DE GRUYTER DOI 10.1515/snde-2012-0065 — Stud. Nonlinear Dyn. E. 2014; 18(3): 339–365

M. Angeles Carnero* and M. Hakan Eratalay

Estimating VAR-MGARCH models in multiple steps

Abstract: This paper analyzes the performance of multiple steps estimators of vector autoregressive multivariate conditional correlation GARCH models by means of Monte Carlo experiments. We show that if innovations are Gaussian, estimating the parameters in multiple steps is a reasonable alternative to the maximization of the full likelihood function. Our results also suggest that for the sample sizes usually encountered in financial econometrics, the differences between the volatility and correlation estimates obtained with the more efficient estimator and the multiple steps estimators are negligible. However, when innovations are distributed as a Student-t, using multiple steps estimators might not be a good idea.

Keywords: financial markets; volatility spillovers.

JEL Code: C32.

***Corresponding author: M. Angeles Carnero,** Dep. Fundamentos del Análisis Económico, Universidad de Alicante, Spain, Phone: +34 965903400 (ext. 3255), Fax: +34 965903898, e-mail: acarnero@ua.es
M. Hakan Eratalay: Dep. Economics, European University at St Petersburg, Russian Federation

1 Introduction

Understanding how stock market returns and volatilities move over time has been of interest to researchers into the time series literature. As the financial crisis has shown, stock markets move together. Evidence of these co-movements can be found, for example, in the fall of several international stock market indices after a very big investment bank in US, Lehman Brothers, declared bankruptcy in September 2008. Therefore, trying to model stock markets in a univariate way ignoring their interactions would be insufficient. In this sense, multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) models have been very popular to capture the volatility and covolatility of assets and markets; see, for example, Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2009) for a survey.

One of the problems with many MGARCH models is the difficulty to verify that the conditional variance-covariance matrix is positive definite. Engle, Granger, and Kraft (1984) provide necessary conditions for the positive definiteness of the variance-covariance matrix in a bivariate ARCH setting. However, extensions of these results to more general models are very complicated. Moreover, imposing restrictions on the log-likelihood function, in order to have the necessary conditions satisfied, is often difficult.

A model that could avoid these problems is the constant conditional correlation GARCH (CCC-GARCH) model proposed by Bollerslev (1990). In this model, the Gaussian maximum likelihood (ML) estimator of the correlation matrix is the sample correlation matrix which is always positive definite. Therefore, the only restrictions needed are the ones for the conditional variances to be positive. On top of that, since the correlation matrix can be concentrated out of the log-likelihood function, the optimization problem becomes simpler. Consequently, the CCC-GARCH model has become very popular in the literature regardless of some limitations such as the constant correlation assumption and the incapability to explain possible volatility interactions. The extension proposed by Jeantheau (1998), the ECCC-GARCH model, addresses the last issue by allowing for volatility spillovers. Relaxing the constant correlation assumption is done by Engle (2002) and Tse and Tsui (2002) who propose the Dynamic Conditional Correlation GARCH (DCC-GARCH) model in which the correlation changes over time. However, since the correlation dynamics require more parameters, the estimation of the DCC-GARCH model can be computationally very heavy. One possible solution is to use the *correlation targeting* approach, see Engle (2009), in which the intercept in the correlation equation is replaced by its sample counterpart. This solution is questioned by Aielli (2008) who suggests a correction to the DCC-GARCH model, denoted by consistent DCC-GARCH (cDCC-GARCH) model.

Alternatively, Pelletier (2006) introduces the regime switching dynamic correlation GARCH (RSDC-GARCH) model in which the correlation is constant over time but changing between different regimes and driven by an unobserved Markov switching chain. This model can be thought as in between the CCC-GARCH model and the DCC-GARCH model, with the problem that the number of correlation parameters increases rapidly with the number of series considered.

When dealing with stock market returns, it is not unusual to find some dynamics in the conditional mean, that could be well approximated by a vector autoregressive moving average (VARMA) model; see, for example, McAleer and da Veiga (2008a,b). One way to estimate the parameters of the VARMA-MGARCH conditional correlation model would be solving the optimization problem of the full log-likelihood function and therefore obtaining the estimates for all the parameters in 1 step. If a Gaussian log-likelihood function is specified and the true data generating process (DGP) is also Gaussian, then it is known that ML estimators are consistent and asymptotically normal. In the case that the true DGP is not Gaussian, then we would be using quasi-maximum likelihood (QML) estimators. Bollerslev and Wooldridge (1992) show that, under quite general conditions, QML estimators are consistent and asymptotically normal. Estimating all parameters in 1 step would be the best we could achieve, however when there are many parameters involved, it is very heavy computationally, when feasible. Bollerslev (1990), Longin and Solnik (1995) and Nakatani and Teräsvirta (2008) are few of the papers using 1-step estimation.

Under the normality assumption, the parameters could also be estimated in 2 steps. First, the mean and variance parameters are estimated assuming no correlation and then, in a second step, the correlation parameters are estimated given the estimates from the first step; see, for example, Engle (2002). However, as Engle and Sheppard (2001) suggest for the DCC-GARCH model, these 2-steps estimators will be consistent and asymptotically normal but not efficient.

The 3-steps estimation method is mentioned in Bauwens, Laurent, and Rombouts (2006). It consists of estimating the mean parameters in a first step, the variance parameters in a second step, given the first step estimates, and finally, given all other parameter estimates, the correlation parameters in the last step. The second and third steps of the procedure will be equivalent to the 2-steps estimation method for a zero-mean series. Therefore, under normal errors, the 3-steps estimators are also consistent and asymptotically normal. Engle and Sheppard (2001) implement the 3-steps estimation procedure in their empirical application.

There are other papers in the literature using multiple steps estimators in multivariate GARCH models; see, for example, Hafner and Reznikova (2012) and Engle, Shephard, and Sheppard (2008) for a deeper discussion on the issue of estimating dynamic conditional correlation models.

Evidence gathered over the past decades shows that stock market returns are often far from having a normal distribution. Consequently, we find interesting to consider the estimation of the models assuming a Student-t distribution. In this case, the 1-step estimator is obtained by maximizing the log-likelihood function based on the multivariate t-distribution; see, for example, Harvey, Ruiz, and Shephard (1992) and Fiorentini, Sentana, and Calzolari (2003). Although there is no theoretical work studying the properties of multiple steps estimation when assuming a Student-t distribution, 2-steps and 3-steps estimators could be analyzed by means of Monte Carlo simulations. In this line of research, Bauwens and Laurent (2005) and Jondeau and Rockinger (2005) analyze 2-steps estimators. However, their first step is performed assuming Gaussian errors while we wonder what would be the behaviour of multiple steps estimators under the assumption that the errors are distributed as a Student-t.

In this paper, we present various Monte Carlo experiments to compare the finite sample performance of the more efficient 1-step estimator with the 2-steps and 3-steps estimators for different vector autoregressive multivariate conditional correlation GARCH models. In particular we consider VAR(1) - CCC, ECCC, DCC, cDCC and RSDC - GARCH(1,1) models and other extensions of these models. When the data is normally distributed, we find that, for the models considered and for the sample sizes usually encountered in financial econometrics, differences between the 1-step and multiple steps estimators are negligible. When we change the assumption on the distribution to a Student-t, we conclude that, for some models, the differences between the estimators could be relevant and therefore, estimating the parameters in multiple steps might not be a good idea.

