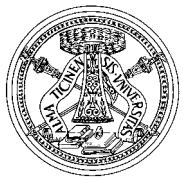

Quaderni di Dipartimento

Model and distribution uncertainty in Multivariate GARCH estimation: a Monte Carlo analysis

Eduardo Rossi
(Università di Pavia)

Filippo Spazzini
(Università di Milano)

208 (11-08)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Novembre 2008

Model and distribution uncertainty in Multivariate GARCH estimation: a Monte Carlo analysis

Eduardo Rossi *
Dipartimento di Economia Politica
e Metodi Quantitativi
Università di Pavia, Italy

F. Spazzini †
Dipartimento di Economia
Università di Milano, Italy

November 26, 2008

Abstract

Multivariate GARCH models are in principle able to accommodate the features of the dynamic conditional correlations processes, although with the drawback, when the number of financial returns series considered increases, that the parameterizations entail too many parameters.In general, the interaction between model parametrization of the second conditional moment and the conditional density of asset returns adopted in the estimation determines the fitting of such models to the observed dynamics of the data. This paper aims to evaluate the interactions between conditional second moment specifications and probability distributions adopted in the likelihood computation, in forecasting volatilities and covolatilities. We measure the relative performances of alternative conditional second moment and probability distributions specifications by means of Monte Carlo simulations, using both statistical and financial forecasting loss functions.

Keywords: Multivariate GARCH models, Model uncertainty, Quasi-maximum likelihood, Monte Carlo methods

*Financial support from PRIN 2006 is gratefully acknowledged. E-mail: `erossi@eco.unipv.it`.
†E-mail: `filippo.spazzini@unimi.it`.

1 Introduction

Predicting the second order moments in assets returns is important for many issues in financial econometrics and in applications like asset pricing, portfolio optimization, risk management, etc. The results on multivariate GARCH models show that this class of multivariate volatility models is able to accommodate the features of dynamic correlation processes. The main drawback is a rapidly increasing number of parameters with the cross-sectional dimension of the dataset considered.

More recently, several models have been proposed that have a fair amount of generality while keeping the number of parameters at a reasonable level. However, simpler conditional covariance models can be highly misspecified, yielding severely distorted forecasts.

Moreover, it is well know that the interaction between the covariance parametrization and the conditional density of asset returns adopted in the quasi-maximum likelihood estimation procedures determines the fitting of such models to the dynamics of the data.

This paper aims to evaluate the interactions, in forecasting volatilities and covolatilities, between conditional second moment specifications and probability distributions adopted in the likelihood computation. We measure, by means of Monte Carlo simulations, the relative performances of alternative conditional second moment and probability distributions specifications using both statistical and financial forecasting loss functions.

In order to compare the competing models, we adopt simulated datasets. The Data Generator is a generalization of the multivariate stochastic volatility model proposed by Danielsson (1998). Our generator is able to generate time-varying dynamic correlations, leverage effects within and between assets, volatility clustering and asset specific kurtosis and skewness. The Monte Carlo analysis is motivated by the fact that the precision of the *ex-post* analysis is affected by the observability of the true volatilities which are indeed unobservable on the markets. As for the choice of the Data Ganerating Process, we choose a process external to the class of models compared in our analysis for two sets of reasons: first, multivariate stochastic volatility models are coherent with the hypothesis that asset prices in a no-arbitrage and frictionless market follow semi-martingale processes (see e.g. Shephard and Andersen (2008)); second, considering as DGP a structure which is external to the class of competing models allows us to fairly compare the alternative forecasts. We limit our analysis to bivariate models, hence our Monte Carlo experiment does not take into account the problems connected with the curse of dimensionality and feasibility considerations in general.

More in details, the results of our Monte Carlo experiment illustrate the effect that the choices of the conditional covariance process, assumed to be multivariate GARCH, and of the log-likelihood used in the estimation (both misspecified by construction) have on the prediction performances, as measured by statistical loss functions and Value-at-risk forecasting.

The models considered are the Dynamic Conditional Correlations (DCC) Model, the Asymmetric DCC-EGARCH, and the BEKK parametrizations using quasi maximum likelihood with Normal, Student's t Copula, multivariate Laplace, multivariate t and the recently introduced Multiple Degrees of Freedom Student's t distribution (Serban et al., 2007).

The forecasts are evaluated through several statistical criteria, among which we also consider the results from the Mincer-Zarnowitz regressions and the related hypothesis testing procedures.

Finally, to rank the models with respect to their asset allocation performances we forecast the Value-at-Risk (VaR) metric at 99% over an holding period of 1 day, and we

evaluate these forecasts by means of the unconditional coverage rate and of the Dynamic Quantiles (DQ) test (Engle and Manganelli (1999)).

First, for what concerns the results, our exercise shows that explicitly modeling the leverage effects present in the data strongly improves the forecasting ability only if the leptokurtosis is correctly accounted for.

While both DCC specifications, with Gaussian disturbances, lead to strong underestimation of the VaR, assuming DCC specifications and allowing for leptokurtic error distributions leads to a sensible enhancement in the risk measurement performances of these models. BEKK models with Normal distributions completely fail to describe the tail behavior of our data set, while assuming Laplace and multivariate t distributions for the innovations leads to a strong overestimation of the VaR. Finally, for the BEKK models coupled with the Multiple Degree of Freedom t, the mean fraction of VaR violations is the best match of the correct one and, the number of DQ rejections is the lowest among all the models considered.

The paper is structured as follows. The models are introduced in Section 2, while Section 3 reports the log-likelihoods for each model and density function considered. The Data Generating Process and the evaluation measures are introduced in Section 4; Section 5 reports the results of the Monte Carlo experiments. Finally, Section 6 concludes the paper.

2 The models

We suppose that the N-vector of asset returns r_t has a conditional distribution:

$$r_t|\Im_{t-1} \sim D(0, \Omega_t) \quad t = 1, \ldots, T \tag{1}$$

where D is a continuous distribution and $\Im_{t-1}$ is the information set at time $t-1$ ($\mathbb{E}_{t-1}[\cdot] \equiv \mathbb{E}[\cdot|\Im_{t-1}]$, hereafter). Ω_t is the conditional variance-covariance matrix, assumed to be time-varying.
We consider the following Multivariate GARCH (MGARCH, hereafter) parameterizations of Ω_t:

- The Dynamic Conditional Correlation GARCH (DCC)
- The Asymmetric DCC EGARCH ($ADCC$)
- The Scalar BEKK ($SBEKK$)
- The Diagonal BEKK ($DBEKK$)

We estimate each of the above models using different densities for the likelihood function. The assumed processes for the conditional covariances are misspecified by construction, as explained in Section 4.

2.1 Dynamic Conditional Correlation Models

2.1.1 DCC-GARCH

In the *Dynamic Conditional Correlation* GARCH, introduced by Engle (2002), as an extension of Bollerslev's Constant Conditional Correlation Model (Bollerslev (1990)), the conditional variance-covariance matrix is written as:

$$\Omega_t = H_t R_t H_t \quad t = 1, \ldots, T$$

where R_t is the conditional correlation matrix, and the standardized returns are defined as:

$$\epsilon_t = H_t^{-1} r_t \quad t = 1, \ldots, T$$

where $H_t = diag\{\sigma_{1t}, \ldots, \sigma_{Nt}\}$ with

$$\sigma_{it} = Var_{t-1}[r_{it}]^{1/2} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$

where the σ_{it}^2 are modeled as univariate GARCH processes. The conditional variance-covariance matrix of ϵ_t is the conditional correlation matrix of the asset returns,

$$\mathbb{E}_{t-1}(\epsilon_t \epsilon_t') = H_t^{-1} \Omega_t H_t^{-1} = R_t = \{\rho_{ij,t}\} \quad t = 1, \ldots, T.$$

