Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

WORKING PAPER 2009/02

Automated Variable Selection in Vector Multiplicative Error Models

Fabrizio Cipollini,
Giampiero M. Gallo

Università degli Studi
di Firenze

Automated Variable Selection in Vector Multiplicative Error Models

Fabrizio Cipollini[1] **Giampiero M. Gallo**[2]

Abstract

Multiplicative Error Models (MEM) can be used to trace the dynamics of non–negative valued processes. Interactions between several such processes are accommodated by the vector MEM and estimated by maximum likelihood (Gamma marginals with copula functions) or by Generalized Method of Moments. In choosing the relevant variables one can follow an automated procedure where the full specification is successively pruned in a general–to–specific approach. An efficient and fast algorithm is presented in this paper and evaluated by means of a simulation and a real world example of volatility spillovers in European markets.
Key words: MEM, Model Selection, Analytic Derivatives, GMM, Copulas.

JEL: C22, C51, C52, C53

1. Introduction

Multiplicative Error Models (MEM) for non-negative processes have experienced a series of recent developments. From the seminal paper by Engle [12], several authors have explored the potential of this class of models in a number of applications (Manganelli [20], Chou [7], Engle and Gallo [13], Ahoniemi and Lanne [1]). The applications in financial econometrics are far reaching and include the modelling and forecasting of a variety of high-frequency phenomena (durations, realized volatilities, ranges, trading intensities, volumes, and so on).

More recently, Cipollini et al. [8] and Cipollini et al. [9] have extended this class of models to a joint specification which tackles the issues of the possible contemporaneous correlation among the innovations and of possible dynamic interdependencies among the conditional expectations (as will be clarified later). In Cipollini et al. [8], a parametric approach is pursued where the unknown joint probability density function is specified through Gamma marginal probability density functions and a copula function (Normal or Student-T); Cipollini et al. [9] suggest a semiparametric specification which avoids the choice of a density function for the innovations altogether and exploits an estimation strategy based on the Generalized Method of Moments.

Irrespective of the estimation procedure chosen (in some cases they deliver approximately the same results), one of the aspects of interest in specifying a vector MEM is to trace the dynamic interdependencies among variables. Model selection techniques (Brownlees and Gallo [5] in the univariate case) allow to detect the importance of predetermined variables which may enrich the dynamic speci-

fication. Examples are the interactions among various measures of volatility for the same asset, or among volatility measured on different markets. Some coefficients in the model matrices may be zero and thus correspond to the absence of a dynamic link from the corresponding lagged variable to the current one. The approach we explore is to check which coefficients are not statistically significant starting from a full specification of the vMEM. Since the computational burden is potentially high, it is paramount to adopt the necessary steps for a fast and reliable estimation. Model selection issues can be considered within the framework of an automated procedure to identify relevant predetermined variables without a user's intervention, in a general to specific approach. In this paper we analyze some of the computational aspects connected with the existing procedures with the goal of highlighting certain choices which turn out to be crucial for these nonlinear models especially when they are estimated on large sets of data. Analytical derivatives and moment conditions are derived in order to speed up calculations. The overall goal is to illustrate such a procedure and what allows a fast and accurate calculation of the results.

The paper is organized as follows. Section 2 introduces the model and the notation adopted throughout the paper. Section 3 describes the inferential procedures, with a special emphasis on computational details. Section 4 discusses the issue of the automated selection of variables within the model specification. A simple simulation exercise investigates the capability of the algorithm to detect the correct model. We adopt realistic parameter values and we show that the size of the coefficients is crucial in managing to tell them apart from zero. Section 5 contains

an empirical application with four volatility measures from European markets between 2000 and 2008 in order to estimate the spillover effects from one market to another. Section 6 concludes. In the Appendix we discuss the software implementation in **R**.

2. The vector Multiplicative Error Model

The Multiplicative Error Model (MEM) extends the GARCH approach to processes x_t with non-negative support (Engle [12], Engle and Gallo [13]). The vector Multiplicative Error Model (vMEM) generalizes the univariate MEM to situations in which the process under analysis is a vector $\mathbf{x}_t$ whose components are non-negative (Cipollini et al. [8], Cipollini et al. [9]).

Let us assume the following: $\mathbf{x}_t$ is a K–dimensional process with non–negative components and follows a vMEM if

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t = \mathrm{diag}(\boldsymbol{\mu}_t)\boldsymbol{\varepsilon}_t, \tag{1}$$

where $\odot$ indicates the Hadamard (element–by–element) product, $\mathrm{diag}(.)$ arranges the elements of a vector as an argument into a diagonal matrix, $\mathcal{F}_{t-1}$ represents the information available for forecasting $\mathbf{x}_t$.[3]

$\boldsymbol{\mu}_t$ is a K-dimensional scale factor assumed to evolve in a deterministic fashion

[3] In 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 $\mathbf{x}_1, \ldots, \mathbf{x}_K$ are (m, n) matrices then $(\mathbf{x}_1; \ldots; \mathbf{x}_K)$ denotes the (mK, n) matrix obtained stacking the $\mathbf{x}_i$'s columnwise and $(\mathbf{x}_1, \ldots, \mathbf{x}_K)$ indicates the (m, nK) matrix obtained linking together the $\mathbf{x}_i$'s rowwise.

conditional upon $\mathcal{F}_{t-1}$: using a general but rather informal notation we can denote this as

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}(\mathcal{F}_{t-1}; \boldsymbol{\theta}), \tag{2}$$

where $\boldsymbol{\theta}$ is a vector of unknown parameters ruling the dynamics of $\boldsymbol{\mu}_t$.

$\boldsymbol{\varepsilon}_t$ is a K–dimensional random innovation term, defined over a $[0, +\infty)^K$ support, following a multivariate distribution D with unit vector $\mathbb{1}$ as expectation and a general variance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1} \sim D(\mathbb{1}, \boldsymbol{\Sigma}). \tag{3}$$

Assumptions (1), (2) and (3) imply that

$$E(\mathbf{x}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \tag{4}$$

$$V(\mathbf{x}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \boldsymbol{\mu}_t' \odot \boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma} \operatorname{diag}(\boldsymbol{\mu}_t), \tag{5}$$

where the latter is a positive definite matrix by construction.

Empirical applications will require to complete the specification of the model with a number of choices detailed in what follows.

2.1. The Scale Factor

A sufficiently general specification for the scale factor can be the following

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \sum_{l=1}^{L} \left[\boldsymbol{\alpha}_l \mathbf{x}_{t-l} + \boldsymbol{\gamma}_l \mathbf{x}_{t-l}^{(-)} + \boldsymbol{\beta}_l \boldsymbol{\mu}_{t-l} \right]. \tag{6}$$

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 does not influence the evolution of $\boldsymbol{\mu}_t$. In this sense, L denotes the maximal lag that corresponds to non-zero elements in at least one of these coefficient matrices. The terms $\boldsymbol{\gamma}_l \mathbf{x}_{t-l}^{(-)}$ aim at capturing possible 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, for example, a return ($0, 1$ values) or a signed trade (buy or sell $1, -1$ values) (see Cipollini et al. [9] for details).

2.2. The Error Term

The conditional distribution of the error term $\boldsymbol{\varepsilon}_t$ can be introduced directly, through a full parametric formulation; alternatively, one can adopt a semiparametric specification based on the first two moments.

2.2.1. The Parametric Formulation Based on Copulas

An approach for specifying the distribution of $\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}$ is to employ *copula functions*.[4] This approach can be seen as modular, since we can decompose the problem of specifying the distribution of a multivariate random variable into:

- the choice of the univariate marginal functions;

[4]The main characteristics of copulas are summarized, among others, in Joe [17] and Nelsen [22]. See also Embrechts et al. [11], Cherubini et al. [6], McNeil et al. [21] and the review of Patton [23] for financial applications.