The main contributions of this paper to the related literature are the following. First, we compare the performance of multiple steps estimators for different conditional correlation models and not only for the dynamic conditional correlation model; see Hafner and Reznikova (2012) and Engle, Shephard, and Sheppard (2008). Second, we analyze multiple steps estimators obtained by maximizing the log-likelihood function based on the normal distribution but also based on the Student-t distribution and we find that if errors follow a Student-t distribution, multiple steps estimators assuming the t distribution do not seem to be consistent for some of the models. Finally, we point out that when the distribution of the errors is skewed, QML estimators based on symmetric distributions could be inconsistent. We also analyze the robustness of our findings to the model misspecification.

One potential problem of our results is their external validity. For the Monte Carlo experiments, we considered bivariate models and in some cases trivariate models. We assume that what we find for two and three time series could be extrapolated for any number $k>3$ of time series.

The rest of the paper is structured as follows. Section 2 introduces the econometric models of interest. 1-step and multiple steps estimators for the different models are discussed in Section 3. Section 4 describes the Monte Carlo experiments and presents a discussion of the results. Finally, Section 5 concludes the paper.

2 Econometric models

For simplicity we consider a k-variate vector autoregressive (VAR) model of order one for the mean equation with the following notation:

$$y_t=\mu+By_{t-1}+\varepsilon_t \tag{1}$$

where $Var(\varepsilon_t|y_{t-1}, \dots y_1)=H_t$, y_t is a $k\times1$ vector of returns, μ is a $k\times1$ vector of constants, B is a $k\times k$ matrix of autoregressive coefficients and ε_t is a $k\times1$ vector of innovations as follows.

$$y_t=[\,y_{1t} \quad y_{2t} \quad \dots \quad y_{kt}\,]', \qquad \mu=[\,\mu_1 \quad \mu_2 \quad \dots \quad \mu_k\,]'$$

$$B=\begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \dots & \beta_{kk} \end{bmatrix}, \qquad \varepsilon_t=[\,\varepsilon_{1t} \quad \varepsilon_{2t} \quad \dots \quad \varepsilon_{kt}\,]'$$

The model is stationary if all values of z solving equation (2) are outside of the unit circle.

$$|I_k-B_z|=0 \tag{2}$$

The number of mean parameters in the matrices μ and B is $k(k+1)$. However, if B is assumed to be diagonal, the number of mean parameters is reduced to $2k$.

The error term ε_t can be written as follows

$$\varepsilon_t=H_t^{1/2}\eta_t$$

where η_t is a $k\times1$ vector with $E(\eta_t)=0$ and $Var(\eta_t)=I_k$.

$$H_t=D_tR_tD_t \tag{3}$$

where $D_t=diag(\,h_{1t}^{1/2},\, h_{2t}^{1/2},\, \dots,\, h_{kt}^{1/2}\,)$ and R_t is the conditional correlation matrix such that

$$H_t = diag(h_{1t}^{1/2}, h_{2t}^{1/2}, \ldots, h_{kt}^{1/2}) \begin{bmatrix} 1 & \rho_{12t} & \cdots & \rho_{1kt} \\ \rho_{12t} & 1 & \cdots & \rho_{2kt} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1kt} & \rho_{2kt} & \cdots & 1 \end{bmatrix} diag(h_{1t}^{1/2}, h_{2t}^{1/2}, \ldots, h_{kt}^{1/2})$$

$$= \begin{bmatrix} h_{1t} & \rho_{12t}\sqrt{h_{1t}h_{2t}} & \cdots & \rho_{1kt}\sqrt{h_{1t}h_{kt}} \\ \rho_{12t}\sqrt{h_{1t}h_{2t}} & h_{2t} & \cdots & \rho_{2kt}\sqrt{h_{2t}h_{kt}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1kt}\sqrt{h_{1t}h_{kt}} & \rho_{2kt}\sqrt{h_{2t}h_{kt}} & \cdots & h_{kt} \end{bmatrix}$$

From previous equations, assuming that the conditional correlation matrix, R_t, is positive definite, it is clear that as long as conditional variances, h_{it}, are positive for any i=1, 2, ..., k, the conditional variance-covariance matrix, H_t, will be also positive definite. The conditional variances h_{it} are assumed to follow a GARCH(1,1) model. Then,

$$h_t = \omega + A\varepsilon_{t-1}^{(2)} + Gh_{t-1} \tag{4}$$

where $h_t = [h_{1t}\ h_{2t} \ldots h_{kt}]'$ and $\varepsilon_t^{(2)} = [\varepsilon_{1t}^2 \quad \varepsilon_{2t}^2 \quad \ldots \quad \varepsilon_{kt}^2]'$ are $k\times 1$ vectors of conditional variances and squared errors respectively and ω is a $k\times 1$ and A and G are $k\times k$ matrices of coefficients. If A and G are restricted to be diagonal, then volatility spillovers cannot be captured; see, for example, Bollerslev (1990) and Engle (2002). Alternatively, if A and G are non-diagonal, then the model allows for volatility spillovers; see, for example, Jeantheau (1998) and Ling and McAleer (2003). In the former case there will be $3k$ variance parameters to estimate, while in the latter that number will be $k(2k+1)$.

Let us denote by $\omega_i = [\omega]_i$, $\alpha_{ij} = [A]_{i,j}$ and $\gamma_{ij} = [G]_{i,j}$. The following conditions, in Jeantheau (1998), are sufficient for the variances to be always positive.

$$\omega_i > 0 \qquad \alpha_{ij} \geqslant 0 \qquad \gamma_{ij} \geqslant 0 \qquad \text{for all } i \text{ and } j.$$

Nakatani and Teräsvirta (2008) provide necessary and sufficient conditions for h_t to have positive elements for all t. They show that off-diagonal elements in G could be negative while H_t is still positive definite. This allows for negative volatility spillovers; see also Conrad and Karanasos (2010). The model is stationary in covariance if the roots of $|I_k-(A+G)z|=0$ are outside of the unit circle. In the diagonal case, this condition is equivalent to

$$\alpha_{ii} + \gamma_{ii} < 1 \qquad \text{for all } i.$$

This paper considers five conditional correlation GARCH models given by different specifications of R_t in (3). The first and simplest one is the *CCC-GARCH* model where the correlations are restricted to be constant over time. Bollerslev (1990) shows that, under this restriction, the Gaussian ML estimator of the correlation matrix, $R_t=R$, is equal to the matrix of sample correlations of the standardized residuals, which is always positive definite, i.e.,

$$|\hat{R}|_{ij} = \hat{\rho}_{ij} = \frac{\sum_t \hat{\nu}_{it}\hat{\nu}_{jt}}{\sqrt{(\sum_t \hat{\nu}_{it}^2)(\sum_t \hat{\nu}_{jt}^2)}} \tag{5}$$

where $\nu_t = D_t^{-1}\varepsilon_t$ are the standardized errors. Notice that, in this case, the number of correlation parameters to be estimated is only $k(k-1)/2$. The *ECCC-GARCH* model of Jeantheau (1998) extends the CCC-GARCH model by allowing for volatility spillovers as A and G in (4) are non-diagonal.

The third model we consider is the *DCC-GARCH* in which $R_t = P_t Q_t P_t$ with $P_t = diag(Q_t)^{-1/2}$ and $Q_t = (1-\delta_1-\delta_2)\bar{Q} + \delta_1\nu_{t-1}\nu_{t-1}' + \delta_2 Q_{t-1}$ where Q_t denotes the covariance matrix and $\bar{Q}$ is the long run covariance (correlation) matrix. The correlation targeting approach suggests replacing $\bar{Q}$ with the sample covariance matrix of the standardized errors ν_t; see Engle (2009). This procedure makes the estimation easier since it reduces the number of correlation parameters from $k(k-1)/2+2$ to only 2: δ_1 and δ_2. If both are non-negative scalars satisfying $\delta_1+\delta_2<1$, then the correlation matrix, R_t, will be positive definite. Hafner and Franses (2009) provide a more general definition of the model where they consider coefficient matrices instead of scalar coefficients allowing for different dynamics on different correlations. However, this increases the number of parameters considerably. For simplicity, we will focus on the set up with the scalar coefficients.