The conditional correlation can be expressed as a function of the conditional covariance, $q_{ijt} = \mathbb{E}_{t-1}[\epsilon_{it}\epsilon_{jt}]$, and of the conditional variances, $q_{iit} = \mathbb{E}_{t-1}[\epsilon_{it}^2]$ that is

$$\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}.$$

The conditional covariances and variances are modeled with a GARCH(1,1) model:

$$q_{ij,t} = \overline{\rho}_{ij} + \alpha(\epsilon_{i,t-1}\epsilon_{j,t-1} - \overline{\rho}_{ij}) + \beta(q_{ij,t-1} - \overline{\rho}_{ij}) \tag{2}$$

where $\overline{\rho}_{ij} = Corr(\epsilon_{it}\epsilon_{jt})$. The unconditional variance of $\epsilon_{i,t}$ is $\overline{\rho}_{ii} = 1$. Finally, it is assumed that

$$\overline{\rho}_{ij} \cong \overline{q}_{ij}.$$

The conditional covariance matrix of ϵ_t, Q_t, is positive definite as long as it is a weighted average of positive definite and semidefinite matrices. In matrix form:

$$Q_t = (1 - A - B)\overline{Q} + A(\epsilon_{t-1}\epsilon_{t-1}') + B(Q_{t-1}) \tag{3}$$

where $\overline{Q}$ is the unconditional covariance matrix of ϵ_t, and A and B are two scalars. The DCC model specification is:

$$\begin{aligned}
& r_t | \Im_{t-1} \sim D(0, \Omega_t) \\
& \epsilon_t = H_t^{-1} r_t \\
& \sigma_{it}^2 = \omega_i + \alpha_i r_{it-1}^2 + \beta_i \sigma_{it-1}^2 && (4) \\
& Q_t = (1 - A - B)\overline{Q} + A(\epsilon_{t-1}\epsilon_{t-1}') + B(Q_{t-1}) && (5) \\
& R_t = \mathrm{diag}\{Q_t\}^{-1/2} Q_t \, \mathrm{diag}\{Q_t\}^{-1/2}. && (6)
\end{aligned}$$

The unconditional correlation matrix $\overline{Q}$ is estimated by the sample correlation matrix $S = \frac{1}{T}\sum_t \widehat{\epsilon}_t \widehat{\epsilon}_t'$.

2.1.2 Asymmetric DCC-EGARCH

The DCC can be extended to include an asymmetric term. The asymmetric DCC specification, with conditional variances modeled as an EGARCH(1,0), is:

$$\begin{aligned}
& \epsilon_t = H_t^{-1} r_t \\
& \sigma_{it}^2 = \exp\left(\omega_i + \alpha_i \epsilon_{it-1} + \delta_i \left(|\epsilon_{it-1}| - E\left[\epsilon_{it-1}\right]\right) + \beta_i \log(\sigma_{it-1}^2)\right) && (7) \\
& Q_t = (\overline{Q} - A\overline{Q} - B\overline{Q} - C\overline{N}) + A(\epsilon_{t-1}\epsilon_{t-1}') + B(Q_{t-1}) + C(\eta_{t-1}\eta_{t-1}') && (8) \\
& R_t = \mathrm{diag}\{Q_t\}^{-1/2} Q_t \, \mathrm{diag}\{Q_t\}^{-1/2}. && (9)
\end{aligned}$$

$$\eta_t = I(\epsilon_t < 0) \odot \epsilon_t$$

($\odot$ is the Hadamard product) with A, B and C scalars and

$$\overline{N} = \mathbb{E}\left[\eta_t \eta_t'\right].$$

The unconditional covariance of negative shocks $\overline{N}$ is estimated by

$$\frac{1}{T}\sum_{t=1}^{T} \widehat{\eta}_t \widehat{\eta}_t'.$$

2.2 BEKK parameterization of MGARCH

2.2.1 Diagonal BEKK

The Multivariate GARCH(1,1) model can be parameterized as follows:

$$\Omega_t = CC' + Ar_{t-1}r_{t-1}'A' + B\Omega_{t-1}B' \tag{10}$$

where this parameterization, called BEKK (from the authors: Baba Engle Kraft and Kroner), guarantees the positive definiteness of the conditional variance matrix Ω_t (see Engle and Kroner (1995)). An obvious restriction is to assume that the matrices A and B are diagonal.

2.2.2 Scalar BEKK

We can further restrict the matrices A and B to be scalars, a and b, so that

$$\Omega_t = CC' + a^2 r_{t-1}r_{t-1}' + b^2\Omega_{t-1} \tag{11}$$

in this way the only difference among the individual conditional variances and covariances lies in the intercept term.

3 The likelihoods

3.1 DCC - Normal log-likelihood

When the conditional density of r_t, $D(0, \Omega_t)$, is a Multivariate Normal density, i.e. $N(0, \Omega_t)$

$$f(r_t; 0, \Omega_t) = \frac{1}{((2\pi)^N |\Omega_t|)^{1/2}} \exp\left(-\frac{(r_t'\Omega_t^{-1}r_t)}{2}\right)$$

the log-likelihood of the DCC-GARCH model can be expressed as:

$$L_T(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left(N\log(2\pi) + \log|H_t|^2 + r_t'H_t^{-2}r_t + \epsilon_t'R_t^{-1}\epsilon_t - \epsilon_t'\epsilon_t + \log|R_t|\right)$$

The parameter vector θ includes the parameters in the conditional variance processes σ_{it}^2, ψ, as well as ϕ, the parameters in the correlation processes R_t. The log-likelihood function can be written as the sum of two components. The first is the *Volatility* component:

$$\begin{aligned}\mathcal{L}_V(\psi) \equiv L_{V,T}(\psi) &= -\frac{1}{2}\sum_{t=1}^{T}(N\log(2\pi) + \log|H_t|^2 + r_t'H_t^{-2}r_t) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\log(2\pi) + \log(\sigma_{it}^2) + \frac{r_{it}^2}{\sigma_{it}^2}\right) \end{aligned} \tag{12}$$

The second is the *Correlation* component:

$$\mathcal{L}_C(\psi,\phi) \equiv L_{C,T}(\psi,\phi) = -\frac{1}{2}\sum_{t=1}^{T}(\epsilon_t' R_t^{-1}\epsilon_t - \epsilon_t'\epsilon_t + \log|R_t|) \tag{13}$$

Once we have estimated the returns conditional variances σ_{it}^2 we can compute the standardized residuals $\widehat{\epsilon}_{it} = r_{it}/\widehat{\sigma}_{it}$ and employ them in the maximization of $\mathcal{L}_C$.

3.2 DCC - Laplace log-likelihood

An alternative assumption for the conditional distribution of the returns is represented by the asymmetric multivariate Laplace (see Cajigas and Urga (2007)). The conditional asymmetric Laplace multivariate density of the vector r_t is given by:

$$f(r_t; m, \Omega_t) = \frac{2\exp(r_t'\Omega_t^{-1}m)}{(2\pi)^{N/2}|\Omega_t|^{1/2}}\left(\frac{r_t'\Omega_t^{-1}r_t}{2+m'\Omega_t^{-1}m}\right)^{\nu/2} K_\nu\left(\sqrt{(2+m'\Omega_t^{-1}m)(r_t'\Omega_t^{-1}r_t)}\right),$$

where $\nu = (2-d)/2$ and $K_\nu(u)$ is the modified Bessel function of the second kind (see the Appendix for details). The log-likelihood of DCC models is

$$\begin{aligned} L_T(\theta) &= \sum_{t=1}^{T}\left\{r_t'\Omega_t^{-1}m - \frac{1}{2}\log|\Omega_t| + \frac{\nu}{2}\left(\log(r_t'\Omega_t^{-1}r_t) - 2\log(2+m'\Omega_t^{-1}m)\right)\right\} \\ &\quad + \sum_{t=1}^{T}\left\{\log K_\nu\left(\sqrt{(2+m'\Omega_t^{-1}m)(r_t'\Omega_t^{-1}r_t)}\right)\right\} \end{aligned}$$

then replacing $\Omega_t = H_t R_t H_t$

$$\begin{aligned} L_T(\theta) =& \sum_{t=1}^{T}\left\{r_t'(H_tR_tH_t)^{-1}m - \frac{1}{2}\log|H_tR_tH_t| + \frac{\nu}{2}\left(\log(r_t'(H_tR_tH_t)^{-1}r_t)\right)\right\} \\ &+ \sum_{t=1}^{T}\left\{-\frac{\nu}{2}\log(2+m'(H_tR_tH_t)^{-1}m)\right. \\ &\left.+\log K_\nu\left(\sqrt{(2+m'(H_tR_tH_t)^{-1}m)(r_t'(H_tR_tH_t)^{-1}r_t)}\right)\right\}. \end{aligned} \tag{14}$$

We assume that $m = 0$ (which corresponds to the symmetric stable Laplace distribution). We can now proceed with the maximization of (14) w.r.t. θ.