- the choice of the copula function linking them,

and therefore express the conditional distribution of the error component of the vector MEM as

$$\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim C(\boldsymbol{\xi}) \times \prod_{i=1}^{K} M_i(\boldsymbol{\phi}_i), \tag{7}$$

where: $C(\boldsymbol{\xi})$ denotes a copula parameterized by a vector $\boldsymbol{\xi}$ and with density $c(\mathbf{u}; \boldsymbol{\xi})$; $M_i(\boldsymbol{\phi}_i)$ indicates the distribution of the i-th marginal (again assumed absolutely continuous, with non-negative support and unit expectation), having probability density function $f_i(x; \boldsymbol{\phi}_i)$ and cumulative density function $F_i(x; \boldsymbol{\phi}_i)$. The conditional probability density function of $\boldsymbol{\varepsilon}_t$ can then be written as

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = c(\mathbf{u}_t; \boldsymbol{\xi}) \prod_{i=1}^{K} f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i),$$

where $\mathbf{u}_t = (u_{t,1}; \ldots; u_{t,K})$ and $u_{t,i} = F_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)$. Specific choices of the univariate marginals and of the copula function are needed in empirical applications. We can consider as natural candidates for the marginal distributions all probability density functions with the characteristics mentioned above: examples are Gamma, Inverse-Gamma, Weibull, Lognormal, mixtures of them. For practical reasons, in the following we will make some references to the Gamma case (cf. the discussion on the flexibility of this choice in Engle and Gallo [13]), but the approach can be adapted to other choices. Restrictions on the parameters have to be imposed in order to satisfy the unit mean constraint in (3). For instance, if a gamma distribution is assumed for the i-th marginal then $\varepsilon_{t,i}|\mathcal{F}_{t-1} \sim Gamma(\phi_i, \phi_i)$ and its

probability density function is given by

$$f_i(\varepsilon_{t,i};\phi_i) = \frac{\phi_i^{\phi_i}}{\Gamma(\phi_i)}\varepsilon_{t,i}^{\phi_i-1}\exp(-\phi_i\varepsilon_{t,i}). \tag{8}$$

A convenient choice for copulas can be represented by some member within the families of *Archimedean* or *Elliptical* copulas. Archimedean copulas (a family including, among others, Frank, Gumbel and Clayton copulas) can model very different dependence profiles and represent a flexible solution for two-dimensional rv's. However, they appear less useful when the dimension K tends to increase (McNeil et al. [21]) On the contrary, Elliptical copulas (a family including, among others, Normal and Student-T copulas) can be employed also for moderately large K and can accommodate tail dependency. However they have an elliptically symmetric behavior that can represent a limitation in some contexts. For practical reasons, we discuss some possible specifications of $C(\boldsymbol{\xi})$ related to the family of Elliptical copulas, but the approach can be easily extended to different choices.
Let us consider a copula generated by an Elliptical distribution whose univariate 'standardized' marginals (intended here with location parameter 0 and dispersion parameter 1) have an absolutely continuous symmetric distribution, centered at zero, with probability density function $g(.;\boldsymbol{\nu})$ and cumulative density function $G(.;\boldsymbol{\nu})$ ($\boldsymbol{\nu}$ represents a vector of shape parameters). The density of the copula can be written as

$$cE(\mathbf{u};\mathbf{R},\boldsymbol{\nu}) = K^*(\boldsymbol{\nu},K)|\mathbf{R}|^{-1/2}\frac{g_1(\mathbf{q}'\mathbf{R}^{-1}\mathbf{q};\boldsymbol{\nu},K)}{\prod_{i=1}^K g_2(q_i;\boldsymbol{\nu})} \tag{9}$$

for suitable choices of $K^*(.,.)$, $g_1(.;.,.)$ and $g_2(.;.)$, where $\mathbf{q} = (q_1; \ldots; q_K)$, $q_i = G^{-1}(u_i; \boldsymbol{\nu})$. Let us discuss two important particular cases in more detail.[5]

The Normal copula, perhaps the most popular member of such family and a frequent choice in applications (see McNeil et al. [21], Cherubini et al. [6], Bouyé et al. [3]), is obtained from (9) by choosing $K^*(K) = 1$, $g_1(x; K) = g_2(x) = \exp(-x/2)$, so that its probability density function is given by

$$cN(\mathbf{u}; \mathbf{R}) = |\mathbf{R}|^{-1/2} \exp\left[-\frac{1}{2}\left(\mathbf{q}'\mathbf{R}^{-1}\mathbf{q} - \mathbf{q}'\mathbf{q}\right)\right], \tag{10}$$

where $\mathbf{q} = (q_1; \ldots; q_K)$, $q_i = \Phi^{-1}(u_i)$ and $\Phi(x)$ denotes the cumulated density function of the standard Normal distribution computed at x. When combined with Gamma marginals, the resulting multivariate distribution is a special case of dispersion distribution generated from a Gaussian copula, as discussed in Song [27]. The Normal copula has a number of attractive features: the capability of capturing a broad range of dependencies (the bivariate Normal copula, according to the value of the correlation parameter, is capable of attaining the lower Fréchet bound, the product copula and the upper Fréchet bound), the analytical tractability, the ease of simulation. However, one of its major drawbacks lies in the asymptotic independence of its tails. Empirically, tail dependence is a behavior frequently observed in financial time series (see McNeil et al. [21], among others): extreme events in different assets tend to be combined.

[5] A deeper discussion of Elliptical copulas is beyond the scope of the paper: see, among others, McNeil et al. [21], Frahm et al. [16], Schmidt [26].

The Student-T copula, obtained from (9) by choosing

$$\begin{aligned} K^*(\nu; K) &= \frac{\Gamma((\nu+K)/2)\Gamma(\nu/2)^{K-1}}{\Gamma((\nu+1)/2)^K}, \\ g_1(x; \nu, K) &= (1 + x/\nu)^{-(\nu+K)/2}, \\ g_2(x; \nu) &= (1 + x/\nu)^{-(\nu+1)/2}, \end{aligned}$$

has a probability density function

$$cT(\mathbf{u}; \mathbf{R}, \nu) = \frac{\Gamma((\nu+K)/2)\Gamma(\nu/2)^{K-1}}{\Gamma((\nu+1)/2)^K} |\mathbf{R}|^{-1/2} \frac{(1 + \mathbf{q}'\mathbf{R}^{-1}\mathbf{q}/\nu)^{-(\nu+K)/2}}{\prod_{i=1}^{K}(1 + q_i^2/\nu)^{-(\nu+1)/2}}, \tag{11}$$

where $\mathbf{q} = (q_1; \ldots; q_K)$, $q_i = T^{-1}(u_i; \nu)$ and $T(x; \nu)$ denotes the cumulated density function of the Student-T distribution computed at x. The Student-T copula shares many of the characteristics of the Normal copula with the main differences in the tails, that are asymptotically dependent. As a further difference relative to the Normal, for $\mathbf{R} = \mathbf{I}$ we do not obtain the independence copula, since uncorrelated multivariate T rv's are not independent (details in McNeil et al. [21]). For a deeper handling of the Student-T copula see Demarta and McNeil [10].

2.2.2. The Semiparametric Formulation

An alternative specification of the error term is obtained by assuming only the first two moments in (3), without an explicit assumption about the shape of the distribution. This approach can be motivated by different arguments (see Cipollini et al. [9]): in particular, a full specification of the distribution of $\boldsymbol{\varepsilon}_t$ may be not crucial if the main focus of the analysis is on the dynamics of the $\boldsymbol{\mu}_t$ component.

3. Inference

The two alternative specifications of the error term lead to different inferential approaches: the parametric assumption of section 2.2.1 gives a fully parametric vMEM that can be estimated via Maximum Likelihood (ML); the semiparametric assumption of section 2.2.2 provides a semiparametric vMEM that can be estimated via Generalized Method of Moments (GMM) or other methods leading, for this model, to identical inferences (see Cipollini et al. [9] about this point).

3.1. ML Inference within the Copula Based vMEM

We assume here a vMEM defined by (1), (6) and (7). Such a model is driven by the following set of parameters: $\boldsymbol{\theta}$ (into the $\boldsymbol{\mu}_t$ equations); $\boldsymbol{\phi}$ (into the marginals); $\boldsymbol{\xi}$ (into the copula). It is useful to recall the relationships and the sequence in which some quantities are computed ($i = 1, \ldots, K$): from $\mu_{t,i}(\boldsymbol{\theta}_i)$'s, the estimated innovations $\varepsilon_{t,i} = x_{t,i}/\mu_{t,i}$ are derived; innovations are then mapped into the corresponding cdf's via $u_{t,i} = F_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)$; finally, $u_{t,i}$'s provides an input for the copula probability density function $c(\mathbf{u}_t; \boldsymbol{\xi})$. The sequence can be summarized as follows:

$$\mu_{t,i}(\boldsymbol{\theta}_i) \rightarrow x_{t,i}/\mu_{t,i} = \varepsilon_{t,i} \rightarrow F_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i) = u_{t,i} \rightarrow c(\mathbf{u}_t; \boldsymbol{\xi}). \tag{12}$$

For each of these quantities, the bolded version without the index i denotes the whole vector of the corresponding quantities at time t.

