



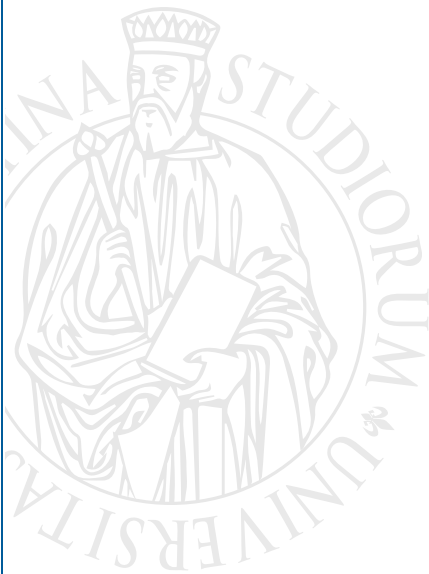
UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DISIA**

DIPARTIMENTO DI STATISTICA,  
INFORMATICA, APPLICAZIONI  
"GIUSEPPE PARENTI"

**Copula–based vMEM Specifications  
versus Alternatives:  
The Case of Trading Activity**

Fabrizio Cipollini, Robert F. Engle,  
Giampiero M. Gallo



**DISIA WORKING PAPER  
2017/02**

© Copyright is held by the author(s).

# Copula-based vMEM Specifications versus Alternatives: The Case of Trading Activity

Fabrizio Cipollini

DiSIA, Università di Firenze, Italy

Robert F. Engle

Department of Finance, Stern School of Business, NYU

Giampiero M. Gallo\*

DiSIA, Università di Firenze, Italy

## Abstract

We discuss several multivariate extensions of the Multiplicative Error Model by Engle (2002) to take into account dynamic interdependence and contemporaneously correlated innovations (vector MEM or vMEM). We suggest copula functions to link Gamma marginals of the innovations, in a specification where past values and conditional expectations of the variables can be simultaneously estimated. Results with realized volatility, volumes and number of trades of the JNJ stock show that significantly superior realized volatility forecasts are delivered with a fully interdependent vMEM relative to a single equation. Alternatives involving log-Normal or semiparametric formulations produce substantially equivalent results.

**Keywords:** GARCH; MEM; Realized Volatility; Trading Volume; Trading Activity; Trades; Copula; Volatility Forecasting.

---

\*Corresponding author: Dipartimento di Statistica, Informatica, Applicazioni “G. Parenti” (DiSIA), Università di Firenze, Viale G.B. Morgagni, 59 - 50134 Firenze – Italy; ph. +39 055 2751 591. The authors thank two anonymous referees, Nikolaus Hautsch, Lorian Mancini, Joseph Noss, Andrew Patton, Kevin Sheppard, David Veredas as well as participants in seminars at CORE, Humboldt Universität zu Berlin, Imperial College, IGIER–Bocconi for helpful comments. The usual disclaimer applies. Financial support from the Italian MIUR under grant PRIN MISURA is gratefully acknowledged.

# 1 Introduction

Dynamics in financial markets can be characterized by many indicators of trading activity such as absolute returns, high-low range, number of trades in a certain interval (possibly labeled as buys or sells), volume, ultra-high frequency based measures of volatility, financial durations, spreads and so on.

Engle (2002) reckons that one striking regularity behind financial time series is that persistence and clustering characterizes the evolution of such processes. As a result, the dynamics of such variables can be specified as the product of a conditionally deterministic scale factor, which evolves according to a GARCH-type equation, and an innovation term which is iid with unit mean. Such models are labeled Multiplicative Error Models (MEM) and can be seen as a generalization of the GARCH (Bollerslev (1986)) and ACD (Engle and Russell (1998)) approaches. 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 volatility, volume and trading intensity Hautsch (2008)).

The model was specified by Engle and Gallo (2006) in a multivariate context (vector MEM or vMEM) allowing just the lagged values of each variable of interest to affect the conditional expectation of the other variables. Such a specification lends itself to producing multi-step ahead forecasts: in Engle and Gallo (2006) three different measures of volatility (absolute returns, daily range and realized volatility), influence each other dynamically. Although such an equation-by-equation estimation ensures consistency of the estimators in a quasi-maximum likelihood context, given the stationarity conditions discussed by Engle (2002), correlation among the innovation terms is not taken into account and leads to a loss in efficiency.

However, full interdependence is not allowed by an equation-by-equation approach both in the form of past conditional expectations influencing the present and a contemporaneous correlations of the innovations. The specification of a multivariate distribution of the innovations is far from trivial: a straightforward extension to a joint Gamma probability distribution is not available except in very special cases. In this paper, we want to compare the features of a novel maximum likelihood estimation strategy adopting copula functions to other parametric and semiparametric alternatives. Copula functions allow to link together marginal probability density functions for individual innovations specified as Gamma as in Engle and Gallo (2006) or as zero-augmented distributions (as in Hautsch *et al.* (2013), distinguishing between the zero occurrences and the strictly positive realizations). Copula functions are used in a Multiplicative Error framework, but in a Dynamic Conditional Correlation context by Bodnar and Hautsch (2016). Within a Maximum Likelihood approach to the vector MEM, we want also to explore the performance of a multivariate log-Normal specification for the joint distribution of the innovations and the semiparametric approach presented in Cipollini *et al.* (2013) resulting in a GMM estimator. The range of potential applications of a vector MEM is quite wide: dynamic interactions among different values of volatility, volatility spillovers across markets (allowing multivariate-multi-step ahead forecasts and impulse response functions, order execution dynamics (Noss (2007)) specifies a MEM for execution depths).

This paper aims at discussing merits of each approach and their performance in forecasting realized volatility with augmented information derived from the dynamic interaction with other measures of market activity (namely, the volumes and the number of trades). We want conditional expectations to depend just on past values (not also on

some contemporary information as in Manganello (2005) and Hautsch (2008)).

What the reader should expect is the following: in Section 2 we lay out the specification of a vector Multiplicative Error Model, discussing the issues arising from the adoption of several types of copula functions linking univariate Gamma marginal distributions. In Section 3 we describe the Maximum Likelihood procedure leading to the inference on the parameters. In Section 4 we discuss the features of a parametric specification with a log–Normal and of a semiparametric approach suggested by Cipollini *et al.* (2013). In Section 5 we present the empirical application to three series of market trading activity, namely realized kernel volatility, traded volumes and the number of trades. The illustration is performed on the JNJ stock over a period between 2007 and 2013. What we find is that specifying the joint distribution of the innovations allowing for contemporaneous correlation dramatically improves the log–likelihood over an independent (i.e. equation–by–equation) approach. Richer specifications (where simultaneous estimation is unavoidable) deliver a better fit, improved serial correlation diagnostics, and a better performance in out–of–sample forecasting. The Student–T copula possesses better features than the Normal copula. Overall, the indication is that we will have significantly superior realized volatility forecasts when other trading activity indicators and contemporaneous correlations are considered. Concluding remarks follow.

## 2 A Copula Approach to MEMs

Let  $\mathbf{x}_t$  be a  $K$ –dimensional process with non–negative components. A vector Multiplicative Error Model (vMEM) for  $\mathbf{x}_t$  is defined as

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

where  $\odot$  indicates the Hadamard (element–by–element) product and  $\text{diag}(\cdot)$  indicates a diagonal matrix with its vector argument on the main diagonal. Conditionally upon the information set  $\mathcal{F}_{t-1}$ , a fairly general specification for  $\boldsymbol{\mu}_t$  is

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}\mathbf{x}_{t-1} + \boldsymbol{\gamma}\mathbf{x}_{t-1}^{(-)} + \boldsymbol{\beta}\boldsymbol{\mu}_{t-1}, \quad (2)$$

where  $\boldsymbol{\omega}$  is  $(K, 1)$  and  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}$  are  $(K, K)$ . The vector  $\mathbf{x}_t^{(-)}$  has a generic element  $x_{t,i}$  multiplied by a function related to a signed variable, be it a positive or negative return (0, 1 values) or a signed trade (buy or sell 1, –1 values), as to capture asymmetric effects. Let the parameters relevant for  $\boldsymbol{\mu}_t$  be collected in a vector  $\boldsymbol{\theta}$ .

Conditions for stationarity of  $\boldsymbol{\mu}_t$  are a simple generalization of those of the univariate case (e.g. Bollerslev (1986); Hamilton (1994)): a vMEM(1,1) with  $\boldsymbol{\mu}_t$  defined as in equation (2) is stationary in mean if all characteristic roots of  $\mathbf{A} = \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma}/2$  are smaller than 1 in modulus. We can think of  $\mathbf{A}$  as the *impact matrix* in the expression

$$E(\mathbf{x}_{t+1}|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_{t+1|t-1} = \boldsymbol{\omega} + \mathbf{A}\boldsymbol{\mu}_{t|t-1}.$$

If more lags are considered, the model is

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \sum_{l=1}^L \left[ \boldsymbol{\alpha}_l \mathbf{x}_{t-l} + \boldsymbol{\gamma}_l \mathbf{x}_{t-l}^{(-)} + \boldsymbol{\beta}_l \boldsymbol{\mu}_{t-l} \right], \quad (3)$$

where  $L$  is the maximum lag occurring in the dynamics. It is often convenient to represent the system (3) in its equivalent *companion form*

$$\boldsymbol{\mu}_{t+L|t-1}^* = \boldsymbol{\omega} + \mathbf{A}^* \boldsymbol{\mu}_{t+L-1|t-1}^*, \quad (4)$$

where  $\boldsymbol{\mu}_{t+L|t-1}^* = (\boldsymbol{\mu}_{t+L|t-1}; \boldsymbol{\mu}_{t+L-1|t-1}; \dots; \boldsymbol{\mu}_{t+1|t-1})$  is a  $KL \times 1$  vector obtained by stacking its elements columnwise and

$$\mathbf{A}^* = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_L \\ & \mathbf{I}_{K(L-1)} & & \mathbf{0}_{K(L-1),K} \end{pmatrix}$$

with  $\mathbf{A}_l = \boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \boldsymbol{\gamma}_l/2$ ,  $l = 1, \dots, L$ ,  $\mathbf{I}_{K(L-1)}$  is a  $K(L-1) \times K(L-1)$  identity matrix and  $\mathbf{0}_{K(L-1),K}$  is a  $K(L-1) \times K$  matrix of zeros. The same stationarity condition holds in terms of eigenvalues of  $\mathbf{A}^*$ .

The innovation vector  $\boldsymbol{\varepsilon}_t$  is a  $K$ -dimensional iid process with probability density function (pdf) defined over a  $[0, +\infty)^K$  support, the unit vector  $\mathbf{1}$  as expectation and a general variance-covariance matrix  $\boldsymbol{\Sigma}$ ,

$$\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1} \sim D^+(\mathbf{1}, \boldsymbol{\Sigma}). \quad (5)$$

The previous conditions guarantee that

$$E(\mathbf{x}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \quad (6)$$

$$V(\mathbf{x}_t | \mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \boldsymbol{\mu}_t' \odot \boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma} \text{diag}(\boldsymbol{\mu}_t), \quad (7)$$

where the latter is a positive definite matrix by construction.

The distribution of the error term  $\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}$  can be specified in different ways. We discuss here a flexible approach using copula functions; different specifications are presented in Section 4.

Using copulas (cf., among others, Joe (1997) and Nelsen (1999), Embrechts *et al.* (2002), Cherubini *et al.* (2004), McNeil *et al.* (2005) and the review of Patton (2013) for financial applications), the conditional pdf of  $\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}$  is given by

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = c(\mathbf{u}_t; \boldsymbol{\xi}) \prod_{i=1}^K f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i), \quad (8)$$

where  $c(\mathbf{u}_t; \boldsymbol{\xi})$  is the pdf of the copula,  $f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)$  and  $u_{t,i} = F_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)$  are the pdf and the cdf, respectively, of the marginals,  $\boldsymbol{\xi}$  and  $\boldsymbol{\phi}_i$  are parameters. A copula approach, hence, requires the specification of two objects: the distribution of the marginals and the copula function.