3.3 DCC - Copula log-likelihood

The multivariate density of r_t is in this case constructed using copulae. There are many types of parametric copulae that can be used to model different dependency relationship. In this paper we adopt a conditional multivariate distribution where each of the conditional univariate margins follows a t distribution linked via a Student's t Copula (see Serban et al. (2007) for a similar application). The Student copula function is defined as

$$C_{R_t,\nu}(u_1,\ldots,u_N) = T_{R_t,\nu}(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_N))$$

where $T_{R_t,\nu}$ is the *c.d.f.* of the multivariate t with conditional correlation matrix R_t and common degree of freedom parameter ν, with t_ν^{-1} being the inverse of the univariate

Student's t distribution. The density function for the Student copula is

$$c(u_1,\ldots,u_N|R_t,\nu_C) = |R_t|^{-1/2}\frac{\Gamma\left(\frac{\nu_C+N}{2}\right)\left[\Gamma\left(\frac{\nu_C}{2}\right)\right]^N\left(1+\frac{1}{\nu_C}\zeta' R_t^{-1}\zeta\right)^{-(\nu_C+k)/2}}{\left[\Gamma\left(\frac{\nu_C+1}{2}\right)\right]^N\Gamma\left(\frac{\nu_C}{2}\right)\prod_{i=1}^N\left(1+\frac{\zeta_i^2}{\nu_C}\right)^{-(\nu_C+1)/2}} \tag{15}$$

where $\zeta_i = t_{\nu_C}^{-1}(u_i)$.

The marginal model for each return series is:

$$r_{it} = \sqrt{\frac{\nu_i-2}{\nu_i}}\sigma_{it}y_{it}$$

$y_{it} \sim i.i.d.t(0,1,\nu_i),\ i=1,\ldots,N$ are i.i.d. Student's t with ν_i degrees of freedom.

$$f_i(r_{it};\nu_i) = \frac{\Gamma((\nu_i+1)/2)}{\sqrt{(\nu_i-2)\pi}\Gamma(\nu_i/2)}\left(1+\frac{r_{it}^2}{(\nu_i-2)\sigma_{it}^2}\right)^{-(\nu_i+1)/2}\sigma_{it}^{-1} \quad \nu_i > 2$$

The σ_{it} are modeled as in Eq. (4) or Eq. (7). The component of the log-likelihood for the t-th observation corresponding to the marginal densities is:

$$\begin{aligned} l_{m,t}(\psi) = \sum_{i=1}^N \Bigg\{ &\log\Gamma\left(\frac{\nu_i+1}{2}\right) - \log\Gamma\left(\frac{\nu_i}{2}\right) - \frac{1}{2}\log\pi - \frac{1}{2}\log(\nu_i-2) \\ &-\frac{\nu_i+1}{2}\log\left(1+\frac{r_{it}^2}{(\nu_i-2)\sigma_{it}^2}\right) - \frac{1}{2}\log\sigma_{it}^2\Bigg\} \end{aligned} \tag{16}$$

while that from the Student copula is:

$$\begin{aligned} l_{c,t}(\phi) = &-\frac{1}{2}\log|R_t| + \log\Gamma\left(\frac{\nu_C+N}{2}\right) - \frac{\nu_C+N}{2}\log\left(1+\frac{1}{\nu_C}\zeta' R_t^{-1}\zeta\right) \\ &+ (N-1)\log\left[\Gamma\left(\frac{\nu_C}{2}\right)\right] - N\log\left[\Gamma\left(\frac{\nu_C+1}{2}\right)\right] - \frac{\nu_C+1}{2}\sum_{i=1}^N\log\left(1+\frac{\zeta_i^2}{\nu_C}\right) \end{aligned} \tag{17}$$

The correlation matrix R_t in the Student copula (15) is modeled as in Eq. (5)-(6) or as in Eq. (8)-(9).

Thus the log-likelihood function is obtained as the sum of the two parts:

$$l_t(\psi,\phi) = l_{m,t}(\psi) + l_{c,t}(\phi)$$

and it is sequentially maximized. We firstly maximize the marginal components with respect to the vector ψ, which includes the parameters of the processes σ_{it}^2 and the degree of freedom parameters ν_i, different in each marginal density, and secondly the copula part, with respect to the parameters in ϕ, that is ν_C and those in the processes of R_t.

A drawback of the proposed copula model is that the copula correlation is not exactly the correlation between returns. Instead, it is the correlation between the transformed returns. In other words, we can see in Eq. (15) that the arguments of the Student copula are the transformations $\zeta_i = t_\nu^{-1}(u_i)$. Thus, R_t is the correlation matrix between these transformed marginals. The actual correlations between returns can be obtained using simulations as in Serban et al. (2007), even if the approximation error is negligible, at least in the bivariate case.

3.4 BEKK - Normal log-likelihood

Let $\theta = (vec(C)', vec(A)', vec(B)')'$ be the parameter vector of the diagonal BEKK. The normal log-likelihood in this case is:

$$l_t(\theta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\Omega_t| - \frac{1}{2}\epsilon_t'\Omega_t^{-1}\epsilon_t. \tag{18}$$

3.5 BEKK - Laplace log-likelihood

The BEKK Symmetric Multivariate Laplace log-likelihood for observation t is:

$$l_t(\theta) = -\frac{1}{2}\log|\Omega_t| + \frac{\nu}{2}\left(\log(r_t'\Omega_t^{-1}r_t) - 2\log(2)\right) + \log K_\nu\left(\sqrt{2(r_t'\Omega_t^{-1}r_t)}\right)$$

3.6 BEKK - Multivariate Student's t log-likelihood

A natural alternative to the multivariate Gaussian density is the Student density, see Harvey et al. (1992) and Fiorentini et al. (2003). The latter has an extra scalar parameter, the degrees of freedom parameter, denoted ν hereafter. When this parameter tends to infinity, the Student density tends to the normal density. When it tends to zero, the tails of the density become thicker and thicker. The value of ν indicates the order of existence of the moments: e.g. if $\nu = 2$, the second-order moments do not exist, but the first-order moments exist. For this reason, it is convenient (although not necessary) to assume that $\nu > 2$, so that Ω_t is always interpretable as a conditional covariance matrix. The multivariate Student's t density for r_t is

$$f\left(r_t; 0, \frac{\nu-2}{\nu}\Omega_t, \nu\right) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)[\pi(\nu-2)]^{N/2}}|\Omega_t|^{-1/2}\left[1 + \frac{r_t'\Omega_t^{-1}r_t}{\nu-2}\right]^{-\frac{N+\nu}{2}}$$

then the log-likelihood function for observation t results to be:

$$\begin{aligned} l_t(\theta) = &\log\Gamma\left(\frac{\nu+N}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{N}{2}\log\pi - \frac{N}{2}\log(\nu-2) \\ &- \frac{N+\nu}{2}\log\left(1 + \frac{r_t'\Omega_t^{-1}r_t}{\nu-2}\right) - \frac{1}{2}\log|\Omega_t|. \end{aligned}$$

3.7 BEKK - Multiple Degrees of Freedom Student's t log-likelihood

To allow different levels of heavy tailedness for different error terms we adopt the Multiple Degrees of Freedom t (MDFt, hereafter) introduced by Serban et al. (2007). Let the returns be generated as:

$$r_t = \Omega_t^{1/2} z_t \quad t = 1, \ldots, T$$

with

$$z_{it} = \sqrt{\frac{\nu_j - 2}{\nu_j}}\, y_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$

where $y_{it} \sim i.i.d.t(0, 1, \nu_i), i = 1, \ldots, N$

$$f(y_{it}; 0, 1, \nu_i) = \frac{\Gamma((\nu_i+1)/2)}{\sqrt{(\nu_i\pi)}\Gamma(\nu_i/2)}\left(1 + \frac{y_{it}^2}{\nu_i}\right)^{-(\nu_i+1)/2} \qquad \nu_i > 2$$

then z_{it} is a standardized student's t random variable with density:

$$g\left(z_{it};0,\frac{\nu_i-2}{\nu_i},\nu_i\right)=\frac{\Gamma((\nu_i+1)/2)}{\sqrt{(\nu_i-2)\pi}\Gamma(\nu_i/2)}\left(1+\frac{z_{it}^2}{\nu_i-2}\right)^{-(\nu_i+1)/2}\qquad \nu_i>2$$

Given the independence assumption of y_{it} the joint density of $z_t=(z_{1t},\ldots,z_{Nt})'$ is given by

$$G(z_t;\nu)=\prod_{j=1}^{N}g\left(z_{it};0,\frac{\nu_i-2}{\nu_i},\nu_i\right)$$

with $\nu=(\nu_1,\ldots,\nu_N)'$, therefore the conditional density of r_t is:

$$D(r_t;\theta)=G(\Omega_t^{-1/2}r_t;\nu)|\det\Omega_t^{-1/2}|$$

with

$$G(\Omega_t^{-1/2}r_t;\nu)=\prod_{j=1}^{N}\sqrt{\frac{\nu_j-2}{\nu_j}}\prod_{j=1}^{N}f\left(\frac{u_j'\Omega_t^{-1/2}r_t}{\sqrt{(\nu_j-2)/\nu_j}};\nu_j\right),$$

or alternatively

$$G(\Omega_t^{-1/2}r_t;\nu)=\prod_{j=1}^{N}\sqrt{\frac{\nu_j-2}{\nu_j}}\prod_{j=1}^{N}g\left(u_j'\Omega_t^{-1/2}r_t;\nu_j\right),$$

with $u_j=(0,\ldots,1,\ldots,0)'$ where 1 is in the j-th position. Finally the t-th contribution to the log-likelihood function is

$$l_t(\theta)=\sum_{j=1}^{K}\log\sqrt{\frac{\nu_j-2}{\nu_j}}+\sum_{j=1}^{K}\log f\left(\frac{u_i'\Omega_t^{-1/2}r_t}{\sqrt{(\nu_j-2)/\nu_j}};\nu_j\right)-\frac{1}{2}\log|\Omega_t|.$$

