Dipartimento di Statistica
"Giuseppe Parenti"

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

WORKING PAPER 2007/16

A Model for Multivariate Non-negative Valued Processes in Financial Econometrics

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

Università degli Studi di Firenze

A Model for Multivariate Non-negative Valued Processes in Financial Econometrics *

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

This version: December 2007

Abstract

The Multiplicative Error Model introduced by Engle (2002) for positive valued processes is specified as the product of a (conditionally autoregressive) scale factor and an innovation process with positive support. In this paper we propose a multivariate extension of such a model, by taking into consideration the possibility that the vector innovation process be contemporaneously correlated. The estimation procedure is hindered by the lack of probability density functions for multivariate positive valued random variables. We suggest the use of copula functions to jointly estimate the parameters of the scale factors and of the correlations of the innovation processes. We illustrate the feasibility of the procedure and the gains over the equation by equation approach using a four variable fully interdependent model with different volatility measures.

*Some of the material presented here circulated in the NBER WP 12690 (2006) "Vector Multiplicative Error Models: Representation and Inference". We had many useful discussions on MEMs and multivariate extensions with Christian T. Brownlees, Marco J. Lombardi and Margherita Velucchi. We thank Nikolaus Hautsch, Loriano Mancini, Joseph Noss, Andrew Patton, Kevin Sheppard as well as participants in seminars at CORE, Humboldt Universität zu Berlin, Imperial College, IGIER–Bocconi and at ESEM 2007 for helpful comments. The usual disclaimer applies. Financial support from the Italian MIUR under grant PRIN 2006131140_004 is gratefully acknowledged.

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

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

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

1 Introduction

The study of financial market behavior is increasingly based on the analysis of the dynamics of nonnegative valued processes, such as exchanged volume, high–low range, absolute returns, financial durations, number of trades, and so on. Generalizing the GARCH (Bollerslev (1986)) and ACD (Engle and Russell (1998)) approaches, 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 i.i.d. with unit mean.

More recently, Engle and Gallo (2006) have investigated three different measures of volatility, namely absolute returns, daily range and realized volatility in a multivariate context in which each lagged indicator is allowed to enter the equation of the scale factor of the other indicators. Model selection techniques are adopted to ascertain the statistical relevance of such variables in explaining the dynamic behavior of each indicator. The model was estimated assuming independence of the innovation terms.

Estimation equation–by–equation ensures consistency of the estimators in a quasi-maximum likelihood context, given the stationarity conditions discussed by Engle (2002). This simple procedure is obviously not efficient, since correlation among the innovation terms is not taken into account: in several cases, especially when predetermined variables are inserted in the specification of the conditional expectation of the variables, it would be advisable to work with estimators with better statistical properties, since model selection and ensuing interpretation of the specification is crucial in the analysis.

In this paper we investigate the problems connected to a multivariate specification and estimation of the MEM. Since joint probability distributions for nonnegative–valued random variables are not available except in very special cases, we resort to the adoption of two different copula functions to link together marginal probability density functions specified as Gamma as in Engle and Gallo (2006). The empirical applications performed on the General Electric stock data show that there are some numerical differences in the estimates obtained by the three methods: assuming innovation dependence gives results which are fairly similar to one another while under the independence assumption we have a substantial departure from the system based results.

2 Multiplicative Error Models

The Multiplicative Error Model (MEM) extends the GARCH approach to processes x_t with non-negative support (Engle (2002), Engle and Gallo (2006)). In a univariate $(1, 1)$ framework, the model is specified as the product of a scale factor conditionally dependent on past information ($\mathcal{F}_{t-1}$) and an iid innovation term

$$x_t = \mu_t \varepsilon_t = (\omega + \alpha x_{t-1} + \beta \mu_{t-1} + \boldsymbol{\gamma}' \mathbf{z}_{t-1}) \, \varepsilon_t, \tag{1}$$

where $\mathbf{z}_{t-1}$ denotes a set of predetermined variables. Typically, a flexible assumption of a Gamma pdf with unit mean for the iid ε_t terms (henceforth $Gamma(\phi_i, \phi_i)$, where $V(\varepsilon_t|\mathcal{F}_{t-1}) = 1/\phi_i$) is adopted, which in turns allows for the interpretation of μ_t as the conditional expectation of x_t. The estimator of the parameters in μ_t is consistent by its Quasi Maximum Likelihood derivation.

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

Extending the specification to a multivariate case, let $\mathbf{x}_t$ be a K–dimensional process with non–negative components;[1] a vector MEM for $\mathbf{x}_t$ is defined as

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t = \mathrm{diag}(\boldsymbol{\mu}_t)\boldsymbol{\varepsilon}_t, \tag{2}$$

where $\odot$ indicates the Hadamard (element–by–element) product. Conditionally on the information set $\mathcal{F}_{t-1}$, $\boldsymbol{\mu}_t$ can be defined as before, except that now we are dealing with a K–dimensional vector depending on a (larger) vector of parameters, say $\boldsymbol{\theta}$. The innovation vector $\boldsymbol{\varepsilon}_t$ is a K–dimensional i.i.d. process with density function defined over a $[0, +\infty)^K$ support, with unit vector $\mathbb{1}$ as expectation and a general variance–covariance matrix $\boldsymbol{\Sigma}$,

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

The previous conditions guarantee that

$$E(\mathbf{x}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \tag{4}$$
$$V(\mathbf{x}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t\boldsymbol{\mu}_t' \odot \boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\mu}_t)\boldsymbol{\Sigma}\,\mathrm{diag}(\boldsymbol{\mu}_t),, \tag{5}$$

where the latter is a positive definite matrix by construction.

A base $(1, 1)$ multivariate specification for $\boldsymbol{\mu}_t$ is

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}\mathbf{x}_{t-1} + \boldsymbol{\beta}\boldsymbol{\mu}_{t-1}, \tag{6}$$

where $\boldsymbol{\omega}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (arranged in a vector $\boldsymbol{\theta}$) have dimensions, respectively, $(K, 1)$, (K, K) and (K, K) (the latter may be assumed diagonal to simplify matters).

In some cases, the specification of $\boldsymbol{\mu}_t$ can be extended to include asymmetric effects

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

associated with the sign of an observed variable:

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}\mathbf{x}_{t-1} + \boldsymbol{\gamma}\mathbf{x}_{t-1}^{(-)} + \boldsymbol{\beta}\boldsymbol{\mu}_{t-1}, \tag{7}$$

where the terms in the vector $\mathbf{x}_t^{(-)}$ contain $x_{t,i}$'s multiplied by a function related to a signed variable, be it a return ($0, 1$ values) or a signed trade (buy or sell $1, -1$ values). For example, when different volatility indicators of the same asset are considered, such an indicator assumes value one when its previous day's return r_{t-1} is negative. In a market volatility spillover study, each market would have its own indicator function built from the sign its own returns $r_{t-1,i}$. Finally, in a microstructure context, we can think of assigning positive or negative values to volumes according to whether the trade was a buy or a sell. The associated parameters $\boldsymbol{\gamma}$ are arranged in a (K, K) matrix.

The parameter space of $\boldsymbol{\theta}$ (including also $\boldsymbol{\gamma}$ if (7) is assumed) must be restricted to ensure $\boldsymbol{\mu}_t \geq \mathbf{0}$ for all t and to ensure stationary distributions for $\mathbf{x}_t$: some technical details are reported in Appendix A.

Before discussing some suitable solutions to handle the multivariate distribution of $\boldsymbol{\varepsilon}_t$, let us provide some examples of cases where a multivariate MEM approach seems appropriate.

Example 1: Volatility Forecasting

There is a huge literature on high–frequency data based measures of volatility. Various forms of realized volatility summarize information on intra–daily activity. Is either of these a sufficient measure of volatility (i.e. depending solely on its own past)?

Question: What are the dynamic interactions among different measures of volatility?

