UNIVERSITÀ DEGLI STUDI FIRENZE
DISIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

Copula–based Specification of vector MEMs

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

DISIA WORKING PAPER
2016/04

© Copyright is held by the author(s).

Copula–based Specification of vector MEMs *

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

Abstract

The Multiplicative Error Model (Engle (2002)) for nonnegative valued processes is specified as the product of a (conditionally autoregressive) scale factor and an innovation process with nonnegative support. A multivariate extension allows for the innovations to be contemporaneously correlated. We overcome the lack of sufficiently flexible probability density functions for such processes by suggesting a copula function approach to estimate the parameters of the scale factors and of the correlations of the innovation processes. We illustrate this vector MEM with an application to the interactions between realized volatility, volume and the number of trades. We show that significantly superior realized volatility forecasts are delivered in the presence of other trading activity indicators and contemporaneous correlations.

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

*Some of the material presented here circulated in the NBER WP 12690 (2006) "Vector Multiplicative Error Models: Representation and Inference". We thank Nikolaus Hautsch, Loriano 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.

†Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", Università di Firenze, Italy. e-mail: cipollini@disia.unifi.it

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

§Corresponding author: Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", Università di Firenze, Viale G.B. Morgagni, 59 - 50134 Firenze – Italy. e-mail: gallog@disia.unifi.it

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, high–low range, ultra-high frequency based measures of volatility, financial durations and so on.

Engle (2002) reckons that one striking regularity of 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 duration, volume and volatility Manganelli (2005); for volatility, volume and trading intensity Hautsch (2008)).

The model can be specified 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 beside its own. Such a specification lends itself to producing multi–step ahead forecasts: for example, Engle and Gallo (2006) specify a multivariate MEM where the dynamics of three different measures of volatility, namely absolute returns, daily range and realized volatility, influence each other temporally, and evaluate the contribution of MEM-based forecasts to the prediction of VIX. Although 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.

The specification of a multivariate distribution of the innovations is far from trivial: joint probability distributions for nonnegative–valued random variables are not available except in very special cases. In this paper, we suggest a maximum likelihood estimation strategy adopting copula functions 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.* (2014) 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). As an alternative for the vector MEM, Cipollini *et al.* (2013) suggest a semiparametric approach resulting in a GMM estimator.

The range of potential applications 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). As an illustration we will concentrate on the interaction of various measures of market activity (volatility, volume and number of trades) in which the conditional expectations depend just on the past values (not also on some

contemporary information as in Manganelli (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. Section 4 presents the application of our model to three series of 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 Multiplicative Error Models

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

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

where $\odot$ indicates the Hadamard (element–by–element) product and diag is 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

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

where $\boldsymbol{\omega}$ is $(K, 1)$ and $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are (K, K). The vector $\boldsymbol{x}_t^{(-)}$ 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 $\boldsymbol{A} = \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma}/2$ are smaller than 1 in modulus. We can think of $\boldsymbol{A}$ as the *impact matrix* in the expression

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

If more lags are considered, the model is

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

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

$$\boldsymbol{\mu}^*_{t+L|t-1} = \boldsymbol{A}^* \boldsymbol{\mu}^*_{t+L-1|t-1}, \tag{4}$$

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

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

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

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

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

The previous conditions guarantee that

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

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

where the latter is a positive definite matrix by construction.

Some alternatives can be considered about the specification of the distribution of the error term $\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}$.

2.1 Multivariate Gamma Formulations

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 probability density functions (pdf). 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 Kowalckzyk and Trycha and by Mathai and Moschopoulos):

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

where $\boldsymbol{\phi} = (\phi_1; \ldots; \phi_K)$ and $0 < \phi_0 < \min(\phi_1, \ldots, \phi_K)$ (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.

2.2 Copula Based Formulations

A different approach to specify the distribution of $\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}$ is to use copula functions (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 (2007) for financial applications). In this case, the conditional pdf of the error component of the vMEM is given by

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

where $c(\boldsymbol{u}_t;\boldsymbol{\xi})$ 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 elements: 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, Lognormal, and mixtures of them). For the second, we discuss some possible specifications within the class of Elliptical copulas.

2.2.1 Normal Copula

The Normal copula is a frequent choice in applications (McNeil *et al.* (2005), Cherubini *et al.* (2004), Bouyé *et al.* (2000)). Its pdf is given by

$$c_N(\boldsymbol{u};\boldsymbol{R}) = |\boldsymbol{R}|^{-1/2}\exp\left[-\frac{1}{2}\left(\boldsymbol{q}'\boldsymbol{R}^{-1}\boldsymbol{q} - \boldsymbol{q}'\boldsymbol{q}\right)\right], \tag{9}$$

where $\boldsymbol{q} = (q_1;\ldots;q_K)$, $q_i = \Phi^{-1}(u_i)$ 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, differently from the Multivariate Gamma in Section 2.1, can assume negative values too.