4 Data Generating Process and Monte Carlo setup

In order to obtain simulated data which are able to mimic as closely as possible the stylized facts observable in real financial time series at daily frequency, we decide to implement a Data Generating Process simulator as in Danielsson (1998). Despite the fact that this Multivariate Stochastic Volatility model is able to accomodate volatility clustering and leverage effects both within and across assets, we further generalize the model in order to introduce different levels of excess kurtosis in the shocks of both the returns and the volatilities. This has been achieved coupling univariate Student's t distributions with $\nu_j,\ j=1,\ldots,4$ degrees of freedom for each innovation with a Student's t copula with ν_C degrees of freedom. In addition, we relaxed the original assumption of constant correlation between the innovations of the returns, allowing this correlation to evolve through time following a function of the cosine of the time index.

In particular, we assume that the bivariate Data Generating Process is defined as:

$$y_t=\Psi_t\epsilon_t \tag{19}$$

where $\Psi_t'\Psi_t=\Omega_t$ is the 2×2 covariance matrix and ϵ_t is a 2×1 vector of independent shocks to returns. The covariance matrix can be further decomposed as follows:

$$\Omega_t=H_tR_tH_t \tag{20}$$

Table 1: Parameters of the Data Generating Process

Univariate volatility parameters					
ω_1	δ_1	γ_1	ω_2	δ_2	γ_2
0.01	0.96	0.23	-0.04	0.95	0.34
Degrees of freedom					
ν_1	ν_2	ν_3	ν_4	ν_C	
7	12	5	10	8	
Leverage parameters and volatility innovations correlation					
L_{11}	L_{22}	L_{12}	ς		
-0.4	-0.3	-0.1	0.32		

where H_t is a diagonal matrix containing the square root of the univariate volatilities, and R_t is the time varying correlation matrix.

The volatility of each asset is modeled via the univariate stochastic volatility process

$$h_{t,i} = \exp[\omega_i + \delta_i \log(h_{t-1,i}) + \gamma_i \eta_{t,i}], \quad i = 1, 2 \; ; \tag{21}$$

while for the dynamic correlation matrix we adopt the specification shown in Eq.(22), which allows the correlation to oscillate between the maximum of 0.7 and the minimum of 0.3 within a period of 1000 observations:

$$\begin{aligned} R_{12,t} &= R_{21,t} = \rho_t \\ \rho_t &= 0.5 + 0.2 \cos\left(\frac{2\pi t}{1000}\right) \; . \end{aligned} \tag{22}$$

As far as the volatility shocks η_t are concerned, we impose unit variance and fixed covariance ς to these shocks. Finally, in order to introduce both within and across assets leverage effects, we impose that the covariance matrix of returns and volatilities shocks $L = \mathbb{E}(\epsilon_t \eta_t') \neq 0$. The diagonal elements of L contain the within asset leverage effects, while the non-diagonal elements induce the across assets leverage.

The values of the parameters chosen for our Monte Carlo experiment are shown in Table 1. Choosing these values for the parameters of the DGP, we are able to closely match the descriptive statistics of the NASDAQ 100 and of the S&P 500 indexes.

4.1 Copula simulation and Monte Carlo setup

Obtaining innovations with a dependence structure described by an elliptical copula is straightforward, even if the computational burden is almost fifteen times bigger than that associated to the simulation of a multivariate normal distribution.

Firstly, we need to draw the innovations from a four-variate t distribution with ν_C degrees of freedom and correlation matrix given by

$$\mathcal{C}_{\mathcal{C}} = \left(\begin{array}{c:cc} R_t & L \\ \hdashline L & \begin{matrix} 1 & \varsigma \\ \varsigma & 1 \end{matrix} \end{array} \right) .$$

There is statistical dependence between these four variables, and each has a Student's t marginal distribution.

Next, we apply the Student's t cumulative distribution function separately to each variable, changing the marginal distributions into Uniform distributions over the interval $(0, 1)$. At this point the statistical dependence is that proper of the multivariate t distribution, but the marginal distributions are Uniforms over the interval $(0, 1)$: we obtained a sample with a t Copula distribution with ν_C degrees of freedom and correlation matrix given by $\mathcal{C}_\mathcal{C}$. Finally, we exploit the fact that applying the inverse *c.d.f.* of any distribution F to a $U(0, 1)$ random variable results in a random variable whose distribution is exactly F; hence, we apply to each variable inverse t *c.d.f.* with d.o.f $\nu_1 \dots \nu_4$ and this concludes our simulation algorithm.

Using the Danielsson model as Data Generating Process, we generate 100 independent datasets with 3000 observations each, discarding the first 1000 observations of each sample to minimize the initialization bias[1].

A typical dataset obtained with this simulation algorithm is depicted in the left panel of Figure 1, as well as the corresponding univariate volatilities. Table 2 reports summary statistics for the same dataset.

Table 2: Descriptive statistics of a sample obtained with the Danielsson Generator

Statistic	Series 1	Series 2
Mean	0.027	-0.016
Variance	1.215	0.602
Max	10.989	4.780
Min	-4.783	-4.006
Skewness	-0.810	0.273
Kurtosis	10.365	7.710
p-Normal	$< 1\%$	$< 1\%$

The first 1500 observations of each sample are used to estimate the 14 competing models. The in-sample performances of the models are compared on the basis of the mean Log-Likelihood, the mean Bayesian Information Criterion, the Mean Absolute Error of the estimated correlation with respect to the true correlation and the mean distance between the estimated conditional variance matrix and the true variance matrix, using the Frobenius norm as decision criterion. These quantities are shown in Table 3.

4.2 Out-of-sample evaluation criteria

In order to assess the out-of-sample performances of the competing models, we use the last 500 observations of each sample to compare the one-step-ahead forecasting capabilities of the models from both statistical and financial points of view. As statistical performance criteria, we compute the Mean Absolute Forecasting Error for each model m over the 100 Monte Carlo replications:

[1]We decided to limit the number of simulations to 100 in order to contain the machine time of the experiment to about 50 hours. All the routines were developed in $Matlab^{©}$ and the experiment run on a PC equipped with an $Intel^{©}$ Q6600 CPU and 4GB of RAM.

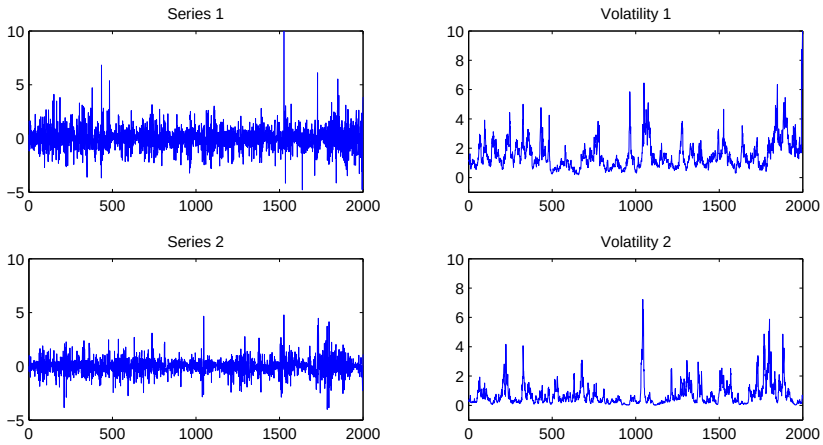

Figure 1: Typical bivariate dataset obtained with the Multivariate Stochastic Volatility Generator with parameters chosen as in Table 1 *(Left Panel)* and corresponding volatilities *(Right Panel)*