MEM Answer: we can build an interdependent model where realized, Bipower, Two-Scale, Daily Range, Absolute Returns and others can engage in a horse race. We may inspect whether there exists a measure depending just on its own past or the significance of cross links. We can forecast volatility based on a full scale multiperiod interdependent forecast and derive density forecasts for returns (VaR).

Example 2: Volatility Spillovers

There is a huge literature on transmission mechanisms (spillovers, contagion across markets). When the attention is devoted to volatility, typically the analysis is hindered by parametric limitations (multivariate GARCH).

Question: What are the dynamic interactions among volatilities in different market indices?

MEM Answer: we can build an interdependent model where one can use a volatility proxy (e.g. daily range) for different markets and analyze interactions through model selection. We may build interdependent forecasts, derive nonlinear impulse response functions as a scenario analysis tool.

Example 3: Order Execution Dynamics

In order-driven markets there is a tradeoff between the potential payoff of placing orders at a better price, against the risk of these orders not executing.

Question: What is the distribution of the quantity of stock that will execute in the next time period at a given distance from the current price? Is there an interaction between such quantities?

MEM Answer: we can specify a MEM for execution depths: here zeros are relevant because there are times when the quantity which could be executed at a certain distance from current price can be zero. Forecasts can be used for a trading strategy (Noss (2007)).

Example 4: Trades, Duration, Volume and Volatility Dynamics

In a ultra-high frequency framework, trades occur at irregular moments but carry with them information about the time elapsed since the last trade, the (possibly signed) volume and the return associated with the trade (as analyzed by Manganelli (2005)).

Question #1: Can the dynamic interrelationship between the available variables reveal the speed (in market and calendar time) at which private information is incorporated into prices?

MEM Answer: we can specify a MEM for durations, volume and volatility in which the conditional expectations depend just on the past values (not also some contemporary information as in Manganelli (2005)) and take into account the contemporaneous correlation of the innovation terms.

Question #2: Are return volatility, average trade size and number of trades per interval driven by a common latent factor together with idiosyncratic components?

MEM Answer: Hautsch (2007) specifies a three variable Stochastic MEM with a latent common AR(1) factor which is estimated by Simulated Maximum Likelihood.

2.1 Specifications for ε_t

In this section we consider some alternatives about the specification of the distribution of the error term $\varepsilon_t|\mathcal{F}_{t-1}$ of the vector MEM defined above.

2.1.1 Multivariate Gamma formulations

An attempt at generalizing the univariate gamma adopted by Engle and Gallo (2006) to a suitable multivariate version is frustrated by the limitations of the multivariate Gamma distributions available in the literature (Johnson *et al.* (2000, chapter 48)): many of them are bivariate versions, not sufficiently general for our purposes, others are defined via the joint characteristic function, and they require tedious numerical inversion formulas to find their pdf's. The only useful versions remain the multivariate Gamma's of Cheriyan and Ramabhadran (in their more general version, henceforth $GammaCR$) (Johnson *et al.*

(2000, 454–470), 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 $0 < \phi_0 < \min(\phi_1, \ldots, \phi_K)$. All univariate marginal probability functions for $\varepsilon_{i,t}$ are, as required, $Gamma(\phi_i, \phi_i)$ distributed, even if the multivariate pdf is expressed in terms of a complicated integral. The conditional variance matrix of $\boldsymbol{\varepsilon}_t$ has elements

$$C(\varepsilon_{i,t}, \varepsilon_{j,t}|\mathcal{F}_{t-1}) = \frac{\phi_0}{\phi_i \phi_j} \tag{8}$$

so that the correlations are

$$\rho(\varepsilon_{t,i}, \varepsilon_{t,j}|\mathcal{F}_{t-1}) = \frac{\phi_0}{\sqrt{\phi_i \phi_j}}. \tag{9}$$

This implies that the $GammaCR$ distribution admits only positive correlation among its components and that the correlation between each couple of elements is strictly linked to the corresponding variances $1/\phi_i$ and $1/\phi_j$. These various drawbacks (the restrictions on the correlation, the very complicated pdf and the constraint $\phi_0 < \min(\phi_1, \ldots, \phi_K)$), suggest to investigate possible alternatives.

2.1.2 Copula based formulations

A different and, in some sense, modular approach for defining the distribution of $\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}$ is to use *copula functions*.[2] Adopting copulas, the definition of the distribution of a multivariate r.v. can be splitted in:

- choice of the univariate marginals;
- choice of the copula linking them.

Within this framework, the conditional distribution of the error component of the vector MEM can be expressed as

$$\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim C(\boldsymbol{\xi}) \times \prod_{i=1}^{K} M_i(\boldsymbol{\phi}_i), \tag{10}$$

where: $C(\boldsymbol{\xi})$ denotes a copula parameterized by a vector $\boldsymbol{\xi}$; $M_i(\boldsymbol{\phi}_i)$ indicates the distribution of the i-th marginal (again assumed absolutely continuous, with non-negative support and unit expectation), having pdf $f_i(x; \boldsymbol{\phi}_i)$ and cdf $F_i(x; \boldsymbol{\phi}_i)$. The conditional pdf of $\boldsymbol{\varepsilon}_t$

[2]The main characteristics of copulas are summarized, among others, in Joe (1997) and Nelsen (1999). See also Embrechts *et al.* (2002), Cherubini and Vecchiato (2004), McNeil *et al.* (2005) and the review of Patton (2007) for financial applications.

can then be written as

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

where $\mathbf{u}_t = (u_{1,t};\ldots;u_{K,t})$ and $u_{t,i} = F_i(\varepsilon_{t,i};\boldsymbol{\phi}_i)$.

In empirical applications, specific choices of the copula and the univariate marginals are needed. In the following sections we discuss some possible specifications of $C(\boldsymbol{\xi})$ related to the class of Elliptical copulas.

As far as the marginal functions are concerned, we can consider as natural candidates all pdf's with the characteristics mentioned above: examples are Gamma, Inverse-Gamma, Weibull, Lognormal, mixtures of them. For practical reasons, in the following we will make some references to the Gamma case (cf. the discussion on the flexibility of this choice in Engle and Gallo (2006)).

2.1.3 Normal copula

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

$$cN(\mathbf{u};\mathbf{R}) = |\mathbf{R}|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{q}'\mathbf{R}^{-1}\mathbf{q}-\mathbf{q}'\mathbf{q}\right)\right], \tag{11}$$

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

The Normal copula possesses many interesting properties: the capability of capturing a broad range of dependencies (the bivariate Normal copula, according to the value of the correlation parameter, is capable of attaining the lower Fréchet bound, the product copula and the upper Fréchet bound.), the analytical tractability, the ease of simulation.

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, as discussed in Song (2000). In such a case, the conditional variance–covariance matrix of $\boldsymbol{\varepsilon}_t$ has a generic element which is approximately equal to (using a first order expansion)

$$C(\varepsilon_{i,t},\varepsilon_{j,t}|\mathcal{F}_{t-1}) \simeq \frac{\mathbf{R}_{ij}}{\sqrt{\phi_i\phi_j}},$$

so that the correlations are, approximately,

$$\rho(\varepsilon_{i,t},\varepsilon_{j,t}|\mathcal{F}_{t-1}) \simeq \mathbf{R}_{ij}.$$

Even if limited to a particular copula with particular marginals, this example shows clearly the advantages of using copulas over a multivariate Gamma (or other similar) specification: the covariance and correlation structures are more flexible (also negative correlations are permitted); the correlations do not depend heavily on the variances of the marginals; there are no complicated constraints on the parameters; the pdf is more easily tractable.

2.1.4 Student-T copula

As an alternative, we consider the Student-T copula. In fact, even if the Normal copula has a number of attractive features, one of its major drawbacks lies in the asymptotic independence of its tails. Empirically, tail dependence is a behaviour observed frequently in financial time series (see McNeil *et al.* (2005), among others): extreme events in different assets tend to be combined. Summarizing, the Student-T copula shares many of the characteristics of the Normal copula with the main differences in the tails, that are asymptotically dependent.