The conditional probability density function of $\mathbf{x}_t$ is then given by

$$f_{\mathbf{x}}(\mathbf{x}_t|\mathcal{F}_{t-1}) = c(\mathbf{u}_t;\boldsymbol{\xi}) \prod_{i=1}^{K} \frac{f_i(\varepsilon_{t,i};\boldsymbol{\phi}_i)}{\mu_{t,i}},$$

so that the log-likelihood of the model can be written as

$$\begin{aligned} l =& \sum_{t=1}^{T} \ln c(\mathbf{u}_t;\boldsymbol{\xi}) + \sum_{t=1}^{T}\sum_{i=1}^{K} \left[\ln\left(\varepsilon_{t,i} f_i(\varepsilon_{t,i};\boldsymbol{\phi}_i)\right) - \ln x_{t,i}\right] \qquad (13) \\ =& [\text{copula contribution}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\xi})] + [\text{marginals contribution}(\boldsymbol{\theta},\boldsymbol{\phi})]. \end{aligned}$$

In general, the above log-likelihood function can be optimized directly by full ML estimation of the three sets of parameters, namely $\boldsymbol{\theta}$, $\boldsymbol{\xi}$ and $\boldsymbol{\phi}$. However, for some choices of the copula, simple estimators of $\boldsymbol{\xi}$ (usually moment estimators) are available and can be computed from current values of residuals $\boldsymbol{\varepsilon}_t$ or of $\mathbf{u}_t$'s. Examples are parameters of Archimedean copulas or of Elliptical copulas derived from Kendall correlations of current estimates of $\boldsymbol{\varepsilon}_t$. When this solution is available, a *pseudo-loglikelihood* can be constructed as

$$l = \sum_{t=1}^{T} \ln c(\mathbf{u}_t;\widehat{\boldsymbol{\xi}}) + \sum_{t=1}^{T}\sum_{i=1}^{K} \left[\ln\left(\varepsilon_{t,i} f_i(\varepsilon_{t,i};\boldsymbol{\phi}_i)\right) - \ln x_{t,i}\right], \qquad (14)$$

where $\widehat{\boldsymbol{\xi}}$ is the current estimate of $\boldsymbol{\xi}$. Invoking asymptotic arguments for its adoption, such a possibility can reduce considerably the amount of computations during optimization because (14) depends only on $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. We will make a further comment about this point below, when we give some details on the estimation of

the correlation matrix appearing in elliptical copulas.

3.1.1. Computational Details with Elliptical Copulas

In this section we illustrate some details of the estimation when the copula belongs to the Elliptical family and all parameters of the vMEM are estimated via full maximum likelihood.

Using the notation given in section 2.2.1, the log-likelihood of the model can be written as

$$
\begin{aligned}
l = & -\sum_{t=1}^{T}\sum_{i=1}^{K}\ln x_{t,i} - \frac{T}{2}|\mathbf{R}| + T\ln K^*(\boldsymbol{\nu};K) \\
& + \sum_{t=1}^{T}\left[\ln g_1(\mathbf{q}_t'\mathbf{R}^{-1}\mathbf{q}_t;\boldsymbol{\nu};K) - \sum_{i=1}^{K}\ln g_2(q_{t,i}^2;\boldsymbol{\nu}) + \sum_{i=1}^{K}\ln\left(\varepsilon_{t,i}f_i(\varepsilon_{t,i};\boldsymbol{\phi}_i)\right)\right],
\end{aligned}
\tag{15}
$$

where

- in the Normal copula case: $\ln K^*(K) = 0$, $\ln g_1(x;K) = \ln g_2(x) = -x/2$;
- in the Student-T copula case:

$$
\begin{aligned}
\ln K^*(\nu;K) &= \ln\left[\frac{\Gamma((\nu+K)/2)\Gamma(\nu/2)^{K-1}}{\Gamma((\nu+1)/2)^K}\right], \\
\ln g_1(x;\nu,K) &= -\frac{\nu+K}{2}\ln\left(1+\frac{x}{\nu}\right), \\
g_2(x;\nu) &= -\frac{\nu+1}{2}\ln\left(1+\frac{x}{\nu}\right).
\end{aligned}
$$

In order to perform full ML estimation of the correlation matrix of the copula, it is

convenient to reparameterize it in an unconstrained way, as illustrated in McNeil et al. [21], p. 235. In fact, any correlation matrix $\mathbf{R}$ can be represented as

$$\mathbf{R} = \mathbf{D}\mathbf{c}'\mathbf{c}\mathbf{D}, \tag{16}$$

where $\mathbf{c}$ is an upper-triangular matrix with ones on the main diagonal and $\mathbf{D}$ is a diagonal matrix with diagonal entries $D_1 = 1$ and $D_j = \left(1 + \sum_{i=1}^{j-1} c_{ij}^2\right)^{-1/2}$ for $j = 2, \ldots, K$. Using this approach, the estimation of $\mathbf{R}$ is transformed in an unconstrained problem, since the $K(K-1)/2$ free elements of $\mathbf{c}$ vary into $\mathbb{R}$. As mentioned before, different methods relative to full ML estimation can be taken into account, in particular for estimating the parameters of the copula: for instance, we can use Kendall correlations for estimating $\boldsymbol{R}$ (Lindskog et al. [19]) or tail dependence indices for estimating the shape parameter $\boldsymbol{\nu}$ (Kostadinov [18]). Using this approach, a pseudo-loglikelihood as in (14) can be constructed and optimized as mentioned in section 3.1. Note that, using (16), $\ln(|\mathbf{R}|) = 2\sum_{i=2}^{K} \ln D_i$.

For speeding up computations and improve numerical stability it is convenient to use analytical derivatives when the optimization algorithm needs the score function. We illustrate it by using the following notation: $\mathbf{C} = \mathbf{c}\mathbf{D}$, $\widetilde{\mathbf{q}}_t = \mathbf{C}'^{-1}\mathbf{q}_t$, $\mathbf{q}_t^* = \mathbf{R}^{-1}\mathbf{q}_t$, $\widetilde{\widetilde{q}}_t = \mathbf{q}_t'\mathbf{R}^{-1}\mathbf{q}_t = \widetilde{\mathbf{q}}_t'\widetilde{\mathbf{q}}_t$, $\mathbf{q}_t = (q_{t,1}; \ldots; q_{t,K})$, $q_{t,i} = G^{-1}(u_{t,i}; \boldsymbol{\nu})$ (see also (12)).

Parameters entering the matrix $\mathbf{c}$

The portion of the score relative to the free parameters of the $\mathbf{c}$ matrix (those above

the main diagonal) has elements

$$\nabla_{c_{ij}} l = \nabla_{c_{ij}} \left[-T \sum_{i=2}^{K} \ln(D_i) + \sum_{t=1}^{T} \ln g_1(\widetilde{\widetilde{q}}_t; \boldsymbol{\nu}; K) \right]. \tag{17}$$

Using some algebra we can show that

$$\nabla_{c_{ij}} \sum_{i=2}^{K} \ln(D_i) = -D_j C_{ij}$$

$$\nabla_{c_{ij}} \ln g_1(\widetilde{\widetilde{q}}_t; \boldsymbol{\nu}; K) = -2 \nabla_{\widetilde{\widetilde{q}}_t} (\ln g_1(\widetilde{\widetilde{q}}_t; \boldsymbol{\nu}; K)) D_j q^*_{t,j} (\widetilde{q}_{t,i} - C_{ij} q_{t,j}).$$

By replacing them into (17) we obtain

$$\nabla_{c_{ij}} l = T D_j C_{ij} + 2 D_j \sum_{t=1}^{T} q^*_{t,j} (C_{ij} q_{t,j} - \widetilde{q}_{t,i}) \nabla_{\widetilde{\widetilde{q}}_t} (\ln g_1(\widetilde{\widetilde{q}}_t; \boldsymbol{\nu}; K)).$$

Parameters entering the vector $\boldsymbol{\nu}$

The portion of the score relative to $\boldsymbol{\nu}$ is

$$\nabla_{\boldsymbol{\nu}} l = \nabla_{\boldsymbol{\nu}} \left[T \ln K^*(\boldsymbol{\nu}; K) + \sum_{t=1}^{T} \ln g_1(\widetilde{\widetilde{q}}_t; \boldsymbol{\nu}; K) - \sum_{t=1}^{T} \sum_{i=1}^{K} \ln g_2(q^2_{t,i}; \boldsymbol{\nu}) \right]$$

The derivative of $\ln K^*(\boldsymbol{\nu}; K)$ can sometimes be computed analytically. For instance, in the Student-T copula we have

$$\nabla_{\nu} \ln K^*(\nu; K) = \frac{1}{2} \left[\psi \left(\frac{\nu + K}{2} \right) + (K-1) \psi \left(\frac{\nu}{2} \right) - K \psi \left(\frac{\nu + 1}{2} \right) \right].$$

For the remaining quantities we suggest numerical derivatives when, as in the

Student-T case, the quantile function $G^{-1}(x;\boldsymbol{\nu})$ cannot be computed analytically.