In view of the flexible properties shown elsewhere (Engle and Gallo (2006)), for the first we adopt Gamma pdf's (but other choices are possible, such as Inverse-Gamma, Weibull, log-Normal, and mixtures of them). This has the important feature of guaranteeing an interpretation of a Quasi Maximum Likelihood Estimator even if the choice of the distribution is not appropriate. For the second, we discuss some possible specifications within the class of Elliptical copulas which provides an unified framework encompassing the Normal, the Student-T and any other member of this family endowed with an explicit pdf.

Elliptical copulas have interesting features and widespread applicability in Finance (for details see McNeil *et al.* (2005), Frahm *et al.* (2003), Schmidt (2002)). They are copulas generated by Elliptical distributions, exactly in the same way as the Normal copula and the Student-T copula stem from the multivariate Normal and Student-T distributions, respectively.<sup>1</sup>

<sup>1</sup>In some applications, especially involving returns, their elliptical symmetry may constitute a limit (cf., for example, Patton (2006), Okimoto (2008), Cerrato *et al.* (2015) and reference therein). Copulas in the Archimedean family (Clayton, Gumbel, Joe-Clayton, etc.) offer a way to bypass such a limitation but suffer from other drawbacks and, in any case, seem to be less relevant for the variables of interest for a vMEM (here, different indicators of trading activity). They will not be pursued in what follows.

We 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 pdf  $g(\cdot; \boldsymbol{\nu})$  and cdf  $G(\cdot; \boldsymbol{\nu})$  ( $\boldsymbol{\nu}$  represents a vector of shape parameters). The density of the copula can then be written as

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

for suitable choices of  $K^*(\cdot, \cdot)$ ,  $g_1(\cdot; \cdot, \cdot)$  and  $g_2(\cdot; \cdot)$ , where  $\mathbf{q} = (q_1; \dots; q_K)$ ,  $q_i = G^{-1}(u_i; \boldsymbol{\nu})$ .

Two notable cases will be considered. The first is the Normal copula (McNeil *et al.* (2005), Cherubini *et al.* (2004), Bouyé *et al.* (2000)); with no explicit shape parameter  $\boldsymbol{\nu}$ , we have  $K^*(K) \equiv 1$ ,  $g_1(x; K) = g_2(x) \equiv \exp(-x/2)$ . The pdf in Equation (9) becomes:

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

where  $\mathbf{q} = (q_1; \dots; q_K)$ ,  $q_i = \Phi^{-1}(u_i)$  and  $\Phi(x)$  denotes the cdf of the standard Normal distribution computed at  $x$ .

The Normal copula has many interesting properties: the ability to reproduce a broad range of dependencies (the bivariate version, 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. When combined with  $Gamma(\phi_i, \phi_i)$  marginals, the resulting multivariate distribution is a special case of dispersion distribution generated from a Gaussian copula (Song (2000)). We note that the conditional correlation matrix of  $\boldsymbol{\varepsilon}_t$  has generic element approximately equal to  $R_{ij}$ , which can assume negative values too.

One of the limitations of the Normal copula is the asymptotic independence of its tails. Empirically, tail dependence is a behavior observed frequently in financial time series (see McNeil *et al.* (2005), among others). Elements of  $\mathbf{x}$  (being different indicators of the same asset or different assets) tend to be affected by the same extreme events. For this reason, as an alternative, we consider the Student-T copula which allows for asymptotically dependent tails. Differently from the Normal copula, when  $\mathbf{R} = \mathbf{I}$  we get uncorrelated but not independent marginals (details in McNeil *et al.* (2005)). Thus, when considering the Student-T copula in Equation (9), with a scalar  $\nu$  shape parameter, we have  $K^*(\nu; K) = \frac{\Gamma((\nu + K)/2) \Gamma(\nu/2)^{K-1}}{\Gamma((\nu + 1)/2)}$ ,  $g_1(x; \nu, K) = (1 + x/\nu)^{-(\nu+K)/2}$ ,  $g_2(x; \nu) = (1 + x/\nu)^{-(\nu+1)/2}$ , and its explicit pdf is

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

where  $\mathbf{q} = (q_1; \dots; q_K)$ ,  $q_i = T^{-1}(u_i; \nu)$  and  $T(x; \nu)$  denotes the cdf of the Student-T distribution with  $\nu$  degrees of freedom computed at  $x$ . Further specifications of the Student-T copula are in Demarta and McNeil (2005).

### 3 Maximum Likelihood Inference

In this section we discuss how to get full Maximum Likelihood (ML) inferences from the vMEM with the parametric specification (3) for  $\boldsymbol{\mu}_t$  (dependent on a parameter vector  $\boldsymbol{\theta}$ ) and a copula-based formulation  $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1})$  of the conditional distribution of the vector

error term (characterized by the parameter vector  $\boldsymbol{\lambda}$ ). Inference on  $\boldsymbol{\theta}$  and  $\boldsymbol{\lambda}$  can be discussed in turn, given that from the model assumptions the log-likelihood function  $l$  is

$$\begin{aligned} l &= \sum_{t=1}^T \ln f_{\mathbf{x}}(\mathbf{x}_t | \mathcal{F}_{t-1}) = \sum_{t=1}^T \ln \left( f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) \prod_{i=1}^K \mu_{t,i}^{-1} \right) \\ &= \sum_{t=1}^T \left[ \ln f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) - \sum_{i=1}^K \ln \mu_{t,i} \right]. \end{aligned} \quad (12)$$

Considering a generic time  $t$ , it is useful to recall the sequence of calculations:

$$\mu_{t,i}(\boldsymbol{\theta}_i) \rightarrow x_{t,i}/\mu_{t,i} = \varepsilon_{t,i} \rightarrow F_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i) = u_{t,i} \rightarrow c(\mathbf{u}_t; \boldsymbol{\xi}) \quad i = 1, \dots, K \quad (13)$$

where  $\boldsymbol{\theta}_i$  is the parameter involved in the  $i$ -th element of the  $\boldsymbol{\mu}_t$  vector.

### 3.1 Parameters in the Conditional Mean

Irrespective of the specification chosen for  $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1})$ , the structure of the vMEM allows to express the portion of the score function corresponding to  $\boldsymbol{\theta}$  as

$$\nabla_{\boldsymbol{\theta}} l = \sum_{t=1}^T \mathbf{A}_t \mathbf{w}_t \quad (14)$$

where

$$\mathbf{A}_t = -\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' \text{diag}(\boldsymbol{\mu}_t)^{-1}. \quad (15)$$

$$\mathbf{w}_t = \boldsymbol{\varepsilon}_t \odot \mathbf{b}_t + \mathbf{1}, \quad (16)$$

$$\mathbf{b}_t = \nabla_{\boldsymbol{\varepsilon}_t} \ln f(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}),$$

In order to have a zero expected score, we need  $E(\mathbf{w}_t | \mathcal{F}_{t-1}) = \mathbf{0}$  or, equivalently,  $E(\boldsymbol{\varepsilon}_t \odot \mathbf{b}_t | \mathcal{F}_{t-1}) = -\mathbf{1}$ . As a consequence, the information matrix and the expected Hessian are given by

$$E \left[ \mathbf{A}_t \boldsymbol{\mathcal{I}}^{(\varepsilon)} \mathbf{A}_t' \right] \quad (17)$$

and

$$E \left[ \mathbf{A}_t \mathbf{H}^{(\varepsilon)} \mathbf{A}_t' \right], \quad (18)$$

respectively, where the matrices

$$\boldsymbol{\mathcal{I}}^{(\varepsilon)} = E [(\boldsymbol{\varepsilon}_t \odot \mathbf{b}_t)(\boldsymbol{\varepsilon}_t \odot \mathbf{b}_t)' | \mathcal{F}_{t-1}] - \mathbf{1}\mathbf{1}'$$

and

$$\mathbf{H}^{(\varepsilon)} = E [\nabla_{\boldsymbol{\varepsilon}_t} \mathbf{b}_t'(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') | \mathcal{F}_{t-1}] - \mathbf{I}$$

depend only on  $\boldsymbol{\lambda}$  but not on  $\boldsymbol{\theta}$ .

For a particular parametric choice of the conditional distribution of  $\boldsymbol{\varepsilon}_t$ , we need to plug the specific expression of  $\ln f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1})$  into  $\mathbf{b}_t$ . For instance, considering the generic copula formulation (8), then

$$\ln f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \ln c(\mathbf{u}_t) + \sum_{i=1}^K \ln f_i(\varepsilon_{t,i}) \quad (19)$$

so that  $\mathbf{b}_t$  has elements

$$b_{t,i} = f_i(\varepsilon_{t,i}) \nabla_{u_{t,i}} \ln c(\mathbf{u}_t) + \nabla_{\varepsilon_{t,i}} \ln f_i(\varepsilon_{t,i}). \quad (20)$$

In what follows we provide specific formulas for the elliptical copula formulation (9) and its main sub-cases.

### 3.2 Parameters in the pdf of the Error Term

Under a copula approach, the portion of the score function corresponding to the term  $\sum_{t=1}^T \ln f(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1})$  (cf. Equation (12)) depends on a vector  $\boldsymbol{\lambda} = (\boldsymbol{\xi}; \boldsymbol{\phi})$  ( $\boldsymbol{\xi}$  and  $\boldsymbol{\phi}$  are the parameters of the copula function and of the marginals, respectively – cf. Section 2),

$$\nabla_{\boldsymbol{\lambda}} l = \sum_{t=1}^T \nabla_{\boldsymbol{\lambda}} \ln f(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \sum_{t=1}^T \nabla_{\boldsymbol{\lambda}} \ln \left( c(\mathbf{u}_t; \boldsymbol{\xi}) \prod_{i=1}^K f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i) \right).$$

Therefore,

$$\nabla_{\boldsymbol{\xi}} l = \sum_{t=1}^T \nabla_{\boldsymbol{\xi}} \ln c(\mathbf{u}_t).$$

and

$$\nabla_{\boldsymbol{\phi}_i} l = \sum_{t=1}^T \left[ \nabla_{\boldsymbol{\phi}_i} F_i(\varepsilon_{t,i}) \nabla_{\mathbf{u}_{t,i}} \ln c(\mathbf{u}_t) + \nabla_{\boldsymbol{\phi}_i} \ln f_i(\varepsilon_{t,i}) \right].$$

As detailed before, beside a possible shape parameter  $\boldsymbol{\nu}$ , elliptical copulas are characterized by a correlation matrix  $\mathbf{R}$  which, in view of its full ML estimation, can be expressed (cf. McNeil *et al.* (2005, p. 235)) as

$$\mathbf{R} = \mathbf{D} \mathbf{c}' \mathbf{c} \mathbf{D}, \quad (21)$$

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  $j = 2, \dots, K$ . By so doing, the estimation of  $\mathbf{R}$  is transformed into an unconstrained problem, since the  $K(K-1)/2$  free elements of  $\mathbf{c}$  can vary into  $\mathbb{R}$ . We can then write  $\boldsymbol{\xi} = (\mathbf{c}; \boldsymbol{\nu})$ .

Let us introduce a compact notation as follows:  $\mathbf{C} = \mathbf{c} \mathbf{D}$ ,  $\mathbf{q}_t = (q_{t,1}; \dots; q_{t,K})$ ,  $q_{t,i} = G^{-1}(u_{t,i}; \boldsymbol{\nu})$ ,  $\tilde{\mathbf{q}}_t = \mathbf{C}'^{-1} \mathbf{q}_t$ ,  $\mathbf{q}_t^* = \mathbf{R}^{-1} \mathbf{q}_t$ ,  $\tilde{\tilde{\mathbf{q}}}_t = \mathbf{q}_t' \mathbf{R}^{-1} \mathbf{q}_t = \tilde{\mathbf{q}}_t' \tilde{\mathbf{q}}_t$ . We can then write

$$\ln c(\mathbf{u}_t) = \ln K^* - \sum_{i=2}^K \ln D_i + \ln g_1(\tilde{\tilde{\mathbf{q}}}_t) - \sum_{i=1}^K \ln g_2(q_{t,i}^2) \quad (22)$$

where we used  $\frac{1}{2} \ln(|\mathbf{R}|) = \sum_{i=2}^K \ln D_i$ .