The density of the Student-T copula is given by

$$cT(\mathbf{u};\mathbf{R},\nu) = \frac{\Gamma((\nu+K)/2)\Gamma(\nu/2)^{K-1}}{\Gamma((\nu+1)/2)}|\mathbf{R}|^{-1/2}\frac{(1+\mathbf{q}'\mathbf{R}^{-1}\mathbf{q}/\nu)^{-(\nu+K)/2}}{\prod_{i=1}^{K}(1+q_i^2/\nu)^{-(\nu+1)/2}}, \qquad (12)$$

where $\mathbf{q} = (q_1;\ldots;q_K)$, $q_i = T^{-1}(u_i;\nu)$ and $T(x;\nu)$ denotes the cdf of the Student-T distribution computed at x.

As a further difference relative to the Normal, for $\mathbf{R} = \mathbf{I}$ we do not obtain the independence copula, since uncorrelated multivariate T r.v.s are not independent (details in McNeil *et al.* (2005)). For a deeper handling of the Student-T copula see Demarta and McNeil (2005).

2.1.5 Elliptical copulas

Both Normal and Student-T copulas are members of the more general family of Elliptical copulas. Elliptical copulas 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. A deeper discussion of this kind of copulas is beyond the scope of the paper: see McNeil *et al.* (2005), Frahm *et al.* (2003), Schmidt (2002). For our purposes, it is sufficient to note that this family provides an unified framework for handling together the Normal, the Student-T and any other member of this family with an explicit density function.

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

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

for suitable choices of $K^*(.,.)$, $g_1(.;.,.)$ and $g_2(.;.)$, where $\mathbf{q} = (q_1;\ldots;q_K)$, $q_i =$

$G^{-1}(u_i;\nu)$. For instance:

- in the Normal copula, where a shape parameter $\boldsymbol{\nu}$ is absent, we have $K^*(K) \equiv 1$, $g_1(x;K) = g_2(x) \equiv \exp(-x/2)$;
- in the Student-T copula, where the shape parameter is a scalar ν, we have $K^*(\nu;K) = \dfrac{\Gamma((\nu+K)/2)\Gamma(\nu/2)^{K-1}}{\Gamma((\nu+1)/2)}$, $g_1(x;\nu,K) = (1+x/\nu)^{-(\nu+K)/2}$, $g_2(x;\nu) = (1+x/\nu)^{-(\nu+1)/2}$.

Elliptical copulas are very interesting and largely applicable. However they have a substantially symmetric behavior (elliptical symmetry) that can constitute a limit in some applications. Different copulas (e.g. Gumbel) can be adopted, following the same scheme proposed here, which may bypass the limitations proposed in the symmetry embedded in the Elliptical copulas.

3 Maximum likelihood inference

We discuss here how to obtain inferences from the vector MEM specified in section 2 assuming that $\boldsymbol{\mu}_t$ is governed by (6) or (7) and, for the moment, the general formulation (10) of the conditional distribution of the error term. A vector MEM specified in this way is driven by the following set of parameters: $\boldsymbol{\theta}$ (into the $\boldsymbol{\mu}_t$ equations); $\boldsymbol{\xi}$ (into the copula); $\boldsymbol{\phi}$ (into the marginals).

It will be useful to recall the sequence in which some objects are computed ($i = 1,\ldots,K$), namely: $\mu_{t,i}(\boldsymbol{\theta}_i) \rightarrow x_{t,i}/\mu_{t,i} = \varepsilon_{t,i} \rightarrow F_i(\varepsilon_{t,i};\boldsymbol{\phi}_i) = u_{t,i} \rightarrow c(\mathbf{u}_t;\boldsymbol{\xi})$. For each of these quantities, the bolded version without the index i denotes the whole vector of the corresponding quantities at time t.

The conditional pdf of $\mathbf{x}_t$ is given by

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

so that the log-likelihood of the model is

$$\begin{aligned} l = & \sum_{t=1}^{T} \ln c(\mathbf{u}_t;\boldsymbol{\xi}) + \sum_{t=1}^{T}\sum_{i=1}^{K} \left[\ln\left(\varepsilon_{t,i} f_i(\varepsilon_{t,i};\boldsymbol{\phi}_i)\right) - \ln x_{t,i}\right] && (13)\\ = & [\text{copula contribution}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\xi})] + [\text{marginals contribution}(\boldsymbol{\theta},\boldsymbol{\phi})]. && (14)\\ & && (15) \end{aligned}$$

In principle, this log-likelihood can be optimized directly using full ML estimators of the three sets of parameters. However, for some choices of the copula, simple estimators of

$\boldsymbol{\xi}$ (usually moment estimators) are available and can be computed from current values of residuals ε_t or of $\mathbf{u}_t$'s. Examples are parameters of Archimedean copulas or of Elliptical copulas derived from Kendall correlations of current estimates of ε_t. When this solution is available, a *pseudo-loglikelihood* can be constructed as

$$l = \sum_{t=1}^{T} \ln c(\mathbf{u}_t; \widehat{\boldsymbol{\xi}}) + \sum_{t=1}^{T} \sum_{i=1}^{K} \left[\ln\left(\varepsilon_{t,i} f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)\right) - \ln x_{t,i},\right] \tag{16}$$

where $\widehat{\boldsymbol{\xi}}$ is the current estimate of $\boldsymbol{\xi}$. Invoking asymptotic arguments for its adoption, such a possibility can reduce considerably the amount of computations during optimization because (16) depends only on $\boldsymbol{\theta}$ and ϕ.

Another solution for speeding up computations and improve numerical stability is to use analytical derivatives when optimization uses the score function. However, details about this point are rather technical and given as a reference in Appendix B.

In the following section we illustrate some details of the estimation when the copula is Normal or Student-T (presenting them within the unified framework of Elliptical copulas) and when the marginals are Gamma distributed. Different choices can be accommodated given the modular approach to the problem.

3.1 The Elliptical copula case

We discuss here full ML estimation of all parameters of the vector MEM when the conditional distribution of the error term is specified as a Normal, Student-T or any other Elliptical copula with explicit density function.

In order to perform full ML estimation of the correlation matrix of the copula, we reparameterize it in an unconstrained way, as illustrated in McNeil *et al.* (2005, p. 235). In fact any correlation matrix $\mathbf{R}$ can be represented as

$$\mathbf{R} = \mathbf{D}\mathbf{c}'\mathbf{c}\mathbf{D}, \tag{17}$$

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 $D_j = \left(1 + \sum_{i=1}^{j} c_{ij}^2\right)^{-1/2}$ for $j = 2, \ldots, K$. Using this approach, the estimation of $\mathbf{R}$ is transformed in an unconstrained problem, since the $K(K-1)/2$ free elements of $\mathbf{c}$ can vary into $\mathbb{R}$.

Using the notation given in section 2.1.5, the log-likelihood of the model can be written as

$$\begin{aligned} l = &-\sum_{t=1}^{T} \ln x_{t,i} + T\left[-\frac{1}{2}|\mathbf{R}| + \ln K^*(\boldsymbol{\nu}; K)\right] \\ &+ \sum_{t=1}^{T}\left[\ln g_1(\mathbf{q}_t'\mathbf{R}^{-1}\mathbf{q}_t; \boldsymbol{\nu}; K) - \sum_{i=1}^{K} \ln g_2(q_{t,i}^2; \boldsymbol{\nu}) + \sum_{i=1}^{K} \ln\left(\varepsilon_{t,i} f_i(\varepsilon_{t,i}; \boldsymbol{\phi}_i)\right)\right], \end{aligned} \tag{18}$$

where

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

We remark that, using (17), $\frac{1}{2}\ln(|\mathbf{R}|) = \sum_{i=2}^{K}\ln D_i$. Furthermore, if $\varepsilon_{t,i}|\mathcal{F}_{t-1} \sim Gamma(\phi_i,\phi_i)$ then $\varepsilon_{t,i}f_i(\varepsilon_{t,i};\boldsymbol{\phi}_i)$ is the pdf of a $Gamma(\phi_i+1,\phi_i)$.

