models, a new momentum is being added to the analysis and forecasting of financial asset volatility by more accurate measurement methods made possible by the availability of ultra-high frequency data. The contributions in the literature on realized volatility (cf. Andersen *et al.* (2003)) have allowed the derivation of daily time series of volatility which are less noisy than the corresponding squared daily returns.

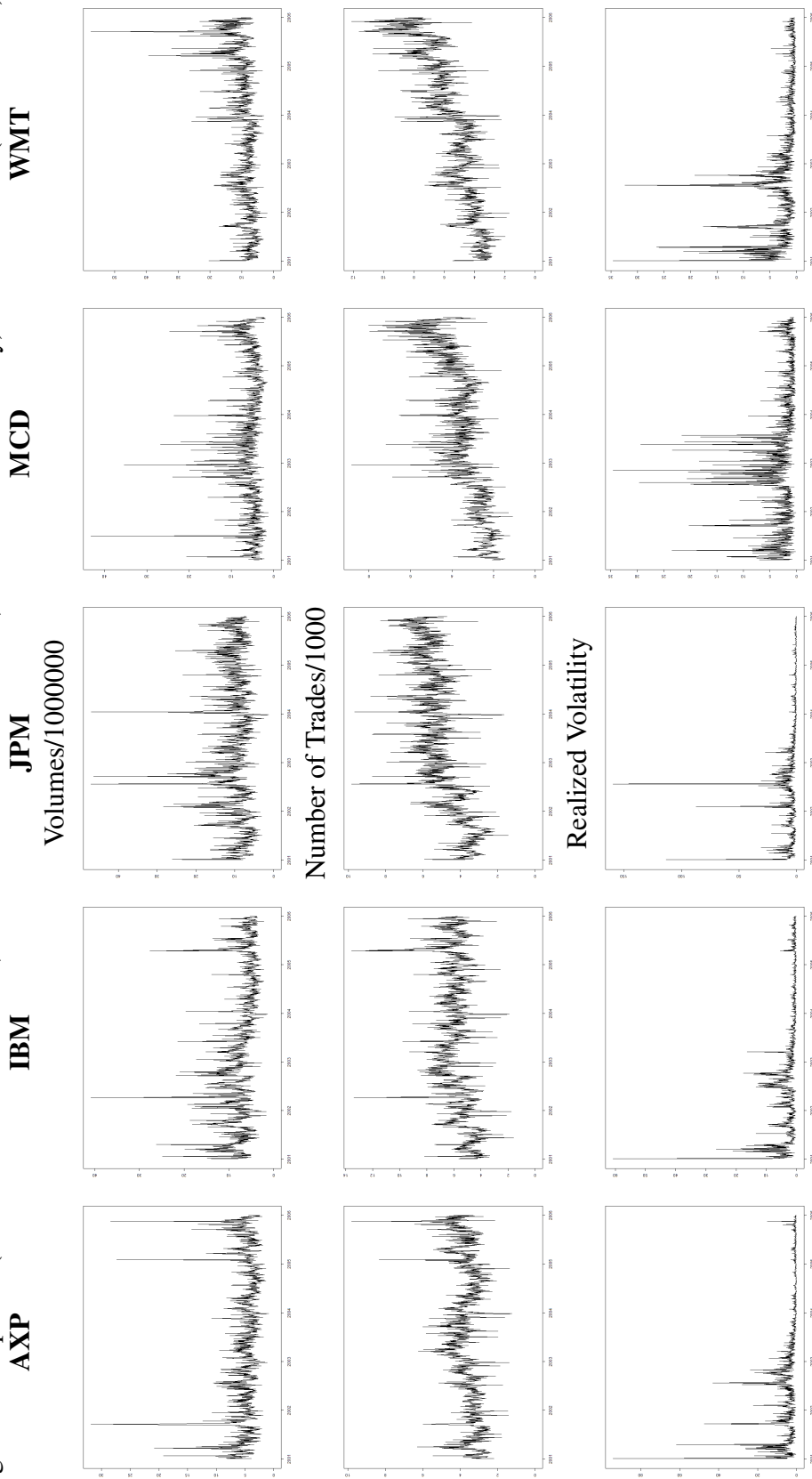
In a modelling framework, it is of interest to investigate whether volatility forecasts can be improved upon by inserting other variables available in the relevant information set. Since the seminal paper by Clark (1973), it has been recognized that stochastic volatility can arise as a consequence of the stochastic number of intra-daily number of trades, and ultimately, the information flow is considered a latent factor which manifests itself as volatility, volume, number of trades, liquidity, and so on. A number of papers deal with the interrelationships among these variables both from a theoretical and an empirical point of view. Without recalling the details of the theoretical contributions, we refer to Andersen (1996) for a microstructure-based model which accounts for the presence of informed and uninformed traders and provides the ground for an econometric model which jointly specifies the dynamics of returns and volumes conditional on an autoregressive process for the flow of information arrival.