In specific cases we get:

- Normal copula:  $\ln K^* = 0$ ,  $\ln g_1(x) = \ln g_2(x) = -x/2$ ;
- Student-T copula:  $\ln K^* = \ln \left[ \frac{\Gamma((\nu+K)/2) \Gamma(\nu/2)^{K-1}}{\Gamma(\nu+1)/2} \right]$ ,  $\ln g_1(x) = -\frac{\nu+K}{2} \ln \left( 1 + \frac{x}{\nu} \right)$ ,  $g_2(x) = -\frac{\nu+1}{2} \ln \left( 1 + \frac{x}{\nu} \right)$ .

#### Parameters entering the matrix $\mathbf{c}$

The portion of the score relative to the free parameters of the  $\mathbf{c}$  matrix has elements

$$\nabla_{c_{ij}} l = \nabla_{c_{ij}} \left[ -T \sum_{i=2}^K \ln D_i + \sum_{t=1}^T \ln g_1(\tilde{\tilde{\mathbf{q}}}_t) \right], \quad i < j. \quad (23)$$

Using some algebra we can show that

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



and

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

By replacing them into (23), we obtain

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

### Parameters entering the vector $\nu$

The portion of the score relative to  $\nu$  is

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

The derivative of  $\ln K^* = \ln K^*(\nu; K)$  can sometimes be computed analytically. For instance, for 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  $G^{-1}(x; \nu)$  cannot be computed analytically.

### Parameters entering the vector $\phi$

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

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

After some algebra we obtain

$$\nabla_{\phi_i} l = \sum_{t=1}^T [\nabla_{\phi_i} F_i(\varepsilon_{t,i}) d_{t,i} + \nabla_{\phi_i} \ln f_i(\varepsilon_{t,i})], \quad (24)$$

where

$$d_{t,i} = \frac{1}{g(q_{t,i})} \left[ 2q_{t,i}^* \nabla_{\tilde{q}_t} \ln g_1(\tilde{q}_t) - \nabla_{q_{t,i}} \ln g_2(q_{t,i}^2) \right]. \quad (25)$$

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

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

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

### Parameters entering the vector $\theta$

By exploiting the notation introduced in this section, we can now detail the structure of  $\mathbf{b}_t$  entering into (16) and then into the portion of the score function relative to  $\theta$ . From (22),  $\varepsilon_t \odot \mathbf{b}_t + \mathbf{1}$  (cf. 16) has elements

$$\varepsilon_{t,i} b_{t,i} + 1 = \varepsilon_{t,i} f_i(\varepsilon_{t,i}) d_{t,i} + \varepsilon_{t,i} \nabla_{\varepsilon_{t,i}} \ln f_i(\varepsilon_{t,i}) + 1 \quad (26)$$

where  $d_{t,i}$  is given in (25). For our choice,  $f_i(\varepsilon_{t,i})$  is the pdf of a  $\text{Gamma}(\phi_i, \phi_i)$  distribution, so that

$$\varepsilon_{t,i} \nabla_{\varepsilon_{t,i}} \ln f_i(\varepsilon_{t,i}) + 1 = \phi_i - \varepsilon_{t,i} \phi_i. \quad (27)$$

### 3.2.1 Expectation Targeting

Assuming weak-stationarity of the process, numerical stability and a reduction in the number of parameters to be estimated can be achieved by expressing  $\boldsymbol{\omega}$  in terms of the unconditional mean of the process, say  $\boldsymbol{\mu}$ , which can be easily estimated by the sample mean (*expectation targeting*<sup>2</sup>). Since  $E(\mathbf{x}_t) = E(\boldsymbol{\mu}_t) = \boldsymbol{\mu}$  is well defined and is equal to

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

Equation (3) becomes

$$\boldsymbol{\mu}_t = \left[ \mathbf{I} - \sum_{l=1}^L \left( \boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \frac{\boldsymbol{\gamma}_l}{2} \right) \right] \boldsymbol{\mu} + \sum_{l=1}^L \left( \boldsymbol{\alpha}_l \mathbf{x}_{t-l} + \boldsymbol{\gamma}_l \mathbf{x}_{t-l}^{(-)} + \boldsymbol{\beta}_l \boldsymbol{\mu}_{t-l} \right). \quad (29)$$

In a GARCH framework, the consequences of this practice have been investigated by Kristensen and Linton (2004) and, more recently, by Francq *et al.* (2011) who show its merits in terms of stability of estimation algorithms and accuracy in estimation of both coefficients and long term volatility (cf. Appendix A for some details in the present context).

From a technical point of view, a trivial replacement of  $\boldsymbol{\mu}$  by the sample mean  $\bar{\mathbf{x}}_T$  followed by a ML estimation of the remaining parameters, preserves consistency but leads to wrong standard errors (Kristensen and Linton (2004)). The issue can be faced by identifying the inference problem as involving a two-step estimator (Newey and McFadden (1994, ch. 6)), namely by rearranging  $(\boldsymbol{\theta}; \boldsymbol{\lambda})$  as  $(\boldsymbol{\mu}; \boldsymbol{\vartheta})$ , where  $\boldsymbol{\vartheta}$  collects all model parameters but  $\boldsymbol{\mu}$ . Under conditions able to guarantee consistency and asymptotic normality of  $(\boldsymbol{\mu}; \boldsymbol{\vartheta})$  (in particular, the existence of  $E(\boldsymbol{\mu}_t \boldsymbol{\mu}_t')$ : see Francq *et al.* (2011)), we can adapt the notation of Newey and McFadden (1994) to write the asymptotic variance of  $\sqrt{T} (\hat{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta})$  as

$$\mathbf{G}_{\boldsymbol{\vartheta}}^{-1} \left[ \mathbf{I} \quad -\mathbf{G}_{\boldsymbol{\mu}} \mathbf{M}^{-1} \right] \begin{bmatrix} \boldsymbol{\Omega}_{\boldsymbol{\vartheta}, \boldsymbol{\vartheta}'} & \boldsymbol{\Omega}_{\boldsymbol{\vartheta}, \boldsymbol{\mu}'} \\ \boldsymbol{\Omega}_{\boldsymbol{\mu}, \boldsymbol{\vartheta}'} & \boldsymbol{\Omega}_{\boldsymbol{\mu}, \boldsymbol{\mu}'} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ -(\mathbf{G}_{\boldsymbol{\mu}} \mathbf{M}^{-1})' \end{bmatrix} \mathbf{G}_{\boldsymbol{\vartheta}}^{-1'} \quad (30)$$

where

$$\begin{aligned} \mathbf{G}_{\boldsymbol{\vartheta}} &= E(\nabla_{\boldsymbol{\vartheta}'} l_t) \\ \mathbf{G}_{\boldsymbol{\mu}} &= E(\nabla_{\boldsymbol{\vartheta}'} l_t) \\ \mathbf{M} &= E(\nabla_{\boldsymbol{\mu}'} \mathbf{m}_t), \end{aligned}$$

and

$$\mathbf{m} = \sum_{t=1}^T \mathbf{m}_t = \sum_{t=1}^T (\mathbf{x}_t - \boldsymbol{\mu})$$

is the moment function giving the sample average  $\bar{\mathbf{x}}_T$  as an estimator of  $\boldsymbol{\mu}$ . The  $\boldsymbol{\Omega}$  matrix denotes the variance-covariance matrix of  $(\nabla_{\boldsymbol{\vartheta}'} l_t; \mathbf{m}_t)$  partitioned in the corresponding blocks.

---

<sup>2</sup>This is equivalent to *variance targeting* in a GARCH context (Engle and Mezrich (1995)), where the constant term of the conditional variance model is assumed to be a function of the sample unconditional variance and of the other parameters. In this context, other than a preference for the term expectation targeting since we are modeling a conditional mean, the main argument stays unchanged.

To give account of the necessary modifications to adopt with expectation targeting, we provide sufficiently general and compact expressions for the parameters  $\boldsymbol{\mu}$  (the unconditional mean) and  $\boldsymbol{\theta}$  (the remaining parameters) in the conditional mean expressed by (29).

$$\begin{aligned}
\mathbf{G}_\theta &= E(\nabla_{\theta\theta'} l_t) = E[\mathbf{A}_t \mathbf{H}^{(\varepsilon)} \mathbf{A}_t'] \\
\mathbf{G}_\mu &= E(\nabla_{\theta\mu'} l_t) = -E(\mathbf{A}_t \mathbf{H}^{(\varepsilon)} \text{diag}(\boldsymbol{\mu}_t)^{-1}) \mathbf{A} \\
\mathbf{M} &= -\mathbf{I} \\
\boldsymbol{\Omega}_{\theta,\theta'} &= E(\nabla_{\theta} l_t \nabla_{\theta'} l_t) = E[\mathbf{A}_t \boldsymbol{\mathcal{I}}^{(\varepsilon)} \mathbf{A}_t'] \\
\boldsymbol{\Omega}_{\theta,\mu'} &= E(\nabla_{\theta} l_t \mathbf{m}_t') = E[\mathbf{A}_t [E[(\mathbf{b}_t \odot \boldsymbol{\varepsilon}_t) \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1}] + \mathbf{1}\mathbf{1}'] \text{diag}(\boldsymbol{\mu}_t)] \mathbf{B} \mathbf{A}^{-1'} \\
\boldsymbol{\Omega}_{\mu,\mu'} &= E(\mathbf{m}_t \mathbf{m}_t') = \mathbf{A}^{-1} (\mathbf{B} \boldsymbol{\Sigma}_v \mathbf{B}' + \mathbf{C} (\boldsymbol{\Sigma}_v \odot \boldsymbol{\Sigma}_I) \mathbf{C}') \mathbf{A}^{-1'}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{A} &= \mathbf{I} - \sum_{l=1}^L \left( \boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \frac{\boldsymbol{\gamma}_l}{2} \right) \\
\mathbf{B} &= \mathbf{I} - \sum_{l=1}^L \boldsymbol{\beta}_l \\
\mathbf{C} &= \sum_{l=1}^L \boldsymbol{\gamma}_l \\
\boldsymbol{\Sigma}_v &= E(\boldsymbol{\mu}_t \boldsymbol{\mu}_t') \odot \boldsymbol{\Sigma}.
\end{aligned}$$

The expression for  $\boldsymbol{\Omega}_{\mu,\mu'}$  is obtained by using the technique in Horváth *et al.* (2006) and Francq *et al.* (2011). In this sense, we extend the cited works to a multivariate formulation including asymmetric effects as well.

Some further simplification is also possible when the model is correctly specified since, in such a case,  $\mathbf{H}^{(\varepsilon)} = -\boldsymbol{\mathcal{I}}^{(\varepsilon)}$  and  $E[(\mathbf{b}_t \odot \boldsymbol{\varepsilon}_t) \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1}] = -\mathbf{1}\mathbf{1}'$  leading to  $\boldsymbol{\Omega}_{\theta,\mu'} = \mathbf{0}$  and to

$$E[\mathbf{A}_t \boldsymbol{\mathcal{I}}^{(\varepsilon)} \mathbf{A}_t'] + E(\mathbf{A}_t \boldsymbol{\mathcal{I}}^{(\varepsilon)} \text{diag}(\boldsymbol{\mu}_t)^{-1}) [\boldsymbol{\Sigma}_v \mathbf{B}' + \mathbf{C} (\boldsymbol{\Sigma}_v \odot \boldsymbol{\Sigma}_I) \mathbf{C}'] E(\text{diag}(\boldsymbol{\mu}_t)^{-1} \boldsymbol{\mathcal{I}}^{(\varepsilon)} \mathbf{A}_t')$$

for what concerns the inner part of (30).

### 3.2.2 Concentrated Log-likelihood

Some further numerical estimation stability and reduction in the number of parameters can be achieved – if needed – within the framework of elliptical copulas: we can use current values of residuals to compute current estimates of  $\mathbf{R}$  (Kendall correlations are suggested by Lindskog *et al.* (2003)) and of the shape parameter  $\boldsymbol{\nu}$  (tail dependence indices are proposed by Kostadinov (2005)). This approach may be fruitful with other copulas as well when sufficiently simple moment conditions can be exploited. A similar strategy can be applied also to the parameters of the marginals. For instance, if they are assumed *Gamma*( $\phi_i, \phi_i$ ) distributed, the relationship  $V(\varepsilon_{t,i} | \mathcal{F}_{t-1}) = 1/\phi_i$  leads to very simple estimator of  $\phi_i$  from current values of residuals. By means of this approach, the remaining parameters can be updated from a sort of pseudo-loglikelihood conditioned on current estimates of the pre-estimated parameters.