2.2.2 Student-T Copula

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 $\boldsymbol{x}$ (be they 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. The pdf of the Student-T copula is given by

$$c_T(\boldsymbol{u}; \boldsymbol{R}, \nu) = \frac{\Gamma((\nu+K)/2)\Gamma(\nu/2)^{K-1}}{\Gamma((\nu+1)/2)} |\boldsymbol{R}|^{-1/2} \frac{(1+\boldsymbol{q}'\boldsymbol{R}^{-1}\boldsymbol{q}/\nu)^{-(\nu+K)/2}}{\prod_{i=1}^{K}(1+q_i^2/\nu)^{-(\nu+1)/2}}, \qquad (10)$$

where $\boldsymbol{q} = (q_1; \ldots; q_K)$, $q_i = T^{-1}(u_i; \nu)$ and $T(x; \nu)$ denotes the cdf of the Student-T distribution with ν degrees of freedom computed at x. Differently from the Normal copula, when $\boldsymbol{R} = \boldsymbol{I}$ we get uncorrelated but not independent marginals (details in McNeil *et al.* (2005)). Further specifications of the Student-T copula are in Demarta and McNeil (2005).

2.2.3 Elliptical Copulas

The Elliptical copulas family provides an unified framework encompassing the Normal, the Student-T and any other member of this family endowed of an explicit pdf. Elliptical copulas (for details see McNeil *et al.* (2005), Frahm *et al.* (2003), Schmidt (2002)) 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. Elliptical copulas have interesting features and widespread applicability, even if their elliptical symmetry may constitute a limit in some applications.[1]

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

$$c_E(\boldsymbol{u}; \boldsymbol{R}, \boldsymbol{\nu}) = K^*(\boldsymbol{\nu}, K)|R|^{-1/2} \frac{g_1(\boldsymbol{q}'\boldsymbol{R}^{-1}\boldsymbol{q}; \boldsymbol{\nu}, K)}{\prod_{i=1}^{K} g_2(q_i^2; \boldsymbol{\nu})} \qquad (11)$$

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

- in the Normal copula, with no explicit shape parameter $\boldsymbol{\nu}$, we have $K^*(K) \equiv 1$, $g_1(x; K) = g_2(x) \equiv \exp(-x/2)$;
- in the Student-T copula, with a scalar ν shape parameter, we have $K^*(\nu; K) = \frac{\Gamma((\nu+K)/2)\Gamma(\nu/2)^{K-1}}{\Gamma((\nu+1)/2)}$, $g_1(x; \nu, K) = (1 + x/\nu)^{-(\nu+K)/2}$, $g_2(x; \nu) = (1 + x/\nu)^{-(\nu+1)/2}$.

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 $\boldsymbol{\theta}$)

[1] Copulas in the Archimedean family offer a way to bypass such a limitation but suffer from other drawbacks and will not be pursued here.

and a generic 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 is

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

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

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

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

$$\nabla_{\boldsymbol{\theta}}\, l = \sum_{t=1}^{T} \boldsymbol{A}_t \boldsymbol{w}_t \tag{13}$$

where

$$\begin{aligned} \boldsymbol{A}_t &= -\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' \,\mathrm{diag}(\boldsymbol{\mu}_t)^{-1}. \\ \boldsymbol{w}_t &= \boldsymbol{\varepsilon}_t \odot \boldsymbol{b}_t + \mathbb{1}, \\ \boldsymbol{b}_t &= \nabla_{\boldsymbol{\varepsilon}_t} \ln f(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}), \end{aligned} \tag{14}$$

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

$$E\left[\boldsymbol{A}_t \boldsymbol{\mathcal{I}}^{(\varepsilon)} \boldsymbol{A}_t'\right] \tag{15}$$

and

$$E\left[\boldsymbol{A}_t \boldsymbol{H}^{(\varepsilon)} \boldsymbol{A}_t'\right], \tag{16}$$

respectively, where the matrices

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

and

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

depend only on $\boldsymbol{\lambda}$ but not on $\boldsymbol{\theta}$. Of course, under a correct specification $\boldsymbol{H}^{(\varepsilon)} = -\boldsymbol{\mathcal{I}}^{(\varepsilon)}$.