The DCC-GARCH model suffers from two problems. First, as Engle and Sheppard (2001) and later Engle, Shephard, and Sheppard (2008) point out, when k is large the correlation targeting approach used in the DCC-GARCH model causes significant biases to estimators of the parameters δ_1 and δ_2. To fix this problem, Engle, Shephard, and Sheppard (2008) suggest a composite likelihood estimator which is based on the sum of the likelihoods obtained from smaller number of series and therefore avoid the trap of high dimensionality. Another solution is proposed by Hafner and Reznikova (2012), where the authors use shrinkage to target methods to eliminate these biases asymptotically. The second problem, as Aielli (2008) argues, is that multiple steps estimators of DCC-GARCH models with correlation targeting are inconsistent since the covariance matrix of the standardized residuals is not a consistent estimator of the long run covariance matrix $\bar{Q}$. As Caporin and McAleer (2009) point out as well, Aielli's conclusion follows from the fact that the unconditional expectations of Q_t could differ from the unconditional expectation of $\nu_{t-1}\nu'_{t-1}$, the former being a covariance matrix while the latter is a correlation matrix by construction. Aielli (2008) therefore suggests a corrected version of the DCC-GARCH model, denoted by *cDCC-GARCH*, in which $Q_t=(1-\delta_1-\delta_2)\bar{Q}+\delta_1\nu^*_{t-1}\nu^{*\prime}_{t-1}+\delta_2 Q_{t-1}$ where $\nu^*_t=diag(Q_t)^{1/2}\nu_t$. He argues that in this model a natural estimator for the long run covariance matrix, $\bar{Q}$, would be the sample covariance matrix of ν^*_t. The number of parameters to be estimated will be then only 2 as in the DCC-GARCH model of Engle (2002).

We will also consider in this paper the *RSDC-GARCH* model introduced by Pelletier (2006). In this model the conditional correlations follow a switching regime driven by an unobserved Markov chain such that they are fixed in each regime but may change across regimes. For simplicity, we assume a two-states Markov process such that R_t, at any time t, could be equal to either R^L or R^H, which stands for low and high state correlation matrices, respectively. The transition probabilities matrix is given by $\Pi=\{\{\pi_{L,L}, \pi_{H,L}\}, \{\pi_{L,H}, \pi_{H,H}\}\}$, where $\pi_{i,j}$ is the probability of moving from state j to state i. Given that $\pi_{j,j}+\pi_{i,j}=1$, the number of correlation parameters is $k(k-1)+2$.

Additionally, the extended versions of the DCC, cDCC and RSDC-GARCH models, namely EDCC, EcDCC and ERSDC-GARCH models are also considered. Like the ECCC-GARCH model, these extended models allow for volatility spillovers letting A and G to be non-diagonal.

In the next section we will discuss how to estimate the parameters of these models.

3 Estimation procedures

Multivariate GARCH models can be estimated using maximum likelihood. However, how the estimation is implemented in practice is one of the main problems. When the number of parameters is large, it is common that optimization procedures fail to find the maximum of the likelihood function. In this section we will describe alternative estimation methods which could be used in practice.

Let us start by introducing some notation. Let $\theta=(\mu', vec(B)')'$ be the vector containing all the mean parameters in equation (1). The vector containing all the variance parameters in (4) will be denoted by $\phi=(\omega', vec(A)', vec(G)')'$ and ψ will be the vector with all the correlation parameters, that will change according to the model considered in each case. For example, $\psi=vech(R)$ for a CCC-GARCH model, while for a cDCC-GARCH model, it will be $\psi=(vech(\bar{Q})', \delta_1, \delta_2)'$.[1]

1 Notice that the *vec* operator stacks the columns of a matrix while the *vech* operator stacks the columns of the lower triangular part of a matrix.

3.1 Vector autoregressive CCC, DCC and cDCC GARCH models and their extensions

In this section we analyze three possible procedures to estimate the parameters in equations (1) and (3), denoted by $\Phi=(\theta', \phi', \psi')'$ when R_t in equation (3) is specified by the CCC-GARCH, DCC-GARCH, cDCC-GARCH models and their extensions, namely ECCC-GARCH, EDCC-GARCH and EcDCC-GARCH models.

3.1.1 One-step estimation

One possibility is to estimate all parameters of the model, $\Phi=(\theta', \phi', \psi')'$ simultaneously. If data is assumed to be normally distributed, this 1-step estimator will be the maximum likelihood estimator of Φ and it can be found by maximizing the multivariate Gaussian log-likelihood function:

$$L(\Phi)=-\frac{Tk}{2}\log\,(2\pi)-\frac{1}{2}\sum_{t=2}^{T}(\log|H_t|+\varepsilon_t' H_t^{-1}\varepsilon_t)$$

From equation (3) we have that

$$\begin{aligned} L(\Phi)&=-\frac{Tk}{2}\log\,(2\pi)-\frac{1}{2}\sum_{t=2}^{T}\log\,|D_tR_tD_t|-\frac{1}{2}\sum_{t=2}^{T}\varepsilon_t'\,(D_tR_tD_t)^{-1}\varepsilon_t= \\ &=-\frac{Tk}{2}\log\,(2\pi)-\frac{1}{2}\sum_{t=2}^{T}\log\,|R_t|-\sum_{t=2}^{T}\log|D_t|-\frac{1}{2}\sum_{t=2}^{T}\nu_t' R_t^{-1}\nu_t \end{aligned} \tag{6}$$

If errors are assumed to follow a Student-t distribution, then the function to be maximized will be the multivariate Student-t log-likelihood as in Fiorentini, Sentana, and Calzolari (2003):

$$\begin{aligned} L(\Phi,\eta)=&T\log\left[\Gamma\left(\frac{\eta k+1}{2\eta}\right)\right]-T\log\left[\Gamma\left(\frac{1}{2\eta}\right)\right]-\frac{Tk}{2}\log\left(\frac{1-2\eta}{\eta}\right)-\frac{Tk}{2}\log(\pi) \\ &-\sum_{t=2}^{T}\left[\frac{1}{2}\log|H_t|+\left(\frac{\eta k+1}{2\eta}\right)\log\left(1+\frac{\eta}{1-2\eta}\nu_t' R_t^{-1}\nu_t\right)\right] \end{aligned} \tag{7}$$

where η is the inverse of the degrees of freedom as a measure of tail thickness. We assume $0<\eta<0.5$ in order to have existence of the second order moments.

As Newey and Steigerwald (1997) pointed out, one concern when maximizing the log-likelihood function based on a Student-t distribution is that estimators can be inconsistent if the data does not follow a Student-t distribution. However, this will not be the case as long as both the true and assumed distributions are symmetric.

Under Gaussianity assumption, 1-step estimators of the parameters, Φ, obtained by maximizing the corresponding likelihood function in (6), are consistent and asymptotically normal. In particular,

$$\sqrt{n}(\hat{\Phi}_n-\Phi_0)\sim^{A} N(0, A_0^{-1}B_0A_0^{-1})$$

where A_0 is the negative expectation of the Hessian matrix evaluated at the true parameter vector Φ_0 and B_0 is the expectation of the outer product of the score vector evaluated at Φ_0 obtained from the likelihood function in (6).

If data is assumed to follow a Student-t distribution, 1-step estimators of the parameters, Φ, computed by maximizing the likelihood function in (7), are consistent and asymptotically normal; see Fiorentini, Sentana, and Calzolari (2003). It is important to note that if the true distribution of the data is Student-t, maximum likelihood (ML) estimators (in this case, 1-step estimators using (7)) are more efficient than Quasi-maximum likelihood (QML) estimators obtained from maximizing the likelihood function under the normality assumption given in (6).