Of course, different methods relative to full ML estimation can be taken into account, in particular for estimating the parameters of the copula: for instance, we can use Kendall correlations for estimating $\boldsymbol{R}$ (Lindskog *et al.* (2003)) or tail dependence indices for estimating the shape parameter $\boldsymbol{\nu}$ (Kostadinov (2005)). Using this approach, a pseudo-loglikelihood as in (16) can be constructed and optimized. We do not explore further this alternative in the paper, even if this approach can be implemented along the lines illustrated.

3.2 The Normal copula case

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

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

where

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

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

$$\frac{\partial l}{\partial \mathbf{R}} = \frac{T}{2}(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Q}\mathbf{R}^{-1}) \Rightarrow \widehat{\mathbf{R}} = \mathbf{Q}.$$

Hence the unconstrained ML estimator of $\mathbf{R}$ has an explicit form. Replacing this estimator of $\mathbf{R}$ in the log-likelihood function we obtain a *concentrated log-likelihood*

$$lc = \frac{T}{2}\left[-\ln|\mathbf{Q}| - K + \text{trace}(\mathbf{Q})\right] + [\text{marginals contribution}], \tag{19}$$

However, the estimator of $\mathbf{R}$ above is obtained without imposing any constraint relative to its nature as a correlation matrix ($\text{diag}(\mathbf{R}) = \mathbb{1}$ and positive definiteness). Computing directly the derivatives with respect to the off–diagonal elements of $\mathbf{R}$ we obtain, after some algebra, that the ML estimate 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.[3] An acceptable compromise which should increase efficiency, although formally it cannot be interpreted as an ML estimator, is to normalize the estimator obtained above, $\widehat{\mathbf{R}}$, in order to transform it in a correlation matrix:

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

where $\mathbf{D}_Q = \mathrm{diag}(Q_{11}, \dots, Q_{KK})$. 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 i.i.d. r.v.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

$$lc = \frac{T}{2}\left[-\ln|\widetilde{\mathbf{R}}| - \mathrm{trace}(\widetilde{\mathbf{R}}^{-1}\mathbf{Q}) + \mathrm{trace}(\mathbf{Q})\right] + (\text{marginals contribution}). \tag{20}$$

It is interesting to note that, as long as (19), (20) also gives a relatively simple structure of the score function (see appendix B), even if the second one provides better estimates of parameters (consideration based on simulations not reported here).

4 An Illustrative Example

The multivariate extension of the MEM model is illustrated on the GE daily stock (time period 01/03/1995-12/29/2001, corresponding to 1515 observations) for which the following four indicators of volatility are considered:

1. one is the *absolute daily return* $|r_t|$, taken to be the absolute value of the open-to-close difference between log-prices;
2. the second is the *realized absolute variation* av_t (Barndorff-Nielsen and Shephard (2004)), computed as the sum of the absolute value of the M intra-daily returns, namely, $arv_t = \frac{1}{\Gamma(0.5)}\sqrt{\frac{2}{M}}\sum_{d=1}^{M}|r_{d,t}|$;
3. the third is the *realized bi-power variation* brv_t (Barndorff-Nielsen and Shephard (2004)), defined as the sum of the $M-1$ products of pairs of subsequent intra-daily

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

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

absolute returns, namely, $bv_t = \frac{1}{\Gamma(0.5)}\sqrt{2\sum_{d=2}^{m}|r_{d,t}||r_{d-1,t}|}$ (the scale transformation is needed to have it comparable with the other indicators)

4. the fourth is *realized volatility* rv_t (Andersen *et al.* (2003)), defined as the sum of the squares of the M intra-daily returns, namely $rv_t = \sqrt{\sum_{d=1}^{M} r_{d,t}^2}$.

Within the last three measures, the $r_{d,t}$'s represent the log differences of the transaction price series recorded at or about five minute intervals; here M is 77, as we are excluding overnight returns.

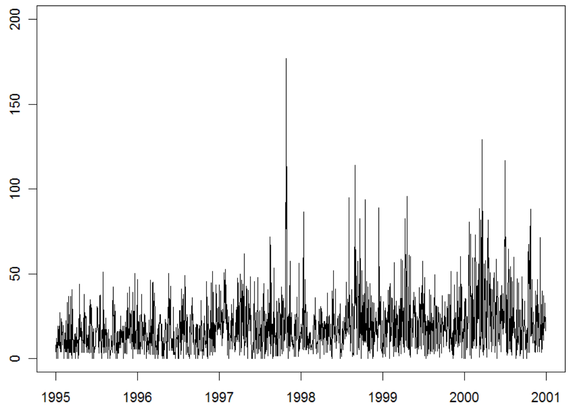
All these variables were rescaled multiplying them by $100\sqrt{252}$. In Table 1 we report some descriptive statistics relative to the variables analyzed, the time series of which are depicted in Figure 1. As one may expect, realized absolute and bi-power variations show a relatively close behavior. Realized volatility, although very similar to the previous ones, has a slightly larger mean and higher kurtosis that, however, depends essentially on an extreme observation located at the end of 1997. As usual, absolute returns have a more erratic behavior, as witnessed by the graph and the larger standard deviation. Furthermore, there are some zero values in absolute daily returns. Unconditional correlations across the whole sample period confirm the fact that the ultra-high frequency based measures share a great deal of common information: the values concerning realized volatilities only are above 0.97, whereas correlations of $|r_t|$ with each one of the remaining are just above 0.4.

Table 1: Some descriptive statistics of the indicators $|r_t|$, arv_t, brv_t rv_t – GE stock, 01/03/1995-12/29/2001 (1515 obs).

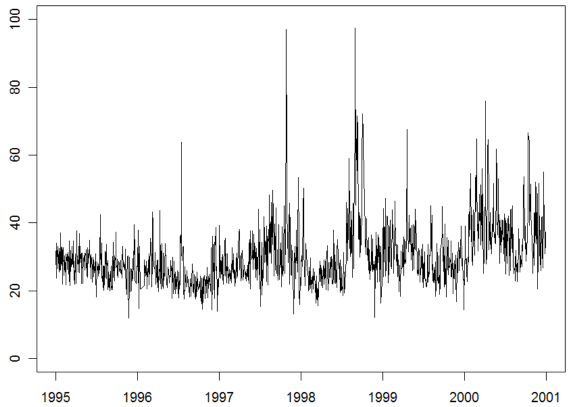
	Indicator			
Statistics	$\|r_t\|$	arv_t	brv_t	rv_t
min	0.00	11.99	15.04	19.05
max	176.77	97.36	99.15	142.83
mean	19.16	30.22	32.48	36.03
sd	16.80	8.96	9.38	9.41
skewness	2.20	1.98	1.95	2.65
kurtosis	12.42	10.05	9.44	18.69
Correlations		arv_t	brv_t	rv_t
$\|r_t\|$		0.4202	0.4207	0.4403
arv_t			0.9832	0.9741
brv_t				0.9751

As an initial benchmark to describe the characteristics of the series examined in isolation, let us consider a *diagonal* model in which all matrices $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are taken to be diagonal and hence no interdependence is allowed among indicators. Such diagonal model is estimated equation by equation and the results are reported in table 2. The familiar GARCH range of values for $\widehat{\gamma}_i$ and $\widehat{\beta}_i$ surfaces for the absolute returns, while quite different values are obtained for the ultra-high frequencies based measures. The shape parameter for the Gamma marginals are reported at the bottom of the table pointing to absolute returns being more noisy as expected.

