Vector Multiplicative Error Models: Representation and Inference*

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

This version: Feb. 2006 – Preliminary and Incomplete

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 and of estimating equations to jointly estimate the parameters of the scale factors and of the correlations of the innovation processes. Empirical applications on volatility indicators are used to illustrate the gains over the equation by equation procedure.

*We thank Christian T. Brownlees, Marco J. Lombardi and Margherita Velucchi for many discussions on MEMs and multivariate extensions. The usual disclaimer applies.

[†]Department of Statistics "G. Parenti", University of 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: Department of Statistics "G. Parenti", University of Firenze, Viale G.B. Morgagni, 59 - 50134 Firenze – Italy. e-mail: gallog@ds.unifi.it

Contents

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 process describing 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. Empirical results (e.g. Chou (2005); Manganelli (2002)) show a good performance of these types of models in capturing the stylized facts of the observed series.

More recently, Engle and Gallo (2006) have investigated three different indicators of volatility, namely absolute returns, daily range and realized volatility in a multivariate context in which each lagged indicator was allowed to enter the equation of the scale factor of the other indicators. Model selection techniques were adopted to ascertain the statistical relevance of such variables in explaining the dynamic behavior of each indicator. The model was estimated assuming a diagonal variance covariance matrix of the innovation terms. Further applications along the same lines may be found in Gallo and Velucchi (2005) for different measures of volatility based on 5-minute returns, Brownlees and Gallo (2005) for hourly returns, volumes and number of trades, Engle *et al.* (2005) for volatility transmission across financial markets.

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 two different strategies in order to manage vector-MEM: the first is to adopt copula functions to link together marginal probability density functions specified as Gamma as in Engle and Gallo (2006); the second is to adopt an Estimating Equation approach (Heyde (1997), Bibby *et al.* (2004)). 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: copula and estimating equations results are fairly similar to one another while the equation–by–equation procedure provides estimates which depart from the system–based ones the more correlated the variables inserted in the analysis.

2 The Univariate MEM Reconsidered

Let us start by recalling the main features of the univariate Multiplicative Error Model (MEM) introduced by Engle (2002) and extended by Engle and Gallo (2006). For ease of reference, some characteristics and properties of the statistical distributions involved are detailed in appendix A.

2.1 Definition and formulations

Let us consider x_t a non–negative univariate process and let $\mathcal{F}_{t-1}$ be the information about the process up to time $t-1$. Then the MEM for x_t is specified as

$$x_t = \mu_t \varepsilon_t, \tag{1}$$

where, conditionally on the information $\mathcal{F}_{t-1}$: μ_t is a nonnegative *conditionally deterministic* (or *predictable*) process, that is the evolution of which depends on a vector of unknown parameters $\boldsymbol{\theta}$,

$$\mu_t = \mu_t(\boldsymbol{\theta}); \tag{2}$$

ε_t is a *conditionally stochastic* i.i.d. process, with density having non–negative support, mean 1 and unknown variance σ^2,

$$\varepsilon_t|\mathcal{F}_{t-1} \sim D(1, \sigma^2). \tag{3}$$

The previous conditions on μ_t and ε_t guarantee that

$$E(x_t|\mathcal{F}_{t-1}) = \mu_t \tag{4}$$
$$V(x_t|\mathcal{F}_{t-1}) = \sigma^2 \mu_t^2. \tag{5}$$

To close the model, we need to adopt a parametric density function for ε_t and to specify an equation for μ_t.

For the former step, we follow Engle and Gallo (2006) in adopting a Gamma distribution which has the usual exponential density function (as in the original Engle and Russell (1998) ACD model) as a special case. We have

$$\varepsilon_t|\mathcal{F}_{t-1} \sim Gamma(\phi, \phi), \tag{6}$$

with $E(\varepsilon_t|\mathcal{F}_{t-1}) = 1$ and $V(\varepsilon_t|\mathcal{F}_{t-1}) = 1/\phi$. Taken jointly, assumptions (1) and (6) can be written compactly as (see appendix A)

$$x_t|\mathcal{F}_{t-1} \sim Gamma(\phi, \phi/\mu_t). \tag{7}$$

Note that a well-known relationship (cf. Appendix A) between the Gamma distribution and the Generalized Error Distribution (GED) suggests an equivalent formulation which may prove useful for estimation purposes. We have:

$$x_t|\mathcal{F}_{t-1} \sim Gamma(\phi, \phi/\mu_t) \Leftrightarrow x_t^{\phi}|\mathcal{F}_{t-1} \sim Half-GED(0, \mu_t^{\phi}, \phi). \tag{8}$$

Thus the conditional density of x_t has a correspondence in a conditional density of x_t^{ϕ}, that is

$$x_t^{\phi} = \mu_t^{\phi}\nu_t \tag{9}$$

where

$$\nu_t|\mathcal{F}_{t-1} \sim Half-GED(0, 1, \phi). \tag{10}$$

This result somewhat generalizes the practice, suggested by Engle and Russell (1998), to estimate an ACD model by estimating the parameters of the second moment of the square root of the durations (imposing mean zero) with a GARCH routine assuming normality of the errors. The result extends to any estimation carried out on observations x_t^{ϕ} (with ϕ known)[1] assuming a GED distribution with mean zero and dispersion parameter μ_t^{ϕ}.

As per the specification of μ_t, following, again, Engle (2002) and Engle and Gallo (2006), let us consider the simplest GARCH–type of order (1,1) formulation, and, for the time being, let us not include predetermined variables in the specification.

The *base* $(1, 1)$ specification of μ_t is

$$\mu_t = \omega + \alpha x_{t-1} + \beta\mu_{t-1}, \tag{11}$$

which is appropriate when, say, x_t is a series of positive valued variables such as financial durations or traded volumes. An extended specification is appropriate when one can refer to the sign of financial returns as relevant extra information: well-known empirical evidence suggests a symmetric distribution for the returns but an asymmetric response of conditional variance to negative innovations (so–called leverage effect). Thus, when considering volatility related variables (absolute or squared returns, realized volatility, range, etc.), asymmetric effects can be inserted in the form:

$$\mu_t = \omega + \alpha(x_{t-1}^{1/2}\,\mathrm{sign}(r_{t-1}) + \gamma_1)^2 + \gamma_2 x_{t-1}\,\mathrm{I}(r_{t-1} < 0) + \beta\mu_{t-1}, \tag{12}$$

where γ_1 and γ_2 are parameters that capture the asymmetry. Following Engle and Gallo (2006), we have inserted in the model two variants of existing specifications for including asymmetric effects in the GARCH literature: an APARCH-like asymmetry effect (the squared term with γ_1 – see Ding *et al.* (1993)) and a GJR-like asymmetry effect (the last term with γ_2 – see Glosten *et al.* (1993)). Computing the square, expression (12) can be rewritten as

$$\mu_t = \omega^* + \alpha x_{t-1} + \gamma_1^* x_{t-1}^{(s)} + \gamma_2 x_{t-1}^{(-)} + \beta\mu_{t-1}, \tag{13}$$

[1] In raising the observations x_t to the ϕ makes the ensuing model more similar to the Ding *et al.* (1993) APARCH specification except that the exponent parameter appears also in the error distribution.

where $x_t^{(s)} = x_t^{1/2}\,\mathrm{sign}(r_t)$, $x_t^{(-)} = x_t\,\mathrm{I}(r_t < 0)$, $\omega^* = \omega + \alpha\gamma_1^2$ and $\gamma_1^* = 2\alpha\gamma_1$.

Both specifications can be written compactly as

$$\mu_t = \omega^* + \boldsymbol{x}_{t-1}^{*\prime}\boldsymbol{\alpha}^* + \beta\mu_{t-1}. \tag{14}$$

where, if we assume $\boldsymbol{x}_t^* = (x_t, x_t^{(s)}, x_t^{(-)})$ and $\boldsymbol{\alpha}^* = (\alpha, \gamma_1^*, \gamma_2)$ we get the asymmetric specification (13). The base specification can be retrieved more simply taking $\omega^* = \omega$, $\boldsymbol{x}_t^* = x_t$ and $\boldsymbol{\alpha}^* = \alpha$ (cf. 11).

The parameter space for $\boldsymbol{\theta} = (\omega^*, \boldsymbol{\alpha}^{*\prime}, \beta)$ must be restricted in order to ensure that $\mu_t \geq 0$ for all t and to ensure stationary distributions for x_t. However restrictions depend on the formulation taken into account: we consider here formulation (13). Sufficient conditions for stationarity can been obtained taking the unconditional expectation of both members of (14) and solving for $E(x_t) = \mu$. This implies x_t stationary if

$$\alpha + \beta + \gamma_2/2 < 1.$$

Sufficient conditions for non–negativity of μ_t can be obtained taking $\beta \geq 0$ and imposing $\omega^* + \alpha x_t + \gamma_1^* x_t^{1/2}\,\mathrm{sign}(r_t) + \gamma_2 x_t\,\mathrm{I}(r_t < 0)$ for all x_t's and r_t's. One can verify that these conditions are satisfied when

$$\alpha \geq 0, \quad \beta \geq 0, \quad 4\alpha\omega^* - \gamma_1^{*2} \geq 0, \quad \alpha + \gamma_2 \geq 0, \quad 4(\alpha + \gamma_2)\omega^* - \gamma_1^{*2} \geq 0.$$

2.2 Estimation and Inference

Let us introduce estimation and inference issues by discussing first the role of a generic observation x_t. From (7), the contribution of x_t to the log–likelihood function l_t is

$$l_t = \ln L_t = \phi\ln\phi - \ln\Gamma(\phi) + (\phi - 1)\ln x_t - \phi(\ln\mu_t + x_t/\mu_t).$$

The contribution of x_t to the score is $\mathbf{s}_t = \begin{pmatrix}\mathbf{s}_{t,\boldsymbol{\theta}}\\ s_{t,\phi}\end{pmatrix}$ with components

$$\mathbf{s}_{t,\boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} l_t = \phi\left(\frac{x_t - \mu_t}{\mu_t^2}\right)\nabla_{\boldsymbol{\theta}}\mu_t$$
$$s_{t,\phi} = \nabla_{\phi} l_t = \ln\phi + 1 - \psi(\phi) + \ln\left(\frac{x_t}{\mu_t}\right) - \frac{x_t}{\mu_t},$$

where $\psi(\phi) = \dfrac{\Gamma'(\phi)}{\Gamma(\phi)}$ is the *digamma* function and the operator $\nabla_{\boldsymbol{\lambda}}$ denotes the derivatives with respect to (the components of) $\boldsymbol{\lambda}$.

The contribution of x_t to the Hessian is $\mathbf{H}_t = \begin{pmatrix} \mathbf{H}_{t,\boldsymbol{\theta}\boldsymbol{\theta}'} & \mathbf{H}_{t,\boldsymbol{\theta}\phi} \\ \mathbf{H}'_{t,\boldsymbol{\theta}\phi} & \mathbf{H}_{t,\phi\phi} \end{pmatrix}$ with components

$$\begin{aligned}
\mathbf{H}_{t,\boldsymbol{\theta}\boldsymbol{\theta}'} &= \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} l_t = \phi \left(\frac{-2x_t + \mu_t}{\mu_t^3} \nabla_{\boldsymbol{\theta}} \mu_t \nabla_{\boldsymbol{\theta}'} \mu_t + \frac{x_t - \mu_t}{\mu_t^2} \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \mu_t \right) \\
\mathbf{H}_{t,\boldsymbol{\theta}\phi} &= \nabla_{\boldsymbol{\theta}\phi} l_t = \frac{x_t - \mu_t}{\mu_t^2} \nabla_{\boldsymbol{\theta}} \mu_t \\
H_{t,\phi\phi} &= \nabla_{\phi\phi} l_t = \frac{1}{\phi} - \psi'(\phi),
\end{aligned}$$

where $\psi'(\phi)$ is the *trigamma* function.

Exploiting the previous results we obtain the following first order conditions for $\boldsymbol{\theta}$ and ϕ:

$$\frac{1}{T} \sum_{t=1}^{T} \frac{x_t - \mu_t}{\mu_t^2} \nabla_{\boldsymbol{\theta}} \mu_t = 0 \tag{16}$$

$$\ln \phi + 1 - \psi(\phi) + \frac{1}{T} \sum_{t=1}^{T} \left[\ln \left(\frac{x_t}{\mu_t} \right) - \frac{x_t}{\mu_t} \right] = 0 \tag{17}$$

As noted by Engle and Gallo (2006), first–order conditions for $\boldsymbol{\theta}$ do not depend on ϕ. As a consequence, whatever value ϕ may take, any Gamma-based MEM or any equivalent GED-based power formulation will provide the same point estimates for θ. The ML estimation of ϕ can then be performed after $\boldsymbol{\theta}$ has been estimated.

Furthermore, so long as $\mu_t = E(x_t|\mathcal{F}_{t-1})$, the expected value of the score of θ evaluated at the true parameters is a vector of zeroes even if the density of $\varepsilon_t|\mathcal{F}_{t-1}$ does not belong to the $Gamma(\phi, \phi)$ family.

Altogether, these considerations strengthen the claim (e.g. Engle (2002) for the case $\phi = 1$) that, whatever the value of ϕ, the log–likelihood functions of any one of these formulations (MEM or equivalent power formulation) can be interpreted as Quasi Likelihood functions and the corresponding estimator $\widehat{\boldsymbol{\theta}}$ is a QML estimator.

After some computations we obtain the asymptotic variance–covariance matrix of the ML estimator

$$V_\infty = \begin{pmatrix} \phi \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\mu_t^2} \nabla_{\boldsymbol{\theta}} \mu_t \nabla_{\boldsymbol{\theta}'} \mu_t & \mathbf{0} \\ \mathbf{0} & \psi'(\phi) - \frac{1}{\phi} \end{pmatrix}^{-1}.$$

Note that even if ϕ is not involved in the estimate of $\boldsymbol{\theta}$ the variance of $\widehat{\boldsymbol{\theta}}$ is proportional to $1/\phi$. We note also that the ML estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\phi}$ are asymptotically uncorrelated.

Alternative estimators of ϕ are available, which exploit the orthogonal nature of the parameters $\boldsymbol{\theta}$ and ϕ. As an example, by defining $v_t = x_t/\mu_t - 1$, we note that[2] $V(v_t|\mathcal{F}_{t-1}) = \phi^{-1}$. Therefore, a simple method of moment estimator is

$$\widehat{\phi^{-1}} = \frac{1}{T}\sum_{t=1}^{T} \hat{v}_t^2. \tag{18}$$

In view of (17), this expression has the advantage of not being affected by the presence of x_t's equal to zero (cf. also the comments below).

Obviously, the inference about $(\boldsymbol{\theta}, \phi)$ must be based on estimates of V_∞, that can be obtained evaluating the average Hessian or the average outer product of the gradients evaluated at the estimates $(\widehat{\boldsymbol{\theta}}, \widehat{\phi})$. The *sandwich* estimator

$$\widehat{V}_\infty = \widehat{\overline{\mathbf{H}}}_T^{-1} \widehat{\overline{\mathbf{OPG}}}_T \widehat{\overline{\mathbf{H}}}_T^{-1},$$

where $\widehat{\overline{\mathbf{H}}}_T = \frac{1}{T}\sum_{t=1}^{T} \widehat{\mathbf{H}}_t$ and $\widehat{\overline{\mathbf{OPG}}}_T = \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{s}}_t \hat{\mathbf{s}}_t'$, get rid of the dependence of the submatrix relative to $\boldsymbol{\theta}$ on ϕ altogether.

2.3 Some Further Comments

2.3.1 MEM as member of a more general family of models

The MEM defined in section 2.1 belongs to the Generalized Linear Autoregressive Moving Average (GLARMA) family of models introduced by Shephard (1995) and extended by Benjamin *et al.* (2003)[3]. In the GLARMA model, the conditional density of x_t is assumed to belong to the same exponential family, with density given by[4]

$$f(x_t|\mathcal{F}_{t-1}) = \exp\left[\phi\left(x_t\vartheta_t - b(\vartheta_t)\right) + d(x_t, \phi)\right]. \tag{19}$$

ϑ_t and ϕ are the canonical and precision parameters respectively, whereas $b(\cdot)$ and $d(\cdot)$ are specific functions that define the particular distribution into the family. The main moments are

$$E(x_t|\mathcal{F}_{t-1}) = b'(\vartheta_t) = \mu_t$$
$$V(x_t|\mathcal{F}_{t-1}) = b''(\vartheta_t)/\phi = v(\mu_t)/\phi.$$

Since μ_t may have a bounded domain, it is useful to define its (p, q) dynamics through a twice differentiable and monotonic *link function* $g(\cdot)$, introduced in general terms

[2] We thank Neil Shephard for pointing this out to us.