In what follows we will illustrate the estimation and model selection procedures for the semiparametric vector MEM in reference to the joint dynamics of three variables measured on single stocks and related to market activity: volumes (vm), number of trades (tr) and realized volatility (rv). In so doing, we adopt an empirically motivated approach without an explicit theoretical model linking the three variables (for a similar analysis on volumes, durations and squared returns with a different model, see Manganelli (2005)).

From the New York Stock Exchange Trades and Quotes (TAQ) database, we constructed the series relative to the tickers American Express (AXP), IBM, J.P. Morgan (JPM), McDonald's (MCD) and Walmart (WMT) during the period 01/02/2001 - 12/30/2005 (1256 observations). In order to avoid numerical problems, we rescaled vm and tr by expressing them, respectively, in million shares and thousand trades; realized volatility is not annualized. The graphs of the variables are reported in Figure 1.

Table 1 reports a number of descriptive statistics and diagnostics on the original series. We start by reporting the minimum, maximum, mean and standard deviation (sd) of the series grouped by ticker. By and large the series exhibit range of values that are comparable with one another, with JPM having some very high values of realized volatility which increase the sample mean. The Ljung-Box statistics computed at various lags show that the autocorrelation in the variables is very high (p-values uniformly below 10^{-5}) and clustering is an empirical feature of the data to be taken into consideration. Finally, we report

Figure 1: Graphs of (vm = volumes / 1000000; tr = number of trades / 1000; rv = realized volatility) for five tickers (2001 - 2005).



the unconditional correlations (both Pearson's and Spearman's) which show that generally the volumes have a high correlation with the number of trades and realized volatility, while the latter two have a correspondingly lower (and in the case of Spearman coefficients, even negative) correlation. The period analyzed is characterized by a decrease of volatility and an increase in the number of trades (with a lower average volume per trade) which accounts for this phenomenon.

We estimated the vMEM on each ticker, starting from a full specification (in the sense outlined below) and removing each time the coefficient with the smallest and not significant (at 5% significance level) Student's t test statistic. This general-to-specific specification search ends up retaining only significant coefficients. As the starting estimated model, we have assumed a vMEM with all coefficients at lag 1 (i.e. $\alpha_{i,j,1}$, $\gamma_{i,j,1}$, $\beta_{i,j,1}$ for $i, j = 1, 2, 3$) and a diagonal lag-2 matrix α_2 , included to capture further significant serial correlation. During the general-to-specific model selection, the elements on the main diagonals of α_1 and β_1 were always kept in the specification.

We report the coefficients of the models selected in Table 2. The format of the table was chosen as to have four matrices reported by each ticker (α_1 , γ_1 , β_1 and α_2). The coefficient corresponds to the estimated effect of the variable indicated in the column (at time $t-1$) on the variable indicated in the row (at time t); below each value, in parentheses, we report the associated Student's t test statistic.

We can initially comment on the fact that the main innovation in this version of vMEM, namely the nondiagonality of β_1 , proves to be useful since there are many significant (albeit small) coefficients. Asymmetric effects are present (mainly for the coefficients of lagged realized volatility on itself and other variables).

The estimation results are complemented by a number of diagnostic test and additional information on the specification of the model (Table 3). In it we report the p-values of the Ljung-Box tests for the whole three-variable system at various lags, showing the generalized improvement of the autocorrelation in the vMEM residuals (some minor problems are still present for JPM, MCD, and WMT).