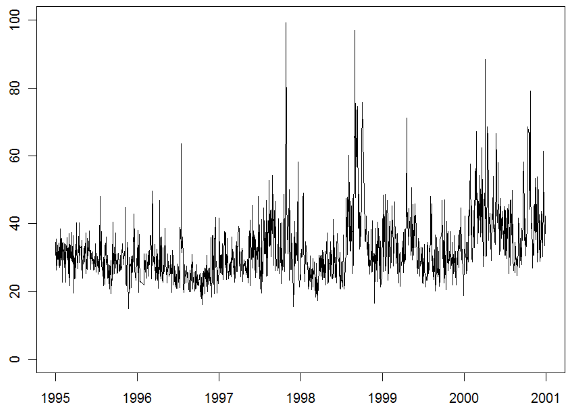
Figure 1: Time series of the indicators $|r_t|$, arv_t, brv_t rv_t – GE stock, 01/03/1995-12/29/2001.



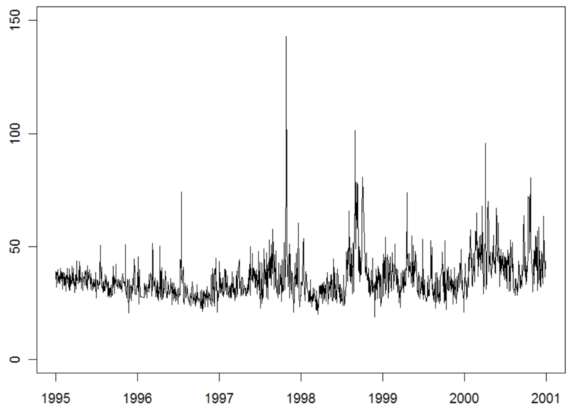
(a) $|r_t|$



(b) arv_t



(c) brv_t



(d) rv_t

Table 2: vMEM on Four Volatility Indicators: Parameter estimates assuming diagonal $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ matrices and the Independent copula (standard errors in small font).

	$\|r_t\|$	arv_t	brv_t	rv_t
ω_i	0.4622	2.6794	2.6320	3.5493
α_i	0	0.4020	0.3726	0.4131
	-	0.0399	0.0402	0.0463
γ_i	0.0637	0	0	0
	0.0212	-	-	-
β_i	0.9441	0.5094	0.5463	0.4884
	0.0239	0.0537	0.0544	0.0560
ϕ_i	1.2852	27.9853	28.6715	36.3381

We then formulated a vMEM for these four measures of volatility assuming that each one of the multiplicative error terms has a Gamma marginal distribution. We compared three different assumptions about the copula linking these marginals: T copula, Normal copula, Independent copula. The vMEM relative to the first assumption has been estimated using full ML, as illustrated in section 3.1; that corresponding to the second one has been estimated using a concentrated ML approach (see section 3.2); that concerning last assumption has been estimated equation by equation.

In Table 3 we compare estimates of the $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\beta}$ matrices corresponding to these different assumptions. In order to make comparisons easier, we report the vMEM in the specification obtaining the smallest BIC when a Student-T copula is assumed. The results show that the T copula and the Normal copula deliver substantially similar results. The equation by equation method, on the contrary, has some trouble in estimating some parameters as they are absorbed by the boundary of zero.

Note that lagged absolute returns enter significantly in all specifications (as an asymmetric term), realized volatility does not spill over onto other indicators and that absolute and bi-power volatility show links with one another and with absolute returns (the former) and with realized volatility (the latter).

A comment on the shape of the conditional distribution of the multiplicative error terms can be based on the estimated parameters of the assumed Gamma marginals and of the copula linking them. Tables 4 and 5 show the estimates for the three formulations investigated. The estimated ϕ_i parameters are quite different from one to another, with the exceptions of arv and brv that are very similar. Residuals from absolute returns display higher variability (we recall that $V(\varepsilon_{t,i}|\mathcal{F}_{t-1}) = 1/\phi_i$) than those from realized volatilities. Furthermore, there some difference depending on the assumed copula. In fact, ϕ_i parameters estimated when the copula is Student-T are smaller than those obtained assuming the Normal or the Independent copula. We speculate that this is due to the tail dependence parameter ν, estimated around 13, that can take into account some combined extreme values of the error terms, otherwise inputed to the marginals. On the contrary, the estimated correlations $\mathbf{R}$ are substantially similar.

A synthetic picture of the total influence of variables can be measured by means of an *Impact Matrix* $\mathbf{A} = \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma}/2$ (see appendix B) displayed in Table 6. We can note that those obtained assuming the Student-T and the Normal copula are substantially similar, and substantially different from those obtained under the Independent copula. Similar comments can be made for the estimated characteristic roots of the impact matrix ruling the dynamics of the system: in this particular case, the estimated persistence is underestimated, relative to the other two cases, assuming the Independent copula.

As a way to illustrate the practical implications of the estimated system by different methods, we can produce dynamic volatility forecasts. Figure 2 shows forecasts of the four indicators choosing some dates as starting values and projecting the model 40 periods ahead. The results for the Student-T copula are superimposed to those of the Normal copula in all cases. The difference over the Independent copula is that not taking into consideration the simultaneous correlation of the innovations, one has a faster reversion to the long run expectation with differences which can be in the order of 3–4 volatility

Table 3: vMEM on Four Volatility Indicators: Estimated Parameters (Robust standard errors in small font).

α matrix

	T copula				Normal copula				Independent copula			
	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}
$\|r_t\|$	0	0	0.0813	0	0	0	0.0982	0	0	0	0.1346	0
	-	-	0.0815	-	-	-	0.0629	-	-	-	0.0706	-
arv_t	0	0.1927	0.0548	0	0	0.1961	0.0529	0	0	0.3379	0.0000	0
	-	0.0426	0.0210	-	-	0.0407	0.0228	-	-	0.0340	0.0000	-
brv_t	0	0.1914	0.0640	0	0	0.1864	0.0694	0	0	0.3439	0.0000	0
	-	0.0434	0.0233	-	-	0.0402	0.0250	-	-	0.0344	0.0000	-
rv_t	0	0.0765	0	0.1787	0	0.0775	0	0.1770	0	0.1940	0	0.1567
	-	0.0313	-	0.0229	-	0.0290	-	0.0228	-	0.0645	-	0.0669

γ matrix

	T copula				Normal copula				Independent copula			
	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}
$\|r_t\|$	0.0782	0	0	0	0.0811	0	0	0	0.0869	0	0	0
	0.0209	-	-	-	0.0203	-	-	-	0.0294	-	-	-
arv_t	0.0655	0	0	0	0.0576	0	0	0	0.0744	0	0	0
	0.0119	-	-	-	0.0121	-	-	-	0.0117	-	-	-
brv_t	0.0657	0	0	0	0.0579	0	0	0	0.0752	0	0	0
	0.0124	-	-	-	0.0128	-	-	-	0.0121	-	-	-
rv_t	0.0683	0	0	0	0.0595	0	0	0	0.0799	0	0	0
	0.0126	-	-	-	0.0123	-	-	-	0.0141	-	-	-

β matrix

	T copula				Normal copula				Independent copula			
	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}
$\|r_t\|$	0.8041	0	0	0	0.7805	0	0	0	0.7139	0	0	0
	0.1377	-	-	-	0.1031	-	-	-	0.1366	-	-	-
arv_t	0	0.6645	0	0	0	0.6675	0	0	0	0.5385	0	0
	-	0.0587	-	-	-	0.0538	-	-	-	0.0465	-	-
brv_t	0	0	0.6660	0	0	0	0.6715	0	0	0	0.5489	0
	-	-	0.0580	-	-	-	0.0523	-	-	-	0.0461	-
rv_t	0	0	0	0.6555	0	0	0	0.6619	0	0	0	0.5200
	-	-	-	0.0583	-	-	-	0.0520	-	-	-	0.0474

Note: For the sake of not overburdening symbols, we partially abused notation by choosing common labels for the rows and columns of the matrices.

Table 4: vMEM on Four Volatility Indicators: Estimated parameters of the Gamma marginals ($\phi_i = 1/V(\varepsilon_{t,i}|\mathcal{F}_{t-1})$).