[3] For other works related to GLARMA models see, among others, Li (1994), Davis *et al.* (2002).

[4] We adapt the formulation of Benjamin *et al.* (2003) to this work.

by Benjamin *et al.* (2003) as

$$g(\mu_t) = z_t'\gamma + \sum_{i=1}^{p} \alpha_i \mathcal{A}(x_{t-i}, z_{t-i}, \gamma) + \sum_{j=1}^{q} \beta_j \mathcal{M}(x_{t-j}, \mu_{t-j}),$$

where $\mathcal{A}(\cdot)$ and $\mathcal{M}(\cdot)$ are functions representing the autoregressive and moving average terms. Such a formulation, admittedly too general for practical purposes, can be replaced by a more *practical version*:

$$g(\mu_t) = z_t'\gamma + \sum_{i=1}^{p} \alpha_i \left[g(x_{t-i}) - z_{t-i}'\gamma\right] + \sum_{j=1}^{q} \beta_j \left[g(x_{t-j}) - g(\mu_{t-j})\right]. \qquad (20)$$

In this formulation, for certain functions g it may be necessary to replace some x_{t-i}'s by x_{t-i}^* to avoid the nonexistence of $g(x_{t-i})$ for certain values of the argument (e.g. zeros, see Benjamin *et al.* (2003) for some examples).

Specifying (19) as a $Gamma(\phi, \phi/\mu_t)$ density (in this case $v(\mu_t) = \mu_t^2$) and $g(\cdot)$ as the identity function, after some simple arrangements of (20) it is easily verified that the MEM is a particular GLARMA model.

2.3.2 Exact zero values in the data

In introducing the MEM for modelling non–negative time series, Engle (2002) states that the structure of the model avoids problems caused by exact zeros in $\{x_t\}$, problems that, instead, are typical of $\log(x_t)$ formulations. This needs to be further clarified here.

Considering the MEM as defined by formula (1), in principle an $x_t = 0$ can be caused by $\varepsilon_t = 0$, by $\mu_t = 0$ or by both. We discuss these possibilities in turn.

- As the process μ_t is supposed deterministic conditionally to the information $\mathcal{F}_{t-1}$ (section 2.1), the value of μ_t cannot be 'altered' from any of the observations after time $t-1$: in practice it is not possible to take $\mu_t = 0$ if $x_t = 0$. Hence: or μ_t is really 0 (very unlikely!) or an $x_t = 0$ cannot be a consequence of $\mu_t = 0$. As a second observation, a possible $\mu_t = 0$ causes serious problems to the inference, as evidenced by the first–order conditions (16).
- The only possibility is then $\varepsilon_t = 0$. However, as we assumed $\varepsilon_t|\mathcal{F}_{t-1} \sim Gamma(\phi, \phi)$, this distribution cannot be defined for 0 values when $\phi < 1$. Furthermore, even restricting the parameter space to $\phi \geq 1$, the first order condition for ϕ in (17) requires the computation of $\ln x_t$, not defined for exact zero values. In practice the ML estimation of ϕ is not feasible in presence of zeros which enforces the usefulness of the method of moments estimator introduced above.

The formulation of the MEM considered by Engle (2002) avoid problems caused by exact zeros because it assume $\varepsilon_t|\mathcal{F}_{t-1} \sim Exponential(1)$, the density of which, $f(\varepsilon_t|\mathcal{F}_{t-1}) = e^{-\varepsilon_t}$, can be defined for $\varepsilon_t \geq 0$. Obviously, any MEM with a *fixed* $\phi \geq 1$ shares this characteristic.

3 The vector MEM

As introduced in Engle (2002), a vector MEM is a simple generalization of the univariate MEM. It is a suitable representation of non–negative valued processes which have dynamic interactions with one another. The application in Engle and Gallo (2006) envisages three indicators of volatility (absolute returns, daily range and realized volatility), while the study in Gallo and Velucchi (2005) takes daily absolute returns on seven East Asian markets to study contagion. In all cases, the model is estimated equation by equation.

3.1 Definition and formulations

Let $\mathbf{x}_t$ a K–dimensional process with non–negative components;[5] a vector MEM for $\mathbf{x}_t$ is defined as

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t = \mathrm{diag}(\boldsymbol{\mu}_t)\boldsymbol{\varepsilon}_t, \tag{21}$$

where $\odot$ indicates the Hadamard (element–by–element) product. Conditional on the information $\mathcal{F}_{t-1}$, $\boldsymbol{\mu}_t$ is defined as in (2) except that now we are dealing with a K– dimensional vector depending on a (larger) vector of parameters $\boldsymbol{\theta}$. The innovation vector $\boldsymbol{\varepsilon}_t$ is a *conditionally stochastic* K–dimensional i.i.d. process. Its density function is defined over a $[0, +\infty)^K$ support, with unit vector $\mathbb{1}$ as expectation and a general variance–covariance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim D(\mathbb{1}, \boldsymbol{\Sigma}). \tag{22}$$

The previous conditions guarantee that

$$E(\mathbf{x}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t \tag{23}$$
$$V(\mathbf{x}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t\boldsymbol{\mu}_t' \odot \boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\mu}_t)\boldsymbol{\Sigma}\,\mathrm{diag}(\boldsymbol{\mu}_t), \tag{24}$$

which is a positive definite matrix by construction, as emphasized by Engle (2002).

3.2 Specifications for $\boldsymbol{\varepsilon}_t$

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

[5] In what follows we will adopt the convention that 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.

extension to be considered is to limit ourselves to the assumption that $\boldsymbol{\varepsilon}_{i,t}|\mathcal{F}_{t-1} \sim Gamma(\phi_i, \phi_i)$, $i = 1, \ldots, K$.

3.2.1 Multivariate Gamma distributions

Johnson *et al.* (2000, chapter 48) describe a number of multivariate generalizations of the univariate Gamma distribution. However, many of them are bivariate versions, not sufficiently general for our purposes. Among the others, we do not consider here distributions defined via the joint characteristic function, as they require numerical inversion formulas to find their pdf's. Hence, the only useful versions remain the multivariate Gammas of Cheriyan and Ramabhadran (in their more general version, henceforth *GammaCR*, see appendix C.1 for the details), of Kowalckzyk and Trycha and of Mathai and Moschopoulos (Johnson *et al.* (2000, 454–470)). If one wants the domain of $\boldsymbol{\varepsilon}_t$ to be defined on $[0, +\infty)^K$, it can be shown that the three mentioned versions are perfectly equivalent.[6] As a consequence, we will consider the following multivariate Gamma assumption for the conditional distribution of the innovation term $\boldsymbol{\varepsilon}_t$

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

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$ and $0 < \phi_0 < \min(\phi_1, \ldots, \phi_K)$. As described in the appendix, all univariate marginal probability functions for $\varepsilon_{i,t}$ are, as required, $Gamma(\phi_i, \phi_i)$, even if the multivariate pdf is expressed in terms of a complicated integral. The conditional variance matrix of $\boldsymbol{\varepsilon}_t$ has elements

$$C(\varepsilon_{i,t}, \varepsilon_{j,t}|\mathcal{F}_{t-1}) = \frac{\phi_0}{\phi_i \phi_j} \tag{25}$$

so that the correlations are

$$\rho(\varepsilon_{i,t}, \varepsilon_{j,t}|\mathcal{F}_{t-1}) = \frac{\phi_0}{\sqrt{\phi_i \phi_j}}. \tag{26}$$

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 better alternatives.

3.2.2 The Normal copula-Gamma marginals distribution

A different way to define the distribution of $\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}$ is to start from the assumption that all univariate marginal probability density functions are $Gamma(\phi_i, \phi_i)$, $i = 1, \ldots, K$ and to use *copula functions*.[7] Adopting copulas, the definition of the

[6] The proof is tedious and is available upon request.

[7] The main characteristics of copulas are summarized in appendix B.

distribution of a multivariate r.v. is completed by defining the copula that represents the structure of dependence among the univariate marginal pdf's. Many copula functions have been proposed in theoretical and applied works (see, among others, Embrechts *et al.* (2001) and Bouyé *et al.* (2000)). In particular, the Normal copula possesses many interesting properties: the capability of capture a broad range of dependencies[8], the analytical tractability, the easy of simulation. For simplicity, we will assume

$$\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim N(\mathbf{R}) - \prod_{i=1}^{K} Gamma(\phi_i, \phi_i). \tag{27}$$

where in the distribution, the first part refers to the copula ($\mathbf{R}$ is a correlation matrix) and the second part to the marginals. As detailed in appendix C.2, this distribution is a special case of multivariate dispersion distributions generated from a Gaussian copula discussed in Song (2000). The conditional variance–covariance matrix of $\boldsymbol{\varepsilon}_t$ has a generic element which is approximately equal to

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

so that the correlations are, approximately,

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

The advantages of using copulas over a multivariate Gamma specification are apparent and suggest their adoption in this context: the covariance and correlation structures are more flexible (also negative correlations are permitted); the correlations do not depend on the variances of the marginals; there are no complicated constraints on the parameters; the pdf is more easily tractable. Furthermore, by adopting copula functions, the framework considered here can be easily extended to different choices of the distribution of the marginal pdf's (Weibull for instance).

3.3 Specification for $\boldsymbol{\mu}_t$

The base $(1,1)$ specification for $\boldsymbol{\mu}_t$ can be written as

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\alpha} x_{t-1} + \boldsymbol{\beta} \boldsymbol{\mu}_{t-1}, \tag{28}$$

where $\boldsymbol{\omega}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have dimensions, respectively, $(K,1)$, (K,K) and (K,K).

The enlarged $(1,1)$ multivariate specification will include lagged cross-effects and may include asymmetric effects, defined as before. For example, when the first component of $\mathbf{x}_t$ is $x_{1,t} = r_t^2$ and the conditional distribution of $r_t = x_{1,t}^{1/2}\,\mathrm{sign}(r_t)$ is

[8] The bivariate Normal copula, according to the value of the correlation parameter is capable of attaining the lower Frechèt bound, the product copula and the upper Frechèt bound.

symmetric with zero mean, generalizing formulation (13) we have

$$\boldsymbol{\mu}_t = \boldsymbol{\omega}^* + \boldsymbol{\alpha}\mathbf{x}_{t-1} + \boldsymbol{\gamma}_2\mathbf{x}_{t-1}^{(-)} + \boldsymbol{\gamma}_1^*\mathbf{x}_{t-1}^{(s)} + \boldsymbol{\beta}\boldsymbol{\mu}_{t-1}, \tag{29}$$

where $\mathbf{x}_t^{(-)} = \mathbf{x}_t \,\mathrm{I}(r_t < 0)$, $\mathbf{x}_t^{(s)} = \mathbf{x}_t^{1/2}\,\mathrm{sign}(r_t)$. Both parameters $\boldsymbol{\gamma}_1^*$ and $\boldsymbol{\gamma}_2$ have dimension (K, K), whence the others are as before.

Another specification can be taken into account when the purpose is to model contagion among volatilities of different markets (Gallo and Velucchi, 2005). For example, we consider the case in which each component of $\boldsymbol{\mu}_t$ is the conditional variance of the corresponding market index, so that $x_{i,t} = r_{i,t}^2$. If we assume that the conditional distribution of each market index $r_{i,t} = x_{i,t}^{1/2}\,\mathrm{sign}(r_{i,t})$ is symmetric with zero mean, the 'contagion' (1,1) formulation can be structured exactly as in (29) where $x_{i,t}^{(-)} = x_{i,t}\,\mathrm{I}(r_{it} < 0)$ and $x_{i,t}^{(s)} = x_{i,t}^{1/2}\,\mathrm{sign}(r_{i,t})$.

The base, the enlarged and the contagion $(1,1)$ specifications can be written compactly as

$$\boldsymbol{\mu}_t = \boldsymbol{\omega}^* + \boldsymbol{\alpha}^*\mathbf{x}_{t-1}^* + \boldsymbol{\beta}\boldsymbol{\mu}_{t-1}. \tag{30}$$

Taking $\boldsymbol{\omega}^* = \boldsymbol{\omega}$, $\mathbf{x}_t^* = \mathbf{x}_t$ and $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}$ we have (28); considering $\mathbf{x}_t^* = (\mathbf{x}_t', \mathbf{x}_t^{(-)\prime}, \mathbf{x}_t^{(s)\prime})'$ (of dimension $(3K, 1)$) and $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}', \boldsymbol{\gamma}_2', \boldsymbol{\gamma}_1^{*\prime})'$ (of dimension $(K, 3K)$), we obtain (29).

Considering formulation (30), we have $\boldsymbol{\theta} = (\boldsymbol{\omega}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta})'$. The parameter space of θ must be restricted to ensure $\boldsymbol{\mu}_t \geq \mathbf{0}$ for all t and to ensure stationary distributions for $\mathbf{x}_t$. To discuss these, we consider the enlarged and the contagion formulations above.

Sufficient conditions the stationarity of $\boldsymbol{\mu}_t$ are a simple generalization of those showed in section 2.1 for the univariate MEM and can be obtained in an analogous way: $\mathbf{x}_t$ is stationary if all characteristic roots of $\mathbf{A} = \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma}_2/2$ are smaller than 1 in modulus. We can think of $\mathbf{A}$ as the *impact matrix* in the expression

$$\boldsymbol{\mu}_t = \mathbf{A}\boldsymbol{\mu}_{t-1}. \tag{31}$$

Sufficient conditions for non–negativity of the components of $\boldsymbol{\mu}_t$ are again a generalization of the corresponding conditions of the univariate model, but the derivation is more involved. As proved in appendix E, the vector MEM defined in equation (30) gives $\boldsymbol{\mu}_t \geq 0$ for all t if all the following conditions are satisfied for all $i, j = 1, \ldots, K$:

- for the enlarged $(1,1)$ formulation:
    1. $\beta_{ij} \geq 0$, $\alpha_{ij} \geq 0$, $\alpha_{ij} + \gamma_{2ij} \geq 0$ for all j;
    2. if $\alpha_{ij} = 0$ then $\gamma_{1ij}^* \geq 0$; if $\alpha_{ij} + \gamma_{2ij} = 0$ then $\gamma_{1ij}^* \leq 0$;
    3. $\omega_i^* - \dfrac{1}{4}\displaystyle\sum_{j=1}^{K} \frac{\gamma_{1ij}^{*2}}{\alpha_{ij}}\,\mathrm{I}(\alpha_{ij} > 0)\,\mathrm{I}(\gamma_{1ij}^* < 0) \geq 0$ and

$$\omega_i^* - \frac{1}{4}\sum_{j=1}^{K} \frac{\gamma_{1ij}^{*2}}{\alpha_{ij}+\gamma_{2ij}} \mathrm{I}(\alpha_{ij}+\gamma_{2ij}>0)\,\mathrm{I}(\gamma_{1ij}^*>0) \geq 0.$$

- for the contagion $(1,1)$ formulation: $\beta_{ij} \geq 0$, $\alpha_{ij}+\gamma_{1ij}^* \geq 0$, $\alpha_{ij}+\gamma_{2ij}-\gamma_{1ij}^* \geq 0$.

3.4 Maximum likelihood inference

Given the assumptions about the conditional distribution of $\boldsymbol{\varepsilon}_t$ (section 3.2.2), we have that $\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}$ has a pdf

$$f(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = c(F_1(\varepsilon_{1,t}),\ldots,F_K(\varepsilon_{K,t}))\prod_{i=1}^{K} f_i(\varepsilon_{i,t})$$

where $f_i(.)$ and $F_i(.)$ indicate, respectively, the pdf and the cdf of the i-th component of $\boldsymbol{\varepsilon}_t$, $c(.)$ is the density function of the chosen copula. In the case considered here,

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

$$F_i(\varepsilon_{i,t}) = \Gamma(\phi_i;\phi_i\varepsilon_{i,t})$$

$$c(\mathbf{u}_t) = |\mathbf{R}|^{-1/2}\exp\left[-\frac{1}{2}\mathbf{q}_t'(\mathbf{R}^{-1}-I)\mathbf{q}_t\right]$$

where $\Gamma(\zeta;x)$ indicates the incomplete Gamma function with parameter ζ computed at x (or, in other words, the cdf of a *Gamma*$(\zeta,1)$ r.v. computed at x), $\mathbf{q}_t = (\Phi^{-1}(F_1(\varepsilon_{1,t})),\ldots,\Phi^{-1}(F_K(\varepsilon_{K,t})))$, $\Phi^{-1}(x)$ indicates the quantile function of the standard normal distribution computed at x.