3.1.2 Two-steps estimation

It is possible to estimate the parameters of the model, $\Phi=(\theta', \phi', \psi')'$ in 2 steps following Engle (2002) and Engle and Sheppard (2001). They proposed to use 2-steps when estimating the parameters of the DCC-GARCH model. The idea is to separate the estimation of the correlation parameters, ψ, from the mean and variance parameters, θ and ϕ, respectively.

In the first step, the mean and variance parameters, θ and ϕ, are estimated by maximizing the Gaussian log-likelihood function in (6) in which the correlation matrix R_t is replaced by the identity matrix. Therefore, in the first step, the function to be maximized is the following:

$$L_1(\theta, \phi)=-\frac{Tk}{2}\log(2\pi)-\sum_{t=2}^{T}\log|D_t|-\frac{1}{2}\sum_{t=2}^{T}\nu_t'\nu_t$$

If volatility spillovers are not allowed, i.e., A and G in equation (4) are restricted to be diagonal, the first step estimation is equivalent to estimating k univariate models separately; see Engle and Sheppard (2001) for details. In this sense, we note that all extended models considered in this paper require multivariate estimation when estimating the variance parameters.

In the second step, given the estimates from the first step, $\hat{\theta}$ and $\hat{\phi}$, the correlation coefficients are estimated by maximizing the following function

$$L_2(\psi|\hat{\theta}, \hat{\phi})=-\frac{1}{2}\sum_{t=2}^{T}(\log|R_t|+\hat{\nu}_t' R_t^{-1}\hat{\nu}_t) \tag{8}$$

where $\hat{\nu}_t$ are the standardized residuals obtained in the first step.

Bollerslev (1990) shows that when the correlations are constant over time, i.e., in the CCC-GARCH model, the correlation coefficients estimator obtained in the second step is equal to the sample correlation matrix of the standardized residuals given in (5).

If data is assumed to follow a normal distribution, 2-steps estimators are also consistent. Furthermore, Engle and Sheppard (2001) give conditions for the DCC-GARCH model under which 2-steps estimators are also asymptotically normal.

Next, we also consider 2-steps estimation using the log-likelihood function based on the Student-t distribution. Accordingly, in the first step the function to be maximized is the multivariate Student-t log-likelihood function in (7) where the correlation matrix R_t has been replaced by I_k. That is

$$\begin{aligned}L_1(\theta, \phi, \eta)=&T\log\left[\Gamma\left(\frac{\eta k+1}{2\eta}\right)\right]-T\log\left[\Gamma\left(\frac{1}{2\eta}\right)\right]-\frac{Tk}{2}\log\left(\frac{1-2\eta}{\eta}\right)-\frac{Tk}{2}\log(\pi)\\&-\sum_{t=2}^{T}\left[\log|D_t|+\left(\frac{\eta k+1}{2\eta}\right)\log\left(1+\frac{\eta}{1-2\eta}\nu_t'\nu_t\right)\right]\end{aligned}$$

Similar to the case of Gaussian innovations, when no volatility spillovers are considered, we employ univariate estimation for each series while when there are volatility spillovers, we solve the multivariate problem. In the second step the correlation coefficients are estimated by maximizing the following function

$$L_2(\psi, \eta|\hat{\theta}, \hat{\phi})=-\sum_{t=2}^{T}\left[\frac{1}{2}\log|R_t|+\left(\frac{\eta k+1}{2\eta}\right)\log\left(1+\frac{\eta}{1-2\eta}\hat{\nu}_t' R_t^{-1}\hat{\nu}_t\right)\right] \tag{9}$$

where $\hat{\nu}_t$ are the standardized residuals obtained in the first step.

3.1.3 Three-steps estimation

An alternative procedure that we will analyze in this paper is the estimation of $\Phi=(\theta', \phi', \psi')'$ in three steps. In the first step, the parameters of the mean equation, θ, are estimated assuming constant variance, i.e., $h_{it}=h_i \forall t$, and assuming that the correlation matrix R_t is equal to the identity matrix for all t. Therefore, the function to be maximized is the following

$$L_1(\theta, h_i)=-\frac{Tk}{2}\log(2\pi)-\sum_{t=2}^{T}\log |D|-\frac{1}{2}\sum_{t=2}^{T}\nu_t' \nu_t$$

where $D=diag(h_1^{1/2}, h_2^{1/2}, ..., h_k^{1/2})$ contains the conditional standard deviations. This is equivalent to OLS estimation for the univariate mean equations, given that the variance-covariance matrix is block diagonal.

In the second step, the parameters of the variance equation, ϕ, are estimated given the estimates of the parameters of the mean equation, $\hat{\theta}$, and substituting the correlation matrix R_t by I_k. This leads to the maximization of the following function:

$$L_2(\phi|\hat{\theta})=-\frac{Tk}{2}\log(2\pi)-\sum_{t=2}^{T}\log |D_t|-\frac{1}{2}\sum_{t=2}^{T}\tilde{\nu}_t' \tilde{\nu}_t$$

where $\tilde{\nu}_t=D_t^{-1}\hat{\varepsilon}_t$ and $\hat{\varepsilon}_t$ are the residuals obtained in the first step. After obtaining $\hat{\theta}$ and $\hat{\phi}$ from the two previous steps, in the last step, the correlation coefficients are estimated by maximizing the following function

$$L_3(\psi|\hat{\theta}, \hat{\phi})=-\frac{1}{2}\sum_{t=2}^{T}(\log |R_t|+\hat{\nu}_t' R_t^{-1}\hat{\nu}_t) \tag{10}$$

where $\hat{\nu}_t$ are the standardized residuals obtained from the second step. When the correlations are constant over time, the correlation coefficients estimator obtained in the third step is, as in the two steps estimation procedure, equal to the sample correlation matrix of the standardized residuals given in (5).

Under the Gaussianity assumption, 3-step estimators are also consistent and their asymptotic distribution is very similar to that of the 2-step estimators; see Engle and Sheppard (2001).

When using the log-likelihood function based on the Student-t distribution, the three steps estimation is performed in a similar manner. In the first step, the mean parameters, θ, are estimated along with the inverse of the degrees of freedom assuming homoscedastic innovations, i.e., $h_{it}=h_i \forall t$. The function to be maximized in the first step is the following

$$L_1(\theta, \eta, h_i)=T\log\left[\Gamma\left(\frac{\eta k+1}{2\eta}\right)\right]-T\log\left[\Gamma\left(\frac{1}{2\eta}\right)\right]-\frac{Tk}{2}\log\left(\frac{1-2\eta}{\eta}\right)-\frac{Tk}{2}\log(\pi)$$
$$-\sum_{t=2}^{T}\left[\log |D|+\left(\frac{\eta k+1}{2\eta}\right)\log\left(1+\frac{\eta}{1-2\eta}\nu_t' \nu_t\right)\right]$$

In the second step, the variance parameters, ϕ, and the inverse of the degrees of freedom, η, are estimated conditional on the mean parameter estimates, $\hat{\theta}$. The function to be maximized is the following

$$L_2(\phi, \eta|\hat{\theta})=T\log\left[\Gamma\left(\frac{\eta k+1}{2\eta}\right)\right]-T\log\left[\Gamma\left(\frac{1}{2\eta}\right)\right]-\frac{Tk}{2}\log\left(\frac{1-2\eta}{\eta}\right)-\frac{Tk}{2}\log(\pi)$$
$$-\sum_{t=2}^{T}\left[\log |D_t|+\left(\frac{\eta k+1}{2\eta}\right)\log\left(1+\frac{\eta}{1-2\eta}\tilde{\nu}_t' \tilde{\nu}_t\right)\right]$$

Finally, in the third step, the correlation coefficients and the inverse of the degrees of freedom are estimated by maximizing the following function

$$L_3(\psi,\eta|\hat{\theta},\hat{\phi})=-\sum_{t=2}^{T}\left[\frac{1}{2}\log|R_t|+\left(\frac{\eta k+1}{2\eta}\right)\log\left(1+\frac{\eta}{1-2\eta}\hat{\nu}_t' R_t^{-1}\hat{\nu}_t\right)\right] \tag{11}$$

where $\hat{\nu}_t$ are the standardized residuals obtained in the second step.