In the case of a Normal copula a different strategy can be followed. The formula of the (unconstrained) ML estimator of the  $\mathbf{R}$  matrix (Cherubini *et al.* (2004, p. 155)), namely

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

where  $\mathbf{q} = (\mathbf{q}'_1; \dots; \mathbf{q}'_T)$  is a  $T \times K$  matrix, can be plugged into the log-likelihood in place of  $\mathbf{R}$  obtaining a sort of *concentrated log-likelihood*

$$\frac{T}{2} [-\ln |\mathbf{Q}| - K + \text{trace}(\mathbf{Q})] + \sum_{t=1}^T \sum_{i=1}^K \ln f_i(\varepsilon_{t,i} | \mathcal{F}_{t-1}) \quad (31)$$

leading to a relatively simple structure of the score function. However, this estimator of  $\mathbf{R}$  is obtained without imposing any constraint relative to its nature as a correlation matrix ( $\text{diag}(\mathbf{R}) = \mathbf{1}$  and positive definiteness). Computing directly the derivatives with respect to the off-diagonal elements of  $\mathbf{R}$  we obtain, after some algebra, that the constrained ML estimator of  $\mathbf{R}$  satisfies the following equations:

$$(\mathbf{R}^{-1})_{ij} - (\mathbf{R}^{-1})_i \frac{\mathbf{q}'\mathbf{q}}{T} (\mathbf{R}^{-1})_j = 0$$

for  $i \neq j = 1, \dots, K$ , where  $\mathbf{R}_i$  and  $\mathbf{R}_j$  indicate, respectively, the  $i$ -th row and the  $j$ -th column of the matrix  $\mathbf{R}$ . Unfortunately, these equations do not have an explicit solution.<sup>3</sup> An acceptable compromise which should increase efficiency, although formally it cannot be interpreted as an ML estimator, is to normalize the estimator  $\mathbf{Q}$  obtained above in order to transform it to a correlation matrix:

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

where  $\mathbf{D}_Q = \text{diag}(Q_{11}, \dots, Q_{KK})$ . This solution can be justified observing that the copula contribution to the likelihood depends on  $\mathbf{R}$  exactly as if it was the correlation matrix of iid rv's  $\mathbf{q}_t$  normally distributed with mean  $\mathbf{0}$  and correlation matrix  $\mathbf{R}$  (see also McNeil *et al.* (2005, p. 235)). Using this constrained estimator of  $\mathbf{R}$ , the concentrated log-likelihood becomes

$$\frac{T}{2} [-\ln |\tilde{\mathbf{R}}| - \text{trace}(\tilde{\mathbf{R}}^{-1} \mathbf{Q} + \text{trace}(\mathbf{Q}))] + \sum_{t=1}^T \sum_{i=1}^K \ln f_i(\varepsilon_{t,i} | \mathcal{F}_{t-1}). \quad (32)$$

It is interesting to note that, as long as (31), Equation (32) too gives a relatively simple structure of the score function. Using some tedious algebra, we can show that the components of the score corresponding to  $\boldsymbol{\theta}$  and  $\boldsymbol{\phi}$  have exactly the same structure as above, with the quantity  $d_{t,i}$  into (25) changed to

$$d_{t,i} = \frac{\mathbf{C}_i \mathbf{q}_t}{\phi(q_{t,i})} \quad (33)$$

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

$$\mathbf{C} = \mathbf{Q}^{-1} \mathbf{D}_Q^{1/2} \mathbf{Q} \mathbf{D}_Q^{1/2} \mathbf{Q}^{-1} - \mathbf{Q}^{-1} + \mathbf{I}_K - \tilde{\mathbf{R}}^{-1} + \mathbf{D}_Q^{-1} - \mathbf{D}_Q^{-1/2} \text{diag}(\mathbf{Q}^{-1} \mathbf{D}_Q^{1/2} \mathbf{Q})$$

and  $\phi(\cdot)$  indicates here the pdf of the standard normal computed at its argument. Of course, also in this case the parameters of the marginals can be updated by means of moment estimators computed from current residuals (instead that via ML) exactly as explained above.

---

<sup>3</sup>Even when  $\mathbf{R}$  is a (2, 2) matrix, the value of  $\mathbf{R}_{12}$  has to satisfy the cubic equation:

$$\mathbf{R}_{12}^3 - \mathbf{R}_{12}^2 \frac{q'_1 q_2}{T} + \mathbf{R}_{12} \left[ \frac{q'_1 q_1}{T} + \frac{q'_2 q_2}{T} - 1 \right] - \frac{q'_1 q_2}{T} = 0.$$

### 3.3 Asymptotic Variance-Covariance Matrix

As customary in ML inference, the asymptotic variance-covariance matrix of parameter estimators can be expressed by using the Outer-Product of Gradients (OPG), the Hessian matrix, or in the sandwich form, more robust toward error distribution misspecification. The last option is what we use in the application of Section 5.

## 4 Alternative Specifications of the Distribution of Errors

### 4.1 Multivariate Gamma

The generalization of the univariate gamma adopted by Engle and Gallo (2006) to a multivariate counterpart is frustrated by the limitations of the multivariate Gamma distributions available in the literature (all references below come from Johnson *et al.* (2000, chapter 48)): many of them are bivariate formulations, not sufficiently general for our purposes; others are defined via the joint characteristic function, so that they require tedious numerical inversion formulas to find their pdf's. The formulation that is closest to our needs (it provides all univariate marginal probability functions for  $\varepsilon_{i,t}$  as  $Gamma(\phi_i, \phi_i)$ ), is a particular version of the multivariate Gamma's by Cheriyan and Ramabhadran (henceforth *GammaCR*, which is equivalent to other versions by Kowalekzyk and Trycha and by Mathai and Moschopoulos):

$$\varepsilon_t | \mathcal{F}_{t-1} \sim GammaCR(\phi_0, \boldsymbol{\phi}, \boldsymbol{\phi}),$$

where  $\boldsymbol{\phi} = (\phi_1; \dots; \phi_K)$  and  $0 < \phi_0 < \min(\phi_1, \dots, \phi_K)$  (Johnson *et al.* (2000, 454–470)). The multivariate pdf is expressed in terms of a cumbersome integral and the conditional correlations matrix of  $\boldsymbol{\varepsilon}_t$  has generic element

$$\rho(\varepsilon_{t,i}, \varepsilon_{t,j} | \mathcal{F}_{t-1}) = \frac{\phi_0}{\sqrt{\phi_i \phi_j}},$$

which is restricted to be positive and is strictly related to the variances  $1/\phi_i$  and  $1/\phi_j$ . Given these drawbacks, Multivariate Gamma's will not be adopted here.

### 4.2 Multivariate Lognormal

A different specification is obtained assuming  $\boldsymbol{\varepsilon}_t$  conditionally log-Normal,

$$\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1} \sim LN(\boldsymbol{m}, \mathbf{V}) \quad \text{with} \quad \boldsymbol{m} = -0.5 \text{diag}(\mathbf{V}) \quad (34)$$

where  $\boldsymbol{m}$  and  $\mathbf{V}$  are, respectively, the mean vector and the variance-covariance matrix of  $\ln \boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}$  and the constraint in (34) is needed in order for  $E(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1})$  to be equal to  $\mathbf{1}$  (cf. Equation (5)).

The log-likelihood function is obtained by replacing the expression of  $\ln f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1})$  coming from (34) into Equation (12), leading to

$$l = - \sum_{t=1}^T \sum_{i=1}^K \ln x_{t,i} - \frac{TK}{2} \ln(2\pi) - \frac{T}{2} \ln |\mathbf{V}| - \frac{1}{2} \sum_{t=1}^T (\ln \boldsymbol{\varepsilon}_t - \boldsymbol{m})' \mathbf{V}^{-1} (\ln \boldsymbol{\varepsilon}_t - \boldsymbol{m}). \quad (35)$$

Accordingly, the portion of the score related to the derivative of  $l$  in  $\boldsymbol{\theta}$  (the parameters in the conditional mean) is given by Equation (14), with  $\mathbf{A}_t$  defined as in (15) and

$$\boldsymbol{w}_t = \mathbf{V}^{-1} (\ln \boldsymbol{\varepsilon}_t - \boldsymbol{m}), \quad (36)$$

leading to

$$\sum_{t=1}^T \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}'_t \text{diag}(\boldsymbol{\mu}_t)^{-1} \mathbf{V}^{-1} (\ln \boldsymbol{\varepsilon}_t - \mathbf{m}) = \mathbf{0}. \quad (37)$$

A full application of the ML principle would require optimization of (35) also in the elements of the matrix  $\mathbf{V}$ . However, differently from the maximization with respect to the same parameters of a Gaussian log-likelihood, this approach is tedious since the diagonal elements of  $\mathbf{V}$  appear also in the term  $\mathbf{m}$ , complicating the constraints to preserve its positive definiteness.

A possible approach to bypass this issue would be to resort to a reparameterization of  $\mathbf{V}$  based on its Cholesky decomposition, as we did in Section 3 to represent the correlation matrix  $\mathbf{R}$  in the copula-based specifications. Here we adopted a simpler approach, based on a method of moment estimator of the  $\mathbf{V}$  elements. The ensuing estimation algorithm follows the sequence:

- We compute log-residuals  $\ln \hat{\boldsymbol{\varepsilon}}_t$  ( $t = 1, \dots, T$ ), according to the first two steps of (13), at the current parameter estimates;
- We then estimate their variance-covariance matrix  $\mathbf{V}$ ;
- We estimate  $\mathbf{m} = -0.5 \text{diag}(\mathbf{V})$ ;
- We use Equation (37) to update the estimate of  $\boldsymbol{\theta}$ ,

iterating until convergence.

Following (Newey and McFadden (1994, ch. 6)) once again, the resulting asymptotic variance-covariance matrix of  $\hat{\boldsymbol{\theta}}$  can be expressed as

$$\text{avar}(\hat{\boldsymbol{\theta}}) = \mathbf{G}_{\boldsymbol{\theta}\boldsymbol{\theta}'}^{-1} \left( \mathbf{I} \quad \mathbf{G}'_{\boldsymbol{\lambda}\boldsymbol{\theta}'} \right) \begin{pmatrix} \boldsymbol{\Omega}_{\boldsymbol{\theta}\boldsymbol{\theta}'} & \boldsymbol{\Omega}_{\boldsymbol{\theta}\boldsymbol{\lambda}'} \\ \boldsymbol{\Omega}'_{\boldsymbol{\theta}\boldsymbol{\lambda}'} & \boldsymbol{\Omega}_{\boldsymbol{\lambda}\boldsymbol{\lambda}'} \end{pmatrix} \begin{pmatrix} \mathbf{I} \\ \mathbf{G}_{\boldsymbol{\lambda}\boldsymbol{\theta}'} \end{pmatrix} \mathbf{G}_{\boldsymbol{\theta}\boldsymbol{\theta}'}^{-1} \quad (38)$$

where

$$\begin{aligned} \mathbf{G}_{\boldsymbol{\theta}\boldsymbol{\theta}'} &= \lim_{T \rightarrow \infty} E \left[ \nabla_{\boldsymbol{\theta}} \left( T^{-1} \sum_{t=1}^T \mathbf{g}_t \right) \right] \\ \mathbf{G}_{\boldsymbol{\lambda}\boldsymbol{\theta}'} &= \lim_{T \rightarrow \infty} E \left[ \nabla_{\boldsymbol{\lambda}} \left( T^{-1} \sum_{t=1}^T \mathbf{g}_t \right) \right] \\ \boldsymbol{\Omega}_{\boldsymbol{\theta}\boldsymbol{\theta}'} &= \lim_{T \rightarrow \infty} V \left( T^{-1/2} \sum_{t=1}^T \mathbf{g}_t \right) \\ \boldsymbol{\Omega}_{\boldsymbol{\theta}\boldsymbol{\lambda}'} &= \lim_{T \rightarrow \infty} \text{Cov} \left( T^{-1/2} \sum_{t=1}^T \mathbf{g}_t, T^{1/2} (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \right) \\ \boldsymbol{\Omega}_{\boldsymbol{\lambda}\boldsymbol{\lambda}'} &= \lim_{T \rightarrow \infty} V \left( T^{1/2} (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \right) \end{aligned}$$