Hence, $\mathbf{x}_t|\mathcal{F}_{t-1}$ has pdf

$$f(\mathbf{x}_t|\mathcal{F}_{t-1}) = c(F_1(x_{1,t}/\mu_{1,t}),\ldots,F_K(x_{K,t}/\mu_{K,t}))\prod_{i=1}^{K}\frac{f_i(x_{i,t}/\mu_{i,t})}{\mu_{i,t}},$$

where

$$\frac{f_i(x_{i,t}/\mu_{i,t})}{\mu_{i,t}}$$

is the pdf of a *Gamma*$(\phi_i,\phi_i/\mu_{i,t})$.

3.4.1 The log–likelihood

The log–likelihood of the model is then

$$l = \sum_{t=1}^{T} l_t = \sum_{t=1}^{T} \ln f(\mathbf{x}_t|\mathcal{F}_{t-1}),$$

where

$$\begin{aligned}
l_t &= \ln c(F_1(x_{1,t}/\mu_{1,t}), \ldots, F_K(x_{K,t}/\mu_{K,t})) + \sum_{i=1}^{K} (\ln f_i(x_{i,t}/\mu_{i,t}) - \ln \mu_{i,t}) \\
&= \frac{1}{2} \ln |\mathbf{R}^{-1}| - \frac{1}{2} \mathbf{q}_t' \mathbf{R}^{-1} \mathbf{q}_t + \frac{1}{2} \mathbf{q}_t' \mathbf{q}_t \\
&+ \sum_{i=1}^{K} \left[\phi_i \ln \phi_i - \ln \Gamma(\phi_i) - \ln x_{i,t} + \phi_i \left(\ln x_{i,t} - \ln \mu_{i,t} - \frac{x_{i,t}}{\mu_{i,t}} \right) \right]. \\
&= (\text{copula contribution})_t + (\text{marginals contribution})_t.
\end{aligned}$$

The contribution of the t-th observation to the inference can then be decomposed in the contribution of the copula plus the contribution of the marginals. The contribution of the marginals depends only on $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, whereas the contribution of the copula depends on $\mathbf{R}$, $\boldsymbol{\theta}$, $\boldsymbol{\phi}$. This implies

$$\frac{\partial l}{\partial \mathbf{R}^{-1}} = \frac{1}{2}(T\mathbf{R} - \mathbf{q}'\mathbf{q}) \Rightarrow \widehat{\mathbf{R}} = \frac{\mathbf{q}'\mathbf{q}}{T},$$

where $\mathbf{q} = (\mathbf{q}_1'; \ldots; \mathbf{q}_T')$ is a $T \times K$ matrix. Hence the unconstrained ML estimator of $\mathbf{R}$ has an explicit form.

Replacing the estimated $\mathbf{R}$ in the log-likelihood function we obtain a *concentrated log-likelihood*

$$\begin{aligned}
lc &= -\frac{T}{2} \ln |\widehat{\mathbf{R}}| - \frac{1}{2} \sum_{t=1}^{T} q_t' (\widehat{\mathbf{R}}^{-1} - I) q_t + (\text{marginals contribution}) \\
&= -\frac{T}{2} \ln |\widehat{\mathbf{R}}| - \frac{T}{2} \left[K - \text{trace}(\widehat{\mathbf{R}}) \right] + (\text{marginals contribution})
\end{aligned}$$

In deriving the concentrated log–likelihood, $\mathbf{R}$ is estimated without imposing any constraint relative to its nature as 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 $i \neq j = 1, \ldots, K$, 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.[9]

An acceptable compromise which should increase efficiency, although formally it cannot be interpreted as an ML estimator, is to adopt the sample correlation matrix of the $\mathbf{q}_t$'s as a constrained estimator of $\mathbf{R}$, that is

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

where

$$\mathbf{Q} = \frac{\mathbf{q}'\mathbf{q}}{T} \qquad \mathbf{D}_Q = \mathrm{diag}(Q_{11}, \ldots, Q_{KK}).$$

This solution can be justified observing that the copula contribution to the likelihood depends on $\mathbf{R}$ exactly as if it were the correlation matrix of i.i.d. r.v. $\mathbf{q}_t$ normally distributed with mean $\mathbf{0}$ and correlation matrix $\mathbf{R}$. Using this new estimate of $\mathbf{R}$ the trace of $\widetilde{\mathbf{R}}$ is now constrained to K and the concentrated log-likelihood simplifies to

$$\begin{aligned} lc &= -\frac{T}{2} \ln |\widetilde{\mathbf{R}}| + (\text{marginals contribution}) \\ &= -\frac{T}{2} \left[\ln |\mathbf{q}'\mathbf{q}| - \sum_{i=1}^{K} \ln(\mathbf{q}'_{.i}\mathbf{q}_{.i}) \right] + (\text{marginals contribution}). \end{aligned}$$

3.4.2 Derivatives of the concentrated log-likelihood

In what follows, let us consider $\boldsymbol{\lambda} = (\boldsymbol{\theta}; \boldsymbol{\phi})$. As

$$\begin{aligned} \nabla_{\boldsymbol{\lambda}} \ln |\mathbf{q}'\mathbf{q}| &= \frac{2}{T} \sum_{t=1}^{T} \nabla_{\boldsymbol{\lambda}} \mathbf{q}'_t \mathbf{Q}^{-1} \mathbf{q}_t \\ \nabla_{\boldsymbol{\lambda}} \sum_{i=1}^{K} \ln(\mathbf{q}'_{.i}\mathbf{q}_{.i}) &= \frac{2}{T} \sum_{t=1}^{T} \nabla_{\boldsymbol{\lambda}} \mathbf{q}'_t \mathbf{D}_Q^{-1} \mathbf{q}_t \end{aligned}$$

the derivative of lc is given by

$$\nabla_{\boldsymbol{\lambda}} lc = \sum_{t=1}^{T} \left[\nabla_{\boldsymbol{\lambda}} \mathbf{q}'_t [\mathbf{D}_Q^{-1} - \mathbf{Q}^{-1}] \mathbf{q}_t + \nabla_{\boldsymbol{\lambda}} (\text{marginals contribution})_t \right].$$

To develop further this formula we need to distinguish the derivatives $\nabla_{\boldsymbol{\lambda}} \mathbf{q}'_t$ and $\nabla_{\boldsymbol{\lambda}}(\text{marginals contribution})_t$ with respect to $\boldsymbol{\theta}$ (the parameters of $\boldsymbol{\mu}_t$) and $\boldsymbol{\phi}$. After

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

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

some algebra we obtain

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} \mathbf{q}_t' &= -\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' \mathbf{D}_{1t} \\ \nabla_{\boldsymbol{\theta}} (\text{marginals contribution})_t &= \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' \mathbf{v}_{1t} \\ \nabla_{\boldsymbol{\phi}} \mathbf{q}_t' &= \mathbf{D}_{2t} \\ \nabla_{\boldsymbol{\phi}} (\text{marginals contribution})_t &= \mathbf{v}_{2t}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{D}_{1t} &= \text{diag}\left(\frac{f_i(\widehat{\varepsilon}_{i,t})\widehat{\varepsilon}_{i,t}}{\phi(q_{i,t})\mu_{i,t}} : i = 1, \ldots, K \right) \\ \mathbf{v}_{1t} &= \left(\phi_i \frac{\widehat{\varepsilon}_{i,t} - 1}{\mu_{i,t}} : i = 1, \ldots, K \right) \\ \mathbf{D}_{2t} &= \text{diag}\left(\frac{1}{\phi(q_{i,t})} \frac{\partial F_i(\widehat{\varepsilon}_{i,t})}{\partial \phi_i} : i = 1, \ldots, K \right) \\ \mathbf{v}_{2t} &= \left(\ln \phi_i - \psi(\phi_i) + \ln(\widehat{\varepsilon}_{i,t}) - \widehat{\varepsilon}_{i,t} + 1 : i = 1, \ldots, K \right). \end{aligned}$$

and $\widehat{\varepsilon}_{i,t} = x_{i,t}/\mu_{i,t}$. Then

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} lc &= \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' \left[-\mathbf{D}_{1t} (\mathbf{D}_Q^{-1} - \mathbf{Q}^{-1}) \mathbf{q}_t + \mathbf{v}_{1t} \right] \\ \nabla_{\boldsymbol{\phi}} lc &= \sum_{t=1}^{T} \left[\mathbf{D}_{2t} (\mathbf{D}_Q^{-1} - \mathbf{Q}^{-1}) \mathbf{q}_t + \mathbf{v}_{2t} \right] \end{aligned}$$

We remark that the derivatives $\dfrac{\partial F_i(\widehat{\varepsilon}_{i,t})}{\partial \phi_i}$ must be computed numerically.

3.4.3 Details about the estimation of $\boldsymbol{\theta}$

The estimation of the parameter $\boldsymbol{\theta}$ involved in the equation of $\boldsymbol{\mu}_t$ requires further details. We explain them considering the more general version of $\boldsymbol{\mu}_t$ defined in section 3.3

$$\begin{aligned} \boldsymbol{\mu}_t &= \boldsymbol{\omega}^* + \boldsymbol{\alpha} \mathbf{x}_{t-1} + \boldsymbol{\gamma}_1^* \mathbf{x}_{t-1}^{(s)} + \boldsymbol{\gamma}_2 \mathbf{x}_{t-1}^{(-)} + \boldsymbol{\beta} \boldsymbol{\mu}_{t-1} \\ &= \boldsymbol{\omega}^* + \boldsymbol{\alpha}^* \mathbf{x}_{t-1}^* + \boldsymbol{\beta} \boldsymbol{\mu}_{t-1} \end{aligned}$$

(see the cited section for details and notation).

Such a structure for $\boldsymbol{\mu}_t$ depends in the general case from $K + 4K^2$ parameters. For instance, when $K = 3$ there are 39 parameters. We think that, in general, data do not provide enough information to capture asymmetries of both types ($\mathbf{x}_{t-1}^{(s)}$ and $\mathbf{x}_{t-1}^{(-)}$) but even removing one of these two components the parameters become $K + 3K^2$ (30 when $K = 3$).

A reduction in the number of parameters can obtained estimating $\boldsymbol{\omega}^*$ from stationary conditions. Imposing that $\boldsymbol{\mu}_t$ is stationary we have

$$\boldsymbol{\omega}^* = [\mathbf{I} - (\boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma}_2/2)]\boldsymbol{\mu},$$

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

$$(\boldsymbol{\mu}_t - \boldsymbol{\mu}) = \boldsymbol{\alpha}(\mathbf{x}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\gamma}_1^* \mathbf{x}_{t-1}^{(s)} + \boldsymbol{\gamma}_2(\mathbf{x}_{t-1}^{(-)} - \boldsymbol{\mu}/2) + \boldsymbol{\beta}(\boldsymbol{\mu}_{t-1} - \boldsymbol{\mu}).$$

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

$$\begin{aligned} \widetilde{\boldsymbol{\mu}}_t &= \boldsymbol{\alpha}\widetilde{\mathbf{x}}_{t-1} + \boldsymbol{\gamma}_1^* \widetilde{\mathbf{x}}_{t-1}^{(s)} + \boldsymbol{\gamma}_2 \widetilde{\mathbf{x}}_{t-1}^{(-)} + \boldsymbol{\beta}\widetilde{\boldsymbol{\mu}}_{t-1} \\ &= \boldsymbol{\alpha}^* \widetilde{\mathbf{x}}_{t-1}^* + \boldsymbol{\beta}\widetilde{\boldsymbol{\mu}}_{t-1} \end{aligned} \tag{32}$$

where the symbol $\widetilde{\mathbf{x}}$ means the demeaned version of $\mathbf{x}$. This solution save K parameters in the iterative estimation and provides better performances than direct ML estimates of $\boldsymbol{\omega}^*$ in simulations.

To save time it is also useful take into account analytic derivatives of $\widetilde{\boldsymbol{\mu}}_t$ with respect to parameters. To this purpose rewrite (32) as

$$\widetilde{\boldsymbol{\mu}}_t = \widetilde{\mathbf{B}}_{t-1}' \operatorname{vec}(\boldsymbol{\beta}') + \widetilde{\mathbf{A}}_{t-1}^{*\prime} \operatorname{vec}(\boldsymbol{\alpha}^{*\prime})$$

where

$$\widetilde{\mathbf{B}}_t = \begin{pmatrix} \widetilde{\boldsymbol{\mu}}_t & 0 & \dots & 0 \\ 0 & \widetilde{\boldsymbol{\mu}}_t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \widetilde{\boldsymbol{\mu}}_t \end{pmatrix} \qquad \widetilde{\mathbf{A}}_t^* = \begin{pmatrix} \widetilde{\mathbf{x}}_t^* & 0 & \dots & 0 \\ 0 & \widetilde{\mathbf{x}}_t^* & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \widetilde{\mathbf{x}}_t^* \end{pmatrix}$$

and the operator vec() stacks the columns of the matrix inside brackets. Taking derivatives with respect to $\operatorname{vec}(\boldsymbol{\beta}')$ and $\operatorname{vec}(\boldsymbol{\alpha}^{*\prime})$ and arranging results we obtain

$$\frac{\partial \widetilde{\boldsymbol{\mu}}_t}{\partial \operatorname{vec}(\boldsymbol{\theta})} = \begin{pmatrix} \widetilde{\mathbf{B}}_{t-1} \\ \widetilde{\mathbf{A}}_{t-1}^* \end{pmatrix} + \frac{\partial \widetilde{\boldsymbol{\mu}}_{t-1}}{\partial \operatorname{vec}(\boldsymbol{\theta})}$$

where $\operatorname{vec}(\boldsymbol{\theta}) = (\operatorname{vec}(\boldsymbol{\beta}'); \operatorname{vec}(\boldsymbol{\alpha}^{*\prime}))$.

3.5 Estimating Functions inference

A different approach to estimation and inference of a vector MEM is provided by *Estimating Functions* (Heyde, 1997) which has less demanding assumptions relative to ML: in the simplest version, requires only the specification of the first two conditional moments of $\mathbf{x}_t$, as in (23) and (24). The advantage of not formulating assumptions about the shape of the conditional distribution of $\mathbf{x}_t$ is balanced by a loss in efficiency with respect to Maximum likelihood under correct specification of the complete model. Given that even in the univariate MEM we had interpreted the ML estimator as QML, the flexibility provided by the Estimating Functions (henceforth EF) seems promising.

3.5.1 Framework

We consider here the EF inference of the vector MEM defined by (21) and (22) or, equivalently, by (23) and (24). In this framework the parameters are collected in the p^*-dimensional vector $\boldsymbol{\lambda} = (\boldsymbol{\theta}, \boldsymbol{\Sigma})$, with $p^* = p + K(K+1)/2$: $\boldsymbol{\theta}$ includes the parameters of primary interest involved in the expression of $\boldsymbol{\mu}_t$; $\boldsymbol{\Sigma}$ is the conditional variance–covariance matrix of the multiplicative innovations $\boldsymbol{\varepsilon}_t$ and represents a nuisance parameter with respect to $\boldsymbol{\theta}$.

Following Bibby *et al.* (2004), an *estimating function* for the p^*-dimensional vector $\boldsymbol{\lambda}$ based on a sample $\mathbf{x}_{(T)}$ is a p^*-dimensional function denoted as

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

The EF estimator $\widehat{\boldsymbol{\lambda}}$ is defined as the solution to the corresponding *estimating equation*:

$$\widehat{\boldsymbol{\lambda}} \text{ such that } \mathbf{g}(\widehat{\boldsymbol{\lambda}}) = 0. \tag{33}$$

To be an useful estimating function, regularity conditions on $\mathbf{g}$ are usually imposed. Following Heyde (1997, sect. 2.6) or Bibby *et al.* (2004, Theorem 2.3), we consider zero mean and square integrable martingale estimating functions that are sufficiently regular to guarantee that a WLLN and a CLT applies. The martingale restriction is often motivated observing that the score function, if it exists, is usually a martingale: so it is quite natural to approximate it using families of martingales estimating functions. Furthermore, in modelling time series processes, estimating functions that are martingales arise quite naturally from the structure of the process. We denote as $\mathcal{M}$ the class of EF satisfying the asserted regularity conditions.

Following the decomposition of $\boldsymbol{\lambda}$ into $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, let us decompose the estimating function $\mathbf{g}(\boldsymbol{\lambda})$ as

$$\mathbf{g}(\boldsymbol{\lambda}) = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\lambda}) \\ \mathbf{g}_2(\boldsymbol{\lambda}) \end{pmatrix} : \tag{34}$$

$\mathbf{g}_1$ has dimension p and refers to $\boldsymbol{\theta}$; $\mathbf{g}_2$ has dimension $K(K+1)/2$ and refers to $\boldsymbol{\Sigma}$.