$$MAFE(m) = \frac{1}{N}\sum_{i=1}^{N}|(e_{m,T+i+1})| \tag{23}$$

where $N = 500$, $T = 1500$ and $e_{m,i}$ is the forecasting error at time i, computed as the difference between the forecasted volatility and the actual volatility. This error statistic presents the shortcoming that the underlying loss function is symmetric, since over-predictions and under-predictions have the same weight. This is not a realistic assumption from a practical point of view. In fact, for example, an under-prediction of stock price volatility will lead to a downward biased estimate of the call option price, given the positive relation between the volatility of underlying stock prices and call option prices. This under-estimate of the price is more likely to be of greater concern to a seller than to a buyer; the reverse is clearly true for over-predictions of stock price volatility. Having said that, as additional performance measures, we include in our analysis the Mean Mixed Error of Underprediction (MMEU) and the Mean Mixed Error of Overprediction (MMEO), as suggested by Brailsford and Faff (1996). The MMEU can be defined as:

$$MMEU(m) = \frac{1}{N}\left(\sum_{i=1}^{N_U}\sqrt{|(e_{m,T+i+1})|} + \sum_{i=1}^{N_O}|(e_{m,T+i+1})|\right) \tag{24}$$

while the MMEO as:

$$MMEO(m) = \frac{1}{N}\left(\sum_{i=1}^{N_O}\sqrt{|(e_{m,T+i+1})|} + \sum_{i=1}^{N_U}|(e_{m,T+i+1})|\right) \tag{25}$$

where N_U is the number of times that the predicted conditional variance is smaller than the actual one and N_O its complementary. While computing these statistics, when the absolute value of the forecast error is larger than unity, we square the error in order to achieve the desired penalty. Finally, a 'biased' forecast model can be viewed as one which systematically over- or under-predicts, whereas an 'unbiased' forecast model, when not providing a perfect forecast, should over-predict 50 percent of the time and under-predict 50 percent of the time.

The last statistic evaluation criterion presented in this work entails the classical Mincer-Zarnowitz regression, which involves the regression of the realization of a variable on its forecast; the forecast is optimal if the null hypothesis of zero intercept and unity slope coefficient cannot be rejected. In our simulation study, we perform this test simply by regressing the true volatilities on the forecasted volatilities[2]. The univariate Mincer-Zarnowitz regression takes the form:

$$\sigma_{i,T+t} = \alpha + \beta\hat{\sigma}_{i,T+t} + \epsilon_{T+t} \tag{26}$$

where $i = 1, 2$ indicates the appropriate volatility, $T = 1500$ is our in-sample threshold, $t = 1, \ldots, 500$ selects the appropriate forecast and observation while $\sigma_{i,T+t}$ and $\hat{\sigma}_{i,T+t}$ are respectively the true variance and its forecast at time $T + t$. The covariance matrix of the estimated regression parameters is constructed using the Newey and West estimator, in order to control for heteroscedasticity and autocorrelation in the residuals. If the null hypothesis $\alpha = 0 \cap \beta = 1$ cannot be rejected by a Wald test, the forecast can be considered optimal with respect to the relevant filtration.

Furthermore, we implement a multivariate version of the Mincer-Zarnowitz test; in this latter case, we adopt a slightly different approach from that suggested in Patton and Sheppard (2008), since we regress the diagonal of the actual covariance matrix on the diagonal of the forecasted covariance matrix using a Seemingly Unrelated Regression Equations approach, rather than a panel estimator as suggested in Patton and Sheppard (2008). Hence in our framework the Multivariate Mincer-Zarnowitz regression takes the form:

$$diag(\Omega_{T+t}) = \alpha + \beta \odot diag(\hat{\Omega}_{T+t}) + \epsilon_{T+t} \tag{27}$$

where the *diag* operator stacks the diagonal elements of a matrix in a column vector and $\odot$ indicates the Hadamard product; α and β are in this case 2×1 vectors. As an indicator of goodness of fit in these SURE regressions we adopt the McElroy systemwide measure:

$$R^2 = 1 - \frac{\hat{\epsilon}'\hat{\Sigma}^{-1}\hat{\epsilon}}{y'\left(\hat{S}^{-1} \otimes \left(I - \frac{\iota\iota'}{N}\right)\right)y} \tag{28}$$

where N is the number of observations in each equation, I is an $N \times N$ identity matrix, ι is a $N \times 1$ vector of ones and $y = vec(\{diag(\Omega_{T+t})'\}_{t=1,\ldots,500})$.

After the comparison between the forecast capabilities of these models measured with statistical indicators, we now focus on an exercise of asset allocation in order to inspect the forecasting abilities of these models when exploited in a financial context.

In order to do so, we forecast the Value at Risk at 99% over an holding period of 1 day using the last 500 observations of each sample as back-testing data, considering an equally weighted portfolio and a rolling window of 1500 observations. This exercise is then repeated for every Monte Carlo iteration.

The vectors of Value at Risk violations and forecasts allow us to perform the Dynamic Quantile test proposed by Engle and Manganelli (1999) to assess the accuracy of these forecasts. Let us define

$$HIT_t \equiv I(r_{p,t} < VaR_t) - \theta$$

where θ is the percentile of interest in the probability distribution of the portfolio returns $r_{p,t}$[3], VaR_t is the VaR forecast for day t, while $I(x)$ is an indicator function that assumes

[2]Since our datasets are simulated, we do not need to rely on volatility proxies as in Patton and Sheppard (2008).

[3]The first percentile when the VaR threshold is 99%.

value 1 when x is true and 0 otherwise. The HIT function assumes value $1-\theta$ every time $r_{p,t} < VaR_t$ (i.e., every time a violation is realized) and $-\theta$ otherwise. Clearly the expected value of HIT is zero. Furthermore, the conditional expectation of HIT given any information known at $t-1$ must also be zero. In particular, HIT_t must be uncorrelated with any lagged HIT_{t-k}, with the forecasted VaR_t and with a constant. If HIT_t satisfies these moment conditions, then there will be no autocorrelation in the violations and there will be the correct fraction of exceedences. A way to test this condition is to construct the artificial regression:

$$HIT_t = \delta_0 + \delta_1 HIT_{t-1} + \ldots + \delta_p HIT_{t-p} + \delta_{p+1} VaR_t + u_t \ .$$

In matrix form:

$$HIT_t = X\delta + u_t$$

where $u_t - \theta \sim Bernoulli(\theta)$. A good forecasting model should produce a sequence of unbiased and uncorrelated VaR violations, therefore the explanatory power of this artificial regression should be zero. In this case the null hypothesis should be $H_0 : \ \delta = 0$. The asymptotic distribution of the OLS estimator under the null is

$$\hat{\delta} = (X'X)^{-1}X'HIT \overset{a}{\sim} N\left(0, \theta(1-\theta)(X'X)^{-1}\right) .$$

Hence, the Dynamic Quantile test statistic is

$$\frac{\hat{\delta}'X'X\hat{\delta}}{\theta(1-\theta)} \overset{a}{\sim} \chi^2_{p+2} \ .$$

5 Results

5.1 In sample results

Table 3 reports some decision criteria based on the in-sample performances of the competing models. The entries of the Table are simply the averages over the 100 Monte Carlo replications of the log-likelihoods and of the Bayesian Information Criteria[4].

Looking at the Table, we can affirm that the more general formulations (ADCC-EGARCH *vs.* DCC-GARCH and Diagonal *vs.* Scalar BEKK) are preferred by the BIC criterion. The direct parametrization of the conditional correlation sensibly increases the likelihood and decreases the fitting errors, for every distribution of the innovations. Furthermore, we compare the estimated correlations with the true correlations by Mean Absolute Error:

$$MAE = \frac{1}{T}\sum |\hat{\rho}_t - \rho_t| \ , \tag{29}$$

and the estimated conditional variances with the true variances by the average Frobenius norm of the difference between the diagonal elements of the estimated covariance matrices and the diagonal elements of the true covariance matrix. These indicators confirm that the DCC specifications are preferred to the BEKK parametrizations.

Looking now at the different performances induced by the choice of the distribution for the innovations, we can observe how the Laplace distribution outperforms both the Gaussian distribution and the t Copula. We expected this result with respect to the

[4]Unfortunately, we cannot compare the likelihood values of the models estimated using the Laplace distribution with the Gaussian and Student's t counterparts because of a constant involved in the numerical integration of the modified Bessel function of the second kind.