Parameters entering the vector $\boldsymbol{\theta}$

The portion of the score relative to $\boldsymbol{\theta}$ is

$$\nabla_{\boldsymbol{\theta}} l = \nabla_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left[\ln g_1(\widetilde{\widetilde{q}}_t; \boldsymbol{\nu}; K) - \sum_{i=1}^{K} \ln g_2(q_{t,i}^2; \boldsymbol{\nu}) + \sum_{i=1}^{K} \ln \left(\varepsilon_{t,i} f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i) \right) \right].$$

After some algebra we obtain that

$$\nabla_{\boldsymbol{\theta}} l = \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' \mathbf{a}_t \tag{18}$$

where $\mathbf{a}_t$ has components

$$a_{t,i} = \frac{f_i^{(1)}(\varepsilon_{t,i}; \boldsymbol{\phi_i}) b_{t,i} + f_i^{(2)}(\varepsilon_{t,i}; \boldsymbol{\phi_i})}{\mu_{t,i}}$$

with

$$b_{t,i} = 2 \frac{q_{t,i} \nabla_{q_{t,i}^2} \ln g_2(q_{t,i}^2; \boldsymbol{\nu}) - q_{t,i}^* \nabla_{\widetilde{\widetilde{q}}_t} \ln g_1(\widetilde{\widetilde{q}}_t; \boldsymbol{\nu}, K)}{g(q_{t,i}; \boldsymbol{\nu})} \tag{19}$$

$$\begin{aligned} f_i^{(1)}(\varepsilon_{t,i}; \boldsymbol{\phi}_i) =& \varepsilon_{t,i} f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i) \\ f_i^{(2)}(\varepsilon_{t,i}; \boldsymbol{\phi}_i) =& - \left[\varepsilon_{t,i} \nabla_{\varepsilon_{t,i}} \ln f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i) + 1 \right]. \end{aligned}$$

For instance, if a marginal has a distribution $Gamma(\phi_i, \phi_i)$ then

$$\begin{aligned} f_i^{(1)}(\varepsilon_{t,i}; \phi_i) =& dGamma(\varepsilon_{t,i}; \phi_i + 1, \phi_i) \\ f_i^{(2)}(\varepsilon_{t,i}; \phi_i) =& \phi_i(\varepsilon_{t,i} - 1). \end{aligned}$$

where $dGamma(x; \alpha, \beta)$ denotes the probability density function of the $Gamma(\alpha, \beta)$ distribution computed at x.

Parameters entering the vector $\boldsymbol{\phi}$

The portion of the score relative to $\boldsymbol{\phi}$ has elements

$$\nabla_{\boldsymbol{\phi_i}} l = \nabla_{\boldsymbol{\phi_i}} \sum_{t=1}^{T} \left[\ln g_1(\widetilde{\widetilde{q}}_t; \boldsymbol{\nu}; K) - \sum_{i=1}^{K} \ln g_2(q_{t,i}^2; \boldsymbol{\nu}) + \sum_{i=1}^{K} \ln f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i) \right].$$

After some algebra we obtain

$$\nabla_{\boldsymbol{\phi_i}} l = \sum_{t=1}^{T} [-\nabla_{\boldsymbol{\phi}_i} F_i(\varepsilon_{t,i}; \boldsymbol{\phi_i}) b_{t,i} + \nabla_{\boldsymbol{\phi}_i} \ln f_i(\varepsilon_{t,i}; \boldsymbol{\phi_i})] \tag{20}$$

where $b_{t,i}$ are given in (19). For instance, if a marginal has a distribution $Gamma(\phi_i, \phi_i)$ then

$$\nabla_{\phi_i} f_i(\varepsilon_{t,i}; \phi_i) = \ln(\phi_i) - \psi(\phi_i) + \ln(\varepsilon_{t,i}) - \varepsilon_{t,i} + 1$$

whereas $\nabla_{\phi_i} F_i(\varepsilon_{t,i}; \phi_i)$ can be computed numerically.

3.1.2. Computational Details with the Normal Copula

Even if the Normal copula can be obtained as a special case of the Elliptical copula presented in the previous section, its particular analytical structure suggests alternative solutions for estimating $\mathbf{R}$ and hence the remaining parameters. In fact, using some matrix algebra the contribution of the copula to the loglikelihood can be rewritten as

$$[\text{copula contribution}] = \frac{T}{2}\left[-\ln|\mathbf{R}| - \operatorname{trace}(\mathbf{R}^{-1}\mathbf{Q}) + \operatorname{trace}(\mathbf{Q})\right], \qquad (21)$$

where

$$\mathbf{Q} = \frac{\mathbf{q}'\mathbf{q}}{T}$$

and $\mathbf{q} = (\mathbf{q}_1'; \ldots; \mathbf{q}_T')$ is a $T \times K$ matrix.

A direct maximization in $\mathbf{R}$ of this quantity, leads to the estimator $\widehat{\mathbf{R}} = \mathbf{Q}$. However this result is obtained without imposing any constraint relative to its nature of correlation matrix ($\operatorname{diag}(\mathbf{R}) = \mathbb{1}$ and positive definiteness). On the other hand, a direct maximization of (21) taking into account these constraints does not return an equation with an explicit solution (see Cipollini et al. [8] for the details).

An acceptable compromise which should increase efficiency, although the result is formally not interpretable as a ML estimator, is to normalize the estimator $\widehat{\mathbf{R}}$, in order to transform it into a correlation matrix:

$$\widetilde{\mathbf{R}} = \mathbf{D}_Q^{-\frac{1}{2}}\mathbf{Q}\mathbf{D}_Q^{-\frac{1}{2}},$$

where $\mathbf{D}_Q = \mathrm{diag}(Q_{11}, \ldots, Q_{KK})$. This solution can be justified observing that the copula contribution to the likelihood depends on $\mathbf{R}$ exactly as if it were the correlation matrix of i.i.d. normally distributed rv's $\mathbf{q}_t$ with mean $\mathbf{0}$ and correlation matrix $\mathbf{R}$ (see also McNeil et al. [21], p. 235). Using this constrained estimator of $\mathbf{R}$ we obtain the following *concentrated log-likelihood*

$$lc = \frac{T}{2}\left[-\ln|\widetilde{\mathbf{R}}| - \mathrm{trace}(\widetilde{\mathbf{R}}^{-1}\mathbf{Q}) + \mathrm{trace}(\mathbf{Q})\right] + [\text{marginals contribution}]. \quad (22)$$

It is interesting to note that (22) gives a relatively simple structure of the score function. Using some tedious algebra, we can show that the components of the score $\nabla_{\boldsymbol{\theta}} lc$ and $\nabla_{\boldsymbol{\phi}} lc$ have exactly the structure in (18) and (20), with the quantity $b_{t,i}$ in (19) changed into

$$b_{t,i} = -\frac{(\mathbf{C}\mathbf{q}_t)_i}{\phi(q_{t,i})}, \quad (23)$$

where the $\mathbf{C}$ matrix is here given by

$$\mathbf{C} = \mathbf{Q}^{-1}\mathbf{D}_Q^{1/2}\mathbf{Q}\mathbf{D}_Q^{1/2}\mathbf{Q}^{-1} - \mathbf{Q}^{-1} + \mathbf{I}_K - \widetilde{\mathbf{R}}^{-1} + \mathbf{D}_Q^{-1} - \mathbf{D}_Q^{-1/2}\,\mathrm{diag}(\mathbf{Q}^{-1}\mathbf{D}_Q^{1/2}\mathbf{Q}).$$