3.5.2 Inference on θ

From (23) we observe immediately that "residuals" $\mathbf{v}_t = \mathbf{x}_t - \boldsymbol{\mu}_t$ have the property to be martingale differences. Using these as basic ingredient, we can then construct the *Hutton–Nelson quasi-score function* (Heyde (1997, p. 32)) as an estimating function for $\boldsymbol{\theta}$:

$$\mathbf{g}_1(\boldsymbol{\lambda}) = \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' [\mathrm{diag}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma}\, \mathrm{diag}(\boldsymbol{\mu}_t)]^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_t). \tag{35}$$

Transposing to the vector MEM considered here the arguments in Heyde (1997), (35) is (for fixed $\boldsymbol{\Sigma}$) the "best" estimating function for $\boldsymbol{\theta}$ in the subclass of $\mathcal{M}$ composed by all estimating functions that are linear in the residuals $\mathbf{v}_t = \mathbf{x}_t - \boldsymbol{\mu}_t$. In fact, more precisely, the Hutton–Nelson quasi score function is *optimal* (both in the *asymptotic* and in the *fixed sample* sense – see Heyde (1997, ch. 2) for details) in the class of estimating functions

$$\mathcal{M}^{(1)} = \left\{ \mathbf{g} \in \mathcal{M} : \mathbf{g}(\boldsymbol{\lambda}) = \sum_{t=1}^{T} \mathbf{a}_t(\boldsymbol{\lambda}) \mathbf{v}_t(\boldsymbol{\lambda}) \right\},$$

where $\mathbf{v}_t(\boldsymbol{\lambda})$ are $(p, 1)$ martingale differences and $\mathbf{a}_t(\boldsymbol{\lambda})$ are (p, p) $\mathcal{F}_{t-1}$–measurable functions.

Very interestingly, in the 1-dimensional case, (35) specializes to

$$\mathbf{g}_1(\boldsymbol{\lambda}) = \sigma^{-2} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \mu_t \frac{x_t - \mu_t}{\mu_t^2},$$

which provides exactly the first order condition of the univariate MEM. Hence, formula (16) can be also justified from an EF perspective. The main substantial difference from the multivariate case, is that in the vector MEM explicit dependence from the nuisance parameter $\boldsymbol{\Sigma}$ cannot be suppressed. About this point, however, we claim that EF inferences of vector MEM are not too seriously affected by the presence of $\boldsymbol{\Sigma}$, as stated in the following section.

3.5.3 Inference on $\boldsymbol{\Sigma}$

A clear discussion of inferential issues with nuisance parameters can be found, among others, in Liang and Zeger (1999). The main inferential problem is that in presence of nuisance parameters some properties of estimating functions are no longer valid if we replace parameters with the corresponding estimators. For instance, unbiasedness of $g_1(\boldsymbol{\lambda})$ is not guaranteed if $\boldsymbol{\lambda}$ is replaced by an estimator $\widehat{\boldsymbol{\lambda}}$. By consequence, optimality properties also are not guaranteed.

An interesting statistical handling of nuisance parameter in the estimating functions framework is provided in Knudsen (1999) and Jørgensen and Knudsen (2004). Their handling parallels in some aspects the notion of *Fisher-orthogonality* or, shortly, *F-orthogonality*, in ML estimation. F-orthogonality is defined by block diagonality of the Fisher information matrix for $\boldsymbol{\lambda} = (\boldsymbol{\theta}, \boldsymbol{\Sigma})$. This particular structure guarantees the following properties of the ML estimator (see the cited reference for details):

1. asymptotic independence of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Sigma}}$;
2. efficiency-stable estimation of $\boldsymbol{\theta}$, in the sense that the asymptotic variance for $\boldsymbol{\theta}$ is the same whether $\boldsymbol{\Sigma}$ is treated as known or unknown;
3. simplification of the estimation algorithm;
4. $\widehat{\boldsymbol{\theta}}(\boldsymbol{\Sigma})$, the estimate of $\boldsymbol{\theta}$ when $\boldsymbol{\Sigma}$ is given, varies only slowly with $\boldsymbol{\Sigma}$.

Starting from this point, Jørgensen and Knudsen (2004) extend F-orthogonality to EFs introducing the concepts of *nuisance parameter insensitivity* (henceforth NPI). NPI is an extended parameter orthogonality notion for estimating function, that guarantees properties 2-4 above. Among these, Jørgensen and Knudsen (2004) state that efficiency stable estimation is the crucial property and prove that, in a certain sense, NPI is a necessary condition for this.

Here we emphasize two major points:

1. Jørgensen and Knudsen (2004) define and prove properties connected to NPI within a *fixed sample* framework. However definitions and theorems can be easily adapted and extended to the *asymptotic* framework, that is the typical reference when stochastic processes are of interest.
2. Estimating function (35) for estimating $\boldsymbol{\theta}$ is nuisance parameter insensitive.

About the first point, key quantities for inferences on estimating functions estimators within the fixed sample framework are the *variance–covariance matrix* $V(\mathbf{g})$ and the *sensitivity matrix* $E(\nabla_{\boldsymbol{\theta}}\mathbf{g}')$ of the EF $\mathbf{g}$. These quantities are replaced in the asymptotic framework by their 'estimable' counterparts: the *quadratic characteristic*

$$\langle \mathbf{g} \rangle = \sum_{t=1}^{T} V(\mathbf{g}_t|\mathcal{F}_{t-1}),$$

and the *compensator*,

$$\overline{\mathbf{g}} = \sum_{t=1}^{T} E\left(\nabla_{\boldsymbol{\theta}}\mathbf{g}_t'|\mathcal{F}_{t-1}\right),$$

of $\mathbf{g}$, where the martingale EF $\mathbf{g}$ is represented as sum of martingale differences $\mathbf{g}_t$:

$$\mathbf{g} = \sum_{t=1}^{T} \mathbf{g}_t. \tag{36}$$

As pointed by Heyde (1997, p. 28), there is a close relation between $V(\mathbf{g})$ and $\langle \mathbf{g} \rangle$ and between $E(\nabla_{\theta}\mathbf{g}')$ and $\overline{\mathbf{g}}$. But, above all, the optimality theory within the asymptotic framework parallels substantially that within the fixed sample framework because these quantities share the same essential properties. These properties can be summarized taking into account two points. The first: under regularity conditions (see Heyde (1997), Bibby *et al.* (2004) and Jørgensen and Knudsen (2004) for details) we have,

$$E(\nabla_{\boldsymbol{\theta}} \mathbf{g}') = C(\mathbf{g}, \mathbf{s})$$

and

$$\overline{\mathbf{g}} = \sum_{t=1}^{T} C(\mathbf{g}_t, \mathbf{s}_t | \mathcal{F}_{t-1}),$$

where $\mathbf{s} = \sum_{t=1}^{T} \mathbf{s}_t$ is the score function expressed as a sum of martingale differences. The second: both the covariance and the sum of conditional covariances of the martingale difference components share the bilinearity properties typical of a scalar product. Just by virtue of these common properties, it is possibile to show that NPI can be defined and guarantees properties 2–4 above also in the asymptotic framework. Hence, we sketch here only the main points and we remaind to Jørgensen and Knudsen (2004) for the original full treatment.

We partition the compensator and the quadratic characteristic conformably to $\mathbf{g}$ as in (34), that is

$$\overline{\mathbf{g}} = \begin{pmatrix} \overline{\mathbf{g}}_{11} & \overline{\mathbf{g}}_{12} \\ \overline{\mathbf{g}}_{21} & \overline{\mathbf{g}}_{22} \end{pmatrix}$$

and

$$\langle \mathbf{g} \rangle = \begin{pmatrix} \langle \mathbf{g} \rangle_{11} & \langle \mathbf{g} \rangle_{12} \\ \langle \mathbf{g} \rangle_{21} & \langle \mathbf{g} \rangle_{22} \end{pmatrix},$$

and we consider *Heyde information*

$$\mathbf{I} = \overline{\mathbf{g}}' \langle \mathbf{g} \rangle^{-1} \overline{\mathbf{g}} \tag{37}$$

as the natural counterpart, within the asymptotic framework, of the *Godambe information*

$$\mathbf{J} = E(\nabla_{\theta'} \mathbf{g}) V(\mathbf{g})^{-1} E(\nabla_{\theta} \mathbf{g}'),$$

that instead is typical of the fixed sample framework. we remember that each of these versions of *information*, in the respective framework, provides the inverse of the asymptotic variance covariance matrix of the EF parameter.

We say that the marginal EF $\mathbf{g}_1$ is *nuisance parameter insensitive* if $\overline{\mathbf{g}}_{12} = 0$. Using block–matrix algebra we check easily that NPI of $\mathbf{g}_1$ implies that the asymptotic variance–covariance matrix of the EF estimator of $\boldsymbol{\theta}$ is given by

$$\overline{\mathbf{g}}_{11}^{-1} \langle \mathbf{g} \rangle_{11} \overline{\mathbf{g}}_{11}^{-1\prime}. \tag{38}$$

Employing concepts and definitions above is possible to show, exactly using the same devices, that the whole theory in Jørgensen and Knudsen (2004), and in particular their *insensitivity theorem*, preserves substantially unaltered.

Returning now to the vector MEM, we can check easily that the EF (35) for estimating $\boldsymbol{\theta}$ is $\boldsymbol{\Sigma}$–insensitive. In fact, expressing (35) as sum of martingale differences

$\mathbf{g}_{1,t} = -\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_t'(\mathrm{diag}(\boldsymbol{\mu}_t)\boldsymbol{\Sigma}\,\mathrm{diag}(\boldsymbol{\mu}_t))^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_t)$ as in (36), we have

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

whose conditional expected value is $\mathbf{0}$. This implies $\overline{\mathbf{g}}_{12} = 0$. As a consequence, the asymptotic variance–covariance matrix of $\boldsymbol{\theta}$ is given by (38), that for the model considered becomes

$$\overline{\mathbf{g}}_{11}^{-1}\langle\mathbf{g}\rangle_{11}\overline{\mathbf{g}}_{11}^{-1\prime} = \langle\mathbf{g}\rangle_{11}^{-1} = \left[\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_t'[\mathrm{diag}(\boldsymbol{\mu}_t)\boldsymbol{\Sigma}\,\mathrm{diag}(\boldsymbol{\mu}_t)]^{-1}\nabla_{\boldsymbol{\theta}'}\boldsymbol{\mu}_t\right]^{-1}. \tag{39}$$

However (35) and (39) require values for $\boldsymbol{\Sigma}$. We suggest to estimate it using the estimating function:

$$\mathbf{g}_2(\lambda) = \sum_{t=1}^{T}(\mathbf{v}_t\mathbf{v}_t' - \boldsymbol{\Sigma}), \tag{40}$$

that implies

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{v}_t\mathbf{v}_t'.$$

This estimator generalized (18) to the multivariate model. We note that (40) is unbiased, but does not satisfy optimality properties. In this sense it is possible to improve (40) but we do not pursue this here because it can be accomplished only adding further assumptions about higher order moments of $\mathbf{x}_t|\mathcal{F}_{t-1}$, thus contrasting against the sprit with which we adopted EF to make inferences in the vector MEM.

Furthermore is possible to show that $\mathbf{g}_2(\boldsymbol{\lambda})$ is $\boldsymbol{\theta}$–insensitive, so that the compensator is diagonal. In fact

$$\nabla_{\theta_i}(\mathbf{v}_t\mathbf{v}_t' - \boldsymbol{\Sigma}) = -(\mathbf{v}_t\nabla_{\theta_i}\boldsymbol{\mu}_t' + \nabla_{\theta_i}\boldsymbol{\mu}_t\mathbf{v}_t'),$$

whose conditional expected value is 0. We remark that block–diagonality of $\overline{\mathbf{g}}$ does not implies asymptotic independency of the EF estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ because, in (37), the off–diagonal components of $\langle\mathbf{g}\rangle$ are in general different from $\mathbf{0}$.

4 Empirical Analysis

The multivariate extension of the MEM model is illustrated on two datasets: the first concerns three variables as in Engle and Gallo (2006), namely absolute returns $|r_t|$, high-low range hl_t and realized volatility rv_t derived from 5-minutes returns as in Andersen *et al.* (2003) measured on the General Electric stock between Jan. 3, 1995 and Dec. 29, 2001 (1515 obs.). The observations were rescaled so to have the same quadratic mean as absolute returns.

In Table 1 we report a few descriptive statistics relative to the variables employed. As one may expect, daily range and realized volatility have high maximum values and there are a few zeroes in the absolute returns. We also report the unconditional correlations across the whole sample period showing that daily range has a relatively high correlation with both absolute returns and realized volatility (which are less correlated with one another). Finally, the time series behavior of the series is depicted in Figure 1.

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

	Indicator		
Statistics	$\|r_t\|$	hl_t	rv_t
min	0	0.266	.429
max	11.123	6.318	6.006
mean	1.194	1.194	1.194
standard deviation	1.058	0.614	0.424
skewness	2.20	2.03	2.65
kurtosis	12.42	10.97	18.69
n. of zeros	60	0	0
correlations	$\|r_t\|$	hl_t	rv_t
$\|r_t\|$		0.767	0.440
hl_t			0.757

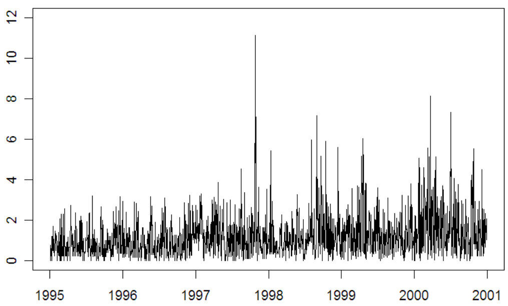
The estimated coefficients of the model selected according to the smallest BIC criterion recorded in the copula model (labelled ML-R) are reported in Table 2 with the associated robust standard errors. For comparison purposes, the other two estimators (equation by equation – labelled ML-I, and estimating equations – labelled EE) have been applied on the same specification. The single subscripts 1, 2, 3 refer, respectively, to the variables $|r_t|$, hl_t, rv_t. A coefficient with double subscript i, j indicates the impact of the j-th variable at time $t-1$ on the i-th variable at time t.

We can note that the selected model confirms the findings of Engle and Gallo (2006) in that the daily range has significant lagged (be they overall or asymmetric) effects on the other variables, while it is not affected by them. The total influence of variables can be measured by means of an *Impact Matrix* $\mathbf{A}$, previously defined in (31). Again, we propose a comparison of the estimated values of the non–zero elements of the matrix $\mathbf{A}$ in Table 3.

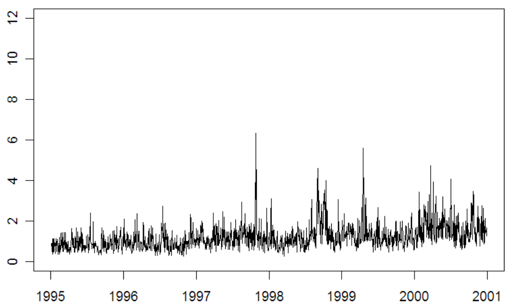
Overall, the estimated coefficient values do not differ too much between the system estimation with copulas and with estimating equations. Somewhat different (but not dramatically so) are the estimates equation by equation.

Similar comments can be had for the estimated characteristic roots of such a matrix which rule the dynamics of the system: the estimated persistence is underestimated

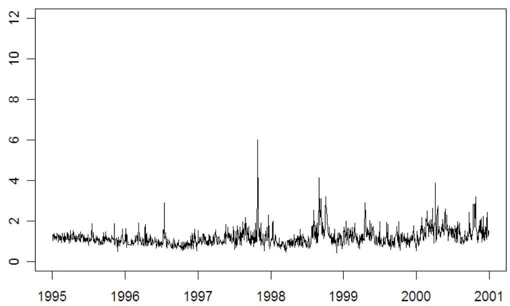
Figure 1: Time series of the indicators $|r_t|$, hl_t, rv_t – GE stock, 01/03/1995-12/29/2001.



(a) $|r_t|$



(b) hl_t



(c) rv_t

Table 2: Coefficient estimation of the vMEM on the indicators $|r_t|$, hl_t, rv_t – GE stock, 01/03/1995-12/29/2001.