Table 3: Mean log-likelihood, Bayesian Information Criterion, Frobenius norm and Mean Absolute Error of the 14 competing models

Model	DCC N	DCC L	DCC C	ADCC N	ADCC L	ADCC C
Log-like	-3376.6	-1065.9	-3235.4	-3352.9	-1049.2	-3208.6
BIC	4.5417	1.4599	4.3670	4.5234	1.4525	4.4367
Frobenius	0.9108	0.8806	0.9088	0.8591	0.8226	0.8530
MAE	0.0881	0.0840	0.0962	0.0804	0.0963	0.1005

Model	SBEKK N	SBEKK L	SBEKK t	DBEKK N	DBEKK L	DBEKK t
Log-like	-3505.2	-2860.7	-3361.3	-3495.1	-2838.3	-3352.1
BIC	4.7078	3.8484	4.5208	4.6845	3.8088	4.4986
Frobenius	0.9147	1.1300	0.8913	0.8980	1.0801	0.8869
MAE	0.1317	0.1264	0.1234	0.1291	0.1239	0.1261

Model	SBEKK MDFt	DBEKK MDFt
Log-like	-3376.7	-3369.7
BIC	4.5364	4.5368
Frobenius	0.8890	0.8971
MAE	0.1205	0.1190

Gaussian distribution, but the relatively poor performance of the Copula model is indeed puzzling, given that the data were drawn from a model with a structure rather close to the structure of a Copula DCC.

Another interesting evidence of our experiment is that, considering the leverage effects in volatilities, we sensibly improve the variance fitting, deteriorating at the same time the correlation fitting.

Finally, estimating the BEKK models allowing for a non standard t Distribution with different values of degrees of freedom for each univariate series, we are able to match the performance of the BEKK models with the classic Multivariate t in terms of fitting of the conditional volatilities, improving at the same time the fitting of the conditional correlations.

5.2 Out of sample results

Table 4 reports the results of the out-of-sample comparison of the competing models on the basis of the loss functions described in Eq. (23)-(24) and (25). The subscript $j = 1, 2$ indicates the relevant volatility while NU and NO are, respectively, the number of underpredictions and of overpredictions over the 500 out-of-sample observations. Every entry in the Tables 4 and 6 is computed as the sample mean over the 100 Monte Carlo replications of the relevant statistic.

We can observe, in general, how the DCC-GARCH and the BEKK specifications have

similar performances as indicated by the MAFE criterion. On the other hand, the ADCC-EGARCH parametrizations, able to capture the leverage effects present in the data, lead to a substantial improvement of the forecasting accuracy. Quite interestingly, looking at the forecasting accuracy of the first series characterized by an higher kurtosis of the shocks of both the return and the volatility, we can observe how the MAFE criterion prefers a Laplace distribution over a Student's t when analyzing DCC models, whereas the opposite holds when considering BEKK models. Focusing now on the second series, less leptokurtic, we can observe how these differences in accuracy fade away, with the simple Normal distribution being preferred by the MAFE criterion when coupled with the DCC-GARCH and Scalar BEKK models.

Looking now more in detail at the BEKK performances, we can observe how the forecasting accuracy of the diagonal specification is lower than the accuracy of the simpler scalar specification for every distribution of the innovations analyzed, with the Multivariate t preferred to the Multivariate Laplace which is, surprisingly, clearly outperformed even by the Gaussian.

Concluding, we obtain some counterintuitive results; it seems, in fact, that increasing the flexibility in the specification of the conditional variances (Diagonal BEKK *vs.* Scalar BEKK) and/or in the distribution of the innovations (Student's t marginal and Copula *vs.* Multivariate Laplace and Multiple Degrees of Freedom t *vs.* Multivariate t) in order to capture the heterogeneity of the 2 series, the forecasting accuracy is not improved, despite the increased numerical complexity of the estimation procedures. On the other hand, explicitly modeling the leverage effects present in the data strongly improves the forecasting results.

Looking now at the asymmetric loss functions, we can observe how every model, with the exceptions of the Scalar and Diagonal BEKK when coupled with Laplace innovations, strongly underpredicts both the volatilities, given that the difference between MMEU and MMEO is positive. Forecasting the second series, characterized by a lower kurtosis in both the return and volatility shocks, is in a certain sense easier, and this is confirmed by the shrinkage of the differences between MMEU and MMEO and by much more balanced numbers of under and over predictions. This situation suggests how, forecasting the second series, we underpredict the volatility in a smaller number of cases but the magnitude of these underpredictions is bigger than that observed forecasting the first series.

If we now focus on the DCC models, we can observe how, with both series and every probability distribution, modeling the leverage effects, both the MMEU and the MMEO are decreased, increasing at the same time the difference between the number of under and over predictions. This means that, despite the increased number of underpredictions, the magnitude of these events has reduced substantially; this can be interpreted as another evidence of the centrality of capturing the leverage effects when comparing forecasting performances with these loss functions. On the other hand, we are not able to improve the forecasting performances neither allowing for flexible distribution functions of the univariate processes nor increasing the dimensionality of the BEKK parametrizations.

The analysis of the correlation between point forecasts reported in Table 5 reveals two main interesting features. First, if we focus on the single class of models, e.g. DCC, we notice that the forecasts are highly correlated across the distributions, suggesting that the error distribution assumed for the estimation has a negligible impact on the description of the dynamics of the data. Second, it seems that correlations across models are mainly driven by the presence of an asymmetric specification, both in the conditional volatilities and in the conditional correlations. For instance, the DCC is highly correlated with BEKK specifications (both classes are symmetric) while the correlation between ADCC

Table 4: Out of sample forecasting accuracy measures

Model	$MAFE_1$	$MMEU_1$	$MMEO_1$	NU_1	NO_1	$MAFE_2$	$MMEU_2$	$MMEO_2$	NU_2	NO_2
DCC N	0.6585	1.3321	0.7637	329	171	0.4179	0.8607	0.6043	243	257
DCC L	0.6378	1.2210	0.7917	302	198	0.4213	0.8237	0.6399	228	272
DCC C	0.6504	1.2210	0.7677	322	178	0.4184	0.8237	0.6341	235	265
ADCC N	0.6191	1.2670	0.6938	346	154	0.3930	0.8853	0.5269	246	254
ADCC L	0.5834	1.1392	0.6810	313	187	0.3939	0.8345	0.5488	226	274
ADCC C	0.6098	1.2347	0.6878	340	160	0.3918	0.8403	0.5354	238	262
SBEKK N	0.6424	1.2093	0.8105	309	192	0.4143	0.9100	0.5508	239	261
SBEKK L	0.8057	1.0848	1.5192	133	368	0.5187	0.8082	0.8834	123	377
SBEKK t	0.6344	1.1824	0.8174	294	206	0.4151	0.8918	0.5603	231	269
DBEKK N	0.6523	1.3362	0.7468	318	182	0.4179	0.8679	0.5960	235	265
DBEKK L	0.7495	1.0924	1.1338	143	357	0.5481	0.7915	1.0878	119	381
DBEKK t	0.6434	1.3010	0.7520	304	196	0.4170	0.8423	0.6037	229	271
SBEKK MDFt	0.6359	1.2163	0.7901	300	200	0.4154	0.9158	0.5487	238	262
DBEKK MDFt	0.6530	1.3530	0.7494	311	189	0.4157	0.8505	0.5915	234	266

and BEKK forecasts is lower. This tendency is less definite in the forecasts of the second series, which is characterized by a milder leverage effect.

We can conclude that while analyzing point forecasts, when the Data Generating Process is characterized by strong asymmetric effects, the explicit description of these effects in the estimated model is far more crucial than the density specification.

The results of the Multivariate Mincer-Zarnowitz regressions are shown in the left panel of Table 6; the performances of the competing models are extremely poor, with rejection rates over 98 % in every case. The explanatory power of the regressions, as described by the McElroy systemwide measure (Eq. 28), is constant over the whole range of models, with the interesting exception of the ADCC-EGARCH specifications. Even in this case, capturing the leverage effects allows to sensibly increase the explanatory power of the regressions, even if the rejection rate fails to lower. Once again we can observe the poor results obtained by the BEKK models when coupled with the Laplace distribution, characterized by an extremely strong overprediction of the volatilities.

Keeping in mind the fact that the first series shows a more extreme behavior than the second series in term of kurtosis of the innovations, we perform the same test in an univariate framework, implementing autonomous regressions for the two forecast series. The results are shown in the right panel of Table 6. With the exception of the Diagonal BEKK model with Laplace innovations, we can conclude that forecasting the second series is less problematic, given the reduced kurtosis in the variance shocks, as confirmed by the lower rejection rate for every model considered.

Interestingly enough, the rejection rate relative to the first series increases with the ADCC-EGARCH models, while the opposite is true if we look at the second series. This confirms that the benefit in forecasting accuracy induced by explicitly modeling the leverage effect present in the data quickly disappears as the probability of extreme shocks increases. Once again, when coupled with DCC models, the Multivariate Laplace Distribution clearly outperfoms the Normal and the t-Copula with Student's t marginals; this result is fascinating, since the Data Generating Process innovations are drawn exactly from a t-Copula with univariate marginal t distributions.