3.2 Vector autoregressive RSDC-GARCH model

The mean, variance and correlation parameters $\Phi=(\theta', \phi', \psi')'$ when R_t in equation (3) is specified by the RSDC-GARCH model can also be estimated in multiple steps.

Let us denote by Ω_{t-1} all previous information up to $t-1$ and let $f(\cdot)$ be the likelihood function obtained under the assumption of either a Gaussian or a Student-t distribution. The 1-step estimator of Φ would be obtained by maximizing the following log-likelihood function:

$$L(\Phi)=\sum_{t=2}^{T}\log f(Y_t|\Omega_{t-1}) \tag{12}$$

where

$$f(Y_t|\Omega_{t-1})=f(Y_t|S_t=L,\Omega_{t-1})\times\Pr(S_t=L|\Omega_{t-1})+f(Y_t|S_t=H,\Omega_{t-1})\times\Pr(S_t=H|\Omega_{t-1})$$

The function $f(Y_t|S_t, \Omega_{t-1})$ is the likelihood function of Y_t conditional on the state S_t, that can be L or H, and all previous information. The function $f(Y_t|\Omega_{t-1})$ is the likelihood when the state is marginalized out. On the other hand, $\Pr(S_t|\Omega_{t-1})$ denotes the probability of being in a certain state, S_t, conditional on previous information. This probability can be computed using Hamilton filter (Hamilton 1994, Chapter 22). In the case of a model with only two states, as the one analyzed in this section, $\Pr(S_t|\Omega_{t-1})$ is given by:

$$\Pr(S_t=L|\Omega_{t-1})=(1-\pi_{H,H})+(\pi_{L,L}+\pi_{H,H}-1)\times$$

$$\times\frac{f(Y_{t-1}|S_{t-1}=L,\Omega_{t-2})\times\Pr(S_{t-1}=L|\Omega_{t-2})}{f(Y_{t-1}|S_{t-1}=L,\Omega_{t-2})\times\Pr(S_{t-1}=L|\Omega_{t-2})+f(Y_{t-1}|S_{t-1}=H,\Omega_{t-2})\times(1-\Pr(S_{t-1}=L|\Omega_{t-2}))}$$

and consequently, $\Pr(S_t=H|\Omega_{t-1})=1-\Pr(S_t=L|\Omega_{t-1})$. The long run probabilities for each state are used as initial conditions for the iterative process.

Alternatively, the estimation of $\Phi=(\theta', \phi', \psi')'$ can be done in two steps. In the first step, estimates of the mean and variance parameters are obtained from maximizing the function in (12) where the correlation matrix R_t is substituted by the identity matrix. In the second step, the estimation of the correlation parameters will be done by maximizing the log-likelihood function taking the mean and variance parameter estimates from previous step as given.

Another alternative is the estimation of $\Phi=(\theta', \phi', \psi')'$ in three steps. In the first step, estimates of the mean parameters are obtained from maximizing the function in (12) where the variance and correlation matrix R_t are assumed to be constant. In the second step, variance parameters are estimated conditional on the mean parameters obtained in the previous step, and finally, the estimation of the correlation parameters will be done by maximizing the log-likelihood function taking the mean and variance parameter estimates from the two previous steps as given.

Pelletier (2006) estimates a RSDC-GARCH model by using data on four exchange rate series. After demeaning the data, the correlation parameters are separately estimated from the variance parameters. This corresponds to what we have called the 3-step estimation procedure without paying much attention to the mean parameters or a 2-step estimation method for a zero mean series.

Finally, the asymptotic properties of the 1-step and multiple steps estimators of the RSDC-GARCH model under the Gaussianity assumption are similar and can be found in Pelletier (2006).

A summary of the well-known theoretical results about ML estimation is shown in the following table

Distribution		Estimator		
True	**Assumed**	**1-step**	**2-steps**	**3-steps**
Gaussian	Gaussian	Consistent	Consistent	Consistent
Student-t	Student-t	Consistent		
Student-t	Gaussian	Consistent	Consistent	Consistent
Gaussian	Student-t	Consistent		

Consistency of the 1-step ML estimator, under certain regularity conditions, is known from the general maximum likelihood theory. In particular, Bollerslev (1990) proves the consistency of the 1-step Gaussian ML estimator for the CCC-GARCH model, while Jeantheau (1998) proves it for the ECCC-GARCH model. Fiorentini, Sentana, and Calzolari (2003) and Newey and Steigerwald (1997) provide the conditions under which the 1-step Student-t ML estimator is consistent. The consistency of the 1-step Gaussian QML estimators is given in Bollerslev and Wooldridge (1992) and for multiple steps estimators is given in Newey and McFadden (1994). Engle and Shephard (2001) prove the consistency of Gaussian QML multiple steps estimators for the DCC-GARCH model, Aielli (2011) proves it for cDCC-GARCH model and Pelletier (2006) proves it for RSDC-GARCH model.

In the next section we will confirm the previous theoretical results in finite samples and study the cases for which no theory is provided, more specifically, what the behavior of multiple steps estimators is when a Student-t distribution is assumed for the innovations.

4 Monte Carlo experiments

In this section we analyze the finite sample performance of 1-step, 2-step and 3-step estimators of first order Vector Autoregressive CCC, DCC, cDCC, RSDC-GARCH models and their extensions to include volatility spillovers, namely ECCC, EDCC, EcDCC, ERSDC-GARCH. To compare different estimators, true parameter values are reported together with the Monte Carlo mean and standard deviation of the parameter estimates. In addition, kernel density estimates of different estimators of each parameter are plotted to compare the performance of multiple steps estimators for each sample size. Since the main interest of practitioners in this area is not only the estimation of the parameters but more importantly, the estimation of the underlying conditional variances and covariances, we will also look at the estimates of volatilities and correlations to compare different estimators. For RSDC-GARCH models the correlations are driven by an unobservable Markov chain and therefore, estimates of the correlation parameters will be analyzed instead of correlation estimates.

We have carried out Monte Carlo experiments in which 1000 time-series vectors of dimension 2 or 3 for sample sizes T=200, 500, 1000 and 5000 are generated according to the relevant model and distribution function for the innovations. Then, the parameters of the model are estimated using 1-step, 2-step and 3-step estimators assuming either a Gaussian or a Student-t distribution for the errors. All simulations are performed by MATLAB computer language.

Next, we describe in detail the three different experiments we have carried out. In the first one, we simulate time series vectors following the vector autoregressive multivariate GARCH models considered assuming first a Gaussian distribution for the innovations and then, a Student-t distribution. Parameters, volatilities and correlations are then estimated assuming the true data generating process and differences between 1-step and multiple steps estimators are analyzed. In a second experiment we anayze the robustness of the results to the error distribution. With this objective, first we simulate data assuming a Gaussian distribution for the innovations and estimate the true model under the assumption that errors follow a Student-t distribution. Second, time series vectors are generated using a Student-t distribution for the errors and then, true models are estimated under the Gaussianity assumption. In addition, we use a skewed Student-t distribution to generate the data and estimate the true model under the assumption that errors follow a symmetric distribution, Gaussian or Student-t. Finally, in the third and last experiment we analyze how good or bad volatilities and correlations generated from a given model can be estimated using a different model.