3.2. GMM Inference in the Semiparametric vMEM

As illustrated in some detail in Cipollini et al. [9], the semiparametric version of the vMEM leads to by far simpler inferences, obtainable via GMM. The moment

equation to be solved for estimating $\boldsymbol{\theta}$ is given by

$$\mathbf{g}(\boldsymbol{\theta}) \equiv \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' [\operatorname{diag}(\boldsymbol{\mu}_t) \widehat{\boldsymbol{\Sigma}}_T \operatorname{diag}(\boldsymbol{\mu}_t)]^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_t) = \mathbf{0} \tag{24}$$

where

$$\widehat{\boldsymbol{\Sigma}}_T = \frac{1}{T} \sum_{t=1}^{T} \mathbf{u}_t \mathbf{u}_t' \tag{25}$$

is a consistent estimate of the variance matrix of the innovation term (3) and $\mathbf{u}_t$ is defined by

$$\mathbf{u}_t = \mathbf{x}_t \oslash \boldsymbol{\mu}_t - \mathbb{1}. \tag{26}$$

The asymptotic variance matrix of the GMM estimator of $\boldsymbol{\theta}$ is estimated by a sandwich-like estimator whose structure is

$$\widehat{\operatorname{Avar}}(\widehat{\boldsymbol{\theta}}_T) = \left[\sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' [\operatorname{diag}(\boldsymbol{\mu}_t) \widehat{\boldsymbol{\Sigma}}_T \operatorname{diag}(\boldsymbol{\mu}_t)]^{-1} \nabla_{\boldsymbol{\theta}'} \boldsymbol{\mu}_t \right]^{-1}. \tag{27}$$

3.3. Some Details about the Estimation of $\boldsymbol{\theta}$

If one assumes that $\boldsymbol{\mu}_t$ evolves following the general equation (6), its dynamic behavior depends in general upon $K+3K^2L$ parameters, which reduce to $K+3K^2$ for a model with 1 lag only. For instance, for $K = 3$ and $L = 1$ there are 30 parameters.

A reduction in the number of free parameters can be obtained estimating $\boldsymbol{\omega}$ from stationary conditions (the equivalent of *variance targeting* in a GARCH context,

see Engle and Mezrich [14]). Imposing that $\boldsymbol{\mu}_t$ is stationary we have

$$\boldsymbol{\omega} = \left[\mathbf{I} - \sum_{l=1}^{L} (\boldsymbol{\alpha}_l + \boldsymbol{\gamma}_l/2 + \boldsymbol{\beta}_l) \right] \boldsymbol{\mu}, \tag{28}$$

where $\boldsymbol{\mu} = E(\mathbf{x}_t)$. By substituting (28) into (6) and replacing $\boldsymbol{\mu}$ with its natural estimate, that is the unconditional average $\overline{\mathbf{x}}$, we obtain

$$\widetilde{\boldsymbol{\mu}}_t = \sum_{l=1}^{L} \left[\boldsymbol{\alpha}_l \widetilde{\mathbf{x}}_{t-l} + \boldsymbol{\gamma}_l \widetilde{\mathbf{x}}_{t-l}^{(-)} + \boldsymbol{\beta}_l \widetilde{\boldsymbol{\mu}}_{t-l}, \right] \tag{29}$$

where the symbol $\widetilde{\mathbf{x}}$ represents the demeaned version of $\mathbf{x}$. This strategy saves K parameters in the iterative estimation, and, from informal experiments not reported here, provides very good performances in comparison with direct ML estimates of $\boldsymbol{\omega}$ and improves the numerical stability of the algorithms.

To save time, it is also useful take into account analytic derivatives of $\widetilde{\boldsymbol{\mu}}_t$ with respect to the elements of $\boldsymbol{\theta}$. To this aim, we rewrite the i-th equation into (29) as

$$\widetilde{\mu}_{t,i} = \boldsymbol{\alpha}'_{(i)} \underline{\widetilde{\mathbf{x}}}_{t-1(i)} + \boldsymbol{\beta}'_{(i)} \underline{\widetilde{\boldsymbol{\mu}}}_{t-1(i)}, \tag{30}$$

where: $\boldsymbol{\alpha}_{(i)}$ is a vector stacking the elements of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ appearing in the i-th equation that are not constrained to zero; $\underline{\widetilde{\mathbf{x}}}_{t-1(i)}$ is the vector of corresponding elements into lagged $\mathbf{x}_{t-l}$ or $\mathbf{x}_{t-l}^{(-)}$; $\boldsymbol{\beta}_{(i)}$ is a vector stacking the elements of $\boldsymbol{\beta}$ appearing in the i-th equation that are not constrained to zero; $\underline{\widetilde{\boldsymbol{\mu}}}_{t-1(i)}$ is the vector

of corresponding elements into lagged $\boldsymbol{\mu}_{t-l}$. Within this framework we have

$$\nabla_{\boldsymbol{\theta}_i}\widetilde{\mu}_{t,i} = \nabla_{\boldsymbol{\theta}_i}\underline{\widetilde{\boldsymbol{\mu}}}'_{t-1(i)}\boldsymbol{\beta}_{(i)} + \begin{pmatrix} \underline{\widetilde{\mathbf{x}}}_{t-1(i)} \\ \underline{\widetilde{\boldsymbol{\mu}}}_{t-1(i)} \end{pmatrix} \tag{31}$$

where $\boldsymbol{\theta}_{(i)} = (\boldsymbol{\alpha}_{(i)}; \boldsymbol{\beta}_{(i)})$.

3.4. Further Computational Details

3.4.1. Optimization

In order to optimize the log-likelihood functions specified by the parametric vMEM's or the moment functions employed in the semiparametric version, we tested different optimization routines, amenable to the following categories: algorithms that use the criterion function only (we label them as *derivative free*); algorithms that employ the criterion function and the gradient (we name them *BFGS-like*); algorithms that needs the gradient and the Hessian (or some surrogates able to approximate the curvature, like the OPG or some other quantity related to the asymptotic variance matrix of coefficients; we label this category as *Newton-like*). It is known that, broadly speaking, derivative free algorithms are computationally less demanding and more stable than the others, but are characterized by relatively low rates of convergence toward the solution; on the opposite, Newton-like algorithms have better rates of convergence but need derivatives and are relatively instable when the current estimate is far from the solution, in particular if there are many parameters and the function to be optimized is relatively flat.

In principle, ML methods can use algorithms belonging to all three categories. On

the contrary, a GMM method based on just identified moment conditions, as in our case, give rise to *pseudo-score* functions (i.e. functions the expected value of which is exactly zero at the true parameter value) and a corresponding sandwich estimation of the variance matrix. Note, though, that the pseudo-score function does not necessarily correspond to the gradient of an underlying criterion function to be optimized. This makes the Newton-like algorithm as the natural solution for this case.

After some heuristic experiments, we found a good mix by combining a derivative free algorithm and a Newton-like algorithm for both types of vMEM's described in the previous sections. The former starts from some (very rough) initial values and iterates for a pre-specified number of steps (or until convergence, if reached before); the latter exploits the refinement performed by the previous one by iterating until convergence (usually quite fast). Using only the first algorithm causes very slow convergence; employing only the second one entails, very frequently, a crash of the estimation session.

In particular, we used the NEWUOA as a derivative free algorithm, described in Powell [24] and compared with other approaches in Powell [25].[6] Even with many parameters, it does a really good job in refining the starting values to more reliable estimates of the parameters for the following algorithm. Note that since GMM lacks an underlying criterion function (see above), we surrogated it with $\mathbf{g}(\boldsymbol{\theta})'\widehat{\mathrm{Avar}}(\boldsymbol{\theta})^{-1}\mathbf{g}(\boldsymbol{\theta})/2$ (see section 3.2). We programmed the relative code of

[6] We thank the author for making available his Fortran code to us.

the Newton-like algorithm. In ML, we used the OPG (instead of the Hessian) for approximating the curvature of the function, in a way to exploit the job done in computing the score. In GMM we used the moment function (24) as a gradient function and $\widehat{\mathrm{Avar}}(\boldsymbol{\theta})^{-1}$ for capturing the curvature (see section 3.2).