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 $\boldsymbol{b}_t$. For instance, considering the generic copula formulation (8), then

$$\ln f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = \ln c(\boldsymbol{u}_t) + \sum_{i=1}^{K} \ln f_i(\varepsilon_{t,i}) \tag{17}$$

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

$$b_{t,i} = f_i(\varepsilon_{t,i})\nabla_{u_{t,i}} \ln c(\boldsymbol{u}_t) + \nabla_{\varepsilon_{t,i}} \ln f_i(\varepsilon_{t,i}). \tag{18}$$

In what follows we provide specific formulas for the elliptical copula formulation (11) 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 2.2),

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

Therefore,

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

and

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

As detailed in Section 2.2.3, beside a possible shape parameter $\boldsymbol{\nu}$, elliptical copulas are characterized by a correlation matrix $\boldsymbol{R}$ which, in view of its full ML estimation, can be expressed (cf. McNeil *et al.* (2005, p. 235)) as

$$\boldsymbol{R} = \boldsymbol{D}\boldsymbol{c}'\boldsymbol{c}\boldsymbol{D}, \tag{19}$$

where $\boldsymbol{c}$ is an upper-triangular matrix with ones on the main diagonal and $\boldsymbol{D}$ 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 $j = 2, \ldots, K$. So doing, the estimation of $\boldsymbol{R}$ is transformed in an unconstrained problem, since the $K(K-1)/2$ free elements of $\boldsymbol{c}$ can vary into $\mathbb{R}$. We can then write $\boldsymbol{\xi} = (\boldsymbol{c}; \boldsymbol{\nu})$.

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

$$\ln c(\boldsymbol{u}_t) = \ln K^* - \sum_{i=2}^{K} \ln D_i + \ln g_1(\widetilde{\widetilde{q}}_t) - \sum_{i=1}^{K} \ln g_2(q_{t,i}^2) \tag{20}$$

where use was made of the fact that $\frac{1}{2}\ln(|\boldsymbol{R}|) = \sum_{i=2}^{K} \ln D_i$.

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 $\boldsymbol{c}$

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

$$\nabla_{c_{ij}} l = \nabla_{c_{ij}}\left[-T\sum_{i=2}^{K}\ln D_i + \sum_{t=1}^{T}\ln g_1(\widetilde{\widetilde{q}}_t)\right],\ i<j. \tag{21}$$

Using some algebra we can show that

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

and

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

By replacing them into (21) we obtain

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

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

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

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

The derivative of $\ln K^* = \ln K^*(\boldsymbol{\nu}; K)$ 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 $\boldsymbol{\phi}$

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

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

After some algebra we obtain

$$\nabla_{\boldsymbol{\phi_i}} l = \sum_{t=1}^{T} \left[\nabla_{\boldsymbol{\phi}_i} F_i(\varepsilon_{t,i}) d_{t,i} + \nabla_{\boldsymbol{\phi}_i} \ln f_i(\varepsilon_{t,i}) \right], \tag{22}$$

where

$$d_{t,i} = \frac{1}{g(q_{t,i})} \left[2q^*_{t,i} \nabla_{\widetilde{\widetilde{q}}_t} \ln g_1(\widetilde{\widetilde{q}}_t) - \nabla_{q_{t,i}} \ln g_2(q^2_{t,i}) \right]. \tag{23}$$

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

By exploiting the notation introduced in this section, we can now detail the structure of $\boldsymbol{b}_t$ entering into (14) and then into the portion of the score function relative to $\boldsymbol{\theta}$. From (20), $\boldsymbol{\varepsilon}_t \odot \boldsymbol{b}_t + \mathbb{1}$ (cf. 14) has elements

$$\varepsilon_{t,i} b_{t,i} + 1 = \varepsilon_{t,i} f_i(\varepsilon_{t,i}) d_{t,i} + \varepsilon_{t,i} \nabla_{\varepsilon_{t,i}} \ln f_i(\varepsilon_{t,i})) + 1 \tag{24}$$

where $d_{t,i}$ is given in (23). For our choice, $f_i(\varepsilon_{t,i})$ is the pdf of a $Gamma(\phi_i, \phi_i)$ distribution, so that

$$\varepsilon_{t,i} \nabla_{\varepsilon_{t,i}} \ln f_i(\varepsilon_{t,i}) + 1 = \phi_i - \varepsilon_{t,i} \phi_i. \tag{25}$$

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*[2]). Since $E(\boldsymbol{x}_t) = E(\boldsymbol{\mu}_t) = \boldsymbol{\mu}$ is well defined and is equal to

$$\boldsymbol{\mu} = \left[\boldsymbol{I} - \sum_{l=1}^{L} \left(\boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \frac{\boldsymbol{\gamma}_l}{2} \right) \right]^{-1} \boldsymbol{\omega}, \tag{26}$$

Equation (3) becomes

$$\boldsymbol{\mu}_t = \left[\boldsymbol{I} - \sum_{l=1}^{L} \left(\boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \frac{\boldsymbol{\gamma}_l}{2} \right) \right] \boldsymbol{\mu} + \sum_{l=1}^{L} \left(\boldsymbol{\alpha}_l \boldsymbol{x}_{t-l} + \boldsymbol{\gamma}_l \boldsymbol{x}^{(-)}_{t-l} + \boldsymbol{\beta}_l \boldsymbol{\mu}_{t-l} \right). \tag{27}$$

In a GARCH framework, the consequences of this practice have been been investigated by Kristensen and Linton (2004) and, more recently, by Francq *et al.* (2011) who show

[2]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.

its merits in terms of stability of estimation algorithms and accuracy in estimation of both coefficients and long term volatility (cf. Appendix 5 for some details in the present context).