4.1 Innovations distributed as a Gaussian or Student-t

We start by considering the case in which data is generated and estimated assuming a normal distribution. Let us consider a bivariate model given by equations (1) to (3) with a diagonal matrix B and $R_t=R$ as given by the CCC-GARCH model. The unconditional mean and variance are fixed to 1. The mean and variance persistences are set to be different from each other but quite high. Therefore, in this basic bivariate model, we have 11 parameters to estimate. The true parameter values as well as Monte Carlo means and standard deviations of 1-step and multiple steps estimators are given in Table 1. Two main patterns, as expected for consistent estimators, emerge from this table. First, the differences between the Monte Carlo means and true parameter values go to zero as the sample size increases. Second, the Monte Carlo standard deviations of the three estimators considered decrease as the sample size increases. It is remarkable the similarities of the Monte Carlo means and standard deviations of the three estimators. In general, it seems that the 1-step estimator provides estimates with Monte Carlo means slightly closer to the parameter values and Monte Carlo standard deviations slightly smaller than the ones obtained for multiple-steps. However, the differences among the three estimators are practically negligible. On the other hand, we cannot conclude that in finite samples, multiple steps estimators over/under estimate the parameters in a systematic manner. In order to graphically illustrate the distribution, in finite samples, of the different estimators, Figure 1 plots kernel density estimates obtained from 1-step, 2-step and 3-step estimators for the parameter values considered in Table 1 and sample size T=500. As the figure shows, the three estimators give very similar results, even for relatively small sample sizes[2].

Table 1 Monte Carlo means and standard deviations of 1-step, 2-step and 3-step estimators of a bivariate Gaussian VAR(1)-CCC-GARCH model.

Parameter	Value	1-step			2-step			3-step		
		T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
μ_1	0.20	0.207	0.204	0.201	0.207	0.204	0.201	0.208	0.204	0.201
		(0.050)	(0.036)	(0.016)	(0.050)	(0.037)	(0.017)	(0.053)	(0.039)	(0.017)
μ_2	0.40	0.403	0.403	0.400	0.403	0.403	0.400	0.403	0.404	0.400
		(0.060)	(0.043)	(0.017)	(0.061)	(0.044)	(0.018)	(0.062)	(0.044)	(0.018)
β_{11}	0.80	0.793	0.796	0.799	0.793	0.796	0.799	0.792	0.796	0.799
		(0.028)	(0.020)	(0.009)	(0.029)	(0.020)	(0.009)	(0.030)	(0.022)	(0.010)
β_{22}	0.60	0.596	0.598	0.600	0.597	0.598	0.600	0.596	0.597	0.600
		(0.038)	(0.026)	(0.011)	(0.039)	(0.027)	(0.012)	(0.039)	(0.027)	(0.012)
ω_1	0.10	0.180	0.124	0.103	0.182	0.123	0.103	0.183	0.124	0.103
		(0.179)	(0.079)	(0.019)	(0.183)	(0.072)	(0.019)	(0.184)	(0.075)	(0.019)
ω_2	0.05	0.270	0.120	0.053	0.273	0.132	0.053	0.290	0.146	0.054
		(0.308)	(0.177)	(0.015)	(0.311)	(0.198)	(0.015)	(0.339)	(0.231)	(0.031)
α_{11}	0.10	0.108	0.103	0.099	0.109	0.103	0.099	0.106	0.102	0.099
		(0.044)	(0.030)	(0.012)	(0.044)	(0.030)	(0.013)	(0.043)	(0.030)	(0.013)
α_{22}	0.05	0.061	0.054	0.050	0.061	0.054	0.050	0.061	0.054	0.050
		(0.036)	(0.023)	(0.009)	(0.037)	(0.024)	(0.009)	(0.035)	(0.023)	(0.009)
γ_{11}	0.80	0.706	0.772	0.796	0.705	0.773	0.796	0.705	0.772	0.797
		(0.203)	(0.096)	(0.027)	(0.206)	(0.089)	(0.027)	(0.208)	(0.093)	(0.027)
γ_{22}	0.90	0.660	0.822	0.897	0.656	0.810	0.897	0.637	0.796	0.896
		(0.322)	(0.192)	(0.021)	(0.325)	(0.212)	(0.021)	(0.355)	(0.243)	(0.035)
ρ_{12}	0.20	0.199	0.201	0.200	0.198	0.199	0.200	0.198	0.199	0.200
		(0.044)	(0.031)	(0.014)	(0.043)	(0.031)	(0.014)	(0.043)	(0.031)	(0.014)

2 In order to check the robustness of the results, we have also considered different scenarios by changing the parameter values in Table 1 and repeated the Monte Carlo experiment. All the results are similar and they are not included in the paper to save space but they are available from the authors upon request.

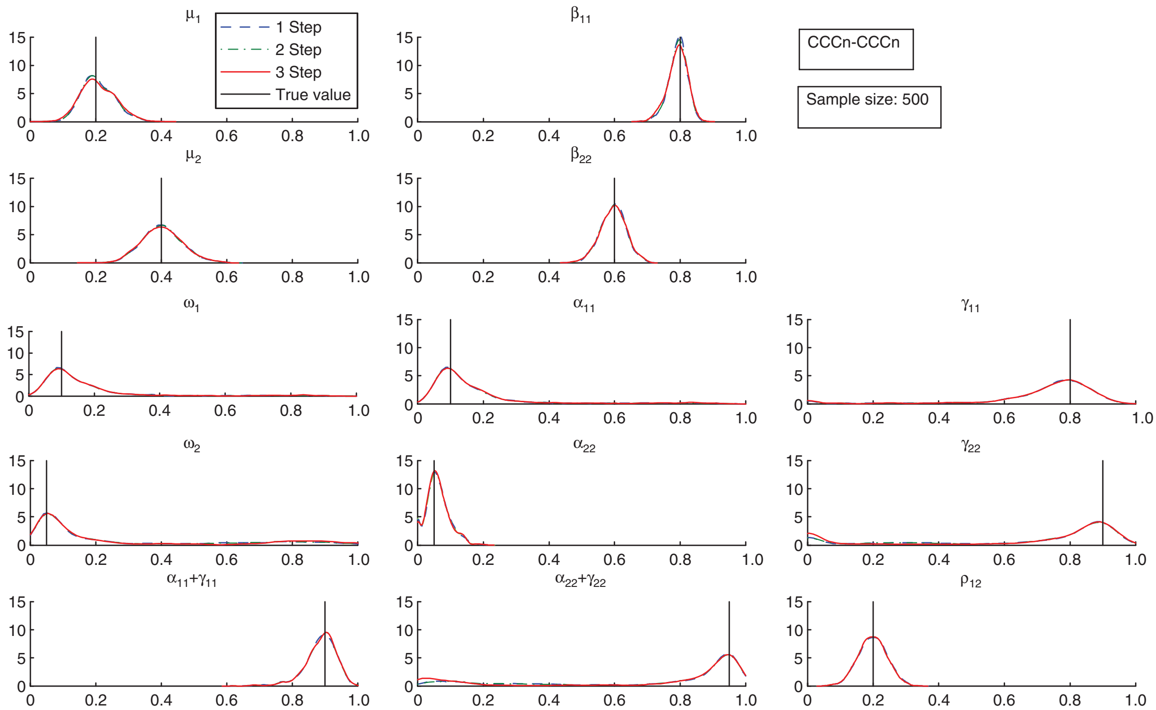

Figure 1 Kernel density estimates for estimated parameters of a Gaussian VAR(1)-CCC-GARCH(1,1) model with T=500.