3.4.2. Computational Tricks with Copulas

Copulas require routinely to compute cumulative density functions corresponding to some input values: in the vMEM, this happens for $F_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)$ (see (12)). Numerical problems arise when the input value $\varepsilon_{t,i}$ is 'extreme', in the sense that, at a computational level, $F_i(\varepsilon_{t,i}) = 1$.[7] Two can be the dangerous consequences of this situation:

1. $G^{-1}(F_i(\varepsilon_{t,i}))$, where G is the symmetric cumulative density function involved in the chosen copula (we omit parameters), returns an 'Inf', causing a crash of the estimation session;
2. numerical derivatives for computing $\nabla_{\boldsymbol{\phi}_i} F_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)$ are relatively unstable.

Among the approaches experimented with, we found that a simple and effective solution is to correct slightly the value of $\varepsilon_{t,i}$ as input for both $G^{-1}(F_i(\varepsilon_{t,i}))$ and $\nabla_{\boldsymbol{\phi}_i} F_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)$. In practice, for some small quantity de that we will specify below, $\varepsilon_{t,i}$ is adjusted to $F^{-1}(1-de)$ whenever $F(\varepsilon_{t,i}) > 1-de$. For reasons of symmetry, we applied a similar adjustment for $\varepsilon_{t,i}$ close to 0, by correcting it to $F^{-1}(de)$ whenever $\varepsilon_{t,i} < de$. The value of de was set to $\sqrt{dneps}$ after some tuning (about

[7] More technically, the problems arise when $F_i(\varepsilon_{t,i}) > 1 - dneps$, where $dneps$ denotes the smallest positive floating-point such that $1 - dneps < 1$.

$dneps$, see note 7).

4. Model Selection Issues

vMEM's are ruled in general by many parameters and we may not have available some theory relative to possible lack of interdependency among the variables under analysis in order to specify a relatively parsimonious formulation of the model. One may thus find it useful to have an automated model selection algorithm that, starting from an initial formulation, is able to select a simple but effective specification. We believe that such a device simplifies considerably the specification of a vMEM for a potential user who can thus avoid a tedious 'manual' selection of the model.

To this aim, we suggest an automated general to specific selection algorithm. Starting from an initial 'large' formulation (for instance a full formulation at lag 1), the algorithm removes at each iteration the component of the model with the smallest not significant Wald test statistic, stopping itself when only significant variables are included in the formulation. In our implementation we constrained the selection, do not allowing to remove parameters on the main diagonal of the $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ matrices. In fact, models with many parameters can experience unreliable estimates of the asymptotic variance matrix, a fact that could produce a precocious removal of components that can be relevant in the model. Moreover, this solution permits to appreciate the additional contribution of non diagonal parameters relative to the results of separated univariate MEM's.

Two comments are in order here. We chose to base the algorithm on the selection

of significant variables since it is relatively simple and applicable to both the parametric and the semiparametric formulations of the vMEM (with semiparametric models we cannot usually refer to information criteria). Second, we pay special attention to a number of statistical and computational details (explained in what follows) which bypass some obstacles potentially responsible for the slowing (or even breaking) down of the procedure. Considering the number of parameters in common applications (with three-four variables), the algorithm is quite fast. In fact, as often it is the case, the main computational burden lies in the estimation of the starting model; subsequent steps are relatively fast because can exploit previous estimates as more precise starting values.

In order to give an idea of the performance of the automated algorithm, we ran a simple simulation experiment. The 'true' model employed in simulations is based on the semiparametric vMEM estimated in Cipollini et al. [9] for modeling the joint dynamics of volumes (in million shares), number of trades (in thousand trades) and realized volatility relative to the IBM stock (period 01/02/2001 - 12/30/2005). As the only modification, we concentrated the dynamic effects in just one lag (table 1). In the spirit of a semiparametric model, the simulated innovations are obtained through resampling the residuals estimated on the real data. Both $\mathbf{x}_0$ and $\boldsymbol{\mu}_0$ were set to unconditional averages of the IBM series.

By using these settings, we simulated 501 series composed by 1200 observations. For each one of them, we selected the model by means of the algorithm described starting from a full model (the one with all coefficients) at lag 1. In order to compare the performances, we estimated also the full model and the model with the

correct formulation. The main results are displayed in table 1, where we compared the Mean Squared Errors (MSE) of the selected and the true model against those of the full model.

We note that the algorithm tends in general to produce more accurate inferences than the full model, particularly for the zero coefficients, even if the correct model performs, as expected, significantly better. We remark that the algorithm has selected the correct formulation in 130 cases over 501, which is a relatively good performance after taking into account the relatively small value of some coefficients inside the true model.

5. Volatility Spillovers

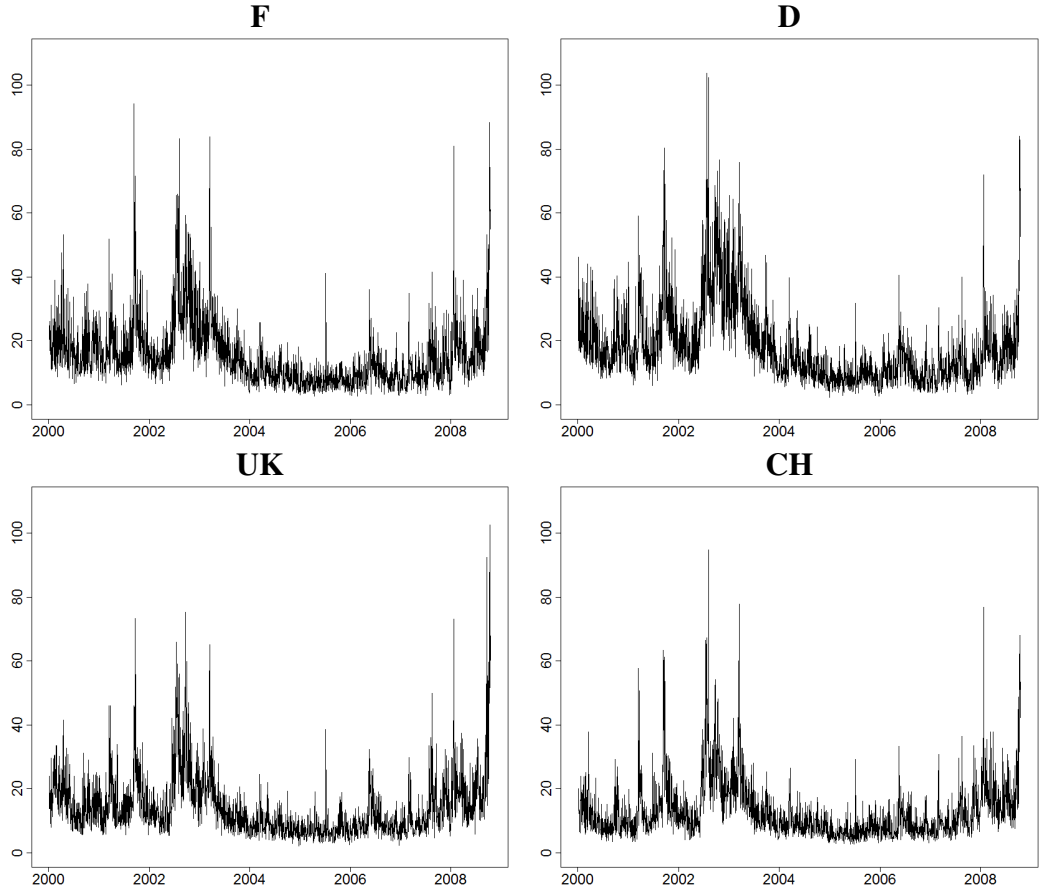
The current (2007-2008) financial crisis is characterized by its global features. Although it may be traced to the money and credit market conditions in the United States, its effects are far reaching and raise the question of how capital market integration is reflected in the transmission of volatility movements from one market to another. In what follows we restrict our attention to some European markets and use our vector MEM to identify significant links in a four variable model for daily ranges in France (CAC40, labeled F), Germany (DAX, labeled D), Great Britain (FTSE100, labeled UK), and Switzerland (SMI, labeled CH). We think that these markets are representative of a number of features of interest for European markets (adoption of the Euro, membership in the EU, presence of globalized transnational corporations with high capitalization, attraction of Eurodeposit in different currency denominations). The choice of the daily range stems from its

Table 1: Simulation experiment: settings and main results. 501 replication by using series of 1200 observations from a vMEM simulated from coefficients displayed below and resampled residuals.