	$\|r_t\|$	arv_t	brv_t	rv_t
T copula	1.2624	27.0821	28.1324	35.8882
Normal copula	1.2946	29.0425	30.1491	38.6049
Independent copula	1.2958	29.0442	30.1346	38.1665

Table 5: vMEM on Four Volatility Indicators: Estimated parameters of the Copulas for the Student-T and the Normal choices.

	T copula		
ν	13.27		
R	arv_t	brv_t	rv_t
$\|r_t\|$	0.2257	0.2246	0.2357
arv_t		0.9691	0.9636
brv_t			0.9588

	Normal copula		
R	arv_t	brv_t	rv_t
$\|r_t\|$	0.2219	0.2219	0.2385
arv_t		0.9634	0.9572
brv_t			0.9542

percentage points (on an annual basis).

5 Conclusions

In this paper we have presented a general discussion of the vector specification of the Multiplicative Error Models 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 multivariate process requires interdependence among the innovation terms. The specification in a multivariate framework cannot exploit multivariate Gamma distributions because they appear too restrictive. The maximum likelihood estimator can be derived by setting the multivariate innovation process in a copula with Gamma marginals framework. The empirical results with daily data on the GE stock show the feasibility of both procedures: the system estimator shows deviations from the equation–by–equation approach. Overall, the applications show the need to incorporate several measures of volatility in a single model

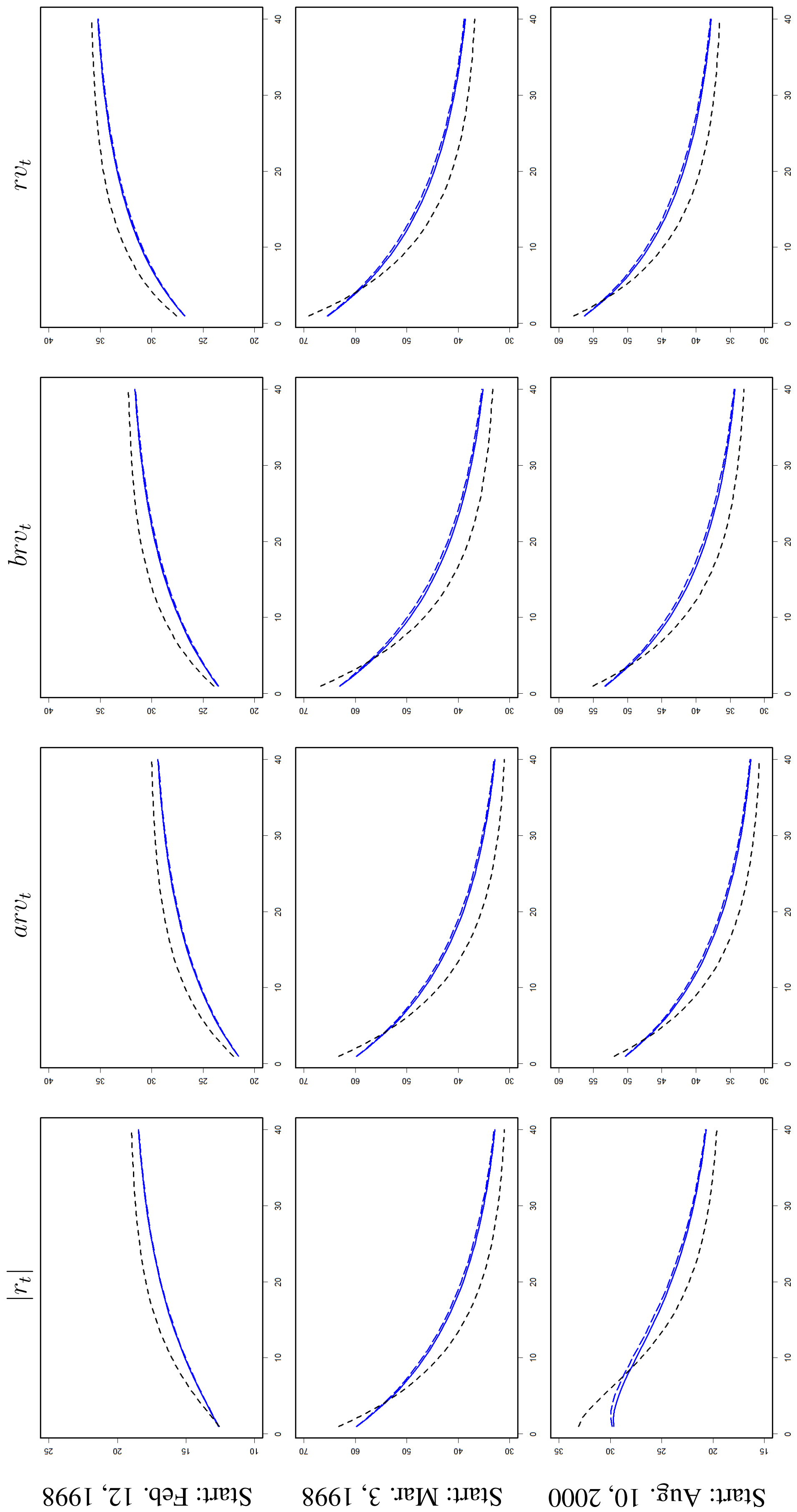


Figure 2: Example of forecasts of $|r_{t+\tau}|$, $arv_{t+\tau}$, $brv_{t+\tau}$, $rv_{t+\tau}$ starting at three different dates and $\tau = 1, \ldots, 40$. (Legend: solid = Student-T copula; long-dashed = Normal copula; dashed = Independent copula. Curves relative to Student-T and Normal copula vMEM are almost indistinguishable.)

Table 6: vMEM on Four Volatility Indicators: Impact matrices and corresponding eigenvalues.

T copula

	$\lvert r_{t-1}\rvert$	arv_{t-1}	brv_{t-1}	rv_{t-1}	eigenvalues
$\lvert r_t\rvert$	0.8432	0	0.0813	0	0.9419
arv_t	0.0328	0.8572	0.0548	0	0.8342
brv_t	0.0329	0.1914	0.7300	0	0.8150
rv_t	0.0342	0.0765	0	0.8342	0.6735

Normal copula

	$\lvert r_{t-1}\rvert$	arv_{t-1}	brv_{t-1}	rv_{t-1}	eigenvalues
$\lvert r_t\rvert$	0.8210	0	0.0982	0	0.9430
arv_t	0.0288	0.8636	0.0529	0	0.8389
brv_t	0.0289	0.1864	0.7409	0	0.7963
rv_t	0.0298	0.0775	0	0.8389	0.6863

Independent copula

	$\lvert r_{t-1}\rvert$	arv_{t-1}	brv_{t-1}	rv_{t-1}	eigenvalues
$\lvert r_t\rvert$	0.7574	0	0.1346	0	0.9106
arv_t	0.0372	0.8765	0.0000	0	0.7222
brv_t	0.0376	0.3439	0.5489	0	0.6767
rv_t	0.0399	0.1940	0	0.6767	0.5499

and illustrate the dynamic impact that such measures have on one another (especially the absolute returns with asymmetric effects). At least for the example at hand, there are no major differences between the two chosen copulas.

Departing from the need of distributional assumptions, we may provide alternative estimators: in Cipollini *et al.* (2006) we suggested the use of estimating equations (cf. Vinod (1998)) which provides results very similar to the system estimation used here.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71**, 579–625.

Barndorff-Nielsen, O. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics*, **2**, 1–48.

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 Opérationelle, Credit Lyonnais, Paris.

Cherubini, U. and Vecchiato, L. (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**(3), 561–582.

Cipollini, F., Engle, R. F., and Gallo, G. M. (2006). Vector multiplicative error models: Representation and inference. Working Paper Series 12690, National Bureau of Economic Research, Inc.