We have also computed the estimated volatilities and correlations obtained from 1-step, 2-step and 3-step estimators. For a sample size T, let us denote by $\hat{h}^s_{i,t}$ the estimated volatilities of series i at time t obtained from estimator s (1-step, 2-step or 3-step) and denote by $h_{i,t}$ the true volatility of series i at time t. Then, the difference between the estimated and the true volatility of series i could be summarized for each estimator s by

$$\Delta\hat{h}^s_i = \frac{1}{T}\sum_{t=1}^{T}(\hat{h}^s_{i,t} - h_{i,t}) \tag{13}$$

Similarly, the difference between the estimated and the true correlation of series i and j could be summarized for each estimator s by

$$\Delta\hat{p}^s_{ij} = \frac{1}{T}\sum_{t=1}^{T}(\hat{p}^s_{ij,t} - p_{ij,t}) \tag{14}$$

Figure 2 plots kernel density estimates of the differences between the estimated and the true volatilities and correlations measured as in (13) and (14) for a VAR(1)-CCC-GARCH(1,1) model with parameter values as in Table 1 and sample sizes T=200, T=500 and T=1000. As the graph illustrates, 1-step, 2-steps and 3-steps estimators provide very similar estimated volatilities and correlations. As the sample size increases, differences between estimated and true volatilities and correlations are becoming closer to zero. Alternatively, we have also computed the relative deviations of the estimated volatilities and correlations from their true values, i.e., $\frac{\hat{h}^s_{i,t} - h_{i,t}}{h_{i,t}}$; $\frac{\hat{p}^s_{ij,t} - p_{ij,t}}{p_{ij,t}}$ and the corresponding plots are very similar to the ones in Figure 2.

We have repeated the Monte Carlo experiments simulating the data from different models. Kernel density estimates of the differences between the estimated and the true volatilities and correlations in VAR(1)-DCC, cDCC and RSDC-GARCH(1,1) models and their extensions were computed. All the graphs are very similar to Figure 2. Consequently, our results suggest that under normal innovations, using multiple steps estimators is

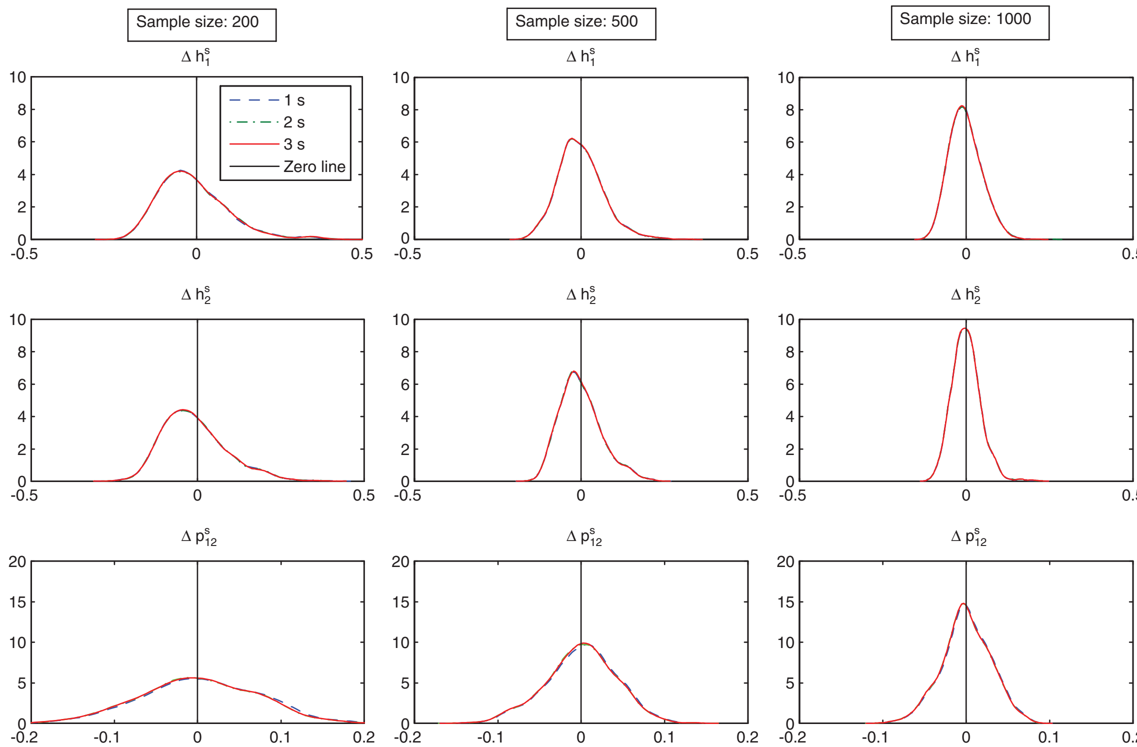

Figure 2 Kernel density estimates of deviations from estimated to true volatility and true correlation in a Gaussian VAR(1)-CCC-GARCH(1,1) model.

a reasonable strategy to estimate volatilities and correlations in all the models considered. This finding supports, for finite samples, the theoretical asymptotic results summarized in Section 3.

Next, we consider the case in which data is generated and estimated assuming a Student-t distribution and we repeat the simulations for all the models. The number of degrees of freedom used in the simulations is $\frac{1}{\eta}=5$. For DCC-GARCH, cDCC-GARCH models and their extended versions, the results are similar to the ones obtained under the normal assumption showing that 1-step, 2-step and 3-step estimators provide volatilities and correlations estimates which are very close. These findings are in line with the results in Bauwens and Laurent (2005) and Jondeau and Rockinger (2005) who show that, for the DCC-GARCH model, estimating mean and variance parameters separately from the correlation parameters provides similar outcomes to 1-step estimation.

However, for the other models considered, namely the VAR(1)-CCC-GARCH(1,1), VAR(1)-ECCC-GARCH(1,1), VAR(1)-RSDC-GARCH(1,1) and VAR(1)-ERSDC-GARCH(1,1) models, important differences appear when estimating the correlations (or correlation parameters and transition probabilities for the RSDC and ERSDC-GARCH model) with different estimators. In this case, 1-step estimator provides the best estimates. Figure 3 plots kernel density estimates of the differences between the estimated and the true volatilities and correlations in the VAR(1)-CCC-GARCH(1,1) model. Volatilities and correlations seem to be underestimated when using multiple steps estimators. The figure corresponding to the VAR(1)-ECCC-GARCH model is very similar to Figure 3. For the RSDC-GARCH model, Figure 4 contains kernel density estimates of the differences between the estimated and the true volatilities and of the correlation parameters and the transition probabilities, instead of differences from estimated to true correlations. As we can see, estimates obtained with multiple steps estimators seem to be far from the ones obtained with the 1-step estimator. The figure corresponding to the VAR(1)-ERSDC-GARCH model is very similar to Figure 4. Therefore, our results suggest that multiple steps estimators obtained assuming that innovations are distributed as a Student-t, do not have good properties, even when the assumption for the innovations is the correct one. This finding points out that multiple steps estimators computed under the assumption that the distribution of the innovations is Student-t, could be inconsistent.