ω	α_1			γ_1			β_1		
True values									
0.6063	0.1010	0.2235	0.0457	0	0	0	0.8478	-0.2781	0
0.3098	0	0.2524	0	0	0	0	-0.0244	0.7177	0.0227
0.1321	0	0	0.1400	0.0213	-0.0222	0.0822	0	-0.0164	0.7929
MSE(estimated selected model)/MSE(estimated full model)									
0.9412	0.5938	0.9214	0.5289	0.0684	0.0668	0.0966	0.8189	1.0634	0.5104
0.9240	0.4534	0.7421	0.2611	0.0547	0.0559	0.0705	0.9108	0.9812	0.6571
1.1870	0.3623	0.3968	1.0347	1.0355	1.0777	1.1886	0.4567	0.5227	0.8371
MSE(estimated correct model)/MSE(estimated full model)									
0.8923	0.1844	0.3857	0.1202	0	0	0	0.2710	0.5167	0
0.9371	0	0.2723	0	0	0	0	0.2509	0.4957	0.1879
0.7620	0	0	0.7010	0.2755	0.3475	0.6423	0	0.0986	0.5665

good properties in estimating daily volatility Alizadeh et al. [2] and the simplicity of its calculation. The time series are depicted in Figure 1 with the fairly common features of high volatility in the first part of the sample (with a further boost in 2003), a period of low volatility from 2003 to the first half of 2007, and, finally, the burst of volatility starting in the last few months of 2007 and in 2008.

Figure 1: Graphs of daily range (in annualized percentage terms) of four indices: CAC40 (F), DAX (D), FTSE100 (UK), SMI (CH) (Jan. 4, 2000 - Oct. 14, 2008).



Our analysis starts from the descriptive statistics of the features of the series, which show similar properties, with the DAX index exhibiting higher average

volatility and volatility of volatility. The serial correlation values of the Ljung-Box statistics are extremely high signaling a high degree of persistence in the series. Finally, unconditional correlations are quite high (more so in the values of Pearson's than in Spearman's, the latter being more robust to common exceptional days) with the clear indication that markets do have a high degree of common movement.

Table 2: Descriptive statistics of daily range (in annualized percentage terms) of four indices: CAC40 (F), DAX (D), FTSE100 (UK), SMI (CH) (Jan. 4, 2000 - Oct. 14, 2008).

	F	D	UK	CH
min	2.76	2.44	2.21	2.70
max	94.05	103.65	102.52	94.65
mean	15.56	18.27	14.24	12.87
sd	10.47	12.54	9.69	8.99
	Ljung-Box statistics			
lag 12	8111.0	10893.3	8049.7	8524.3
lag 22	12428.3	17367.7	11771.5	12531.3
lag 32	15625.4	22738.6	14282.1	15378.3
	Pearson correlations			
F		0.8841	0.8299	0.8430
D			0.7721	0.8037
UK				0.7879
	Spearman correlations			
F		0.8814	0.7923	0.7524
D			0.7497	0.7279
UK				0.7213

The results of the final selected model are presented in table 3. We have inserted a dummy variable over the period Jan. 5, 2000 to July 1, 2003 and then, again, from July 2, 2007 to the end of the sample in order to mitigate the problem of residual autocorrelation when the time span is characterized by a higher average

level of volatility. The insertion of a flexible function modelling the dynamics of a smooth component of the series (as in, e.g. Engle and Rangel [15], Brownlees and Gallo [4] in the univariate case) is beyond the scope of this paper. We group the coefficients by matrices, also to visually impress the result that the automated procedure is selecting a model where many off diagonal elements are statistically significant. A few comments are in order: the values of the coefficients are in line with volatility clustering (relatively low α and γ coefficients, and values of β's on the main diagonal around 0.9). Interestingly, no asymmetric coefficient for the French index is statistically significant (while all others are). The off diagonal elements of the matrix $\boldsymbol{\beta}$ all have a negative sign. We interpret this finding taking as an example the German index volatility. At a given time say that we have a higher value of the range (and that the corresponding returns are negative). This has an immediate effect on the own conditional expectation through the own diagonal coefficients α and γ and a spillover to other markets as well (in the direction of an increase in the conditional expected volatility). Two periods after (isolating just the subsequent effects) the relatively high own β coefficient gets decreased by the other lagged conditional expectations through the negative coefficients. In other words, the model seems to signal that the speed of absorption of a shock is higher in this model than in one where $\boldsymbol{\beta}$ is diagonal, with a richer dynamics of effects.

Table 3: Estimated coefficients and t statistics (within brackets) of the selected vMEM applied to the daily range (in annualized percentage terms) of four indices: CAC40 (F), DAX (D), FTSE100 (UK), SMI (CH) (Jan. 4, 2000 - Oct. 14, 2008).

		F	D	UK	CH
	ω_1	0.00388	0.00400	0.00429	0.00374
	ω_2	0.00287	0.00242	0.00321	0.00229
α_1	F	0.0345	0.0156	0.0365	-
		(4.82)	(2.04)	(3.49)	
	D	-	0.0515	-	-
			(5.1)		
	UK	0.0408	-	0.0497	-
		(5.25)		(4.48)	
	CH	-	-	0.043	0.051
				(4.35)	(5.28)
γ_1	F	-	0.0462	0.0344	0.0213
			(4.71)	(2.98)	(2.17)
	D	-	0.065	0.0341	0.0278
			(3.8)	(2.72)	(2.27)
	UK	-	0.0352	0.0511	0.0302
			(4.52)	(4.19)	(2.9)
	CH	-	0.0321	0.0249	0.0457
			(4.7)	(2.23)	(4.07)
β_1	F	0.9113	-0.0226	-0.0632	-
		(54.88)	(-2.34)	(-4.37)	
	D	-	0.8544	-0.0467	-
			(64.11)	(-3.76)	
	UK	-0.0934	-	0.9033	-
		(-5.93)		(53.28)	
	CH	-0.0449	-	-0.0573	0.9209
		(-3.73)		(-3.56)	(88.6)
α_2	F	-	-	-	-
	D	-	0.0458	-	-
			(2.84)		
	UK	-	-	-	-
	CH	-	-	-	-

Some diagnostics complete our illustrative analysis and are reported in Table 4. We have an overall portmanteau statistics at different lags which still pinpoints to the difficulty of reining in the autocorrelation. Judging from the original values of autocorrelation in each series discussed in Table 2 the problem has been noticeably reduced with our vMEM. The issue of spillovers is addressed by testing the joint significance of all off-diagonal elements (null hypothesis strongly rejected) and of coefficients from one market to all others. In the latter case, there is always dynamic interdependence with the exception of a direct link from the French index to the German index.

Table 4: Some diagnostics of the selected vMEM applied to the daily range (in annualized percentage terms) of four indices: CAC40 (F), DAX (D), FTSE100 (UK), SMI (CH) (Jan. 4, 2000 - Oct. 14, 2008).