From a technical point of view, a trivial replacement of $\boldsymbol{\mu}$ by the sample mean $\overline{\boldsymbol{x}}_T$ followed by a ML estimation of the remaining parameters, preserves consistency but leads to wrong standard errors (Kristensen and Linton, 2004). The issue can 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 $(\boldsymbol{\vartheta};\boldsymbol{\mu})$ (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 $\sqrt{T}\left(\widehat{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}\right)$ as

$$\boldsymbol{G}_{\boldsymbol{\vartheta}}^{-1}\left[\begin{array}{cc}\boldsymbol{I} & -\boldsymbol{G}_{\boldsymbol{\mu}}\boldsymbol{M}^{-1}\end{array}\right]\left[\begin{array}{cc}\boldsymbol{\Omega}_{\boldsymbol{\vartheta},\boldsymbol{\vartheta}'} & \boldsymbol{\Omega}_{\boldsymbol{\vartheta},\boldsymbol{\mu}'}\\ \boldsymbol{\Omega}_{\boldsymbol{\mu},\boldsymbol{\vartheta}'} & \boldsymbol{\Omega}_{\boldsymbol{\mu},\boldsymbol{\mu}'}\end{array}\right]\left[\begin{array}{c}\boldsymbol{I}\\ -\left(\boldsymbol{G}_{\boldsymbol{\mu}}\boldsymbol{M}^{-1}\right)'\end{array}\right]\boldsymbol{G}_{\boldsymbol{\vartheta}}^{-1\prime} \tag{28}$$

where

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

and

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

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

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 (27).

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

where

$$\boldsymbol{A} = \boldsymbol{I} - \sum_{l=1}^{L} \left(\boldsymbol{\alpha}_l + \boldsymbol{\beta}_l + \frac{\boldsymbol{\gamma}_l}{2} \right)$$

$$\boldsymbol{B} = \boldsymbol{I} - \sum_{l=1}^{L} \boldsymbol{\beta}_l$$

$$\boldsymbol{C} = \sum_{l=1}^{L} \boldsymbol{\gamma}_l$$

$$\boldsymbol{\Sigma}_v = E(\boldsymbol{\mu}_t \boldsymbol{\mu}_t') \odot \boldsymbol{\Sigma}.$$

The expression for $\boldsymbol{\Omega}_{\boldsymbol{\mu},\boldsymbol{\mu}'}$ 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, $\boldsymbol{H}^{(\varepsilon)} = -\boldsymbol{\mathcal{I}}^{(\varepsilon)}$ and $E\left[(\boldsymbol{b}_t \odot \boldsymbol{\varepsilon}_t) \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1} \right] = -\mathbb{1}\mathbb{1}'$ leading to $\boldsymbol{\Omega}_{\boldsymbol{\theta},\boldsymbol{\mu}'} = \boldsymbol{0}$ and to

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

for what concerns the inner part of (28).

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 $\boldsymbol{R}$ (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 relation $V(\varepsilon_{t,i} | \mathcal{F}_{t-1}) = 1/\phi_i$ leads to very simple estimator of ϕ_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 $\boldsymbol{R}$ matrix (Cherubini *et al.* (2004, p. 155)), namely

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

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

$$\frac{T}{2} \left[-\ln |\boldsymbol{Q}| - K + \operatorname{trace}(\boldsymbol{Q}) \right] + \sum_{t=1}^{T} \sum_{i=1}^{K} \ln f_i(\varepsilon_{t,i} | \mathcal{F}_{t-1}) \tag{29}$$

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

$$(\boldsymbol{R}^{-1})_{ij} - (\boldsymbol{R}^{-1})_{i\cdot}\frac{\boldsymbol{q}'\boldsymbol{q}}{T}(\boldsymbol{R}^{-1})_{\cdot j} = 0$$

for $i \neq j = 1, \ldots, K$, where $\boldsymbol{R}_{i\cdot}$ and $\boldsymbol{R}_{\cdot j}$ indicate, respectively, the i–th row and the j–th column of the matrix $\boldsymbol{R}$. Unfortunately, these equations do not have an explicit solution.[3] An acceptable compromise which should increase efficiency, although formally it cannot be interpreted as an ML estimator, is to normalize the estimator $\boldsymbol{Q}$ obtained above in order to transform it to a correlation matrix:

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

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

$$\frac{T}{2}\left[-\ln|\widetilde{\boldsymbol{R}}| - \mathrm{trace}(\widetilde{\boldsymbol{R}}^{-1}\boldsymbol{Q} + \mathrm{trace}(\boldsymbol{Q})\right] + \sum_{t=1}^{T}\sum_{i=1}^{K}\ln f_i(\varepsilon_{t,i}|\mathcal{F}_{t-1}). \tag{30}$$

It is interesting to note that, as long as (29), Equation (30) 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 (23) changed to

$$d_{t,i} = \frac{\boldsymbol{C}_{i\cdot}\boldsymbol{q}_t}{\phi(q_{t,i})} \tag{31}$$

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

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

and $\phi(.)$ 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.