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

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**(1–2), 3–27.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127–62.

Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: Applicability and limitations. *Statistics & Probability Letters*, **63**, 275–286.

Hautsch, N. (2007). Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model. Working paper, Humboldt Universitaet zu Berlin, SFB 649 Discussion Paper 2007-052.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

Johnson, N., Kotz, S., and Balakrishnan, N. (2000). *Continuous Multivariate Distributions*. John Wiley & Sons, New York.

Kostadinov, K. (2005). Non-parametric estimation of elliptical copulae with application to credit risk. Working paper, Munich University of Technology.

Lindskog, F., McNeil, A. J., and Schmock, U. (2003). Kendall's tau for elliptical distributions. In Bol, Nakhaeizadeh, Rachev, Ridder, and 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.

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. Andersen, R. 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.-K. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, **27**, 305–320.

Vinod, H. D. (1998). Using Godambe-Durbin estimating functions in econometrics. In I. W. Basawa, V. P. Godambe, and R. L. Taylor, editors, *Selected Proceedings of the Symposium on Estimating Functions*, volume 32 of *IMS Lecture Notes Series*. Institute of Mathematical Statistics.

A Some Details on the VMEM Properties

A.1 Stationarity conditions

Sufficient conditions for stationarity of $\boldsymbol{\mu}_t$ are a simple generalization of those of the univariate case: a vector MEM with $\boldsymbol{\mu}_t$ defined as in equation (7) 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

$$\boldsymbol{\mu}_t = \mathbf{A}\boldsymbol{\mu}_{t-1}. \tag{21}$$

If more lags are considered, a similar result can be obtained replacing $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ with the sum of the corresponding matrices $\boldsymbol{\alpha}_l$, $\boldsymbol{\beta}_l$ and $\boldsymbol{\gamma}_l$ for all lags considered.

A.2 Non-negativity conditions

Sufficient conditions for non–negativity of the components of $\boldsymbol{\mu}_t$ are again a generalization of the corresponding conditions of the univariate model: the vector MEM with $\boldsymbol{\mu}_t$ defined as in equation (7) gives $\boldsymbol{\mu}_t \geq 0$ for all t if $\omega_i \geq 0$, $\beta_{ij} \geq 0$, $\alpha_{ij} \geq 0$, $\alpha_{ij} + \gamma_{2ij} \geq 0$ for all $i, j = 1, \ldots, K$.

A more involved results can be obtained when a different type of asymmetric effect is considered. We provide here some details.

Proposition 1 *The relation*

$$\sum_{i=1}^{n} \left(a_i x_i^2 + b_i x_i\right) + c \geq 0 \tag{22}$$

is satisfied for all $x_i \geq 0$ $(i = 1, \ldots, n)$ *if and only if the coefficients* a_i, b_i *and* c *satisfy all the following conditions:*

1. $a_i \geq 0$ *for all* $i \in S_n$*;*
2. $b_i \geq 0$ *for all* $i \in S_n$ *such that* $a_i = 0$*;*
3. $c - \dfrac{1}{4}\displaystyle\sum_{i=1}^{n} \frac{b_i^2}{a_i}\,\mathrm{I}(b_i < 0)\,\mathrm{I}(a_i > 0) \geq 0,$

where $S_n = \{1, \ldots, n\}$.

Proof:

The minimum of

$$\sum_{i=1}^{n}(a_i x_i^2 + b_i x_i) + c \tag{23}$$

with respect to the x_i's is simply the sum of c and the minima of the additive quantities $(a_i x_i^2 + b_i x_i)$. Hence, considering assumption $x_i \geq 0$:

1. if $a_i < 0$, the minimum of $(a_i x_i^2 + b_i x_i)$ is always $-\infty$;
2. if $a_i = 0$, the minimum of $(a_i x_i^2 + b_i x_i) = b_i x_i$ is nonnegative (0) only if $b_i \geq 0$;
3. if $a_i > 0$, the minimum of $(a_i x_i^2 + b_i x_i)$ for $x_i \geq 0$ is reached for $x_i = -\frac{b_i}{2a_i} I(b_i < 0)$ and is given by $-\frac{b_i^2}{4a_i} I(b_i < 0)$.

Requiring that (23) has a nonnegative minimum, from the above conditions those in the proposition follow immediately.

□

As a corollary of the previous proposition, we prove sufficient conditions for nonnegativity of the components of $\boldsymbol{\mu}_t$ in a more general formulation.

Corollary 1 *Let*

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \sum_{l=1}^{L}\left[\boldsymbol{\beta}_l \boldsymbol{\mu}_{t-l} + \boldsymbol{\alpha}_l \mathbf{x}_{t-l} + \boldsymbol{\gamma}_l \mathbf{x}_{t-l}^{(-)} + \boldsymbol{\delta}_l \mathbf{x}_{t-l}^{(s)}\right], \tag{24}$$

be the equation that describes the evolution of $\boldsymbol{\mu}_t$ in the vector MEM, where:

- *$\mathbf{x}_t$, $\boldsymbol{\mu}_t$ ($t = 1, \ldots, T$) and $\boldsymbol{\omega}$ are $(K, 1)$-vectors;*
- *$\boldsymbol{\beta}_l$, $\boldsymbol{\alpha}_l$, $\boldsymbol{\gamma}_l$ and $\boldsymbol{\delta}_l$ ($l = 1, \ldots, L$) are (K, K) matrices;*
- *in the 'volatility indicators' formulation: $x_{i,t}^{(-)} = x_{i,t}\,\mathrm{I}(r_t < 0)$, $x_{i,t}^{(s)} = x_{i,t}^{1/2}\,\mathrm{sign}(r_t)$;*
- *in the 'splillover' formulation: $x_{i,t}^{(-)} = x_{i,t}\,\mathrm{I}(r_{i,t} < 0)$, $x_{i,t}^{(s)} = x_{i,t}^{1/2}\,\mathrm{sign}(r_{i,t})$.*

We assume that $\mathbf{x}_t$ and $\boldsymbol{\mu}_t$, for $t = 1, \ldots, L$, have non-negative components.

Then $\boldsymbol{\mu}_t$ has non-negative components if all the following conditions are satisfied ($i, j = 1, \ldots, K$, $l = 1, \ldots, L$):

1. *$\beta_{ijl} \geq 0$, $\alpha_{ijl} \geq 0$, $\alpha_{ijl} + \gamma_{ijl} \geq 0$;*

2. *if* $\alpha_{ijl} = 0$ *then* $\delta_{ijl} \geq 0$*; if* $\alpha_{ijl} + \gamma_{ijl} = 0$ *then* $\delta_{ijl} \leq 0$*;*

3. $\omega_i - \frac{1}{4} \sum_{j=1}^{K} \sum_{l=1}^{L} \delta_{ijl}^2 \left[\frac{\mathrm{I}(\delta_{ijl} < 0)\,\mathrm{I}(\alpha_{ijl} > 0)}{\alpha_{ijl}} + \frac{\mathrm{I}(\delta_{ijl} > 0)\,\mathrm{I}(\alpha_{ijl} + \gamma_{ijl} > 0)}{\alpha_{ijl} + \gamma_{ijl}} \right] \geq 0.$

Proof:

The proof is by induction on t. Assume that $\mathbf{x}_{t-l}, \boldsymbol{\mu}_{t-l} \geq 0$ for $l = 1, \ldots, L$. Then the i-th subequation of (24) can be rewritten taking:

$$
\begin{aligned}
(x_1; \ldots; x_n) &= (\boldsymbol{\mu}_{t-1}; \ldots; \boldsymbol{\mu}_{t-L}; \mathbf{x}_{t-1}; \ldots; \mathbf{x}_{t-L})^{1/2} \\
(a_1, \ldots, a_n) &= (\boldsymbol{\beta}_{i.1}, \ldots, \boldsymbol{\beta}_{i.l}, \boldsymbol{\alpha}_{i.1} + \boldsymbol{\gamma}_{i.1}\,\mathrm{I}(r_{t-1} < 0), \ldots, \boldsymbol{\alpha}_{i.L} + \boldsymbol{\gamma}_{i.L}\,\mathrm{I}(r_{t-L} < 0)) \\
(b_1, \ldots, b_n) &= (\mathbf{0}', \ldots, \mathbf{0}', \boldsymbol{\delta}_{i.1}\,\mathrm{sign}(r_{t-1}), \ldots, \boldsymbol{\delta}_{i.L}\,\mathrm{sign}(r_{t-L})) \\
c &= \omega_i,
\end{aligned}
$$

where $\boldsymbol{\beta}_{i.l}$, $\boldsymbol{\alpha}_{i.l}$, $\gamma_{i.l}$, and $\boldsymbol{\delta}_{i.l}$ denote the i-th row of the corresponding matrix of coefficients at lag l.