$\mathbf{g}_t$  is the  $t$ -th addend in (37) and  $\boldsymbol{\lambda} = \text{vech}(\mathbf{V})$  is estimated with the corresponding portion of the variance-covariance matrix of log-residuals  $\ln \hat{\boldsymbol{\varepsilon}}_t$  ( $t = 1, \dots, T$ ). After some algebra

we get

$$\mathbf{G}_{\theta\theta'} = -\mathbf{\Omega}_{\theta\theta'} = -\lim_{T \rightarrow \infty} \left[ T^{-1} \sum_{t=1}^T E(\mathbf{B}_t \mathbf{V}^{-1} \mathbf{B}_t') \right] \quad (39)$$

$$\begin{aligned} \mathbf{G}_{\lambda\theta'} &= -\frac{1}{2} \lim_{T \rightarrow \infty} \left[ T^{-1} \sum_{t=1}^T \nabla_{\lambda} \text{diag}(\mathbf{V}) \mathbf{V}^{-1} E(\mathbf{B}_t) \right] \\ \mathbf{\Omega}_{\theta\lambda'} &= \mathbf{0} \end{aligned} \quad (40)$$

$$\mathbf{\Omega}_{\lambda\lambda'} \left\{ i + (j-1) \left( K - \frac{j}{2} \right), h + (k-1) \left( K - \frac{k}{2} \right) \right\} = V_{i,j} V_{h,k} - V_{i,h} V_{j,k} \quad (41)$$

$i = 1, \dots, K; j = i, \dots, K$

where  $\mathbf{B}_t = \nabla_{\theta'} \boldsymbol{\mu}_t \text{diag}(\boldsymbol{\mu}_t)^{-1}$  and  $\mathbf{\Omega}_{\lambda\lambda'} \{u, v\}$  indicates the  $(u, v)$ -th element of  $\mathbf{\Omega}_{\lambda\lambda'}$ .<sup>4</sup> Because of (39) and (40), Equation (38) simplifies to

$$\text{avar}(\hat{\boldsymbol{\theta}}) = \mathbf{\Omega}_{\theta\theta'}^{-1} + \mathbf{\Omega}_{\theta\theta'}^{-1} \mathbf{G}'_{\lambda\theta'} \mathbf{\Omega}_{\lambda\lambda'} \mathbf{G}_{\lambda\theta'} \mathbf{\Omega}_{\theta\theta'}^{-1}$$

### 4.3 Semiparametric

As an alternative, we can use the semiparametric approach by Cipollini *et al.* (2013), where the distribution of the error term  $\boldsymbol{\varepsilon}_t$  is left unspecified. In this case the pseudo-score equation corresponding to the  $\boldsymbol{\theta}$  parameter is again given by Equation (14), with  $\mathbf{A}_t$  defined as in (15) and

$$\mathbf{w}_t = \boldsymbol{\Sigma}^{-1} (\boldsymbol{\varepsilon}_t - \mathbf{1}). \quad (42)$$

leading to the equation

$$\sum_{t=1}^T \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' \text{diag}(\boldsymbol{\mu}_t)^{-1} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\varepsilon}_t - \mathbf{1}) = \mathbf{0}. \quad (43)$$

The ensuing estimation algorithm works similarly to the log-Normal case iterating until convergence the following sequence:

- We compute residuals  $\hat{\boldsymbol{\varepsilon}}_t$  ( $t = 1, \dots, T$ ), according the first two steps of (13), at the current parameter estimates;
- We estimate  $\boldsymbol{\Sigma}$  as their variance-covariance matrix;
- We update the estimate of  $\boldsymbol{\theta}$  using Equation (43).

The asymptotic variance-covariance matrix of  $\hat{\boldsymbol{\theta}}$  is given by

$$\text{avar}(\hat{\boldsymbol{\theta}}) = \left[ \lim_{T \rightarrow \infty} \left[ T^{-1} \sum_{t=1}^T E(\mathbf{B}_t \boldsymbol{\Sigma}^{-1} \mathbf{B}_t') \right] \right]^{-1}.$$

the form of which looks very similar to  $\mathbf{\Omega}_{\theta\theta'}^{-1}$  in Equation (39).

---

<sup>4</sup>(40) and (41) come from standard properties of the multivariate Normal distribution: the former is implied by the fact that all centered odd moments of  $\ln \boldsymbol{\varepsilon}_t$  are zero; the latter is implied by the structure of fourth centered moments of  $\ln \boldsymbol{\varepsilon}_t$ .

The (pseudo-)score Equations (37) and (43) look quite similar: beside a simpler expression of the GMM asymptotic variance-covariance matrix of  $\hat{\boldsymbol{\theta}}$ , the essential difference rests in the nature of the martingale difference (MD)  $\boldsymbol{w}_t$  ‘driving’ the equation. The GMM  $\boldsymbol{w}_t$  relies on minimal model assumptions and does not require logs; as a consequence, it is feasible also in case of zeros in the data. Apart from the need to take logs, the expression of  $\boldsymbol{w}_t$  in the log–Normal case is a MD only if  $\ln \boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}$  has  $-0.5 \text{diag}(\mathbf{V})$  as its mean; if this is not true, we would have a corresponding bias in estimating the  $\boldsymbol{\omega}$  coefficient appearing in the conditional mean  $\boldsymbol{\mu}_t$ .

## 5 Trading Activity and Volatility within a vMEM

Trading activity produces a lot of indicators characterized by both simultaneous and dynamic interdependence. In this application, we concentrate on the joint dynamics of three series stemming from such trading activity, namely volatility (measured as realized kernel volatility, cf. Barndorff-Nielsen *et al.* (2011), and references therein), volume of shares traded and number of shares per day. The relationship between volatility and volume as relating to trading activity was documented, for example, in the early contribution by Andersen (1996). Ever since, the evolution of the structure of financial markets, industry innovation, the increasing participation of institutional investors and the adoption of automated trading practices have strengthened such a relationship, and the number of trades clearly reflects an important aspect of trading strategies. For the sake of parsimony, we choose to focus on the realized volatility forecasts as the main product of interest from the multivariate model exploiting the extra information in the other indicators, as well as the contemporaneous correlation in the error terms.

The availability of ultra high frequency data allows us to construct daily series of the variables exploiting the most recent development in the volatility measurement literature.

As a leading example, we consider Johnson & Johnson (JNJ) between January 3, 2007 to June 30, 2013 (1886 observations). Such a stock has desirable properties of liquidity and a limited riskiness represented by a market beta generally smaller than 1. Raw trade data from TAQ are cleaned according to the Brownlees and Gallo (2006) algorithm. Subsequently, we build the daily series of realized kernel volatility, following Barndorff-Nielsen *et al.* (2011), computing the value at day  $t$  as

$$rkv = \sqrt{\sum_{h=-H}^H k\left(\frac{h}{H}\right) \gamma_h} \quad \gamma_h = \sum_{j=|h|+1}^n x_j x_{j-|h|}$$

where  $k(x)$  is the Parzen kernel

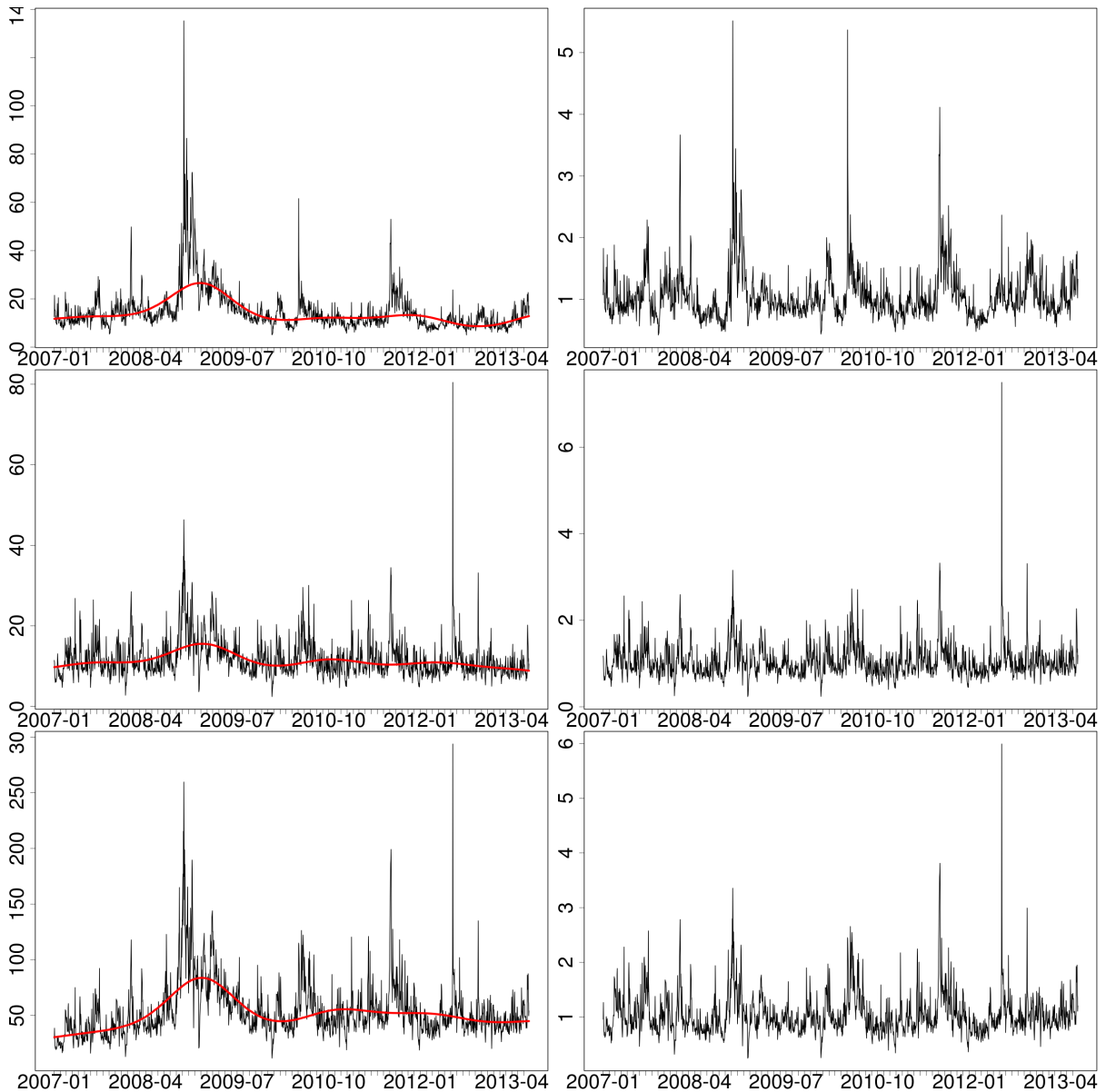
$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & \text{if } x \in [0, 1/2] \\ 2(1 - x)^3 & \text{if } x \in (1/2, 1] \\ 0 & \text{otherwise} \end{cases} ,$$

$$H = 3.51 \cdot n^{3/5} \left( \frac{\sum_{j=1}^n x_j^2 / (2n)}{\sum_{j=1}^{\tilde{n}} \tilde{x}_j^2} \right)^{2/5} ,$$

$x_j$  is the  $j$ -th high frequency return computed according to Barndorff-Nielsen *et al.* (2009, Section 2.2)) and  $\tilde{x}_j$  is the intradaily return of the  $j$ -th bin (equally spaced on 15 minute intervals). For volumes ( $vol$ ), that is the total number of shares traded during the day, and the number of trades ( $nt$ ) executed during the day (irrespective of their size), we



Figure 1: Time series of the trading activity indices for JNJ (Jan. 3, 2007 – June 30, 2013). Left: original data; Right: data after removing the low frequency component. Top: realized kernel volatility (annualized percentage); Middle: volumes (millions); Bottom: number of trades (thousands). The spike on June 13, 2012 corresponds to an important acquisition and a buy back of some of its common stock by a subsidiary.



simply aggregate the data (sum of intradaily volumes and count of the number of trades, respectively).