[3] Even when $\boldsymbol{R}$ is a $(2,2)$ matrix, the value of $\boldsymbol{R}_{12}$ has to satisfy the cubic equation:

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

4 Trading Activity and Volatility within a vMEM

Trading activity produces a lot of indicators which are therefore characterized by interdependence, even in a dynamic sense. In this application, we concentrate on the joint dynamics of three series proxying 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 intensity. To be clear at the outset, for the sake of parsimony, we choose the realized volatility forecasts as the main object of interest of the multivariate effort; we take univariate modeling of volatility by itself as a benchmark against specifications which explore 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 July 31, 2013 (1656 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

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

where $k(x)$ is the Parzen kernel

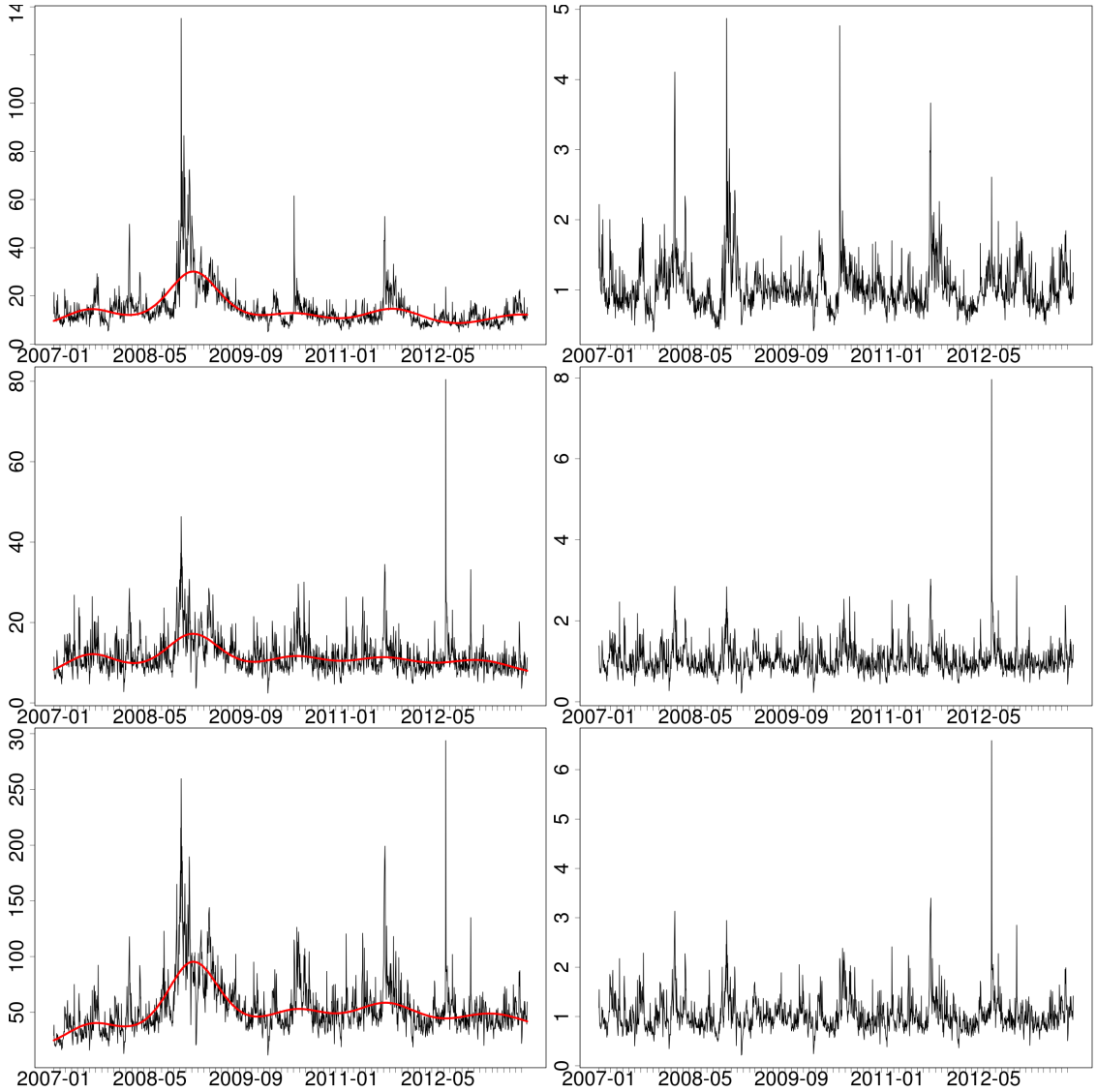
$$k(x) = \begin{cases} 1-6x^2+6x^3 & \text{if } x \in [0,1/2] \\ 2(1-x)^3 & \text{if } x \in (1/2,1] \\ 0 & \text{otherwise} \end{cases},$$