	ML-R		ML-I		EE	
Parameter	coeff	s.err.	coeff	s.err.	coeff	s.err.
Equation for $\mathbf{|r_t|}$						
ω_1^*	0.5827	.	0.7618	.	0.5948	.
$|r_{t-1}|^{(-)}$	0.0459	0.0182	0.064	0.0282	0.0444	0.018
μ_{t-1}^r	0.7777	0.0432	0.7575	0.0796	0.779	0.0448
hl_{t-1}	0.1506	0.0336	0.1468	0.048	0.149	0.035
Equation for $\mathbf{hl_t}$						
ω_2^*	0.4283	.	0.5746	.	0.486	.
hl_{t-1}	0.1554	0.0192	0.1614	0.027	0.1544	0.016
$hl_{t-1}^{(-)}$	0.0493	0.012	0.0592	0.0164	0.0479	0.0106
μ_{t-1}^h	0.784	0.0266	0.7609	0.0436	0.7809	0.0199
Equation for $\mathbf{rv_t}$						
ω_3^*	0.8524	.	1.2213	.	0.8706	.
rv_{t-1}	0.2544	0.0276	0.329	0.0355	0.2551	0.0173
$rv_{t-1}^{(-)}$	-0.0569	0.0172	-0.0693	0.0226	-0.0706	0.0184
μ_{t-1}^v	0.6508	0.0379	0.5418	0.0483	0.6498	0.0212
hl_{t-1}	0.1038	0.0179	0.1232	0.0228	0.115	0.0169

by the equation–by–equation estimator relative to the other two methods.

A comment on the shape of the distribution as it emerges from the estimation methods based on the Gamma distribution is in order (cf. Table 5). The estimated ϕ parameters are quite different from one another (there is no major difference according to whether the system is estimated equation–by–equation or jointly) reflecting the different characteristics of each variable.

The estimator equation–by–equation provides an estimate of the correlation across residuals which is then captured well by the system estimators (cf. Table 6).

The second set of results is obtained by analyzing four measures of volatility based on intra-daily data:

1. one is absolute daily returns, $|r_t|$;
2. the second is the *realized absolute variation* (Barndorff-Nielsen and Shephard, 2004) av_t computed as the sum of the absolute value of the M intra-daily returns, namely, $av_t = \frac{1}{\Gamma(0.5)}\sqrt{\frac{2}{M}}\sum_{d=1}^{M}|r_{d,t}|$. M in our case is 77, as we are taking five–minute intervals but we are excluding overnight returns;
3. the third is the *bi-power variation* (Barndorff-Nielsen and Shephard, 2004) bv_t defined as the sum of the $M-1$ products of pairs of subsequent intra-daily

Table 3: Estimated Impact Matrices of the vMEM on the indicators $|r_t|$, hl_t, rv_t – GE stock, 01/03/1995-12/29/2001.

	$\|r_{t-1}\|$	hl_{t-1}	rv_{t-1}
ML-R			
$\|r_t\|$	0.80064	0.15057	0
hl_t	0	0.96413	0
rv_t	0	0.05189	0.87673
ML-I			
$\|r_t\|$	0.78945	0.14675	0
hl_t	0	0.95188	0
rv_t	0	0.0616	0.83613
EE			
$\|r_t\|$	0.80118	0.14901	0
hl_t	0	0.9593	0
rv_t	0	0.05751	0.86959

Table 4: Characteristic roots of the estimated Impact Matrices of the vMEM on the indicators $|r_t|$, hl_t, rv_t – GE stock, 01/03/1995-12/29/2001.

	Characteristic roots of $\mathbf{A}$		
ML-R	0.80064	0.85259	0.98827
ML-I	0.78945	0.80948	0.97853
EE	0.80118	0.84151	0.98738

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 the *realized volatility* rv_t defined as before as the sum of the squares of the M intra-daily returns $r_{d,t}$ computed as the log differences of the transaction price series recorded at or about five minute intervals: $rv_t = \sqrt{\sum_{d=1}^{M} r_{d,t}^2}$.

As before there is a mild correlation between the absolute returns and the intradaily volatility measures, whereas there is quite a large correlation between the latter, in excess of 0.97. This is of interest for comparing the methodologies proposed, since it can show what impact it may have on the estimation results. The time series are reported in Figure 2.

Table 5: Estimated parameters of the Gamma marginals of the vMEM on the indicators $|r_t|$, hl_t, rv_t – GE stock, 01/03/1995-12/29/2001 (1515 obs.)

	ML-R	ML-I
ϕ_1	1.2808	1.272
ϕ_2	6.6435	6.689
ϕ_3	20.6836	20.84

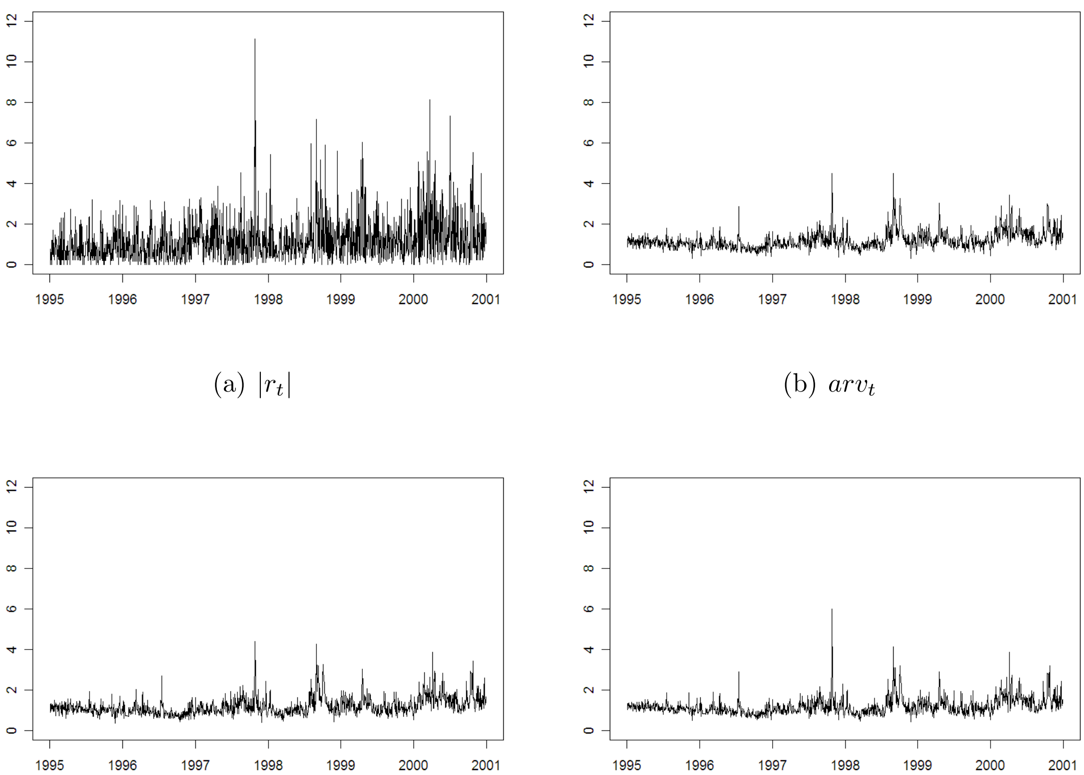
Table 6: Correlation matrices of the residuals of the vMEM on the indicators $|r_t|$, hl_t, rv_t – GE stock, 01/03/1995-12/29/2001.

	ML-R		
	$\|r_t\|$	hl_t	rv_t
$\|r_t\|$	1	0.74945	0.31387
hl_t	0.74945	1	0.62916
rv_t	0.31387	0.62916	1
	ML-I		
	$\|r_t\|$	hl_t	rv_t
$\|r_t\|$	1	0.74893	0.30798
hl_t	0.74893	1	0.62016
rv_t	0.30798	0.62016	1
	EE		
	$\|r_t\|$	hl_t	rv_t
$\|r_t\|$	1	0.74879	0.31459
hl_t	0.74879	1	0.63088
rv_t	0.31459	0.63088	1

Table 7: 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	.296	.360	.429
max	11.123	4.505	4.384	6.006
mean	1.194	1.194	1.194	1.194
standard deviation	1.058	.442	.449	.424
skewness	2.20	1.98	1.95	2.65
kurtosis	12.42	10.05	9.44	18.69
n. of zeros	60	0	0	0
correlations	$\|r_t\|$	arv_t	brv_t	rv_t
$\|r_t\|$		0.420	0.421	0.440
arv_t			0.983	0.974
brv_t				0.975

Figure 2: 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

As before, the estimated coefficients of the model selected according to the smallest BIC value recorded in the copula model (labelled ML-R) are reported in Table 8 with the associated robust standard errors. Also in this case, we have favored comparison by applying the same specification on the other two estimators (equation by equation – labelled ML-I, and estimating equations – labelled EE).

Table 8: Coefficient estimation of the vMEM on the indicators $|r_t|$, arv_t, brv_t, rv_t – GE stock, 01/03/1995-12/29/2001.

	ML-R		ML-I		EE	
Parameter	coeff	std. err.	coeff	std. err.	coeff	std. err.
ω_1^*	0.2879	.	0.2887	.	0.2806	.
ω_2^*	0.7029	.	1.0297	.	0.7409	.
ω_3^*	0.6624	.	0.928	.	0.6986	.
ω_4^*	0.7991	.	1.1608	.	0.8231	.
$\beta_{1,1}$	0.9446	0.0138	0.9438	0.0241	0.9456	0.0114
$\beta_{2,2}$	0.6649	0.0192	0.5422	0.0494	0.6746	0.0203
$\beta_{3,3}$	0.6723	0.0189	0.5793	0.0498	0.6829	0.0205
$\beta_{4,4}$	0.662	0.0176	0.5359	0.0533	0.6726	0.0208
$\alpha_{2,2}$	0.0563	0.0172	0.3465	0.0353	0.0582	0.0169
$\alpha_{2,3}$	0.196	0.0214	0	.	0.1792	0.0219
$\alpha_{3,3}$	0.2492	0.0163	0.3192	0.0354	0.2342	0.0176
$\alpha_{4,3}$	0.0849	0.0156	0.0876	0.0568	0.0759	0.0162
$\alpha_{4,4}$	0.1614	0.013	0.2545	0.0763	0.158	0.0158
$\gamma_{1,1}$	0.0625	0.0134	0.064	0.0214	0.0618	0.0113
$\gamma_{2,2}$	0.048	0.0084	0.0503	0.0111	0.0518	0.0086
$\gamma_{3,2}$	0.0461	0.0084	0.0477	0.0111	0.0489	0.0088
$\gamma_{4,4}$	0.0494	0.0073	0.0497	0.0107	0.0493	0.0083

The results can be synthesized within the estimated Impact Matrix **A** (Table 9) which shows that neither the absolute returns nor the realized volatility feed into other measures, while the realized absolute variation and the bipower feed into one another (the latter more strongly than the former) and the bipower also into the realized volatility.

5 Conclusions

In this paper we have presented a general discussion of the vector specification of the Multiplicative Error Models introduced by Engle (2002). The specification in a multivariate framework cannot exploit multivariate Gamma distributions because they appear too restrictive. The maximum likelihood estimator can be derived by framing the multivariate innovation process in a copula with Gamma marginals framework. As an alternative, which provides estimators which are not based on any distributional assumptions, we also suggested the use of estimating equations.

Table 9: Estimated Impact Matrices of the vMEM on the indicators $|r_t|$, arv_t, brv_t, rv_t – GE stock, 01/03/1995-12/29/2001.

	ML-R			
	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}
$\|r_t\|$	0.97589	0	0	0
arv_t	0	0.74517	0.19597	0
brv_t	0	0.02304	0.92149	0
rv_t	0	0	0.08494	0.84814
	ML-I			
	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}
$\|r_t\|$	0.97583	0	0	0
arv_t	0	0.91377	0	0
brv_t	0	0.02385	0.89843	0
rv_t	0	0	0.08763	0.81517
	EE			
	$\|r_{t-1}\|$	arv_{t-1}	brv_{t-1}	rv_{t-1}
$\|r_t\|$	0.9765	0	0	0
arv_t	0	0.75873	0.17923	0
brv_t	0	0.02447	0.91703	0
rv_t	0	0	0.07587	0.8552

Table 10: Characteristic roots of the estimated Impact Matrices of the vMEM on the indicators $|r_t|$, arv_t, brv_t, rv_t – GE stock, 01/03/1995-12/29/2001.

	Characteristic roots of **A**			
ML-R	0.74051	0.79536	0.97589	0.97893
ML-I	0.75872	0.90632	0.96233	0.97583
EE	0.75316	0.80779	0.97001	0.9765

The results are quite comforting given that the system estimation provides fairly similar values.

Table 11: Estimated parameters of the Gamma marginals of the vMEM on the indicators $|r_t|$, arv_t, brv_t, rv_t – GE stock, 01/03/1995-12/29/2001.

	ML-R	ML-I
ϕ_1	1.2592	1.2559
ϕ_2	17.7402	18.0384
ϕ_3	17.0704	17.1381
ϕ_4	20.1848	20.4209

A Useful univariate statistical distributions

In this section we summarize the properties of the statistical distributions employed in the work.

A.1 The Gamma distribution

A r. v. X follows a $Gamma(\phi, \beta)$ distribution if its pdf if given by

$$f(x) = \frac{\beta^{\phi}}{\Gamma(\phi)} x^{\phi-1} e^{-\beta x}$$

for $x > 0$ and 0 otherwise, where $\phi, \beta > 0$. The main moments are

$$E(X) = \phi/\beta$$
$$V(X) = \phi/\beta^2.$$

An useful property of a Gamma r.v. is the following:

$$X \sim Gamma(\phi, \beta) \text{ and } c > 0 \Rightarrow Y = cX \sim Gamma(\phi, \beta/c).$$

Some interesting special cases of a *Gamma* distributions are the Exponential distribution and the χ^2 distribution:

$$Exponential(\beta) = Gamma(1, \beta)$$
$$\chi^2(n) = Gamma(n/2, 1/2).$$

In this work we often employ *Gamma* distributions restricted to have mean 1, that implies $\beta = \phi$.

Table 12: Correlation matrices of the residuals of the vMEM on the indicators $|r_t|$, arv_t, brv_t, rv_t – GE stock, 01/03/1995-12/29/2001.

		ML-R		
	$\|r_t\|$	arv_t	brv_t	rv_t
$\|r_t\|$	1	0.27991	0.28106	0.31301
arv_t	0.27991	1	0.96501	0.95538
brv_t	0.28106	0.96501	1	0.95124
rv_t	0.31301	0.95538	0.95124	1
		ML-I		
	$\|r_t\|$	arv_t	brv_t	rv_t
$\|r_t\|$	1	0.27428	0.2725	0.30258
arv_t	0.27428	1	0.95868	0.9529
brv_t	0.2725	0.95868	1	0.94992
rv_t	0.30258	0.9529	0.94992	1
		EE		
	$\|r_t\|$	arv_t	brv_t	rv_t
$\|r_t\|$	1	0.28186	0.28306	0.31487
arv_t	0.28186	1	0.96498	0.95527
brv_t	0.28306	0.96498	1	0.95115
rv_t	0.31487	0.95527	0.95115	1

A.2 The GED distribution

The Generalized Error distribution (GED) is also known as Exponential Power distribution or as Subbotin distribution.

A random variable X follow a $GED(\mu, \sigma, \phi)$ distribution distribution if its pdf if given by

$$f(x) = \frac{\phi^{\phi-1}}{2\Gamma(\phi)\sigma} \exp\left(-\phi \left|\frac{x-\mu}{\sigma}\right|^{1/\phi}\right)$$

for $x \in \mathbb{R}$, where $\mu \in \mathbb{R}$, $\sigma, \phi > 0$. The main moments are

$$E(X) = \mu$$
$$V(X) = \frac{\Gamma(3\phi)}{\phi^{2\phi}\Gamma(\phi)}\sigma^2.$$

An useful property of a GED r.v. is the following:

$$X \sim GED(\mu, \sigma, \phi) \text{ and } c_1, c_2 \in \mathbb{R} \Rightarrow Y = c_1 + c_2 X \sim GED(c_1 + c_2\mu, |c_2|\sigma, \phi).$$

This property qualifies the GED as a location–scale distribution.

Some interesting special cases of a GED distributions are the Laplace distribution and the Normal distribution:

$$Laplace(\mu, \sigma) = GED(\mu, \sigma, 1)$$
$$N(\mu, \sigma) = GED(\mu, \sigma, 1/2).$$

In the work we consider in particular GED distributions with $\mu = 0$.

A.3 Relations GED–Gamma

The following relations links the GED and Gamma distributions:

$$X \sim GED(0, \sigma, \phi) \Rightarrow Y = \left|\frac{X - \mu}{\sigma}\right|^{1/\phi} \sim Gamma(\phi, \phi)$$

Useful particular cases are:

$$X \sim GED(0, 1, \phi) \Rightarrow Y = |X|^{1/\phi} \sim Gamma(\phi, \phi)$$
$$X \sim Laplace(0, 1) \Rightarrow Y = |X| \sim Exponential(1)$$
$$X \sim N(0, 1) \Rightarrow Y = X^2 \sim \chi^2(1).$$

B A summary on copulas

In this section we summarize main concepts and properties of copulas in a rather informal way and without proofs. For a more detailed and rigorous exposition see, among others, Embrechts *et al.* (2001) and Bouyé *et al.* (2000). For a simpler treatment, we assume that all r.v. considered in this section are absolutely continuous.