According to Figure 1, the turmoil originating with the subprime mortgage crisis is clearly affecting the profile of the series with an underlying low frequency component. For volatility, the presence of a changing average level was analyzed by Gallo and Otranto (2015) with a number of non-linear MEM's. Without adopting their approach, in what follows we implement a filter (separately for each indicator) in order to identify short term interactions among series apart from lower frequency movements. The presence of an upward slope in the first portion of the sample is apparent and is reminiscent of the evidence produced by Andersen (1996) for volumes on an earlier period. Remarkably, this upward movement is interrupted after the peak of the crisis in October 2008, with a substantial and progressive reduction of the average level of the variables. To remove

this pattern, we adopt a solution similar to Andersen (1996), that is, a flexible function of time which smooths out the series. Extracting low frequency movements in a financial market activity series with a spline is reminiscent of the stream of literature initiated by Engle and Rangel (2008) with a spline-GARCH and carried out in Brownlees and Gallo (2010) within a MEM context.

In detail, assuming that this component is multiplicative, we remove it in each indicator as follows:

- we take the log of the original series;
- we fit on each log-series a spline based regression with additive errors, using time  $t$  (a progressive counter from the first to the last observation in the sample) as an independent variable;<sup>5</sup>
- the residuals of the previous regression are then exponentiated to get the adjusted series.

When used to produce out-of-sample forecasts of the original quantities, the described approach is applied assuming that the low frequency component remains constant at the last in-sample estimate. This strategy is simple to implement and fairly reliable for forecasting at moderate horizons.

Table 1 shows some similarities across original and adjusted series, with some expected increase in the correlation between volumes and number of trades in the adjusted series (due to a smaller correlation in the corresponding low frequency components). Although still quite large overall, after adjustment, correlations are more spread apart: as mentioned, the value concerning volumes and the number of trades, above 0.9, is the highest one;  $cor(rkv, vol)$  is about 0.6, whereas the  $cor(rkv, nt)$  is still above 0.7, all confirming the intuition that the information contained in other variables related to trading activity relevant.

Table 1: Correlations for JNJ (Jan. 3, 2007 – June 30, 2013).  $rkv$  = realized kernel volatility;  $vol$  = volume;  $nt$  = number of trades.

|       | Original |       | Low Freq. Comp. |       | Adjusted |       |
|-------|----------|-------|-----------------|-------|----------|-------|
|       | $vol$    | $nt$  | $vol$           | $nt$  | $vol$    | $nt$  |
| $rkv$ | 0.646    | 0.778 | 0.800           | 0.847 | 0.609    | 0.743 |
| $vol$ |          | 0.890 |                 | 0.780 |          | 0.932 |

## 5.1 Modeling Results

In the application, we consider a vMEM on adjusted data, where the conditional expectation  $\mu_t$  has the form (cf. Equation (3))

$$\mu_t = \omega + \alpha_1 \mathbf{x}_{t-1} + \alpha_2 \mathbf{x}_{t-2} + \gamma_1 \mathbf{x}_{t-1}^{(-)} + \beta_1 \mu_{t-1}. \quad (44)$$

In order to appreciate the contribution of the different model variables, we consider alternative specifications for the coefficient matrices  $\alpha_1$  and  $\beta_1$  ( $\alpha_2$  and  $\gamma_1$  are kept diagonal in all specifications), and for the error term.

<sup>5</sup>Alternative methods, such as a moving average of fixed length (centered or uncentered), can be used but in practice they deliver very similar results and will not be discussed in detail here. The spline regression is estimated with the `gam()` function in the R package `mgcv` by using default settings.

As of the former, we consider specifications with both  $\alpha_1$  and  $\beta_1$  diagonal (labeled  $D$ );  $\alpha_1$  full and  $\beta_1$  diagonal (labeled  $A$ );  $\alpha_1$  diagonal and  $\beta_1$  full (labeled  $B$ ); both  $\alpha_1$  and  $\beta_1$  full (labeled  $AB$ ). For the joint distribution of the errors, we adopt a Student–T, a Normal and an Independent copula ( $T$ ,  $N$  and  $I$  as respective labels), in all cases with Gamma distributed marginals. As a comparison, the  $AB$  specification is also estimated by ML with log–Normal (AB-LN) and with GMM in the semiparametric AB-S. The estimated specifications are summarized in Table 2. When coupled with the conditional means in the table, the specifications with the Independent copula can be estimated equation–by–equation.

Table 2: Estimated specifications of the vMEM defined by (1) and (44), with error distribution specified, alternatively, by (8), or (34), or (5). Notation:  $I$  for Independent copula with Gamma marginals;  $N$  for Normal copula with Gamma marginals;  $T$  for Student-T copula with Gamma marginals;  $LN$  for log–Normal;  $S$  for semiparametric.

| Conditional Mean (parameters)                        | Error Distribution |      |      |       |      |
|--|--------------------|------|------|-------|------|
|  | $I$                | $N$  | $T$  | $LN$  | $S$  |
| $D$ : $\alpha_1, \beta_1, \gamma_1$ , diagonal       | D-I                |      |      |       |      |
| $A$ : $\alpha_1$ full; $\beta_1, \gamma_1$ diagonal  | A-I                | A-N  | A-T  |       |      |
| $B$ : $\beta_1$ full; $\alpha_1, \gamma_1$ diagonal  |                    | B-N  | B-T  |       |      |
| $AB$ : $\alpha_1, \beta_1$ full; $\gamma_1$ diagonal |                    | AB-N | AB-T | AB-LN | AB-S |

Estimation results are reported in Table 3, limiting ourselves to the equation for the realized volatility.<sup>6</sup> Among copula-based specifications, the Student-T copula turns out to be the favorite version, judging upon a significantly higher log–likelihood function value, and lower information criteria;<sup>7</sup> the equation–by–equation approach (Independent copula) is dominated in both respects.<sup>8</sup>

An estimation of the Diagonal model with the Normal/Student-T copula function (details not reported) shows log-likelihood values of 1947.91, respectively, 2034.07 (relative to the base value of 205.53 in the first column), pointing to both a substantial improvement coming from the contemporaneous correlation of the innovations and to the joint significance of the other indicators when the  $A$  specification is adopted. The comparison of the  $A$  ( $\alpha_1$  full;  $\beta_1$  and  $\gamma_1$  diagonal) versus the  $B$  ( $\beta_1$  full;  $\alpha_1$  and  $\gamma_1$  diagonal) formulations, seems to indicate a dominance of the former, at least judging upon the overall log-likelihood values. It is also interesting to note that model residual autocorrelation is substantially reduced only in the case of richer parameterizations ( $AB$ ), where non–diagonality in both  $\alpha_1$  and  $\beta_1$  captures possible interdependencies, with a sharp improvement in the LB diagnostics (although only marginally satisfying).

In general, it seems that the more relevant contribution in determining realized volatility comes from the lagged number of trades, but only in the  $AB$  models (some collinearity

<sup>6</sup>All models are estimated using Expectation Targeting (Section 3.2.1). The Normal copula based specifications are estimated resorting to the concentrated log-likelihood approach (Section 3.2.2). We omit estimates of the constant term  $\omega$ .

<sup>7</sup>The estimated degrees of freedom are 8.53 (s.e. 1.16) and 8.18 (s.e. 1.05), respectively, in the A-T and AB-T formulations. We also tried full ML estimation of the AB-N specification getting a value of the log-likelihood equal to 2046.57, very close to the concentrated log-likelihood approach (Section 3.2.2) used in Table 3.

<sup>8</sup>The stark improvement in the likelihood functions, coming from the explicit consideration of the correlation structure, does not guarantee similar gains in the forecasting ability of the same variables. This is similar to what happens in modeling returns, when the likelihood function of ARMA models is improved by superimposing a GARCH structure on the conditional variances: no substantive change in the fit of the conditional mean and no better predictive ability.

effect, diminishing individual significance, is to be expected); the relevance of such information is also highlighted by the Granger Causality tests, to check whether lagged model components attributable to either *vol* or *nt* have joint significance. In detail, using three indices where the first refers to the LHS variable, the second to the conditioning variable, and the third the lag, we test  $H_0 : \alpha_{1,j,1} = 0$  ( $j = 2, 3$ ) in the *A*-based formulations;  $H_0 : \beta_{1,j,1} = 0$  ( $j = 2, 3$ ) in the *B*-based formulations;  $H_0 : \alpha_{1,j,1} = \beta_{1,j,1} = 0$  ( $j = 2, 3$ ) in the *AB*-based formulations.

The own volatility coefficients at lag 2 are always significant (with negative signs); surprisingly, the leverage effect related to the sign of the returns is not.

The Normal and the Student-T specifications appear to provide similar point estimates, except for the non-diagonal  $\beta$  coefficients: for the latter, once again, the picture may be clouded by collinearity, since a formal log-likelihood test does indicate joint significance.<sup>9</sup>

The overview of the estimation results is complete by examining the Table 4, where we report the square root of the estimated elements on the main diagonal of  $\Sigma$  for all specifications, showing substantial similarity.

Table 3: Estimated coefficients of the realized volatility equation for different model formulations (cf. Table 2) for JNJ (Jan. 3, 2007 – June 30, 2013). Robust t-stats in parentheses. An empty space indicates that the specification did not include the corresponding coefficient. Rows labeled *vol* and *nt* report the p-values for causality tests (see the text for details). Overall model diagnostics report Log-likelihood values, Akaike and Bayesian Information Criteria, and p-values of a joint Ljung–Box test of no autocorrelation at various lags. Last row reports the estimation time (in seconds).

|  | D-I                | A-I                | A-N                | A-T                | B-N                | B-T                | AB-N               | AB-T               | AB-LN              | AB-S               |
|--|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| <i>rkvt</i> <sub><i>t</i>-1</sub>                | 0.4967<br>(58.10)  | 0.4575<br>(51.94)  | 0.4194<br>(48.57)  | 0.4165<br>(48.72)  | 0.4048<br>(45.25)  | 0.4019<br>(47.42)  | 0.3687<br>(39.13)  | 0.3724<br>(40.77)  | 0.3630<br>(50.76)  | 0.3777<br>(45.77)  |
| <i>vol</i> <sub><i>t</i>-1</sub>                 |                    | -0.0300<br>(-1.11) | -0.0339<br>(-0.97) | -0.0265<br>(-0.75) |                    |                    | -0.0230<br>(-0.54) | -0.0337<br>(-0.78) | -0.0318<br>(-0.93) | -0.0298<br>(-0.75) |
| <i>nt</i> <sub><i>t</i>-1</sub>                  |                    | 0.0668<br>(1.77)   | 0.0726<br>(1.57)   | 0.0606<br>(1.36)   |                    |                    | 0.1897<br>(2.83)   | 0.1890<br>(2.82)   | 0.1691<br>(3.95)   | 0.1726<br>(3.51)   |
| <i>rkvt</i> <sub><i>t</i>-2</sub>                | -0.2672<br>(-5.12) | -0.2485<br>(-3.82) | -0.1878<br>(-2.91) | -0.1923<br>(-3.54) | -0.1950<br>(-4.64) | -0.2054<br>(-5.45) | -0.1778<br>(-3.75) | -0.1827<br>(-4.19) | -0.1811<br>(-5.46) | -0.1790<br>(-4.64) |
| <i>rkvt</i> <sub><i>t</i>-1</sub> <sup>(-)</sup> | 0.0252<br>(0.79)   | 0.0266<br>(0.79)   | 0.0265<br>(0.83)   | 0.0308<br>(1.00)   | 0.0255<br>(0.94)   | 0.0297<br>(1.13)   | 0.0171<br>(0.41)   | 0.0223<br>(0.53)   | 0.0211<br>(0.73)   | 0.0222<br>(0.67)   |
| $\mu_{t-1}^{(rkv)}$                              | 0.7223<br>(14.19)  | 0.7067<br>(9.94)   | 0.6793<br>(9.49)   | 0.6918<br>(11.41)  | 0.7235<br>(17.23)  | 0.7410<br>(20.73)  | 0.7588<br>(14.13)  | 0.7675<br>(16.46)  | 0.7626<br>(19.22)  | 0.7428<br>(15.50)  |
| $\mu_{t-1}^{(vol)}$                              |                    |                    |                    |                    | -0.0317<br>(-0.68) | -0.0139<br>(-0.32) | -0.1138<br>(-0.96) | -0.0466<br>(-0.44) | -0.0909<br>(-0.95) | -0.0938<br>(-0.86) |
| $\mu_{t-1}^{(nt)}$                               |                    |                    |                    |                    | 0.0492<br>(0.99)   | 0.0297<br>(0.66)   | -0.0984<br>(-0.76) | -0.1568<br>(-1.26) | -0.0867<br>(-0.87) | -0.0839<br>(-0.73) |
| <i>vol</i>                                       |                    | 0.2658             | 0.3296             | 0.4558             | 0.4970             | 0.7459             | 0.1705             | 0.2717             | 0.0853             | 0.1550             |
| <i>nt</i>  |                    | 0.0766             | 0.1153             | 0.1724             | 0.3202             | 0.5098             | 0.0051             | 0.0077             | 0.0001             | 0.0004             |
| logLik   | 205.53             | 239.75             | 2012.39            | 2086.31            | 1967.02            | 2048.80            | 2045.81            | 2125.12            | 2153.79            |                    |
| AIC  | -375.07            | -431.49            | -3970.78           | -4116.61           | -3880.05           | -4041.60           | -4025.62           | -4182.23           | -4241.57           |                    |
| BIC  | -258.11            | -275.55            | -3795.35           | -3934.69           | -3704.62           | -3859.67           | -3811.20           | -3961.32           | -4027.16           |                    |
| LB(12)   | 0.0000             | 0.0000             | 0.0001             | 0.0001             | 0.0000             | 0.0000             | 0.0388             | 0.0141             | 0.0450             | 0.0596             |
| LB(22)   | 0.0000             | 0.0000             | 0.0000             | 0.0000             | 0.0000             | 0.0000             | 0.0178             | 0.0109             | 0.0229             | 0.0264             |
| LB(32)   | 0.0000             | 0.0000             | 0.0003             | 0.0003             | 0.0000             | 0.0000             | 0.0192             | 0.0132             | 0.0291             | 0.0267             |
| time   | 0.342              | 0.753              | 2.628              | 15.044             | 2.485              | 14.707             | 2.714              | 14.859             | 0.155              | 0.132              |