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

x_j is the j-th high frequency return computed according to Barndorff-Nielsen *et al.* (2009, Section 2.2)) and $\widetilde{x}_j$ is the intradaily return of the j-th bin (equally spaced on 15 minute intervals). For volumes (vol) and the number of trades (nt) we 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 trend. 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

Figure 1: Time series of the trading activity indices for JNJ (Jan. 3, 2007 – July 31, 2013). Left: original data; Right: detrended data; 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.



strategy of trend removal (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 in line with the evidence produced by Andersen (1996) for volumes. Remarkably, this upward trend 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 the trend, we adopt a solution similar to Andersen (1996), that is, a flexible function of time which smooths out the series. In detail, assuming that the trend 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;[4]
- the residuals of the previous regression are then exponentiated to get the detrended series.

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

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. Table 1 shows very similar correlations between the estimated trends, all around 0.87. Although quite large, correlations among the detrended series are less homogeneous: as expected, the value concerning volumes and the number of trades, above 0.9, is the highest one; $cor(rkv, vol)$ is below 0.6, whereas the $cor(rkv, nt)$ is around 0.7, confirming the intuition that the enlargement of the variables involved is warranted by the data. Note also that removing trends tends to widen the differences among correlations when compared with the original series.

Table 1: Correlations for JNJ (Jan. 3, 2007 – July 31, 2013). rkv = realized kernel volatility; vol = volume; ntr = number of trades.

	Original		Trend		Detrended	
	vol	ntr	vol	ntr	vol	ntr
rkv	0.645	0.779	0.875	0.877	0.579	0.707
vol		0.903		0.874		0.936

[4] 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.

4.1 Modeling Results

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

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}_1 \boldsymbol{x}_{t-1} + \boldsymbol{\alpha}_2 \boldsymbol{x}_{t-2} + \boldsymbol{\gamma}_1 \boldsymbol{x}_{t-1}^{(-)} + \boldsymbol{\beta}_1 \boldsymbol{\mu}_{t-1}. \tag{32}$$

In order to appreciate the contribution of the different model components, in this conditional mean we consider alternative specifications for the coefficient matrices $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ ($\boldsymbol{\alpha}_2$ and $\boldsymbol{\gamma}_1$ are kept diagonal in all specifications), and for the error term. As of the former, we consider formulations with both $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ diagonal (labeled D); $\boldsymbol{\alpha}_1$ full and $\boldsymbol{\beta}_1$ diagonal (labeled A); both $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ 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. 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), (32) and (8) with Gamma marginals.

	Error Distribution (copula)		
Conditional Mean (parameters)	I: Independent	N: Normal	T: Student-T
D: $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\alpha}_1$ diagonal	D-I		
A: $\boldsymbol{\alpha}_1$ full; $\boldsymbol{\beta}_1$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\alpha}_1$ diagonal	A-I	A-N	A-T
AB: $\boldsymbol{\alpha}_1$, $\boldsymbol{\beta}_1$ full; $\boldsymbol{\gamma}_1$, $\boldsymbol{\alpha}_1$ diagonal		AB-N	AB-T

Estimation results are reported in Table 3, limiting ourselves to the equation for the realized volatility.[5] The Student-T copula turns out to be the favorite specification, judging upon a significantly higher log–likelihood function value, and lower information criteria;[6] the equation–by–equation approach (Independent copula) is dominated in both respects. Without reporting the parameter details, an estimation of the Diagonal model with the Normal/Student-T copula function shows log-likelihood values of 1993.77, respectively, 2076.51 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. It is interesting to note that residual autocorrelation is substantially reduced only in the case of richer parameterizations (AB), where both non–diagonal $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are allowed to capture possible interdependencies.

The impact of volumes and trades on the realized volatility is present but generally individually not significant, probably due to collinearity (as noted, a log–likelihood test would

[5] 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 ω.

[6] The estimated degrees of freedom are 9.10 (s.e. 1.24) and 8.72 (s.e. 1.13), 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 2093.52, very close to the concentrated log-likelihood approach (Section 3.2.2) used in Table 3.

reject the joint hypothesis of the relevant coefficients being equal to zero). In general, it seems that the contribution of the number of trades is more discernible in the presence of a full $\boldsymbol{\beta}_1$; such a significance is also highlighted by the causality tests. The coefficients at lag 2 are always significant with negative signs. The Normal and the Student-T specifications appear to provide similar point estimates, except for the non-diagonal β coefficients: for the latter, once again, the picture is clouded by collinearity, since a formal log–likelihood test does indicate joint significance.