Statistics	
Portmanteau (p-values)	
lag 12	0.0007
lag 22	0.0164
lag 32	0.0469
Restrictions (p-values)	
off diagonal = 0	0.0000
$D_{t-1} \nrightarrow F_t$	0.0000
$D_{t-1} \nrightarrow UK_t$	0.0000
$UK_{t-1} \nrightarrow D_t$	0.0005
$D_{t-1} \nrightarrow CH_t$	0.0000
$CH_{t-1} \nrightarrow D_t$	0.0232
$F_{t-1} \nrightarrow UK_t$	0.0000
$UK_{t-1} \nrightarrow F_t$	0.0000
$F_{t-1} \nrightarrow CH_t$	0.0002
$CH_{t-1} \nrightarrow F_t$	0.0299
$UK_{t-1} \nrightarrow CH_t$	0.0000
$CH_{t-1} \nrightarrow UK_t$	0.0037
Residual Correlations	
$\widehat{\rho}_{F,D}$	0.7730
$\widehat{\rho}_{F,UK}$	0.6252
$\widehat{\rho}_{D,UK}$	0.6190
$\widehat{\rho}_{F,CH}$	0.5969
$\widehat{\rho}_{D,CH}$	0.5717
$\widehat{\rho}_{UK,CH}$	0.5259

Further insight on the long–term dynamics can be gained by looking at the forecast equation

$$\boldsymbol{\mu}_{T+\tau} = \boldsymbol{\omega} + \mathbf{A}_1\boldsymbol{\mu}_{T+\tau-1} + \mathbf{A}_2\boldsymbol{\mu}_{T+\tau-2}$$

for $\tau \geq 2$, where $\mathbf{A}_1 = \boldsymbol{\alpha}_1 + \boldsymbol{\gamma}_1/2 + \boldsymbol{\beta}_1$ collects the combined one-lag effects and $\mathbf{A}_2 = \boldsymbol{\alpha}_2$ the diagonal two-lag effects. This equation can be rewritten in a companion form as

$$\boldsymbol{\mu}^*_{T+\tau} = \boldsymbol{\omega}^* + \mathbf{A}\boldsymbol{\mu}^*_{T+\tau-1}$$

where

$$\boldsymbol{\mu}^*_{T+\tau} = \begin{pmatrix} \boldsymbol{\mu}_{T+\tau} \\ \boldsymbol{\mu}_{T+\tau-1} \end{pmatrix},$$

$$\boldsymbol{\omega}^* = \begin{pmatrix} \boldsymbol{\omega} \\ \mathbf{0} \end{pmatrix},$$

and

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I} & \mathbf{0} \end{pmatrix}. \tag{32}$$

The identification of relevant links is complemented by the estimation and inference on the $\mathbf{A}_1$ matrix, reported in Table 5 with the corresponding t-stats in parentheses. We notice that some individually significant effects, when combined together for multi–period forecasts become statistically insignificant, namely, the link from UK to France and from UK to Switzerland. We can also notice that the effects from France to UK and to Switzerland, as well as from UK to Germany

are negative, the former being more sizeable than other coefficients. Finally, to confirm that the model is stationary, we report the two largest eigenvalues of the companion matrix $\mathbf{A}$, equal, respectively, to 0.9925, and 0.9841.

Table 5: Estimated $\mathbf{A}_1$ matrix in expression (32). The rows and columns refer to the four indices in the following order: CAC40 (F), DAX (D), FTSE100 (UK), SMI (CH) (Jan. 4, 2000 - Oct. 14, 2008).

	t-1			
t	F	D	UK	CH
F	0.9458	0.0161	-0.0095	0.0106
	69.9953	2.0419	-0.5813	2.1442
D	–	0.9384	-0.0296	0.0139
	–	57.0620	-2.6752	2.2702
UK	-0.0525	0.0176	0.9785	0.0151
	-3.8703	4.5091	54.222	2.9077
CH	-0.0449	0.0161	-0.0019	0.9946
	-3.7209	4.6731	-0.1101	106.76

6. Concluding Remarks

Vector Multiplicative Error Models allow for the estimation of a dynamic autoregressive model for non–negative processes which finds many interesting applications in financial econometrics (volatility, volumes, and trading activity in general). We allow the model to have a full interdependent structure in its general form, but we recognize the need for a more parsimonious specification both because of estimation performances and the interpretability of the results. Given the articulated estimation procedure there are various cross-roads at which appropriate choices have to be made from a computational point of view given the highly nonlinear nature of the optimization process involved. We have provided some

computational details showing how they can simplify the burden of repeated estimation of the model in selecting significant links. The simulation exercise with realistic parameter values shows that the performance is quite satisfactory. The application to daily ranges as volatility proxies for four main European markets show that volatility interdependence is adequately captured by this class of models with significant asymmetric effects which increase volatility when returns are negative.

References

[1] Ahoniemi, K., Lanne, M., November 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.
URL \url{www.hecer.fi}

[2] Alizadeh, S., Brandt, M. W., Diebold, F. X., 2002. Range-based estimation of stochastic volatility models. The Journal of Finance 57, 1047–1091.

[3] Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T., July 2000. Copulas for finance: A reading guide and some applications. Technical report, Groupe de Recherche Opérationelle, Credit Lyonnais, Paris.

[4] Brownlees, C. T., Gallo, G. M., 2008. Comparison of volatility measures: a risk management perspective. Tech. rep., Dipartimento di Statistica 'G.Parenti', Università di Firenze, wP 2008-03.

[5] Brownlees, C. T., Gallo, G. M., 2008. On variable selection for volatility forecasting: The role of focused selection criteria. Journal of Financial Econometrics 6, 513–539.

[6] Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance. Wiley.

[7] 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.

[8] Cipollini, F., Engle, R. F., Gallo, G. M., 2007. A model for multivariate positive valued processes in financial econometrics. Working Paper 2007/16, Università di Firenze, Dipartimento di Statistica.

[9] Cipollini, F., Engle, R. F., Gallo, G. M., 2008. Semiparametric vector MEM. Working Paper Series 1282155, SSRN eLibrary.

[10] Demarta, S., McNeil, A. J., 2005. The t copula and related copulas. International Statistical Review 73 (1), 111–129.

[11] Embrechts, P., McNeil, A. J., Straumann, D., 2002. Correlation and dependence in risk management: properties and pitfalls. In: Dempster, M. A. H. (Ed.), Risk Management: Value at Risk and Beyond. Cambridge University Press, Cambridge, pp. 176–223.

[12] Engle, R. F., 2002. New frontiers for ARCH models. Journal of Applied Econometrics 17, 425–446.

[13] Engle, R. F., Gallo, G. M., 2006. A multiple indicators model for volatility using intra-daily data. Journal of Econometrics 131 (1–2), 3–27.

[14] Engle, R. F., Mezrich, J., 1995. Grappling with garch. Risk 8, 112–117.

[15] Engle, R. F., Rangel, J. G., 2008. The spline-garch model for low frequency volatility and its global macroeconomic causes. Review of Financial Studies 21, 1187–1222.

[16] Frahm, G., Junker, M., Szimayer, A., 2003. Elliptical copulas: Applicability and limitations. Statistics & Probability Letters 63, 275–286.

[17] Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman and Hall, London.

[18] Kostadinov, K., 2005. Non-parametric estimation of elliptical copulae with application to credit risk. Working paper, Munich University of Technology. URL \url{citeseer.ist.psu.edu/755744.html}

[19] Lindskog, F., McNeil, A. J., Schmock, U., 2003. Kendall's tau for elliptical distributions. In: Bol, Nakhaeizadeh, Rachev, Ridder, Vollmer (Eds.), Credit Risk - Measurement, Evaluation and Management. Physica-Verlag Heidelberg.

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

[21] McNeil, A. J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press, Princeton.

[22] Nelsen, R. B., 1999. An introduction to Copulas. Lecture Notes in Statistics. Springer Verlag, New York.

[23] Patton, A., 2007. Copula-based models for financial time series. In: Andersen, T., Davis, R., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag.

[24] Powell, M. J. D., 2006. The NEWUOA software for unconstrained optimization without derivatives. In: Pillo, G. D., Roma, M. (Eds.), Large-Scale Nonlinear Optimization. Springer Verlag, New York, pp. 255–297.

[25] Powell, M. J. D., April 2007. A view of algorithms for optimization without derivatives. William Benter Distinguished Lecture, City University of Hong Kong DAMTP 2007/NA03, Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, Cambridge, England.

[26] Schmidt, R., 2002. Tail dependence for elliptically contoured distributions. Mathematical Methods of Operations Research 55, 301–327.

[27] Song, P. X.-K., 2000. Multivariate dispersion models generated from gaussian copula. Scandinavian Journal of Statistics 27, 305–320.

A. Appendix: Software Implementation

The software implementation of vMEM's is constructed with the aim of combining two requirements: a (relatively) user-friendly *front-end* and a (relatively) fast *engine* in computations. The front-end, which handle data, input the model and manage the results, is programmed in **R** (`http://www.r-project.org/`); the computational routines (optimization algorithms, log-likelihoods, scores, Hessians, OPG's, Sandwich estimators, moment functions, etc.) has been programmed in **Fortran 77**; only some probability density functions and cumulative density functions are written in **C**, in a way to exploit some native C routines (named 'standalone math library') available from R sources.