Looking now at the BEKK models, we have another confirmation of the hypothesis that the use of the more general specification does not imply an increase in the forecasting performances when additional forms of heterogeneity are present in the data beside the persistence of the innovations.

Table 5: Average correlations between forecasts

Average Correlations between 1st series volatility forecasts

	DCC			ADCC			SBEKK				DBEKK			
	N	L	C	N	L	C	N	L	t	MDFt	N	L	t	MDFt
DCC N	1,000	0,994	0,982	0,846	0,856	0,845	0,984	0,990	0,984	0,985	0,982	0,979	0,982	0,971
L		1,000	0,985	0,844	0,854	0,843	0,982	0,990	0,982	0,986	0,982	0,980	0,982	0,974
C			1,000	0,834	0,844	0,834	0,970	0,975	0,970	0,975	0,969	0,968	0,969	0,963
ADCC N				1,000	0,993	0,993	0,844	0,848	0,844	0,841	0,841	0,845	0,841	0,836
L					1,000	0,997	0,855	0,859	0,855	0,853	0,850	0,851	0,850	0,845
C						1,000	0,842	0,846	0,842	0,841	0,839	0,842	0,839	0,836
SBEKK N							1,000	0,990	1,000	0,986	0,982	0,972	0,982	0,964
L								1,000	0,990	0,989	0,985	0,980	0,985	0,970
t									1,000	0,986	0,982	0,972	0,982	0,964
MDFt										1,000	0,986	0,980	0,986	0,981
DBEKK N											1,000	0,984	1,000	0,968
L												1,000	0,984	0,988
t													1,000	0,968
MDFt														1,000

Average Correlations between 2nd series volatility forecasts

	DCC			ADCC			SBEKK				DBEKK			
	N	L	C	N	L	C	N	L	t	MDFt	N	L	t	MDFt
DCC N	1,000	0,998	0,998	0,922	0,922	0,921	0,977	0,976	0,977	0,970	0,985	0,959	0,985	0,986
L		1,000	0,999	0,923	0,923	0,923	0,976	0,975	0,976	0,969	0,985	0,956	0,985	0,986
C			1,000	0,923	0,923	0,923	0,975	0,973	0,975	0,968	0,986	0,956	0,986	0,986
ADCC N				1,000	0,996	0,995	0,913	0,915	0,913	0,911	0,921	0,902	0,921	0,922
L					1,000	0,998	0,912	0,913	0,912	0,908	0,919	0,899	0,919	0,921
C						1,000	0,912	0,912	0,912	0,908	0,918	0,897	0,918	0,921
SBEKK N							1,000	0,991	1,000	0,986	0,985	0,984	0,985	0,984
L								1,000	0,991	0,990	0,983	0,989	0,983	0,986
t									1,000	0,986	0,985	0,984	0,985	0,984
MDFt										1,000	0,980	0,985	0,980	0,991
DBEKK N											1,000	0,981	1,000	0,986
L												1,000	0,981	0,975
t													1,000	0,986
MDFt														1,000

Table 6: Mincer-Zarnowitz tests results

Model	MZ	% Rej.	R^2	MZ_1	% Rej.$_1$	R^2_1	MZ_2	% Rej.$_2$	R^2_2
DCC N	259.24	98	0.34	42.62	94	0.31	22.73	82	0.35
DCC L	244.44	100	0.35	34.18	88	0.32	22.32	80	0.35
DCC C	253.89	98	0.35	40.96	90	0.32	22.59	82	0.35
ADCC N	281.58	100	0.42	50.48	98	0.41	16.87	76	0.40
ADCC L	221.76	100	0.42	34.46	90	0.42	15.17	72	0.40
ADCC C	271.24	100	0.42	48.62	96	0.42	15.93	76	0.40
SBEKK N	190.37	98	0.34	38.90	94	0.32	14.25	74	0.35
SBEKK L	757.88	100	0.34	83.80	96	0.32	46.64	80	0.35
SBEKK t	183.33	98	0.34	34.03	92	0.32	13.31	66	0.35
DBEKK N	239.00	98	0.34	37.11	94	0.31	19.51	80	0.35
DBEKK L	698.22	98	0.34	50.45	80	0.31	69.59	92	0.35
DBEKK t	212.61	100	0.34	32.03	88	0.31	17.60	74	0.35
SBEKK MDFt	178.68	100	0.34	34.40	90	0.32	13.46	74	0.35
DBEKK MDFt	219.13	100	0.34	34.97	92	0.31	17.26	76	0.35

5.3 VaR forecasting results

Looking now at the results of the backtesting of the 99% Value at Risk depicted in Table 7, we can draw some general conclusions about the overall performances of the competing models when coupled with different distributions for the innovations. First of all, when coupled with Gaussian disturbances, both DCC specifications lead to a strong underestimation of the Value at Risk, resulting in about 50% rejection rate in the Dynamic Quantile test. Analyzing the causes of these rejections, we observe how in all the cases the rejections are due to a number of Value at Risk violations higher than the theoretical coverage rate. It is worth noting that in a real application these rejections could lead to regulatory consequences, as provided for in the Basel Protocol, for the financial institution that used these models as risk measuring tools.

Assuming DCC specifications and allowing for leptokurtic error distributions leads to a sensible enhancement in the asset allocation performances of these models. In fact, both Laplace and Copula t distributions allow to obtain a fraction of VaR violations much closer to the correct value of 1% and to register a number of DQ test rejections much lower than the rejection rate observed with the Gaussian distribution. In addition to this, when a leptokurtic distribution is assumed, taking into account the leverage effect with an ADCC-EGARCH specification leads to an improvement in the performances of the models, given that the distribution assumed for the innovations is able to correctly describe the tail behavior of the dataset. In fact, estimating an ADCC-EGARCH with Laplace and t Copula distributions sensibly reduces the number of rejections of the DQ test and leads to an unconditional coverage closer to the correct value. When the more general specification is coupled with Gaussian errors, instead, we are not able to detect any of these improvements. Another interesting evidence that emerges from our experiment is that, taken as given the variance specification, the Laplace models outperform the models estimated assuming the t Copula, even if the t Copula is the distribution of the Data Generating Process. This result is even more important if we take into account the addi-

tional difficulties introduced in the estimation of the model by considering a multivariate distribution built via copula functions and the fact that the Value at Risk for a portfolio built with assets distributed as a multivariate Laplace is available in closed form (at least when considering one step ahead forecasts).

Looking at the results of the BEKK models, we can observe once again how the Normal distribution completely fails to describe the tail behavior of our data, leading to a rejection rate of the DQ test of 56% with both the BEKK specifications. It is worth noting that even in this case all the rejections are explained by a number of VaR violations higher than the correct one.

On the other hand, both Laplace and multivariate t BEKK models largely overestimate the Value at Risk. In particular, when we assume a multivariate t distribution for the innovations, we registered a rejection rate in the DQ test of 74% and 52% for the scalar and diagonal BEKK respectively. All these rejections (and those of the Laplace BEKK models) are coupled with zero VaR violations over the 500 observations. The two Laplace BEKK perform better than the Multivariate t ones, but even in this cases the Value at Risk overestimation is significant, especially considering the scalar specification. As Table 4 suggests, the overestimation of the Value at Risk shown by BEKK models has two different natures. In the Laplace case, the models greatly overestimate the conditional variance, while the Multivariate t distribution seems to underestimate the value of the degrees of freedom of the distribution, since the overestimation of the Value at Risk is coupled with a substantial underprediction of the conditional variance.

Even considering BEKK models, allowing for a more general variance specification induces benefits in VaR forecasting only when leptokurtic distributions are considered.

Finally, we look at the results for the BEKK models coupled with Multiple Degrees of Freedom t. The mean fraction of VaR violations is the best match of the correct one and the number of DQ rejections is the lowest among all the models considered. This is very interesting, considering that these BEKK specifications are not able to capture the leverage effects that we imposed to the data.

The drawback related to this class of models is that we can not exploit the two stage estimation as in DCC models, so the computational burden will become unmanageable as we let the cross-sectional dimension of the sample increase. Another difficulty arises in the estimation of the Value at Risk, given that we do not know in closed form the distribution of the return of the portfolio if the assets returns are drawn from a Multiple Degrees of Freedom t. Hence, we need to simulate it in order to compute the Value at Risk. This simulation is, to tell the truth, numerically much simpler than the Value at Risk simulation when we assume a multivariate distribution built via copula functions. This is true because the Multiple Degrees of Freedom t VaR simulation does not involve the computation of Cumulative Distribution Functions and the relative inverses.