Finally, the correlation coefficients implied by the various specifications are reported

<sup>9</sup>No attempt at pruning the structure of the model following the automated procedure suggested in Cipollini and Gallo (2010) was performed.

Table 4: Square root of the estimated elements on the main diagonal of  $\Sigma$  (cf. Equation (5))

|            | D-I   | A-I   | A-N   | A-T   | B-N   | B-T   | AB-N  | AB-T  | AB-LN | AB-S  |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $\sigma_1$ | 0.209 | 0.208 | 0.207 | 0.212 | 0.209 | 0.214 | 0.206 | 0.210 | 0.204 | 0.226 |
| $\sigma_2$ | 0.254 | 0.251 | 0.252 | 0.256 | 0.257 | 0.262 | 0.252 | 0.256 | 0.252 | 0.272 |
| $\sigma_3$ | 0.227 | 0.227 | 0.229 | 0.236 | 0.232 | 0.240 | 0.228 | 0.235 | 0.228 | 0.242 |

in Table 5, showing that a strong correlation among innovations further supports the need for taking simultaneity into account.

Table 5: Estimated correlation matrices of the copula functions (specifications -N and -T), and of  $\varepsilon_t$  (specifications -LN and -S).

|         | A-N     |        | A-T     |        | AB-N    |        | AB-T    |        | AB-LN   |        | AB-S    |        |
|---------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|
|         | $vol_t$ | $nt_t$ | $vol_t$ | $nt_t$ | $vol_t$ | $nt_t$ | $vol_t$ | $nt_t$ | $vol_t$ | $nt_t$ | $vol_t$ | $nt_t$ |
| $rkvt$  | 0.483   | 0.610  | 0.505   | 0.626  | 0.485   | 0.612  | 0.505   | 0.628  | 0.469   | 0.596  | 0.481   | 0.605  |
| $vol_t$ |         | 0.902  |         | 0.917  |         | 0.903  |         | 0.918  |         | 0.902  |         | 0.906  |

The alternative log-Normal and Semiparametric formulations deliver parameter inferences qualitatively similar to the corresponding versions based on copulas. The AIC and BIC statistics of the log-Normal are better than the copula based formulations (for the semiparametric they are not available); also its Ljung-Box diagnostics are marginally better.

A comparison of the estimation times across specifications<sup>10</sup> reveals some differences. Student-T copula based formulations are the slowest; the Normal copula based estimated with the concentrated likelihood approach takes about 1/6 of time; finally, the log-Normal and the Semiparametric are a lot faster (about 1/20 of the Normal based). Differences notwithstanding, the largest times involved are very low on an absolute scale.

Figure 2 shows a comparison between the estimated residuals against their theoretical counterparts in the case of the Gamma and log-Normal distributions. In both cases, the fit seems to break apart for the tail of the distribution, pointing to the need for some care about the evolution of volatility of volatility (a mixture of distribution hypothesis is contemplated in, say, De Luca and Gallo (2009) in a MEM model).

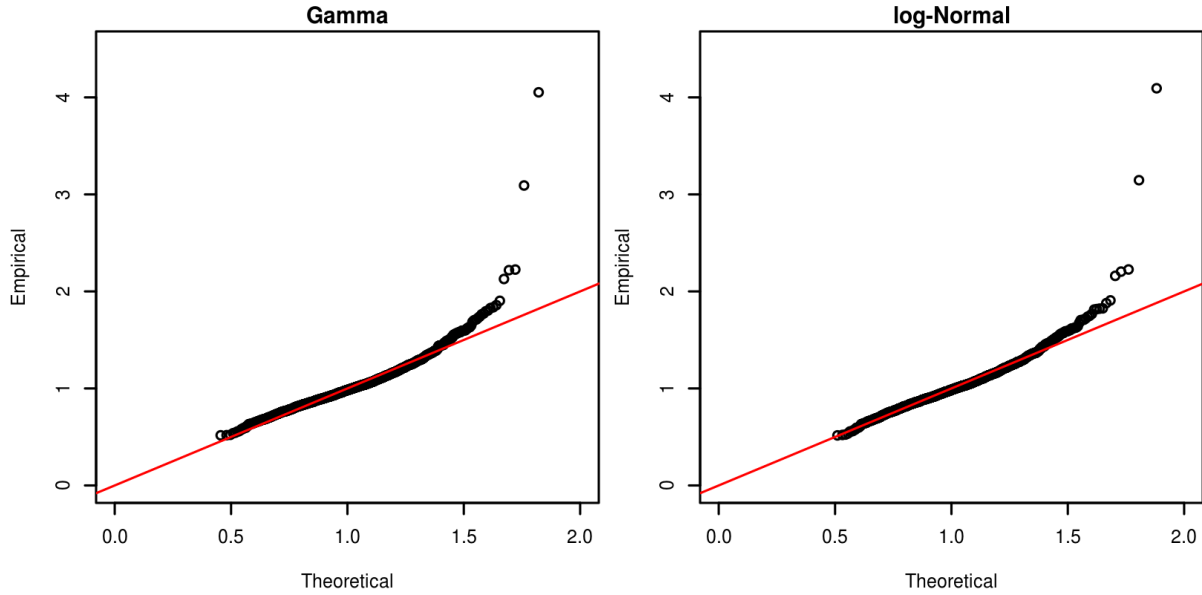
## 5.2 Forecasting

We left the period July 1 – December 31, 2013 (128 observations) for out-of-sample forecasting comparisons. We adopt a Diebold and Mariano (1995) test statistic for superior predictive ability using the error measures

$$e_{N,t} = \frac{1}{2}(x_t - \mu_t)^2 \quad e_{G,t} = \frac{x_t}{\mu_t} - 1 - \ln\left(\frac{x_t}{\mu_t}\right). \quad (45)$$

<sup>10</sup>Calculations with our routines written in R were performed on an Intel i7-5500U 2.4Ghz processor. We did not perform an extensive comparison of estimation times and we did not optimize performance by removing tracing of intermediate results. Moreover, the copula-based and the alternative specifications are optimized resorting to different algorithms: the first ones, more cumbersome to optimize, are estimated toward a combination of NEWUOA (Powell (2006)) and Newton-Raphson; for the second ones we used a dogleg algorithm (Nocedal and Wright (2006, ch. 4)).

Figure 2: QQ-plot between theoretical and empirical residuals for the realized volatility of JNJ. Comparison with Gamma and log-Normal distributions. Estimation period: Jan. 3, 2007 – June 30, 2013.



where  $x_t$  and  $\mu_t$  denote here the observed and the predicted values, respectively.  $e_{N,t}$  is the squared error, and can be interpreted as the loss behind an  $x_t$  Normally distributed with mean  $\mu_t$ ; similarly,  $e_{G,t}$  can be interpreted as the loss we can derive considering  $x_t$  as Gamma distributed with mean  $\mu_t$  and variance proportional to  $\mu_t^2$ . Comparisons are performed on both Original and Adjusted (i.e. removing the low frequency component) series.

Table 6 reports the values of the Diebold-Mariano test statistic of different model formulations, against the AB-T specification (which has the lowest values of the information criteria among copula-based specifications, cf. Table 3), considering one-step ahead predictions.

Table 6: Diebold-Mariano test statistics for unidirectional comparisons, against the AB-T specification, considering 1-step ahead forecasts (out-of-sample period July 1 – December 31, 2013). The error measures are defined as  $e_{N,t} = 0.5(x_t - \mu_t)^2$  (Normal loss) and  $e_{G,t} = x_t/\mu_t - 1 - \ln(x_t/\mu_t)$  (Gamma loss), where  $x_t$  and  $\mu_t$  denote the observed and the predicted values, respectively (cf. Section 5.2 for the interpretation). Boldface (italic) indicates 5% (10%) significant statistics.

| Specification | $e_{N,t}$     |               | $e_{G,t}$     |               |
|---------------|---------------|---------------|---------------|---------------|
|               | Adjusted      | Original      | Adjusted      | Original      |
| D-I           | <b>-2.490</b> | <b>-2.107</b> | <b>-2.428</b> | <b>-2.132</b> |
| A-I           | <i>-1.358</i> | -1.200        | -1.277        | -1.173        |
| A-N           | -0.668        | -0.621        | -0.563        | -0.591        |
| A-T           | -0.510        | -0.449        | -0.422        | -0.431        |
| B-N           | -0.553        | -0.557        | -0.449        | -0.560        |
| B-T           | -0.474        | -0.479        | -0.389        | -0.490        |
| AB-N          | -0.768        | -0.757        | -0.737        | -0.708        |
| AB-LN         | 0.868         | 0.720         | 0.980         | 0.742         |
| AB-S          | -0.277        | -0.230        | -0.180        | -0.164        |

There are no substantial differences in performance between the results on the original and the adjusted series.<sup>11</sup> We notice a definite improvement of the specification involving simultaneous innovations over the equation-by-equation specifications (first two rows). The Student-T Copula improves upon the Normal formulation (marginally, line AB-N), but is substantially equivalent to the AB-LN (slightly better) and the AB-S (slightly worse).

## 6 Conclusions

The Multiplicative Error Model was originally introduced by Engle (2002) as a positive valued product process between a scale factor following a GARCH type dynamics and a unit mean innovation process. In this paper we have presented a general discussion of a *vector* extension of such a model with a dynamically interdependent formulation for the scale factors where lagged variables and conditional expectations may be both allowed to impact each other's conditional mean. Engle and Gallo (2006) estimate a vMEM where the innovations are Gamma-distributed with a diagonal variance-covariance matrix (estimated equation-by-equation). The extension to a truly multivariate process requires a complete treatment of the interdependence among the innovation terms. One possibility is to avoid the specification of the distribution and adopt a semiparametric GMM approach as in Cipollini *et al.* (2013). In this paper we have derived a maximum likelihood estimator by framing the innovation vector as a copula function linking Gamma marginals; a parametric alternative with a multivariate log-Normal distribution is discussed, while we show that the specification using a multivariate Gamma distribution has severe shortcomings.