The second set of results in Table 3 pertain to the Wald test statistics computed on coefficients restrictions in a spirit very similar to Granger causality tests. The first row reports the p-values of such tests for the null hypothesis that all off-diagonal coefficients are jointly equal to zero. All tests are highly significant confirming the relevance of the interaction. Moreover, these links maintain their significance even when tested in a pairwise fashion (a total of $3 * 2 = 6$ pairs). When a '-' is present the effect turned out to be not significant at the estimation stage. Results vary from ticker to ticker: no feedbacks are present from realized volatility to the other variables for AXP; full interdependence among the three variables is exhibited by IBM only; JPM has all links significant, with the exception of volume to number of trades; MCD shows some peculiar behavior, whereby the only significant link is between lagged realized volatility and current number of trades; finally WMT has interdependence between volumes and trades and unilateral influence of realized volatility on the other two variables.

There is high persistence in the estimated system of equations as represented by the 2

Table 1: Descriptive statistics of (vm = volumes / 1000000; tr = number of trades / 1000; rv = realized volatility) for five tickers (2001 - 2005).

	AXP			IBM			JPM			MCD			WMT		
	vm	tr	rv	vm	tr	rv	vm	tr	rv	vm	tr	rv	vm	tr	rv
min	0.894	1.231	0.053	1.413	1.597	0.024	1.347	1.451	0.066	1.285	1.092	0.130	2.038	1.707	0.076
max	31.868	9.805	94.202	40.888	13.574	60.754	47.039	9.834	159.345	43.257	8.840	34.555	57.457	12.110	34.644
mean	4.903	3.672	3.015	7.236	5.843	2.014	9.834	5.067	3.797	5.622	3.478	2.543	9.394	5.408	2.081
sd	2.403	0.753	5.300	3.356	1.171	3.151	4.232	1.188	8.973	3.148	1.076	3.194	4.088	1.660	2.884
Ljung-Box statistics															
lag 12	1280.7	2706.2	1643.5	2315.4	2736.4	2292.7	1264.4	5691.3	1047.7	549.7	6344.0	1013.0	2242.0	8763.7	2839.2
lag 22	1416.2	3761.4	2300.8	3021.8	3310.0	3335.8	1424.6	9148.6	1357.2	681.4	10586.1	1501.0	3103.6	14801.0	4113.6
lag 32	1434.3	4265.8	2682.5	3383.0	3573.7	3993.7	1470.4	11954.9	1526.8	899.6	14788.6	1970.8	3560.9	19905.6	4849.6
Pearson correlations															
vm	0.576	0.463		0.580	0.518	0.510	0.592	0.510	0.499	0.557	0.499	0.752	0.090		
tr		0.131			0.122	0.104		0.104	0.132		0.132		-0.167		
Spearman correlations															
vm	0.515	0.485		0.496	0.707	0.265	0.609	0.265	0.408	0.541	0.408	0.738	0.003		
tr		0.146			0.189	-0.196		-0.196	-0.002		-0.002		-0.348		

largest eigenvalues of the companion matrix associated with the estimated coefficients. Finally, the correlation coefficients computed on the residuals are fairly similar across tickers and so are their estimated standard deviations.

5 Concluding Remarks

In this paper we have presented an extension to the specification of the vector Multiplicative Error Model introduced by Cipollini *et al.* (2006), specifying all matrices of coefficients to be full matrices. In such a specification, the approach proposed in Cipollini *et al.* (2007) to estimate the system with univariate model estimation adopting Gamma marginals, and then using copula functions to retrieve the contemporaneous correlation among innovations would not be applicable. The non-diagonality of β_1 is shown in the empirical applications to be supported by the data. Moreover, the semiparametric approach we have followed here relieves the analysis from the need to specify any distributional assumption or specific choices of a copula function.

The empirical application that is presented in this paper is merely illustrative of the potential importance of the method. We have chosen to estimate a three-dimensional model on variables that are related to the information flow and end up signalling different aspects of market activity. We believe that the dynamic interdependence that is exhibited by the results points to the importance of modelling all available information about volatility in a joint fashion.