The overview on the results is complete by examining the Table 4 where we report the estimated coefficient $\widehat{\phi}$ for all three Gamma marginals, showing that the estimated unconditional distributions of the estimated residuals have slightly different shapes.

Table 3: Estimated coefficients of the realized volatility equation for different model formulations (cf. Table 2) for JNJ (Jan. 3, 2007 – July 31, 2013). Robust t-stats in parentheses. An empty space indicates that the specification did not include the corresponding coefficient. Causality tests (rows with the arrows) report p-values for the hypothesis $H_0 : \alpha_{1,j} = \beta_{1,j} = 0$ $(j = 2, 3)$. 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.

	D-I	A-I	A-N	A-T	AB-N	AB-T
rkv_{t-1}	0.4929	0.4601	0.4263	0.4239	0.3718	0.3757
	(58.78)	(53.00)	(50.39)	(50.30)	(39.53)	(41.37)
vol_{t-1}		−0.0244	−0.0281	−0.0241	−0.0209	−0.0305
		(-0.89)	(-0.83)	(-0.68)	(-0.49)	(-0.72)
nt_{t-1}		0.0568	0.0615	0.0533	0.1846	0.1821
		(1.53)	(1.37)	(1.22)	(2.70)	(2.74)
rkv_{t-2}	−0.2647	−0.2453	−0.1961	−0.2007	−0.1785	−0.1843
	(-5.16)	(-3.91)	(-3.19)	(-3.70)	(-3.84)	(-4.40)
$rkv_{t-1}^{(-)}$	0.0265	0.0276	0.0272	0.0304	0.0182	0.0219
	(0.84)	(0.83)	(0.87)	(1.00)	(0.44)	(0.54)
$\mu_{t-1}^{(rkv)}$	0.7172	0.6990	0.6821	0.6942	0.7533	0.7638
	(13.94)	(9.96)	(10.03)	(11.75)	(14.18)	(16.49)
$\mu_{t-1}^{(vol)}$					−0.1233	−0.0586
					(-0.91)	(-0.49)
$\mu_{t-1}^{(nt)}$					−0.0918	−0.1463
					(-0.64)	(-1.11)
$rkv_t \leftarrow vol_{t-1}$		0.3733	0.4083	0.4979	0.2530	0.3187
$rkv_t \leftarrow nt_{t-1}$		0.1268	0.1703	0.2224	0.0124	0.0133
logLik	238.24	270.87	2062.12	2133.07	2092.77	2168.15
AIC	-440.48	-493.73	-4070.25	-4210.14	-4119.53	-4268.30
BIC	-323.28	-337.47	-3894.46	-4027.83	-3904.68	-4046.94
LB(12)	0.0000	0.0000	0.0005	0.0002	0.0552	0.0239
LB(22)	0.0000	0.0000	0.0003	0.0002	0.0291	0.0195
LB(32)	0.0000	0.0001	0.0015	0.0012	0.0357	0.0261

Table 4: Estimated ϕ parameters of the Gamma marginal distributions.

	D-I	A-I	A-N	A-T	AB-N	AB-T
ϕ_1	23.08	23.22	23.31	22.34	23.70	22.62
ϕ_2	15.56	16.01	15.82	15.38	15.89	15.46
ϕ_3	19.55	19.63	19.23	18.23	19.34	18.36

Finally, the correlation coefficients implied by the copula–based specifications are reported 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.

	A-N		A-T		AB-N		AB-T	
	vol_t	nt_t	vol_t	nt_t	vol_t	nt_t	vol_t	nt_t
rkv_t	0.479	0.606	0.499	0.621	0.481	0.609	0.502	0.625
vol_t		0.902		0.917		0.903		0.917

4.2 Forecasting

We left the period August 1 – December 31, 2013 (106 observations) for out–of–sample forecasting comparisons. We adopt a Diebold and Mariano (1995) test statistic for superior predictive ability using the error measures

$$e_{N,t} = \frac{1}{2}(x_t - \mu_t)^2 \qquad e_{G,t} = \ln \frac{x_t}{\mu_t} - \frac{x_t}{\mu_t} - 1. \tag{33}$$

where x_t and μ_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 μ_t; similarly, $e_{G,t}$ can be interpreted as the loss we can derive considering x_t as Gamma distributed with mean μ_t and variance proportional to μ_t^2.

Table 6 reports the values of the Diebold-Mariano test statistic of different model formulations, against the D-I specification, considering one-step ahead predictions. We notice a progressive and significant improvement of the specifications allowing for more interdependencies, for both the detrended and the original series,[7] when the Normal-based error is considered. As far as the Gamma-based error, significant statistics emerge for the specifications with the Student-T copula, although also those involving the Normal copula are borderline.

[7] One-step ahead predictions at time t for the original series are computed multiplying the corresponding forecast of the detrended indicator by the value of the trend at $t-1$.