B.1 Definitions

Copula. Defined "operationally", a K–dimensional *copula* C is the cdf of a continuous uniform r.v. defined on the unit hypercube $[0, 1]^K$, in the sense that each of the univariate components of the r.v. has $Uniform(0, 1)$ marginals, whenever they may be not independent.

Copula density. A copula C is then a continuous cdf with particular characteristics. For some purposes is however useful the associated *copula density* c, defined as

$$c(u) = \frac{\partial^K C(u)}{\partial u_1 \dots \partial u_K}. \quad (41)$$

Ordering. A copula C_1 is *smaller* then C_2, in symbols $C_1 \prec C_2$ if $C_1(u) \le C_2(u)$ $\forall u \in [0, 1]^K$. In practice this means that the picture of C_1 stay below that of C_2.

Some particular copulas. Among copulas, three play an important role: the *lower Frèchet bound* C^-, the *product copula* $C^\perp$ and the *upper Frèchet bound* C^+, defined as follows [10]

$$C^-(u) = \max\left(\sum_{i=1}^{K} u_i - K + 1, 0\right)$$
$$C^\perp(u) = \prod_{i=1}^{K} u_i$$
$$C^+(u) = \min(u_1, \ldots, u_K).$$

We could show that

$$C^-(u) \prec C^\perp(u) \prec C^+(u).$$

More in general, for any copula C,

$$C^-(u) \prec C(u) \prec C^+(u).$$

The concept of ordering is interesting because is connected with measures of concordance (like the correlation coefficient). Considering two r.v., the product copula $C^\perp(u)$ identifies the situation of independence; copulas smaller then $C^\perp(u)$ are connected with situations of discordance; copulas greater then $C^\perp(u)$ describe situations of concordance.

The ordering on the set of copulas, called *concordance ordering*, is however only a partial ordering, since not every pair of copulas is comparable in this way. However many important parametric families of copulas are totally ordered on the base of the values of the parameters (examples are the Frank copula and the Normal copula).

B.2 Understanding copulas

We attempt to explain the usefulness of copulas starting from two important results.

The first one are these well known relations concerning univariate r.v.:

$$X \sim F \Rightarrow U = F(X) \sim Uniform(0, 1)$$

and, conversely,

$$U \sim Uniform(0, 1) \Rightarrow F^{-1}(U) \sim F$$

where with the symbol $X \sim F$ we mean 'X distributed with cdf F'. F^{-1} is often called the *quantile function* of X.

The second is Sklar's theorem. Sklar's theorem is perhaps the most important result regarding copulas.

[10] C^- is not a copula for $K \geq 3$ but we use this notation for convenience.

Theorem 1 ***(Sklar's theorem)***. *Let F a K–dimensional cdf with univariate continuous marginals $F_1, \dots, F_K$. Then there exist an unique K–dimensional copula C such that, $\forall x \in \mathbb{R}^K$,*

$$F(x_1, \dots, x_K) = C(F_1(x_1), \dots, F_K(x_K)).$$

Conversely, if C is a K–dimensional copula and $F_1, \dots, F_K$ are univariate cdf , then the function F defined above is a K–dimensional cdf with marginals $F_1, \dots, F_K$.

With these results in mind, one can verify that if F is a cdf with univariate marginals $F_1, \dots, F_K$, and we take $u_i = F_i(x_i)$ then $F(F_1^{-1}(u_1), \dots, F_K^{-1}(u_K))$ is a copula that guarantees the representation in the Sklar's theorem. We write then

$$C(u_1, \dots, u_K) = F(F_1^{-1}(u_1), \dots, F_K^{-1}(u_K)). \tag{42}$$

In this framework the main result of Sklar's theorem it is then, not the existence of a copula representation, but the fact that this representation is unique. We note also that given a particular cdf and its univariate marginals, (42) represent a direct way to find functions that are copulas.

Conversely, if in equation (42) we replace each u_i by its probability representation $F_i(x_i)$ we obtain

$$C(F_1(x_1), \dots, F_K(x_K)) = F(x_1, \dots, x_K) \tag{43}$$

that is exactly the formula of the Sklar's theorem.

In practice, the copula of a continuous K–dimensional r.v. X (unique by the Sklar's theorem) is the representation of the distribution of the r.v. in the 'probability domain' (those of u_i's), rather then in the 'quantile domain' (those of x_i's, in practice the original scale). It is however clear these two ways are not equivalent: the former loss completely the information of the marginals but isolates the structure of dependence of among the components of X. This implies that in the study of a multivariate r.v. it is possible to proceed in 2 steps:

1. identification of the marginal distributions;
2. definition of the appropriate copula in order to represent the structure of dependence among the components of the r.v.

This two step analysis is showed also by the form of the pdf of the K–variate r.v. X that follows from the copula representation (43):

$$f(x_1, \dots, x_K) = c(F_1(x_1), \dots, F_K(x_K)) \prod_{i=1}^{K} f_i(x_i). \tag{44}$$

where $f_i(x_i)$ is the pdf of the univariate margin X_i and c is the copula density. It is a very interesting formulation: the product of the marginal densities, which corresponds to the multivariate density in case of independence among the X_i's, is corrected by the structure of dependence isolated by the copula's density c. We note that also (44) is a way to find functions representing copulas.

B.3 The Normal copula

Amongst the copulas presented in letterature an important role is played by the Normal copula, whose density is given by

$$c(u) = |\mathbf{R}|^{-1/2} \exp\left[-\frac{1}{2}q'(\mathbf{R}^{-1} - I)q\right], \tag{45}$$

where $\mathbf{R}$ is a correlation matrix, $q_i = \Phi^{-1}(u_i)$, $\Phi(.)$ means the cdf of the standard Normal distribution. Normal copulas are interesting because possess interesting characteristics: the capability of capture a broad range of dependencies (the bivariate Normal copula, according to the value of the correlation parameter is capable of attaining the lower Frechèt bound, the product copula and the upper Frechèt bound), the analytical tractability, the simple simulation.

C Useful multivariate statistical distributions

C.1 Multivariate Gamma distributions

The Cheriyan and Ramabhadran multivariate Gamma (shortly *GammaCR*) is defined as follows[11]. Let $y_0, y_1, \ldots, y_K$ independent random variables with the following distributions:

$$y_0 \sim Gamma(\phi_0, 1) \qquad y_i \sim Gamma(\phi_i - \phi_0, 1) \text{ for } i = 1, \ldots, K,$$

where $0 < \phi_0 < \phi_i$ for $i = 1, \ldots, K$. Then, the K–dimensional r.v. $x = (x_1, \ldots, x_K)'$, where

$$x_i = \frac{1}{\beta_i}(y_i + y_0), \tag{46}$$

and $\beta_i > 0$, has *GammaCR* distribution with parameters ϕ_0, $\phi = (\phi_1, \ldots, \phi_K)$ and $\beta = (\beta_1, \ldots, \beta_K)$ that is

$$x \sim GammaCR(\phi_0, \phi, \beta).$$

The pdf is given by

$$f(x) = \frac{1}{\Gamma(\phi_0)} \prod_{i=1}^{K} \frac{\beta_i}{\Gamma(\phi_i - \phi_0)} e^{-\sum_{i=1}^{K} \beta_i x_i} \int_0^v y_0^{\phi_0 - 1} e^{(K-1)y_0} \prod_{i=1}^{K} (\beta_i x_i - y_0)^{\phi_i - \phi_0 - 1} dy_0,$$

[11] We adapt the definition in (Johnson *et al.*, 2000, chapter 48) to our framework.

where $v = \min(\beta_1 x_1, \ldots, \beta_K x_K)$ and the integral leads in general to very complicated expressions. Each univariate r.v. x_i has however a $Gamma(\phi_i, \beta_i)$ distribution. The main conjoint moments can be computed via formula (46); in particular the variance matrix has elements

$$C(x_i, x_j) = \frac{\phi_0}{\beta_i \beta_j}$$

and then

$$\rho(x_i, x_j) = \frac{\phi_0}{\sqrt{\phi_i \phi_j}}.$$

This implies that parameters ϕ_0 plays an important role in determining the correlation among the x_i's and that the $GammaCR$ distribution admits only positive correlations among the univariate components.

C.2 Normal copula – Gamma marginals distribution

The *Normal copula – Gamma marginals* distribution can be defined as a multivariate distribution whose univariate marginals are Gamma distributions and the structure of dependence is represented by a Normal copula. In symbols we write

$$X \sim N(\mathbf{R}) - \prod_{l=1}^{K} Gamma(\phi_i, \beta_i)$$

where $\phi = (\phi_1, \ldots, \phi_K)$, $\beta = (\beta_1, \ldots, \beta_K)$, and $\mathbf{R}$ is a correlation matrix.

Song (2000) defines and discusses multivariate distributions whose univariate marginals are dispersion distributions and the structure of dependence is represented by a Normal copula. Since the Gamma distribution is a particular case of dispersion distribution, the Normal copula – Gamma margin case is covered by the cited work. The author discusses various properties of multivariate dispersion models generated from Normal copula. In particular, exploiting first order approximations in Song (2000, pages 433–4) and adapting notation we obtain that the variance matrix has approximated elements

$$C(x_i, x_j) \simeq \mathbf{R}_{ij} \frac{\sqrt{\phi_i \phi_j}}{\beta_i \beta_j}$$

and then

$$\rho(x_i, x_j) \simeq \mathbf{R}_{ij}.$$

With respect to the $GammaCR$ distribution in section C.1, the copula approach admits negative correlations also.

D A summary on estimating functions

We provide a summary on estimating functions for inference on time series precesses. Quite recent references about the topics handled in this section can be found, among others, in Liang and Zeger (1999), Heyde (1997), Knudsen (1999), Bibby *et al.* (2004), Jørgensen and Knudsen (2004).

D.1 Set-up and notation

Let $\{x_t\}_{t=1}^T$ a discrete time K-dimensional time series process, and let $x_{(T)} = \{x_1, \ldots, x_T\}$ a sample from $\{x_t\}$. We assume that the possible probability measures for $x_{(T)}$ are $\mathcal{P} = \{P_\lambda : \lambda \in \Theta\}$, a union of families of parametric models indexed by λ, a p-dimensional parameter belonging to an open set $\Theta \subseteq \mathbb{R}^p$. We assume also that each $(\Omega, \mathcal{F}, P_\lambda)$ is a complete probability space.

Denoting with $\mathcal{F}_t$ the σ-algebra generated from $(x_1, \ldots, x_t)$, let us recall that a stochastic process x_t:

- is *predictable* if
$$E(x_t|\mathcal{F}_{t-1}) = x_t;$$
- is a *martingale* if
$$E(x_t|\mathcal{F}_{t-1}) = x_{t-1};$$
- is a *martingale difference* if
$$E(x_t|\mathcal{F}_{t-1}) = 0.$$

We note that a martingale x_t can be always represented as a sum of martingale differences $\Delta x_t = x_t - x_{t-1}$:
$$x_t = \sum_{s=1}^{t} \Delta x_s.$$

We will use also the following notations:

- if $f(\lambda)$ is a function of the parameter we denote $\dfrac{\partial f(\lambda)}{\partial \lambda'}$ as $\dot{f}(\lambda)$;
- given two square, symmetric and non negative definite (nnd) matrices A and B, we say A not greater than B (or A not smaller than B) in the Löwner ordering if
$$B - A$$
is nnd and we write
$$A \preceq B.$$

D.2 The likelihood framework

The likelihood framework provides an important reference for the theory of estimating functions. In fact, even if they can be usefully employed just in situations where the likelihood is unknown or is difficult to compute, estimating functions generalize the score function in many senses and also share many of its properties. Moreover, the score function plays an important role in many aspects of the estimating functions theory.

For this purpose we assume P_λ absolutely continuous with respect to some σ-finite measure: in this situation a likelihood is defined. Remembering that with the symbol $x_{(T)} = \{x_1, \ldots, x_T\}$ we mean a sample from x_t, we use the following notation:

- $L_{(T)}(\lambda) = f(x_{(T)}; \lambda)$ is the *likelihood function* (the density of $x_{(T)}$);
- $l_{(T)}(\lambda) = \ln L_T(\lambda)$ is the *log-likelihood function*;
- $u_{(T)}(\lambda) = \dot{l}_{(T)}(\lambda)$ (provided that $l_{(T)}(\lambda) \in \mathcal{C}^1(\Theta)$ a.s.) is the *score function.*

Usually the likelihood function for time series samples is expressed as product of the conditional densities $L_t(\lambda) = f(x_t|\mathcal{F}_{t-1}; \lambda)$ of the single observations in the sample. In this case

$$\begin{aligned} L_{(T)}(\lambda) &= \prod_{t=1}^{T} L_t(\lambda) \\ l_{(T)}(\lambda) &= \sum_{t=1}^{T} l_t(\lambda) \\ u_{(T)}(\lambda) &= \sum_{t=1}^{T} u_t(\lambda) \end{aligned} \tag{47}$$

where $l_t(\lambda) = \ln L_t(\lambda)$ and $u_t(\lambda) = \dot{l}_t(\lambda)$.

We assume that the score function $u_{(T)}$[12] satisfies the usual regularity conditions: for all $\lambda \in \Theta$,

1. $u_{(T)}$ is unbiased, that is $E\left(u_{(T)}\right) = 0$, and with finite variance matrix $V\left(u_{(T)}\right)$[13];
2. $u_{(T)}$ is regular, in the sense that $u_{(T)} \in \mathcal{C}^1(\Theta)$ a.s.;
3. $u_{(T)}$ is smooth, in the sense that differentiation and integration can be interchanged in differencing $E\left(u_{(T)}\right)$ with respect to λ.

[12] In many circumstances we suppress explicit dependence from the parameter λ.

[13] We say that a matrix A is finite if its Frobenius norm is finite, that is if $\|A\|^2 = \sum_{i,j} A_{ij}^2 < \infty$.

Using representation (47), we observe that, since u_t is a martingale difference, $u_{(T)}$ is a martingale and condition 1 above follows.

Under the above regularity conditions, the ML estimator is the solution of the score equation

$$u_{(T)}(\lambda) = 0.$$

The *Fisher information* is defined by

$$I_{(T)}(\lambda) = V\left(u_{(T)}(\lambda)\right) = -E\left(\dot{u}_{(T)}\,(\lambda)\right),$$

where the second equality is named *second Bartlett identity.*

D.3 Estimating functions

Within this framework, an *estimating function* (EF) for λ is a p-dimensional function of the parameter λ and of the sample $x_{(T)}$:

$$g_{(T)}(\lambda; x_{(T)}).$$

Usually we suppress explicit dependence from the observations and sometimes from the parameter also. In these cases we indicate compactly the estimating function as[14]

$$g_{(T)}(\lambda).$$

An estimate for λ, is the value $\widehat{\lambda}_T$ that solves the corresponding *estimating equation* (EE)

$$g_{(T)}(\lambda) = 0. \tag{48}$$

The score function is then a particular case of estimating function. However there is a remarkable difference between the score function and a generic EF. In fact $u_{(T)}(\lambda)$ has $l_T(\lambda)$ as its potential function, since $u_{(T)}(\lambda) = \dot{l}_{(T)}\,(\lambda)$. On the contrary, in general an EF does not possesses a potential function $G_{(T)}(\lambda)$ such that $g_{(T)}(\lambda) = \dot{G}_{(T)}\,(\lambda)$ (Knudsen (1999, p. 2)).

As in the likelihood theory, some regularity conditions on EFs $g_{(T)}(\lambda)$ are usually imposed. In particular it is required that, $\forall\lambda \in \Theta$,

1. $g_{(T)}$ is *unbiased*, that is $E\left(g_{(T)}\right) = 0$, and with finite variance matrix $V\left(g_{(T)}\right)$;
2. $g_{(T)}$ is *regular*, that is $g_{(T)} \in C^1(\Theta)$ a.s. and $E\left(\dot{g}_{(T)}\right)$, called *sensitivity matrix*, is non-singular; Knudsen (1999, p. 4) remarks that this condition implies positive-definiteness of the variance matrix $V\left(g_{(T)}\right)$;
3. $g_{(T)}$ is *smooth*, that is differentiation and integration can be interchanged in differencing $E\left(g_{(T)}\right)$ with respect to λ.

[14] Again, sometimes we suppress explicit dependence from the parameter λ.

(We remark the analogies with the corresponding regularity conditions for score functions in section D.2). We denote as $\mathcal{G}_T$ the family of estimating functions satisfying conditions 1, 2, 3 above and we name this as the family of *regular and smooth* EFs.