Table 2: Estimated coefficients and t statistics (within brackets) of the selected vMEM applied to (vm = volumes / 1000000; tr = number of trades / 1000; rv = realized volatility) of five tickers (2001 - 2005).

	α_1			γ_1			β_1			α_2		
	vm_{t-1}	tr_{t-1}	rv_{t-1}	vm_{t-1}	tr_{t-1}	rv_{t-1}	vm_{t-1}	tr_{t-1}	rv_{t-1}	vm_{t-1}	tr_{t-1}	rv_{t-1}
AXP	vm_t	0.3769 (13.1)	-	-	-	-	0.7429 (18.89)	-	-	-0.1727 (-4.35)	-	-
	tr_t	0.0206 (3.53)	0.2818 (11.46)	-	-	-	-0.0301 (-3.94)	0.8818 (52.23)	-	-	-0.1704 (-5.88)	-
	rv_t	0.009 (2.04)	-	0.1031 (3.78)	-0.0233 (-2.98)	0.1303 (5.69)	-	-	0.8234 (46.32)	-	-	-
IBM	vm_t	0.3203 (11.32)	0.2235 (4.54)	0.0457 (3.24)	-	-	0.8478 (43.64)	-0.2781 (-4.68)	-	-0.2193 (-6.75)	-	-
	tr_t	-	0.4025 (17.23)	-	-	-	-0.0244 (-3.21)	0.7177 (19.64)	0.0227 (2.62)	-	-0.1501 (-4.25)	-
	rv_t	-	-	0.14 (5.34)	0.0213 (2.68)	-0.0222 (-2.88)	0.0822 (3.49)	-	-0.0164 (-2.35)	0.7929 (38.92)	-	-
JPM	vm_t	0.3493 (13.09)	0.0839 (3.12)	-	-	0.0375 (3.39)	0.8104 (25.83)	-	-	-0.2342 (-6.89)	-	-
	tr_t	-	0.3785 (17.16)	-	-	0.0042 (2.72)	-0.007 (-2.42)	0.8522 (50.28)	-	-	-0.2252 (-7.71)	-
	rv_t	-	-	0.1732 (6.74)	-	0.0769 (3.15)	-	-0.0192 (-2.06)	0.7794 (35.46)	-	-	-
MCD	vm_t	0.3733 (11.64)	-	-	-	-	0.9218 (46.14)	-	-	-0.307 (-8.31)	-	-
	tr_t	-	0.4338 (19.12)	-0.0057 (-2.32)	-	0.0071 (5.45)	-	0.8915 (60.28)	-	-	-0.3267 (-11.57)	-
	rv_t	-	-	0.2652 (7.9)	-	-	-	-	0.9218 (66.84)	-	-	-0.191 (-5.14)
WMT	vm_t	0.3541 (11.47)	-	0.0802 (3.33)	-	-0.0611 (-3.55)	0.6057 (8.61)	0.2786 (4.27)	-	-0.1508 (-3.53)	-	-
	tr_t	0.0285 (3.98)	0.2878 (10.79)	-	-	-	-0.0657 (-4.51)	0.9383 (39.12)	0.0101 (2.64)	-	-0.1731 (-5.67)	-
	rv_t	-	-	0.2165 (10.68)	-	0.0749 (2.52)	-	-	0.8399 (37.24)	-	-	-0.1044 (-2.95)

Table 3: Selected diagnostics of the selected vMEM applied to (vm = volumes / 1000000; tr = number of trades / 1000; rv = realized volatility) of five tickers (2001 - 2005).