Table 6: Diebold-Mariano test statistics for unidirectional comparisons, against the D-I formulation, considering 1-step ahead forecasts (out–of–sample period August 1 – December 31, 2013). The error measures are defined as $e_{N,t} = 0.5(x_t - \mu_t)^2$ and $e_{G,t} = \ln(x_t/\mu_t) - x_t/\mu_t - 1$, where x_t and μ_t denote the observed and the predicted values, respectively (cf. Section 4.2 for the interpretation). Boldface indicates 5% significant statistics.

	$e_{N,t}$		$e_{G,t}$	
Formulation	detrended	original	detrended	original
A-I	0.924	0.949	1.043	1.050
A-N	1.644	**1.754**	1.564	1.571
A-T	**1.821**	**1.950**	**1.763**	**1.768**
AB-N	**2.004**	**1.940**	1.641	1.649
AB-T	**2.128**	**2.088**	**1.807**	**1.814**

5 Conclusions

In this paper we have presented a general discussion of the vector specification of the Multiplicative Error Model introduced by Engle (2002): a positive valued process is seen as the product of a scale factor which follows a GARCH type specification and a unit mean innovation process. Engle and Gallo (2006) estimate a system version of the MEM by adopting a dynamically interdependent specification for the scale factors (each variable enters other variables' specifications with a lag) but keeping a diagonal variance–covariance matrix for the Gamma–distributed innovations. The extension to a truly multivariate process requires a complete treatment of the interdependence among the innovation terms; in this respect, the specification using multivariate Gamma distributions is too restrictive because of their limitations. One possibility is to avoid the specification of the distribution and adopt a semiparametric GMM approach as in Cipollini *et al.* (2013). Alternatively, and it is the avenue pursued here, we can derive a maximum likelihood estimator by framing the innovation vector as a copula function linking Gamma marginals.

We illustrate the procedure on three indicators related to market 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 can be removed before the rest of the analysis is performed. The three trends are highly correlated with one another, but interestingly, their removal does not have a substantial impact on the correlation among the detrended series. The specifications adopted start from the consideration of a diagonal structure where no dynamic interaction is allowed and an Independent copula (*de facto* an equation–by–equation specification) as a benchmark. 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. Although hindered by the presence of 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 log–likelihood, better information criteria and improved autocorrelation diagnostics. The

results of an out–of–sample forecasting exercise confirm the superiority of richer specifications and a slight preference for the Student-T copula. From a substantive point of view, we interpret the results as showing that the past of realized volatility by itself is not enough information to reconstruct the dynamics and establish its forecastability as such a variable is influenced by other indicators of market activity.

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 Oprationelle, 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.

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., Engle, R. F., and Gallo, G. M. (2013). Semiparametric vector MEM. *Journal of Applied Econometrics*, **28**, 1067–1086.

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

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, **13**, 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. (2014). 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.

Noss, J. (2007). The econometrics of optimal execution in an order-driven market. Technical report, Nuffield College, Oxford University, Master's Thesis.

Patton, A. (2007). Copula-based models for financial time series. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*. Springer Verlag.

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.

Appendix: Expectation Targeting

We show how to obtain the asymptotic distribution of the estimator of the parameters of the vMEM when the constant $\boldsymbol{\omega}$ 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 $\overline{\boldsymbol{x}}_T$ 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 $\boldsymbol{x}_t^{(-)}$'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*

$$\boldsymbol{v}_t = \boldsymbol{x}_t - \boldsymbol{\mu}_t \tag{34}$$

and

$$\widetilde{\boldsymbol{x}}_t = \boldsymbol{x}_t^{(-)} - \boldsymbol{x}_t/2. \tag{35}$$

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

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

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

$$\sqrt{T}\begin{pmatrix} \overline{\boldsymbol{v}}_T \\ \overline{\widetilde{\boldsymbol{x}}}_T \end{pmatrix} \xrightarrow{d} N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_v & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_v \odot \boldsymbol{\Sigma}_I \end{pmatrix}\right]. \tag{36}$$

The asymptotic distribution of $\overline{\boldsymbol{x}}_T$.

By replacing (34) and (35) into equation (3) and arranging it we get

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

so that, averaging both sides,

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

Deriving $\overline{\boldsymbol{x}}_T$ we get

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

where $\boldsymbol{\mu}$ is given in (26).

By means of (36), the asymptotic distribution of $\overline{\boldsymbol{x}}_T$ follows immediately as

$$\sqrt{T}\left(\overline{\boldsymbol{x}}_T - \boldsymbol{\mu}\right) \xrightarrow{d} N\left[\boldsymbol{0}, \boldsymbol{A}^{-1}\left(\boldsymbol{B}\boldsymbol{\Sigma}_v\boldsymbol{B}' + \boldsymbol{C}\left(\boldsymbol{\Sigma}_v \odot \boldsymbol{\Sigma}_I\right)\boldsymbol{C}'\right)\boldsymbol{A}^{-1\prime}\right] \tag{37}$$