We illustrate the procedure on three indicators related to market trading activity: realized volatility, volumes, and number of trades. The empirical results are presented in reference to daily data on the Johnson and Johnson (JNJ) stock. The data on the three variables show a (slowly moving) time varying local average which was removed before estimating the parameters. The three components are highly correlated with one another, but interestingly, their removal does not have a substantial impact on the correlation among the adjusted series. The specifications adopted start from a diagonal structure where no dynamic interaction is allowed, and an Independent copula (*de facto* an equation-by-equation specification). Refinements are obtained by inserting a Normal copula and a Student-T copula (which dramatically improve the estimated log-likelihood function values) and then allowing for the presence of dynamic interdependence in the form of links either between lagged variables, or between lagged conditional means, or both. Although hindered by the presence of some collinearity, the results clearly show a significant improvement for the fit of the equation for realized volatility when volumes and number of trades are considered. This is highlighted by significantly better model log-likelihoods, lower overall information criteria and improved autocorrelation diagnostics. The results of an out-of-sample forecasting exercise confirm the need for a simultaneous approach to model specification and estimation, with a substantial preference for the Student-T copula results within the copula-based formulations. From an interpretive point of view, the results show that the past realized volatility is not enough information by itself to reconstruct the dynamics: its modeling and forecastability are clearly influenced by other indicators of market trading activity, and simultaneous correlations in the innovations must be accounted for.

Under equal models for conditional means, the copula approach is substantially equiv-

---

<sup>11</sup>One-step ahead predictions at time  $t$  for the original series are computed multiplying the corresponding forecast of the adjusted indicator by the value of the low frequency component at  $t - 1$ .

alent to other forms of specifications, be they parametric with a multivariate log-Normal error, or semiparametric (with GMM estimation). The advantage of a slightly more expensive (from a computational point of view) procedure lies in gaining flexibility in the tail dependence offered by the Student–T copula, which can be a plus when reconstructing joint conditional distributions, or when reproducing more realistic behavior in simulation is desirable.

## References

- Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, **51**, 169–204.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realised kernels in practice: trades and quotes. *Econometrics Journal*, **12**, 1–32.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, **162**, 149–169.
- Bodnar, T. and Hautsch, N. (2016). Dynamic conditional correlation multiplicative error processes. *Journal of Empirical Finance*, **36**, 41 – 67.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327.
- Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2000). Copulas for finance: A reading guide and some applications. Technical report, Groupe de Recherche Orationnelle, Credit Lyonnais, Paris.
- Brownlees, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics and Data Analysis*, **51**, 2232–2245.
- Brownlees, C. T. and Gallo, G. M. (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics*, **8**, 29–56.
- Cerrato, M., Crosby, J., Kim, M., and Zhao, Y. (2015). Modeling dependence structure and forecasting market risk with dynamic asymmetric copula. Technical report, Adam Smith Business School, University of Glasgows.
- Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley.
- Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking*, **37**, 561–582.
- Cipollini, F. and Gallo, G. M. (2010). Automated variable selection in vector multiplicative error models. *Computational Statistics & Data Analysis*, **54**(11), 2470–2486.
- Cipollini, F., Engle, R. F., and Gallo, G. M. (2013). Semiparametric vector mem. *Journal of Applied Econometrics*, **28**, 1067–1086.
- De Luca, G. and Gallo, G. M. (2009). Time-varying mixing weights in mixture autoregressive conditional duration models.



- Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, **73**(1), 111–129.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, **13**(3), 253–263.
- Embrechts, P., McNeil, A. J., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In M. A. H. Dempster, editor, *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press, Cambridge.
- 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**, 3–27.
- Engle, R. F. and Mezrich, J. (1995). Grappling with GARCH. *Risk*, **8**, 112–117.
- Engle, R. F. and Rangel, J. G. (2008). The spline-GARCH model for low frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, **21**, 1187–1222.
- Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127–62.
- Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: Applicability and limitations. *Statistics & Probability Letters*, **63**, 275–286.
- Francq, C., Horvath, L., and Zakoïan, J.-M. (2011). Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics*, **9**, 619–656.
- Gallo, G. M. and Otranto, E. (2015). Forecasting realized volatility with changing average levels. *International Journal of Forecasting*, **31**, 620–634.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Hautsch, N. (2008). Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model. *Journal of Economic Dynamics and Control*, **32**, 3978 – 4015.
- Hautsch, N., Malec, P., and Schienle, M. (2013). Capturing the zero: A new class of zero-augmented distributions and multiplicative error processes. *Journal of Financial Econometrics*, **12**(1), 89–121.
- Horváth, L., Kokoszka, P., and Zitikis, R. (2006). Sample and implied volatility in GARCH models. *Journal of Financial Econometrics*, **4**, 617–635.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (2000). *Continuous Multivariate Distributions*, volume 1. John Wiley & Sons, New York.
- Kostadinov, K. (2005). Non-parametric estimation of elliptical copulae with application to credit risk. Technical report, Munich University of Technology.

- Kristensen, D. and Linton, O. (2004). Consistent standard errors for target variance approach to GARCH estimation. *Econometric Theory*, **20**, 990–993.
- Lindskog, F., McNeil, A. J., and Schmock, U. (2003). Kendall’s tau for elliptical distributions. In G. Bol, G. Nakhaeizadeh, S. Rachev, T. Ridder, and K. Vollmer, editors, *Credit Risk - Measurement, Evaluation and Management*. Physica-Verlag Heidelberg.
- 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.
- Nelsen, R. B. (1999). *An introduction to Copulas*. Lecture Notes in Statistics. Springer Verlag, New York.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier.
- Nocedal, J. and Wright, S. J. (2006). *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, Berlin. NEOS guide <http://www-fp.mcs.anl.gov/otc/Guide/>.
- Noss, J. (2007). The econometrics of optimal execution in an order-driven market. Technical report, Nuffield College, Oxford University, Master’s Thesis.
- Okimoto, T. (2008). New evidence of asymmetric dependence structures in international equity markets. *The Journal of Financial and Quantitative Analysis*, **43**, 787–815.
- Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, **47**, 527 – 556.
- Patton, A. (2013). Chapter 16 - copula methods for forecasting multivariate time series. In G. Elliott and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 2, Part B of *Handbook of Economic Forecasting*, pages 899–960. Elsevier.
- Powell, M. J. D. (2006). The newuoa software for unconstrained optimization without derivatives. In G. D. Pillo and M. Roma, editors, *Large-Scale Nonlinear Optimization*, pages 255–297. Springer Verlag, New York.
- Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research*, **55**, 301–327.
- Song, P. X. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, **27**, 305–320.

## A Expectation Targeting

We show how to obtain the asymptotic distribution of the estimator of the parameters of the vMEM when the constant  $\omega$  model is reparameterized, via *expectation targeting* (Section 3.2.1), by exploiting the assumption of weak-stationarity of the process (Section 2). Under assumptions detailed in what follows, the distribution of  $\bar{\mathbf{x}}_T$  can be found by extending the results in Francq *et al.* (2011) and Horváth *et al.* (2006) to a multivariate framework and to asymmetric effects.

### Framework

Let us assume a model defined by (1), (5) and (3) where the  $\mathbf{x}_t^{(-)}$ 's are associated with negative returns on the basis of assumptions detailed in Section 2

Besides mean-stationarity, in order to get asymptotic normality also, we assume the stronger condition that  $E(\boldsymbol{\mu}_t \boldsymbol{\mu}_t')$  exists (a similar condition on existence of the unconditional squared moment of the conditional variance is assumed in the cited papers).

### Auxiliary results

In order to simplify the exposition we introduce two quantities employed in the following, namely the *zero mean residual*

$$\mathbf{v}_t = \mathbf{x}_t - \boldsymbol{\mu}_t \quad (46)$$

and

$$\tilde{\mathbf{x}}_t = \mathbf{x}_t^{(-)} - \mathbf{x}_t/2. \quad (47)$$

Since  $\mathbf{v}_t = \boldsymbol{\mu}_t \odot (\boldsymbol{\varepsilon}_t - \mathbf{1})$ , and  $\tilde{\mathbf{x}}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t \odot (\mathbf{I}_t - \mathbf{1}/2)$ , we can easily check that

$$\begin{aligned} E(\mathbf{v}_t) &= \mathbf{0} \\ V(\mathbf{v}_t) &= E(\boldsymbol{\mu}_t \boldsymbol{\mu}_t') \odot \boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}_v \\ C(\mathbf{v}_s, \mathbf{v}_t) &= \mathbf{0} \quad s \neq t \\ E(\tilde{\mathbf{x}}_t) &= \mathbf{0} \\ V(\tilde{\mathbf{x}}_t) &= \boldsymbol{\Sigma}_v \odot \boldsymbol{\Sigma}_I \\ C(\tilde{\mathbf{x}}_s, \tilde{\mathbf{x}}_t) &= \mathbf{0} \quad s \neq t \\ C(\mathbf{v}_s, \tilde{\mathbf{x}}_t) &= \mathbf{0} \quad s, t. \end{aligned}$$

We remark that  $\boldsymbol{\Sigma}_v$  represents also the unconditional average of  $\mathbf{x}_t \mathbf{x}_t'$ . By consequence, the sample averages  $\bar{\mathbf{v}}_T = T^{-1} \sum_{t=1}^T \mathbf{v}_t$  and  $\bar{\tilde{\mathbf{x}}}_T = T^{-1} \sum_{t=1}^T \tilde{\mathbf{x}}_t$  are such that

$$\sqrt{T} \begin{pmatrix} \bar{\mathbf{v}}_T \\ \bar{\tilde{\mathbf{x}}}_T \end{pmatrix} \xrightarrow{d} N \left[ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_v & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_v \odot \boldsymbol{\Sigma}_I \end{pmatrix} \right]. \quad (48)$$

### The asymptotic distribution of $\bar{\mathbf{x}}_T$

By replacing (46) and (47) into equation (3) and arranging it we get

$$\mathbf{x}_t - \sum_{l=1}^L \left( \boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \frac{\boldsymbol{\gamma}_l}{2} \right) \mathbf{x}_{t-l} = \boldsymbol{\omega} + \mathbf{v}_t - \sum_{l=1}^L \boldsymbol{\beta}_l \mathbf{v}_{t-l} + \sum_{l=1}^L \boldsymbol{\gamma}_l \tilde{\mathbf{x}}_{t-l}^{(-)}$$

so that, averaging both sides,

$$\left[ \mathbf{I} - \sum_{l=1}^L \left( \boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \frac{\boldsymbol{\gamma}_l}{2} \right) \right] \bar{\mathbf{x}}_T = \boldsymbol{\omega} + \left[ \mathbf{I} - \sum_{l=1}^L \boldsymbol{\beta}_l \right] \bar{\mathbf{v}}_T + \sum_{l=1}^L \boldsymbol{\gamma}_l \bar{\tilde{\mathbf{x}}}_T + O_p(T^{-1}).$$

Deriving  $\bar{\mathbf{x}}_T$  we get

$$\begin{aligned}\bar{\mathbf{x}}_T &= \boldsymbol{\mu} + \left[ \mathbf{I} - \sum_{l=1}^L \left( \boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \frac{\boldsymbol{\gamma}_l}{2} \right) \right]^{-1} \left[ \left( \mathbf{I} - \sum_{l=1}^L \boldsymbol{\beta}_l \right) \bar{\mathbf{v}}_T + \sum_{l=1}^L \boldsymbol{\gamma}_l \bar{\mathbf{x}}_T \right] + O_p(T^{-1}) \\ &= \boldsymbol{\mu} + \mathbf{A}^{-1} \left[ \mathbf{B} \bar{\mathbf{v}}_T + \mathbf{C} \bar{\mathbf{x}}_T \right] + O_p(T^{-1}),\end{aligned}$$

where  $\boldsymbol{\mu}$  is given in (28).

By means of (48), the asymptotic distribution of  $\bar{\mathbf{x}}_T$  follows immediately as

$$\sqrt{T} (\bar{\mathbf{x}}_T - \boldsymbol{\mu}) \xrightarrow{d} N \left[ \mathbf{0}, \mathbf{A}^{-1} (\mathbf{B} \boldsymbol{\Sigma}_v \mathbf{B}' + \mathbf{C} (\boldsymbol{\Sigma}_v \odot \boldsymbol{\Sigma}_I) \mathbf{C}') \mathbf{A}^{-1'} \right] \quad (49)$$