Statistics	AXP	IBM	JPM	MCD	WMT
Ljung–Box Test (p-values)					
lag 12	0.0766	0.1451	0.0071	0.1997	0.0130
lag 22	0.1016	0.1498	0.0336	0.0234	0.0908
lag 32	0.1545	0.1657	0.0503	0.1366	0.1393
Restrictions (p-values)					
off diagonal = 0	0.0001	0.0000	0.0025	0.0000	0.0000
$vm_{t-1} \leftrightarrow tr_t$	0.0004	0.0013	0.0156	-	0.0000
$tr_{t-1} \leftrightarrow vm_t$	-	0.0000	0.0018	-	0.0000
$vm_{t-1} \leftrightarrow rv_t$	0.0413	0.0073	-	-	-
$rv_{t-1} \leftrightarrow vm_t$	-	0.0012	0.0007	-	0.0005
$tr_{t-1} \leftrightarrow rv_t$	0.0029	0.0008	0.0393	-	-
$rv_{t-1} \leftrightarrow tr_t$	-	0.0089	0.0065	0.0000	0.0082
Companion matrix eigenvalues					
largest	0.9917	0.9971	0.9944	0.9978	0.9976
2-nd largest	0.9916	0.9722	0.9926	0.9950	0.9882
Residuals statistics					
$\hat{\sigma}_{vm}$	0.1416	0.1075	0.1193	0.2688	0.1018
$\hat{\sigma}_{tr}$	0.0224	0.0213	0.0206	0.0306	0.0211
$\hat{\sigma}_{rv}$	0.7022	0.4437	0.7882	0.8079	0.4692
$\hat{\rho}_{vm,tr}$	0.6927	0.7996	0.7073	0.6728	0.7255
$\hat{\rho}_{vm,rv}$	0.5321	0.4904	0.5390	0.4972	0.3486
$\hat{\rho}_{tr,rv}$	0.4851	0.4700	0.4975	0.5411	0.3556

References

- Ahoniemi, K. and Lanne, M. (2007). Joint modeling of call and put implied volatility. Discussion paper 198, HECER - Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI00014 University of Helsinki, FINLAND.
- Andersen, T. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, **51**, 169–204.
- Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71**, 529–626.
- Bera, A. K. and Biliias, Y. (2002). The mm, me, ml, el, ef and gmm approaches to estimation: a synthesis. *Journal of Econometrics*, **107**, 51–86.
- Bera, A. K., Biliias, Y., and Simlai, P. (2006). Estimating functions and equations: An essay on historical developments with applications to econometrics. In T. C. Mills and K. Patterson, editors, *Palgrave Handbook of Econometrics*, volume 1. Palgrave MacMillan.
- Bibby, B. M., Jacobsen, M., and Sørensen, M. (2004). Estimating functions for discretely sampled diffusion-type models. Working paper, University of Copenhagen, Department of Applied Mathematics and Statistics.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327.
- Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (carr) model. *Journal of Money, Credit and Banking*, **37**(3), 561–582.
- Cipollini, F., Engle, R. F., and Gallo, G. M. (2006). Vector multiplicative error models: Representation and inference. Working Paper Series 12690, National Bureau of Economic Research, Inc.
- Cipollini, F., Engle, R. F., and Gallo, G. M. (2007). A model for multivariate positive valued processes in financial econometrics. Working Paper 2007/16, Università di Firenze, Dipartimento di Statistica.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, **41**, 135–155.
- Durbin, J. (1960). Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society, Series B*, **22**, 139–153.
- Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, **17**, 425–446.
- Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, **131**(1–2), 3–27.

- Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127–62.
- Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, **31**, 1208–1212.
- Heyde, C. (1997). *Quasi-Likelihood and Its Applications*. Springer-Verlag, New York.
- Jørgensen, B. and Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, **31**, 93–114.
- Liang, K.-Y. and Zeger, S. L. (1995). Inference based on estimating functions in presence of nuisance parameters (with discussion). *Statistical Science*, **10**, 158–199.
- Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, **8**, 377–399.
- McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, Princeton.
- Newey, W. K. and McFadden, D. (1994). *Large sample estimation and hypothesis testing*, chapter 36, pages 2111–2245. Elsevier B.V., doi:10.1016/S1573-4412(05)80005-4.
- Wooldridge, J. M. (1994). *Estimation and inference for dependent processes*, chapter 45, pages 2639–2738. Elsevier B.V., doi:10.1016/S1573-4412(05)80014-5.

A Connections with other inferential methods

As stated in section 3.1, GMM inferences about θ that are efficient in some sense can be derived by solving equation (22). However, the same condition (and, by consequence, the same inferences) can also be obtained by means of other inferential methods.