Table 7: VaR forecasting results

Model	DCC N	DCC L	DCC C	E-ADCC N	E-ADCC L	E-ADCC C
Fraction HIT	0.0189	0.0091	0.0123	0.0163	0.0080	0.0099
Rej. DQ	48%	22%	32%	48%	18%	26%
Num. failures	0%	0%	0%	9%	0%	0%
Model	SBEKK N	SBEKK L	SBEKK t	DBEKK N	DBEKK L	DBEKK t
Fraction HIT	0.0160	0.0033	0.0006	0.0164	0.0035	0.0010
Rej. DQ	56%	26%	74%	56%	16%	52%
Num. failures	10%	0%	0%	0%	1%	0%
Model	SBEKK MDFt			DBEKK MDFt		
Fraction HIT	0.0099			0.0103		
Rej. DQ	13%			11%		
Num. failures	2%			1%		

6 Conclusions

In this work we have applied a selection of widely adopted Multivariate GARCH models to simulated data able to mimic the principal stylized facts of real financial data such as volatility clustering and different levels of excess kurtosis across the time series.

The aim of our work is to understand the beneficial effects induced by more flexible volatility structures and probability distributions of the innovations in a context where both the conditional variance and the error distribution are misspecified.

Our main contributions can be summarized as follows. First of all, directly modeling the conditional correlation matrix allows to obtain an increase in both in-sample and out-of-sample performances. Secondly, comparing the models by means of forecasting accuracy measures, we are able to show the paramount importance of explicitly modeling the leverage effects present in the data; furthermore, adopting BEKK specifications we are not able to identify any benefits induced by the Diagonal parametrization over the simpler Scalar specification. Thirdly, all the models considered tend to strongly underestimate the variance of the data, with the notable exception of the BEKK models when coupled with the Laplace distribution. Analyzing the results of the Value at Risk forecasting, we can observe how the Normal distribution is completely inadequate in our framework, leading to strong and persistent under-predictions of the Value at Risk with both BEKK and DCC class models.

More interestingly, modeling the leverage effect in the data presents substantial benefits only if the excess kurtosis of the data is taken into account. These two results together stress further the importance of capturing with suitable probability distributions the excess kurtosis shown by the data.

In our experiments we also show that explicitly modeling the different levels of heavy tailedness of the data with a Copula-DCC GARCH model leads to results very similar to

those obtained with the much computationally simpler Multivariate Laplace DCC model. On the other hand, in the BEKK framework, this flexibility in modeling the tail behavior of the different series greatly improves both in-sample and out-of-sample results.

A natural extension of this work can be the measurement of the costs associated with a misspecification of the skewness of the Data Generating process.

Appendix: The Multivariate Laplace distribution

In the following pages, we review a few important elements of the multivariate asymmetric Laplace distribution. A random vector $Y \in \mathbb{R}^d$ is said to have a multivariate asymmetric Laplace distribution (hereafter AML) if its characteristic function is given by:

$$\Psi(t) = \frac{1}{1 + \frac{1}{2}t'\Sigma t - im't}$$

where $m \in \mathbb{R}^d$ and Σ is a $d \times d$ nonnegative definite symmetric matrix.

The vector m is the location parameter and the matrix Σ is the scale parameter of this distribution. The distribution is unimodal with the mode equal to zero. As a consequence the parameter m also determines the level of asymmetry. The density of $Y \sim AML(m, \Sigma)$ can be expressed as:

$$g(y) = \frac{2e^{y'\Sigma^{-1}m}}{(2\pi)^{d/2}|\Sigma|^{1/2}} \left(\frac{y'\Sigma^{-1}y}{2 + m'\Sigma^{-1}m} \right)^{v/2} K_\nu \left(\sqrt{(2 + m'\Sigma^{-1}m)\,(y'\Sigma^{-1}y)} \right)$$

where $\nu = (2 - d)/2$ and $K_\nu(u)$ is the modified Bessel function of the second kind as described in Abramowitz and Stegun (1965):

$$K_\nu(u) = \frac{(u/2)^v \Gamma(1/2)}{\Gamma(v + 1/2)} \int_1^\infty e^{-ut}(t^2 - 1)^{v-1/2} dt, \quad v \geq -1/2.$$

The class of AML is not closed under summation of independent r.v.'s: if X and Y are independent AML r.v.'s, then in general $X + Y$ does not possess an AML law.

Let $X \sim N(0, \Sigma)$ and W be an exponentially distributed r.v. with mean 1, independent of X. Then

$$mW + W^{1/2}X \to Y \sim AML(m, \Sigma).$$

The first two moments of the asymmetric multivariate Laplace distribution are given by:

$$E[Y] = m$$

$$Cov(Y) = \Sigma + mm'$$

Finally it is worth noting the fact that linear transformations of AML-distributed vectors have an AML distribution. Let $Y = (Y_1, \ldots, T_d)' \sim AML(m, \Sigma)$ and A be an $l \times d$ real matrix. Then the r.v. $AY \sim AML(m_A, \Sigma_A)$ where $m_A = Am$ and $\Sigma_A = A\Sigma A'$. Let $(Y_1, \ldots, Y_n) \sim AML(m, \Sigma)$, where $\Sigma = (\sigma_{ij})_{i,j=1}^d$, then:

- For all $n \leq d$, $(Y_1, \ldots, Y_n) \sim AML(\tilde{m}, \tilde{\Sigma})$, where $\tilde{m} = (m_1, \ldots, m_n)'$ and $\tilde{\Sigma}$ is an $n \times n$ matrix with $\tilde{\sigma}_{ij} = \sigma_{ij}$ for $i, j = 1, \ldots, n$.
- For any $b = (b_{1,\ldots,b_d})' \in \mathbb{R}^d$, the r.v. $Y_b = \sum_{k=1}^d b_k Y_k$ is univariate $AL(\mu, \sigma)$ with $\sigma = \sqrt{b'\Sigma b}$ and $\mu = m'b$. Further, if Y is symmetric Laplace, then so is Y_b.
- For all $k \leq d$, $Y_k \sim AL(\mu, \sigma)$ with $\sigma = \sqrt{\sigma_{kk}}$ and $\mu = m_k$.

References

Abramowitz, M., Stegun, I.A., 1965. Handbook of Mathematical Functions. National Bureau of Standards, Applied Mathematics Series 55, Dover Publications, sections 9.1.1, 9.1.89, and 9.12, formulas 9.1.10 and 9.2.5.

Bauwens, L., Laurent, S., 2005. A New Class of Multivariate Skew Densities, With Application to Generalized Autoregressive Conditional Heteroscedasticity Models. Journal of Business & Economic Statistics, 23:346-354.

Bollerslev, T., 1990. Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach. Review of Economics and Statistics, 72:498-505.

Brailsford, T., Faff, R., 1996. An evaluation of volatility forecasting techniques. Journal of Banking & Finance, 20(3) 419-438.

Cajigas, J.,Urga, G., Dynamic Conditional Correlation Models with Asymmetric Multivariate Laplace Innovations. Mimeo. (Cass Business School, City University, 2007).

Danielsson, J., 1998. Multivariate Stochastic Volatility Models: Estimation and Comparison with VGARCH models. Journal of Empirical Finance, 5 155-173.

Engle, R., 2002. Dynamic Conditional Correlation-A simple class of multivariate GARCH models. Journal of Business and Economic Statistics, 20(3) 339-350(12).

Engle, R., Kroner, K., 1995. Multivariate simultaneous GARCH, Econometric theory, 11 122-150.

Engle, R., Manganelli, S., 1999. CAViaR: Conditional Value at Risk by Quantile Regression. NBER Working Papers, 7341.

Fiorentini, G., Sentana, E., Calzolari, G., 2003. Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with Student t innovations. Journal of Business and Economic Statistics, 21 532546.

Harvey, A.C., Ruiz, E., Shephard, N., 1992. Unobservable component time series models with ARCH disturbances. Journal of Econometrics, 52 129158.

Lee, P., 2003. The generalized lambda distribution applied to spot exchange rates, PhD Thesis,Department of Statistics, Carnegie Mellon University.

Patton, A., Sheppard, K., 2008. Evaluating Volatility and Correlation Forecasts, Oxford Financial Research Centre, OFRC Working Papers Series.

Serban, M., Brockwell, A., Lehoczky, J., Srivastava, S., 2007. Modelling the Dynamic Dependence Structure in Multivariate Financial Time Series. Journal of Time Series Analysis, 28(5) 763-782.

Shephard, N., Andersen, T., 2008. Stochastic Volatility: Origins and Overview. In T. Andersen (Ed.), Handbook of financial time series, Springer Verlag.

Sklar A., 1959. Fonctions de repartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Université de Paris, 8 229-231.