Regular and smooth EFs satisfy some interesting properties. Among these (details in Knudsen (1999, ch. 1)):

- if $A(\lambda)$ is a (p,p) non-random, non-singular matrix belonging to $C^1(\Theta)$ a.s., then also the linear transformation
$$A(\lambda)g_{(T)}(\lambda) \tag{49}$$
is an EF belonging to $\mathcal{G}_T$;
- the sensitivity matrix is strictly linked to the correlation of the EF with the score function, since
$$E\left(\dot{g}_{(T)}\right) = -C(g_{(T)}, u_{(T)}). \tag{50}$$

Combining these two results, from any EF $g_{(T)} \in \mathcal{G}_T$ we can derive another EF

$$g^{(s)}_{(T)} = -E\left(\dot{g}'_{(T)}\right) V\left(g_{(T)}\right)^{-1} g_{(T)}, \tag{51}$$

called *standardized version* of $g_{(T)}$. We can check easily that, under the above regularity conditions, $g_{(T)}$ and $g^{(s)}_{(T)}$ are *equivalent*, since they produce the same estimators and have the same variance matrix. However this EF is more directly comparable with the score function than $g_{(T)}$. In fact, replacing (50) into (51), we can check that $g^{(s)}_{(T)}$ can be interpreted as the orthogonal projection of the score function along the direction of $g_{(T)}$. This means that, for any $g_{(T)} \in \mathcal{G}_T$, $g^{(s)}_{(T)}$ is the version of $g_{(T)}$ "closest" to the score function, in the sense that

$$\left(u_{(T)} - g^{(s)}_{(T)}\right)\left(u_{(T)} - g^{(s)}_{(T)}\right)' \preceq \left(u_{(T)} - Ag_{(T)}\right)\left(u_{(T)} - Ag_{(T)}\right)'$$

for any linear transformation of $g_{(T)}$ as in (49) (we omitted dependence from λ). Moreover, we can show that $g^{(s)}_{(T)}$ satisfies a second Bartlett identity

$$V\left(g^{(s)}_{(T)}\right) = -E\left(\dot{g}^{(s)}_{(T)}\right).$$

analogous to that of the score function.

Finally, another interesting result connects EF and score functions: the *invariance property*. In fact if $g_{(T)}(\lambda)$ is a regular and smooth EF and $\xi(\lambda)$ is a 1-to-1 reparameterization, then $g_{(T)}(\lambda(\xi))$ is a regular and smooth EF for ξ (Knudsen (1999, p. 7)).

D.4 Fixed sample criteria

A central part of the theory is devoted to find estimating functions that produce good estimators. To give an intuitive sketch of the role played by quantities and assumptions mentioned above[15], we note as statistical theory rests on first order expansions of the estimating function around the true value λ of the parameter like

$$\begin{aligned} g_{(T)}(\widehat{\lambda}_T) &\simeq g_{(T)}(\lambda_0) + \dot{g}_{(T)}(\lambda_0)(\widehat{\lambda}_T - \lambda_0) \\ &\simeq g_{(T)}(\lambda_0) + E\left(\dot{g}_{(T)}(\lambda_0)\right)(\widehat{\lambda}_T - \lambda_0), \end{aligned} \tag{52}$$

where in the final equation $\dot{g}_{(T)}(\lambda_0)$ is replaced by its expected value (the sensitivity matrix).

Solving for $\widehat{\lambda}_T$ we obtain

$$\widehat{\lambda}_T \simeq \lambda_0 - E\left(\dot{g}_{(T)}(\lambda_0)\right)^{-1} g_{(T)}(\lambda_0). \tag{53}$$

Taking expectations of both sides of (53) we can deduce that unbiasedness of $g_T(\lambda)$ is a quite natural condition to obtain consistency of the EE estimator. Furthermore, taking variances of both sides of (53) we can reasonably guess that the asymptotic variance matrix of the EE estimator is given by

$$V_\infty\left(\widehat{\lambda}_T\right) = E\left(\dot{g}_{(T)}(\lambda_0)\right)^{-1} V\left(g_{(T)}(\lambda_0)\right) E\left(\dot{g}'_{(T)}(\lambda_0)\right)^{-1} = I\left(g_{(T)}(\lambda_0)\right)^{-1}, \tag{54}$$

that has a *sandwich* structure. An heuristic interpretation is that a small asymptotic variance requires 'small' variability and 'large' sensitivity. The random matrix

$$I\left(g_{(T)}(\lambda)\right) = E\left(\dot{g}'_{(T)}(\lambda)\right) V\left(g_{(T)}(\lambda)\right)^{-1} E\left(\dot{g}_{(T)}(\lambda)\right)$$

can then be interpreted as an information matrix and is usually called *Godambe information*. This interpretation and its link with the asymptotic variance of $\widehat{\lambda}_T$ suggest the criterion usually employed to choose optimal EF: "maximize" the Godambe information. More precisely, once selected a subfamily of EF $\mathcal{H}_T \subseteq \mathcal{G}_T$ we have:

Definition 1 (**F-optimality**) $g^*_{(T)}(\lambda)$ *is* F-optimal *(that is optimal in the fixed sample sense) into* $\mathcal{H}_T \subseteq \mathcal{G}_T$ *if*

$$I\left(g^*_{(T)}(\lambda)\right) \succeq I\left(g_{(T)}(\lambda)\right)$$

for all $\lambda \in \Theta$ *and for all* $g_{(T)} \in \mathcal{H}_T$*. An F-optimal estimating function* $g^*_{(T)}(\lambda)$ *is called a* quasi score function*, whereas the corresponding estimator* $\widehat{\lambda}_T$ *is called a*

[15] Along this subsection we omit regularity conditions about $g_{(T)}(\lambda)$ needed to obtain the described results.

quasi likelihood estimator[16].

Returning for a moment to the Godambe information, we observe that $I(g_{(T)})$ can also be interpreted as the variance of the *standardized* version $g_{(T)}^{(s)}$ of the EF. Very interestingly, it can be shown that, when the score function exists, F-optimality as defined above is equivalent to find into $\mathcal{H}_T$ the standardized EF $g_{(T)}^{*(s)}$ "closest" to the score $u_{(T)}$ in the sense that

$$E\left[\left(u_{(T)}-g_{(T)}^{*(s)}\right)\left(u_T-g_{(T)}^{*(s)}\right)'\right] \preceq E\left[\left(u_{(T)}-g_{(T)}^{(s)}\right)\left(u_T-g_{(T)}^{(s)}\right)'\right] \tag{55}$$

for any other $g_{(T)}^{(s)}$ into $\mathcal{H}_T$. Furthermore, (55) reveals that if $\mathcal{H}_T$ encloses $u_{(T)}$, just the score function is optimal with respect to any other $g_{(T)}$.

As stated in Heyde (1997), the criterion in the definition of F-optimality (or the equivalent formulation of equation 55) is hard to apply directly. In practice the following theorem can be employed:

Theorem 2 (Heyde (1997, p. 14)) *If $g_{(T)}^* \in \mathcal{H}_T \subseteq \mathcal{G}_T$ satisfies*

$$E\left(\dot{g}_{(T)}^*\right)^{-1} V\left(g_{(T)}^*\right) = E\left(\dot{g}_{(T)}\right)^{-1} E\left(g_{(T)} g_{(T)}^{*\prime}\right) \tag{56}$$

or, equivalently,

$$E\left(\dot{g}_T\right)^{-1} E\left(g_T g_T^{*\prime}\right) \tag{57}$$

is a constant matrix for all $g_{(T)} \in \mathcal{H}_T$, then $g_{(T)}^*$ is F-optimal in $\mathcal{H}_T$.

Conversely, if $\mathcal{H}_T$ is closed under addition and $g_{(T)}^*$ is F-optimal in $\mathcal{H}_T$ then (56) and (57) holds.

D.5 Asymptotic criteria

As stated in Heyde (1997, p. 26), since the score function, if it exists, is usually a martingale, it is quite natural to approximate it using families of martingales estimating functions. Furthermore, in modelling time series processes, estimating functions that are martingales arise quite naturally from the structure of the process. In this spirit we stress that the optimality criterion discussed in this section is particular to martingale EFs.

Hence we consider here the family $\mathcal{M}_T^{(1)} \in \mathcal{G}_T$ of martingale EF that are also square integrable. We represent members of $\mathcal{M}_T^{(1)}$ as sums *martingale differences* g_t in this

[16] Despite a quasi likelihood function in the sense of potential function of $g_{(T)}$ may not exist (see section D.3)

way:

$$g_{(T)} = \sum_{t=1}^{T} g_t$$

where $g_t = g_{(T)} - g_{(T-1)}$.

Using this kind of representation, for all $g_{(T)}, g_{1(T)}, g_{2(T)} \in \mathcal{M}_T^{(1)}$ the following quantities can be defined:

- $\langle g_1, g_2' \rangle_T$, the *mutual quadratic characteristic* between $g_{1(T)}$ and $g_{2(T)}$:

$$\langle g_1, g_2' \rangle_T = \sum_{t=1}^{T} C(g_{1t}, g_{2t} | \mathcal{F}_{t-1}),$$

  where $g_{1t} = g_{1(t)} - g_{1(t-1)}$ and $g_{2t} = g_{2(t)} - g_{2(t-1)}$;

- $\langle g \rangle_T$, the *quadratic characteristic* of $g_{(T)}$:

$$\langle g \rangle_T = \sum_{t=1}^{T} V(g_t | \mathcal{F}_{t-1}),$$

- $\overline{g}_T$, the *compensator* of $\dot{g}_{(T)}$, in practice a predictable version of the corresponding quantity:

$$\overline{g}_T = \sum_{t=1}^{T} E\left(\dot{g}_t \, | \mathcal{F}_{t-1}\right).$$

  We note that if g_t is smooth, that is if differentiation and integration can be interchanged in differencing $E\left(g_t | \mathcal{F}_{t-1}\right)$ with respect to λ, we have $E\left(\dot{g}_t \, | \mathcal{F}_{t-1}\right) = -C\left(g_t, u_t | \mathcal{F}_{t-1}\right)$. This implies that

$$\overline{g}_T = \langle g, u \rangle_T.$$

Heyde (1997, ch. 12) and Bibby *et al.* (2004, theorem 2.3) provide conditions for consistency and asymptotic normality of $g_{(T)}$, that is conditions under which

$$g_{(T)}/T \xrightarrow{P} 0$$

and

$$\langle g \rangle_T^{-1/2} g_{(T)} \overset{A}{\sim} N(0, I_p).$$

Using a Taylor expansion of $g_{(T)}(\widehat{\lambda}_T) = 0$ in a neighborhood of the true parameter value λ_0, under further regularity conditions (Heyde (1997, p. 27)) we have

$$\langle g \rangle_T^{-1/2} \, \dot{g}_{(T)} \, (\widehat{\lambda}_T - \lambda_0) \overset{A}{\sim} N(0, I_p). \tag{58}$$

Finally, adding additional regularity conditions (Heyde (1997, p. 27)) the quantity

$\dot{g}_T$ in (58) can be replaced by its compensator:

$$\langle g \rangle_T^{-1/2} \overline{g}_T (\widehat{\lambda}_T - \lambda_0) \overset{A}{\sim} N(0, I_p). \tag{59}$$

We call $\mathcal{M}_T^{(2)}$ ($\supseteq \mathcal{M}_T^{(1)} \supseteq \mathcal{G}_T$) the class of martingale EFs $g_{(T)}$ that belong to $\mathcal{M}_T^{(1)}$ and satisfy the additional regularity conditions needed for (59).

The random matrix

$$J\left(g_{(T)}\right) = \overline{g}_T' \langle g \rangle_T^{-1} \overline{g}_T, \tag{60}$$

the inverse of the asymptotic variance of the EF estimator $\widehat{\lambda}_T$, can then be interpreted as an information matrix: it is called *martingale information* (Heyde, 1997, p. 28) or *Heyde information* (Bibby *et al.*, 2004, p. 30).

About this one can note the relations

$$E\left(\overline{g}_T\right) = E\left(\dot{g}_{(T)}\right)$$
$$E\left(\langle g \rangle_T\right) = V\left(g_{(T)}\right)$$

among the entries involved in the Godambe and in the Heyde informations: following Bibby *et al.* (2004), we can think $\overline{g}_T$ and $\langle g \rangle_T$ as estimates, respectively, of the sensitivity matrix $E\left(\dot{g}_{(T)}\right)$ and of the variance $V\left(g_{(T)}\right)$. Hence, in some sense the Heyde information is a stochastic or estimated version of the Godambe information.

This suggests immediately the criterion usually employed to choose optimal EF: "minimize" the variance of the asymptotic distribution of $\widehat{\lambda}_T$ or, conversely, "maximize" the corresponding information. More precisely:

Definition 2 (**A-optimality**) $g_{(T)}^*$ *is* A-optimal *(optimal in the asymptotic sense) into* $\mathcal{M}_T \subseteq \mathcal{M}_T^{(2)}$ *if*

$$J\left(g_{(T)}^*\right) \succeq J\left(g_{(T)}\right)$$

a.s. for all $\lambda \in \Theta$*, for all* $g_{(T)} \in \mathcal{M}_T$ *and for all* $T \in \mathbb{N}$*. An A-optimal estimating function* $g_{(T)}^*$ *is called a* quasi score function*, whereas the corresponding estimator* $\widehat{\lambda}_T$ *is called a* quasi likelihood estimator[17].

The above criterion is hard to apply directly. In practice the following theorem can be employed:

Theorem 3 (Heyde (1997, p. 29)) *If* $g_{(T)}^* \in \mathcal{M}_T \subseteq \mathcal{M}_T^{(2)}$ *satisfies*

$$\overline{g}_T^{-1} \langle g, g^* \rangle_T = \overline{g}_T^{*-1} \langle g^* \rangle_T, \tag{61}$$

for all $\lambda \in \Theta$*, for all* $g_{(T)} \in \mathcal{M}_T$ *and for all* $T \in \mathbb{N}$*, then* $g_{(T)}^*$ *is A-optimal in* $\mathcal{M}_T^{(2)}$*.*

[17] See note 16.

Conversely, if $\mathcal{M}_T$ is closed under addition and $g^*_{(T)}$ is A-optimal in $\mathcal{M}_T$ then (61) holds.

The relation between A-optimality and F-optimality, both restricted to the same family of EFs, is very close. In fact, as showed in Heyde (1997), under some conditions A-optimality implies F-optimality. More precisely:

Theorem 4 (Heyde (1997, p. 29)) *If $\mathcal{M}_T \subseteq \mathcal{M}_T^{(2)}$ is closed under addition and $g^*_{(T)}$ is an A-optimal EF into $\mathcal{M}_T$, then $\overline{g}_T^{*-1}\langle g^* \rangle_T$ non-random for all $T \in \mathbb{N}$ implies $g^*_{(T)}$ also F-optimal into $\mathcal{M}_T$.*

D.6 The Hutton-Nelson quasi score function

As we always work on subsets of $\mathcal{G}_T$, a crucial point in applications is to make a good choice of the family ($\mathcal{H}_T$ in the language of section D.4 or $\mathcal{M}_T$ if we consider section D.5) of the EFs considered. About this, an important family of EF that often can be usefully employed in time series processes is

$$\mathcal{M}_T = \left\{ g_{(T)} \in \mathcal{G}_T : g_{(T)}(\lambda) = \sum_{t=1}^{T} \alpha_t(\lambda) v_t(\lambda) \right\},$$

where $v_t(\lambda)$ are $(p, 1)$ martingale differences and $\alpha_t(\lambda)$ are (p, p) $\mathcal{F}_{t-1}$-measurable functions. Tacking

$$\alpha_t^* = -E\left(\dot{v}_t' \,|\mathcal{F}_{t-1}\right) V\left(v_t|\mathcal{F}_{t-1}\right)^{-1}$$

we can check immediately that condition (61) of theorem 3 is satisfied and then

$$g^*_{(T)} = -\sum_{t=1}^{T} E\left(\dot{v}_t' \,|\mathcal{F}_{t-1}\right) V\left(v_t|\mathcal{F}_{t-1}\right)^{-1} v_t \qquad (62)$$

is A-optimal into $\mathcal{M}_T$. Furthermore, since $\overline{g}_T^{*-1}\langle g^* \rangle_T = -I_p$ is constant, then $g^*_{(T)}$ is also F-optimal into $\mathcal{M}_T$ and its Heyde information is given by

$$J\left(g^*_{(T)}\right) = -\overline{g}_T^* = \langle g_t^* \rangle_T = \sum_{t=1}^{T} E\left(\dot{v}_t' \,|\mathcal{F}_{t-1}\right) V\left(v_t|\mathcal{F}_{t-1}\right)^{-1} E\left(\dot{v}_t \,|\mathcal{F}_{t-1}\right).$$

(62) is known as the *Hutton-Nelson quasi score function* (see Heyde (1997, sect. 2.6) for a deeper discussion).