At this point we can apply proposition 1 to the formulation obtained. Rewriting conditions 1, 2 and 3 of the proposition in the model notation and considering that, in the whole time series, the returns r_{t-l} take both positive and negative signs, the result follows.

□

B Computational details

We provide details on computing the score function depending on the choice of the copula used in specifying the conditional distribution of $\boldsymbol{\varepsilon}_t$. We consider full ML estimation of all parameters in these two cases: Elliptical copula (in order to handle together the Normal and the Student-T copulas); Normal copula with concentration of the $\mathbf{R}$ parameter. The general case in (13) and the situations in which the parameters of the copula can be estimated from moment conditions can be trivially obtained from the cases presented.

B.1 The score function with Elliptical copulas

We assume that the loglikelihood of the model is given in (18). We consider separately the different set of parameters and we will use the following symbols: $\mathbf{C} = \mathbf{cD}$, $\widetilde{\mathbf{q}}_t = \mathbf{C}'^{-1}\mathbf{q}_t$, $\mathbf{q}_t^* = \mathbf{R}^{-1}\mathbf{q}_t$, $\widetilde{\widetilde{q}}_t = \mathbf{q}_t'\mathbf{R}^{-1}\mathbf{q}_t = \widetilde{\mathbf{q}}_t'\widetilde{\mathbf{q}}_t$.

Parameters into c

The portion of the score relative to the free parameters of the $\mathbf{c}$ matrix (those above the main diagonal) has elements

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

Using some algebra we can show that

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

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

Parameters into $\boldsymbol{\nu}$

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

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

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

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

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 into $\boldsymbol{\theta}$

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

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

After some algebra we obtain that

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

where $\mathbf{a}_t$ has components

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

with

$$b_{t,i} = 2\frac{q_{t,i}\nabla_{q^2_{t,i}} \ln g_2(q^2_{t,i};\boldsymbol{\nu}) - q^*_{t,i}\nabla_{\widetilde{\widetilde{q}}_t} \ln g_1(\widetilde{\widetilde{q}}_t;\boldsymbol{\nu},K)}{g(q_{t,i};\boldsymbol{\nu})} \tag{25}$$

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

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

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

where $dgamma(x;\alpha,\beta)$ denotes the pdf of the $Gamma(\alpha,\beta)$ distribution computed at x.

Parameters into $\boldsymbol{\phi}$

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

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

After some algebra we obtain

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

where $b_{t,i}$ are given in (25). For instance, if a marginal has a distribution $Gamma(\phi_i,\phi_i)$ then

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

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

B.2 The score function with Normal copula and concentrated log-likelihood

Using some tedious algebra, we can show that the components of the score $\nabla_{\boldsymbol{\theta}} lc$ and $\nabla_{\boldsymbol{\phi}} lc$ have exactly the same structure as before, with the quantity $b_{t,i}$ into (25) changed to

$$b_{t,i} = -\frac{(\mathbf{C}\mathbf{q}_t)_i}{\phi(q_{t,i})} \tag{26}$$

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

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

B.3 Details about the estimation of θ

If one assumes that $\boldsymbol{\mu}_t$ evolves following equation (7), its dynamical behavior depends in general from $K + 2K^2$ parameters, which reduce to $2K + K^2$ if we assume a diagonal $\boldsymbol{\beta}$. For instance, when $K = 3$ there are 21 parameters in the general case and 15 if $\boldsymbol{\beta}$ is diagonal.

A further reduction in the number of free parameters can be obtained estimating $\boldsymbol{\omega}$ from stationary conditions (*variance targeting*). Imposing that $\boldsymbol{\mu}_t$ is stationary we have

$$\boldsymbol{\omega} = [\mathbf{I} - (\boldsymbol{\alpha} + \boldsymbol{\beta} + \gamma/2)]\boldsymbol{\mu},$$

where $\boldsymbol{\mu} = E(\mathbf{x}_t)$, and then

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

Replacing $\boldsymbol{\mu}$ with its natural estimate, that is the unconditional average $\overline{\mathbf{x}}$, we obtain

$$\begin{aligned}\widetilde{\boldsymbol{\mu}}_t &= \boldsymbol{\alpha}\widetilde{\mathbf{x}}_{t-1} + \boldsymbol{\gamma}\widetilde{\mathbf{x}}_{t-1}^{(-)} + \boldsymbol{\beta}\widetilde{\boldsymbol{\mu}}_{t-1} \\ &= \boldsymbol{\alpha}^*\widetilde{\mathbf{x}}_{t-1}^* + \boldsymbol{\beta}\widetilde{\boldsymbol{\mu}}_{t-1}\end{aligned} \tag{27}$$

where the symbol $\widetilde{\mathbf{x}}$ means the demeaned version of $\mathbf{x}$, $\widetilde{\mathbf{x}}_t^* = (\widetilde{\mathbf{x}}_t; \widetilde{\mathbf{x}}_t^{(-)})$ and $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}; \boldsymbol{\gamma})$. This strategy saves K parameters in the iterative estimation, provides very good performances in comparison with direct ML estimates of $\boldsymbol{\omega}$ and improves numerical stability of the algorithms (the last two consideration are based on simulations not reported here).

To save time, it is also useful take into account analytic derivatives of $\widetilde{\boldsymbol{\mu}}_t$ with respect to the parameters. They can be expressed as

$$\frac{\partial \widetilde{\boldsymbol{\mu}}_t'}{\partial \boldsymbol{\theta}} = \left[\frac{\partial \widetilde{\boldsymbol{\mu}}_{t-1,(1)}'}{\partial \boldsymbol{\theta}}\boldsymbol{\beta}_{(1)\cdot}', \ldots, \frac{\partial \widetilde{\boldsymbol{\mu}}_{t-1,(K)}'}{\partial \boldsymbol{\theta}}\boldsymbol{\beta}_{(K)\cdot}'\right] + \left(\begin{matrix}\begin{pmatrix}\widetilde{\boldsymbol{\mu}}_{t-1,(1)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \widetilde{\boldsymbol{\mu}}_{t-1,(K)}\end{pmatrix} \\ \begin{pmatrix}\widetilde{\mathbf{x}}_{t-1,(1)}^* & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \widetilde{\mathbf{x}}_{t-1,(K)}^*\end{pmatrix}\end{matrix}\right)$$

where: $\boldsymbol{\theta} = (\mathrm{vec}(\boldsymbol{\beta})', \mathrm{vec}(\boldsymbol{\alpha}^{*\prime}))$ (the operator $\mathrm{vec}(\cdot)$ stacks the columns of the matrix inside brackets); $\boldsymbol{\beta}_{i\cdot}$ denotes the i-th row of $\boldsymbol{\beta}$; the notation $\boldsymbol{v}_{(i)}$ applied to a vector of parameters means that only those elements that are unconstrained (to zero) are selected; the notation $\boldsymbol{v}_{(i)}$ applied to $\widetilde{\boldsymbol{\mu}}_t$ or $\widetilde{\mathbf{x}}_t^*$ means that only elements corresponding to unconstrained parameters are selected.