The first one is *Estimating Functions* (EFs). Proposed originally by Godambe (1960) and Durbin (1960) in a more empirical context, EFs have gained a considerable popularity in some applied domains, as some Biometrics models (see Liang and Zeger (1995) and references therein) and some Stochastic Processes models (see Bibby *et al.* (2004) and references therein). Despite this, EFs have remained relatively unknown to econometricians, at least until the recent surveys of Bera and Biliias (2002) and Bera *et al.* (2006). An excellent reference about EFs for dependent data is provided by Heyde (1997).

Adapting notation to the framework of this paper, an *estimating function* for a p -dimensional vector θ based on a sample $\mathbf{x}_{(T)}$ is a p -dimensional function denoted as

$$\bar{\mathbf{g}}(\theta; \mathbf{x}_{(T)}) \quad \text{in short} \quad \bar{\mathbf{g}}.$$

The EF estimator $\hat{\theta}_T$ is defined as the solution to the corresponding *estimating equation*:

$$\hat{\theta}_T \quad \text{such that} \quad \bar{\mathbf{g}} = \mathbf{0}. \quad (34)$$

To be a useful estimating function, regularity conditions on $\bar{\mathbf{g}}$ are usually imposed. Among these, a central requirement is the $\mathbf{0}$ -unbiasedness of $\bar{\mathbf{g}}$, that is $E(\bar{\mathbf{g}}) = \mathbf{0}$, an assumption that looks like a MM condition. In applying EFs to time series models, when a martingale difference \mathbf{u}_t depending on θ is available a frequently considered class of EFs is

$$\left\{ \bar{\mathbf{g}} : \bar{\mathbf{g}} = \frac{1}{T} \sum_{t=1}^T \mathbf{G}_t \mathbf{u}_t \right\}, \quad (35)$$

where \mathbf{G}_t are (p, K) matrix functions defined exactly as in section 3.1. Writing down the corresponding estimating equation, it can be easily checked that the EFs approach leads exactly to the MM condition (15), hence to the same asymptotic variance of the estimator that, in turn, is 'minimized' by the same efficient instrument (20). In the EF's literature, the efficient EF in the class (35), is sometimes named *Hutton–Nelson quasi score function*.⁵

A second inferential approach that leads to the same inferential results is Quasi Maximum Likelihood (QML) in its Weighted Nonlinear Least Squares (WNLS) declination (Wooldridge (1994, sect. 6)). Such an approach can be usefully employed when the main interest lies in estimating the parameters of the conditional mean of \mathbf{x}_t (μ_t in the vector

⁵A careful comparison of Newey and McFadden (1994, sect. 5) and Heyde (1997, ch. 2) suggests that connections among GMM and EFs are not restricted to formulations considered in (35) and (15). In particular, the optimality theorem in Heyde (1997, p. 29) looks very similar to theorem 5.3 in Newey and McFadden (1994).

MEM) assumed correctly specified. The WNLS estimator of θ solves

$$\min_{\theta} \frac{1}{2T} \sum_{t=1}^T (\mathbf{x}_t - \boldsymbol{\mu}_t)' \widehat{\mathbf{W}}_t^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_t), \quad (36)$$

where \mathbf{W}_t is a (K, K) symmetric, positive definite matrix depending deterministically on \mathcal{F}_{t-1} and, possibly, on nuisance parameters (i.e. exactly as the \mathbf{G}_t matrix of section 3.1). Since, under conditions (7) and (8), the optimal WNLS estimator of θ is obtained exactly by choosing $\mathbf{W}_t = V(\mathbf{x}_t | \mathcal{F}_{t-1})$ (Wooldridge (1994, p. 2700)), \mathbf{W}_t is thought to be a consistent estimator of $V(\mathbf{x}_t | \mathcal{F}_{t-1})$ based on the current values of the parameters. Deriving from (36) the first order condition for $\widehat{\theta}_T$ and replacing \mathbf{W}_t with (8), as indicated above, we obtain equation (22).

Wooldridge (1994) describes also a different declination of QML. Such a different approach can be used whenever, as in the vector MEM, both the mean and the variance function are jointly parameterized by a vector θ . Under correct specification of these two conditional moments, a QML estimator can be obtained under the nominal assumption of $\mathbf{x}_t | \mathcal{F}_{t-1}$ normally distributed with first two moments as in (7) and (8). Such an approach leads to different inferences about θ and we do not pursue further this here.