D.7 Estimating functions in presence of nuisance parameters

Sometimes, the p-dimensional parameter λ that appears in the likelihood can be partitioned as $\lambda = (\boldsymbol{\theta}, \lambda)$, where:

- $\boldsymbol{\theta}$ is a p_1-dimensional parameter of scientific interest called *interest parameter*;
- λ is a p_2-dimensional ($p_2 = p - p_1$) parameter which is not of scientific interest and is called *nuisance parameter.*

In this case the score function for λ can be partitioned as

$$u_{(T)}(\lambda) = \begin{pmatrix} u_{1(T)}(\lambda) \\ u_{2(T)}(\lambda), \end{pmatrix}$$

where $u_{1(T)}(\lambda) = \dfrac{\partial l_T(\lambda)}{\partial \boldsymbol{\theta}}$ and $u_{2(T)}(\lambda) = \dfrac{\partial l_T(\lambda)}{\partial \lambda}$ are *marginal score functions* for $\boldsymbol{\theta}$ and λ respectively.

This structure can be replicated if instead of a score function we consider an estimating function $g_{(T)}(\lambda)$ partitioned as

$$g_{(T)}(\lambda) = \begin{pmatrix} g_{1(T)}(\lambda) \\ g_{2(T)}(\lambda) \end{pmatrix}$$

where $g_{1(T)}(\lambda)$ and $g_{2(T)}(\lambda)$ are the *marginal estimating functions* for $\boldsymbol{\theta}$ and λ respectively. In particular, $g_{1(T)}(\lambda)$ is meant for estimating $\boldsymbol{\theta}$ when λ is known whereas $g_{2(T)}(\lambda)$ is meant for estimating λ when $\boldsymbol{\theta}$ is known.

A clear discussion of inferential issues with nuisance parameters can be found, among others, in Liang and Zeger (1999). The main inferential problem is that in presence of nuisance parameters some properties of estimating functions are no longer valid if we replace parameters with the corresponding estimators. For instance, unbiasedness of $g_{1(T)}(\lambda)$ is not guaranteed if λ is replaced by an estimator $\widehat{\lambda}$. By consequence, optimality properties also are not guaranteed.

An interesting statistical handling of nuisance parameter in the estimating functions framework is provided in Knudsen (1999) and Jørgensen and Knudsen (2004). Their handling parallels in some aspects the notion of *Fisher-orthogonality* or, shortly, *F-orthogonality*, in ML estimation. F-orthogonality is defined by block diagonality of the Fisher information matrix for $\lambda = (\boldsymbol{\theta}, \lambda)$. This particular structure guarantees the following properties of the ML estimator:

1. asymptotic independence of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\lambda}$;
2. efficiency-stable estimation of $\boldsymbol{\theta}$, in the sense that the asymptotic variance for $\boldsymbol{\theta}$ is the same whether λ is treated as known or unknown;

3. simplification of the estimation algorithm;
4. $\widehat{\boldsymbol{\theta}}(\lambda)$, the estimate of $\boldsymbol{\theta}$ when λ is given, varies only slowly with λ.

Among these, Jørgensen and Knudsen (2004) state that efficiency-stable estimation is the crucial property.

Starting from this point, Jørgensen and Knudsen (2004) extend F-orthogonality to EFs introducing the concepts of *nuisance parameter insensitivity*. Even if these authors define and employ these concepts within the F-optimality framework, it can be easily extended to the framework of asymptotic criteria. We discuss this in the following. See the above reference for the original treatment.

We start our exposition partitioning the compensator $\overline{g}_T$ and the quadratic characteristic $\langle g \rangle_T$ of as

$$\overline{g}_T = \begin{pmatrix} \overline{g}_{11T} & \overline{g}_{12T} \\ \overline{g}_{21T} & \overline{g}_{22T} \end{pmatrix} \qquad \langle g \rangle_T = \begin{pmatrix} \langle g \rangle_{11T} & \langle g \rangle_{12T} \\ \langle g \rangle_{21T} & \langle g \rangle_{22T} \end{pmatrix} \tag{63}$$

conformably to the parameters $\boldsymbol{\theta}$ and λ. Using this decomposition, the Heyde information $J(g_{(T)})$ and the asymptotic variance matrix $J(g_{(T)})^{-1}$ have a block structure whose components can be derived from (63).

We say that the marginal EF $g_{1(T)}$ is λ-*insensitive* if

$$\overline{g}_{12(T)} = 0.$$

Using some block matrices algebra, we can show easily that λ-insensitivity of $g_{1(T)}$ is a sufficient condition for

$$[J(g_{(T)})^{-1}]_{11} = \overline{g}'^{-1}_{11T} \langle g \rangle_{11T} \overline{g}^{-1}_{11T}. \tag{64}$$

This implies that under λ-insensitivity, the asymptotic variance of $\boldsymbol{\theta}$ may be calculated based on the sensitivity and the variability of $g_{1(T)}$ only.

Interestingly,

D.8 Algorithms

Let $g_T \in \mathcal{G}$, the class of regular and smooth EF. As anticipated in section D.4, algorithms for computing EF estimates generalizes Fisher-scoring algorithm in the following way

$$\widehat{\lambda}_{i+1} = \widehat{\lambda}_i - E[\dot{g}_T\,(\widehat{\lambda}_i)]^{-1} g_T(\widehat{\lambda}_i),$$

As stated in Knudsen (1999, p. 9), Fisher-scoring (like) algorithms generally behaves better then Newton-scoring (like) algorithms when $\widehat{\lambda}_i$ is far from the solution. Moreover, the sensitivity $E[\dot{g}_T\,(\lambda)]$ is often more easier calculated than $\dot{g}_T\,(\lambda)$ itself.

The previous algorithm can be generalized as

$$\widehat{\lambda}_{i+1} = \widehat{\lambda}_i + s_i W_i g_T(\widehat{\lambda}_i),$$

where $s_i > 0$ is a scalar step length and W_i is a non-random matrix. Often $s_i = 1$ is chosen. However a less crude choice of s_i can performed along the direction $E[\dot{g}_T\ (\widehat{\lambda}_i)]$ until an approximate root of g_T is found.

An obvious stop criterion for the above algorithms is when each component of $g_T(\widehat{\lambda}_i)$ is reasonably close to zero. However Knudsen (1999, p. 10), considering the asymptotic distribution of EF estimator, suggest to complement this criterion checking also that $(\widehat{\lambda}_{i+1} - \widehat{\lambda}_i)' I_{g_T}(\widehat{\lambda}_i)(\widehat{\lambda}_{i+1} - \widehat{\lambda}_i)$ is small.

we note as both statistical theory and algorithms for solving (48) rest on first order expansions of the estimating function like

$$\begin{aligned} g_T(\lambda) &\simeq g_T(\lambda_0) + \dot{g}_T\ (\lambda_0)(\lambda - \lambda_0) \\ &\simeq g_T(\lambda_0) + E[\dot{g}_T\ (\lambda_0)](\lambda - \lambda_0), \end{aligned} \tag{65}$$

where in the final equation $\dot{g}_T\ (\lambda_0)$ is replaced by its expected value (the sensitivity matrix).

If in (65) we interpret λ_0 as an attempt value (say $\widehat{\lambda}_0$) and λ as the updated estimate (say $\widehat{\lambda}$) we obtain, remembering that $g(\widehat{\lambda}) = 0$,

$$\widehat{\lambda} \simeq \widehat{\lambda}_0 - E[\dot{g}_T\ (\widehat{\lambda}_0)]^{-1} g_T(\widehat{\lambda}_0),$$

that generalizes the Fisher-scoring algorithm. We discuss some aspects of algorithms in a later section.

E Mathematical appendix

We prove sufficient conditions for nonnegativity of the components of $\boldsymbol{\mu}_t$. We use the following notation: if $\mathbf{x}$ is a vector or a matrix and a is a scalar, then the symbol $\mathbf{x} \geq 0$ means the corresponding relation for each element of $\mathbf{x}$ and $\mathbf{x}^a$ means that all components of $\mathbf{x}$ are exponentiated by a.

Proposition 1

$$\mathbf{x}' \operatorname{diag}(\mathbf{a})\mathbf{x} + \mathbf{b}'\mathbf{x} + c \geq 0 \qquad \forall \mathbf{x} \geq \mathbf{0}, \tag{66}$$

where $\mathbf{a}$ *and* $\mathbf{b}$ *are* $(n, 1)$ *vectors and* c *is a scalar, if and only if all the following conditions on* $\mathbf{a}$, $\mathbf{b}$ *and* c *are satisfied:*

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. $c - \frac{1}{4}\sum_{i=1}^{n}\frac{b_i^2}{a_i}\,\mathrm{I}(b_i < 0)\,\mathrm{I}(a_i > 0) \geq 0.$

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

Proof:

The minimum of

$$\mathbf{x}'\,\mathrm{diag}(\mathbf{a})\mathbf{x} + \mathbf{b}'\mathbf{x} + c = \sum_{i=1}^{n}(a_i x_i^2 + b_i x_i) + c, \tag{67}$$

with respect to $\mathbf{x}$ is simply the sum of c and the minima of the contributions $(a_i x_i^2 + b_i x_i)$ of each component x_i. Then, considering $x_i \geq 0$:

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

□

As a corollary of the previous proposition we can prove sufficient conditions for nonnegativity of μ_t in the vector MEM.

Corollary 1 *Let*

$$\boldsymbol{\mu}_{it} = \omega_i^* + \boldsymbol{\beta}_{i.}\boldsymbol{\mu}_{t-1} + \boldsymbol{\alpha}_{i.}\mathbf{x}_{t-1} + \boldsymbol{\gamma}_{2i.}\mathbf{x}_{t-1}^{(-)} + \boldsymbol{\gamma}_{1i.}^*\mathbf{x}_{t-1}^{(s)}, \tag{68}$$

the l-th equation that describes the evolution of $\boldsymbol{\mu}_t$ in (30) where ω_i^* is the i-th element of ω^*, the index $i.$ means the i-th row of the corresponding matrix ($i = 1, \ldots, K$) and

- *in the 'enriched' formulation:* $\mathbf{x}_t^{(-)} = \mathbf{x}_t\,\mathrm{I}(r_t < 0)$, $\mathbf{x}_t^{(s)} = \mathbf{x}_t^{1/2}\,\mathrm{sign}(r_t)$;
- *in the 'contagion' formulation:* $\mathbf{x}_t^{(-)} = \mathbf{x}_t \odot \mathrm{I}(\mathbf{r}_t < 0)$, $\mathbf{x}_t^{(s)} = \mathbf{x}_t^{1/2} \odot \mathrm{sign}(\mathbf{r}_{it})$.

Then $\boldsymbol{\mu}_{tl} \geq 0$ for all $\boldsymbol{\mu}_{t-1}$ and $\mathbf{x}_{t-1}$ if all the following conditions are satisfied:

- *for the 'enlarged' formulation:*

  1. $\beta_{ij} \geq 0$, $\alpha_{ij} \geq 0$, $\alpha_{ij} + \gamma_{2ij} \geq 0$ *for all* i;
  2. *if* $\alpha_{ij} = 0$ *then* $\gamma^*_{1ij} \geq 0$; *if* $\alpha_{ij} + \gamma_{2ij} = 0$ *then* $\gamma^*_{1ij} \leq 0$;
  3. $\omega^*_i - \frac{1}{4}\sum_{j=1}^{K} \frac{\gamma^{*2}_{1ij}}{\alpha_{ij}} \mathrm{I}(\alpha_{ij} > 0)\,\mathrm{I}(\gamma^*_{1ij} < 0) \geq 0$ *and*

     $\omega^*_i - \frac{1}{4}\sum_{j=1}^{K} \frac{\gamma^{*2}_{1ij}}{\alpha_{ij} + \gamma_{2ij}} \mathrm{I}(\alpha_{ij} + \gamma_{2ij} > 0)\,\mathrm{I}(\gamma^*_{1ij} > 0) \geq 0.$

- *for the 'contagion' formulation:* $\beta_{ij} \geq 0$, $\alpha_{ij} + \gamma^*_{1ij} \geq 0$, $\alpha_{ij} + \gamma_{2ij} - \gamma^*_{1ij} \geq 0$.

Proof:

Apply proposition 1 with $\mathbf{x} = (\boldsymbol{\mu}^{1/2}_{t-1}, \mathbf{x}^{1/2}_{t-1})$, $c = \omega^*_i$ and:

- in the 'enlarged' formulation: $\mathbf{b} = (0', s\boldsymbol{\gamma}^{*\prime}_{1i.})'$, $\mathbf{a} = (\boldsymbol{\beta}'_{i.}, (\boldsymbol{\alpha}_{i.} + d^- \boldsymbol{\gamma}_{2i.})')'$, $s = \pm 1$ and $d^- = \frac{1-s}{2}$ (s and d^- represent, respectively, the sign and the indicator of negative value of the lagged return);
- in the 'contagion' formulation: $\mathbf{b} = \mathbf{0}$, $\mathbf{a} = (\boldsymbol{\beta}'_{i.}, (\boldsymbol{\alpha}_{i.} + \mathbf{d}^- \odot \boldsymbol{\gamma}_{2i.} + \mathbf{s} \odot \boldsymbol{\gamma}^*_{1i.})')'$, $\mathbf{s} = \{s_i\} = \{\pm 1\}$ and $\mathbf{d}^- = \{d^-_i\} = \left\{\frac{1-s_i}{2}\right\}$ (s_i and d^-_i represent, respectively, the sign and the indicator of negative value of the lagged i-th return).

$\square$

References

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

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2005). Volatility and correlation forecasting. In G. Elliott, C. W. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, page forthcoming. North Holland, Amsterdam.

Barndorff-Nielsen, O. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, **series B 64**, 253–280.

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.

Benjamin, M. A., Rigby, R. A., and Stasinopoulos, M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, **98**(461), 214–223.

Bibby, B. M., Jacobsen, M., and Sørensen, M. (2004). Estimating functions for discretely sampled diffusion-type models. Working paper, University of Copenagen, Department of Applied Mathematics and Statistics.

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.

Brownlees, C. T. and Gallo, G. M. (2005). Ultra high-frequency data management. In C. Provasi, editor, *Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione. Atti del Convegno S.Co. 2005*, pages 431–436. CLUEP, Padova.

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.

Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (1999). Modelling time series of count data. In S. Ghosh, editor, *Asymptotics, Nonparametrics and Time series*, pages 63–113. Marcel Dekker, New York.

Davis, R. A., Dunsmuir, W. T. M., and Strett, S. B. (2002). Observation driven model for poisson counts. Working paper, Department of Statistics, Colorado State University.

Davis, R. A., Dunsmuir, W. T. M., and Strett, S. B. (2003). Maximum likelihood estimation for an observation driven model for poisson counts. Working paper, Department of Statistics, Colorado State University.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**(1), 83–106.

Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling dependence with copulas. Technical report, Department of Mathematics, ETHZ, CH-8092 Zurich Switzerland, `www.math.ethz.ch/finance`.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, **68**(1), 1–22.

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*, **forthcoming**.

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

Engle, R. F., Gallo, G. M., and Velucchi, M. (2005). An mem–based investigation of dominance and interdependence across markets. In *IASC World Congress*. Limassol, Cyprus.

Gallo, G. M. and Velucchi, M. (2005). Ultra-high frequency measures of volatility: An mem-based investigation. In C. Provasi, editor, *Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione. Atti del Convegno S.Co. 2005*, pages 269–274. CLUEP, Padova.

Glosten, L., Jaganathan, R., and Runkle, D. (1993). On the relation between the expected value and volatility of the nominal excess return on stocks. *Journal of Finance*, **48**, 1779–1801.

Heyde, C. (1997). *Quasi-Likelihood and Its Applications*. Springer-Verlag, New York.

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

Jørgensen, B. and Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, **31**, 93–114.

Knudsen, S. J. (1999). Estimating functions and separate inference. Monographs, Vol.1, Department of Statistics and Demography, University of Southern Denmark. Available from http://www.stat.sdu.dk/oldpublications/monographs.html.

Li, W. K. (1994). Time series models based on generalized linear models: Some further results. *Biometrics*, **50**, 506–511.

Liang, K.-Y. and Zeger, S. L. (1999). Inference based on estimating functions in presence of nuisance parameters (with discussion). *Statistical Science*, **10**, 158–199.

Manganelli, S. (2002). Duration, volume and volatility impact of trades. Working paper, European Central Bank, WP ECB 125 http://ssrn.com/abstract=300200.

Shephard, N. (1995). Generalized linear autoregressions. Working paper, Nuffield College, Oxford OX1 1NF, UK.

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