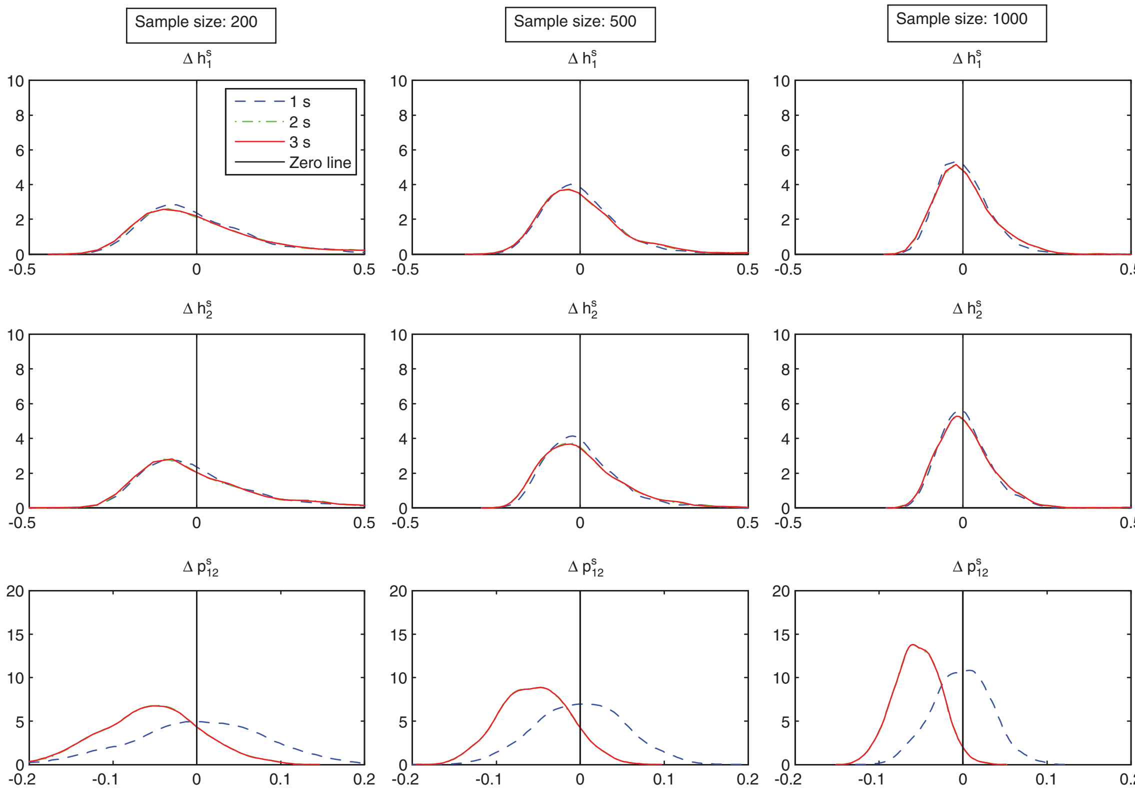

Figure 3 Kernel density estimates of deviations from estimated to true volatility and true correlation in a VAR(1)-CCC-GARCH(1,1) model with Student-t innovations.

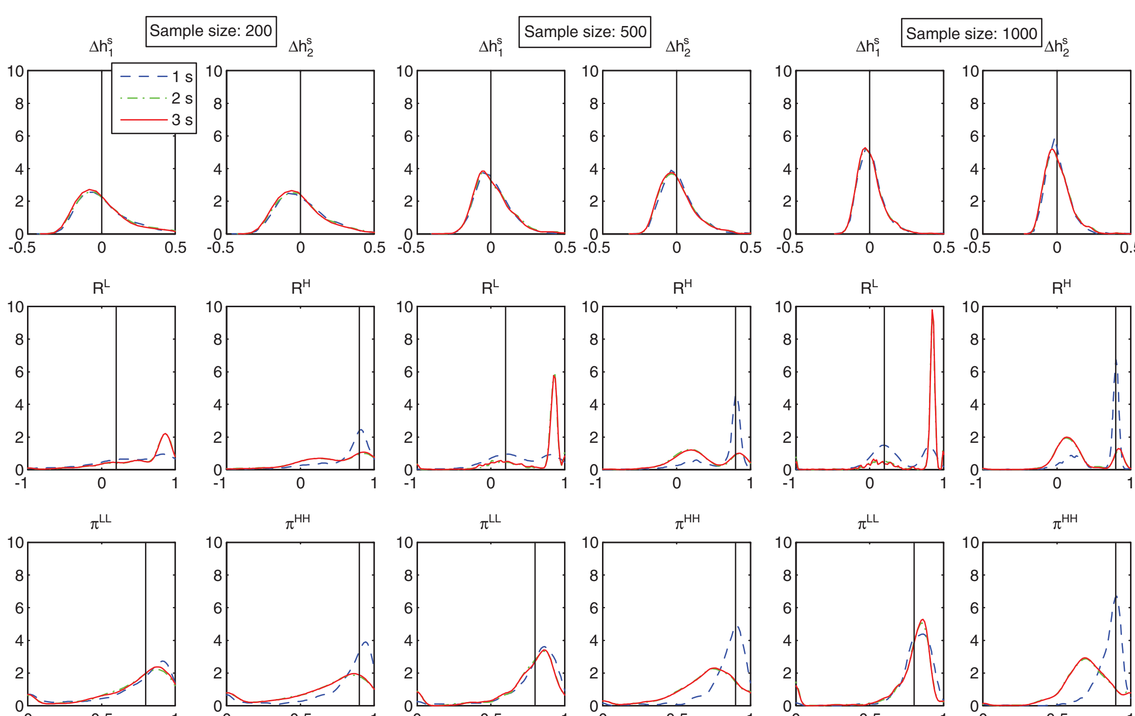

Figure 4 Kernel density estimates of deviations from estimated to true volatility, of estimated correlation parameters and of estimated transition probabilities in a VAR(1)-RSDC-GARCH(1,1) model with Student-t innovations.

The comparison between 1-step and 2-step estimators helps us to measure the efficiency loss when estimating the correlation parameters separately from the mean and variance parameters; see Engle (2002) and Engle and Sheppard (2001). As we have seen, when the errors are assumed to be Gaussian, the small sample behavior of 1-step and 2-step estimators is very similar but, when the estimation is based on the Student-t distribution, in some cases 2-step estimators deviate from 1-step estimators.

Comparing 2-step and 3-step estimators helps us to analyze the effects of separating the estimation of mean and variance parameters; see Bauwens, Laurent, and Rombouts (2006). Our results show that, when the errors are assumed to be Gaussian or Student-t, the small sample behavior of 2-step and 3-step estimators is also very similar.

4.2 Robustness to the error distribution

We are also interested in analyzing how robust the different models and estimators are to the distribution of innovations. In that sense, we have carried out an experiment which consists of generating data from the models considered with errors following a Gaussian distribution and estimating the true model assuming a Student-t distribution for the innovations. Also, we simulate data in which innovations follow a Student-t distribution and estimate the true model assuming Gaussian errors. Finally, we generate the data using a skewed Student-t distribution and estimate the true model under the assumption that errors follow a symmetric distribution, Gaussian or Student-t. The results of the first two experiments are the expected ones. When the data is generated using a Student-t distribution for the errors and estimated assuming Gaussian errors, differences between 1-step and multiple steps seem to be negligible. Comparing this case to the case in which the true and assumed error distributions are both normal, the estimated densities now have fatter tails. When we simulate the data with Gaussian errors and estimate the model under the Student-t distribution assumption, similar results are obtained to the case when the true and the assumed distribution are both Student-t. This was expected since a Student-t distribution approximates a Gaussian distribution when the parameter for the degrees of freedom is sufficiently large. In fact, in the experiments for this case, we obtained very large estimates for the degrees of freedom of the Student-t distribution.

Next, we analyze the case in which innovations follow a skewed distribution. For this purpose, we generate random vectors from a skewed multivariate Student-t distribution following Bauwens and Laurent (2005). At each time t, a k dimensional random vector η_t^* is given by:

$$\eta_t^* = \lambda(\tau)|x_t|$$
$$|x_t| = (|x_{1t}|, |x_{2t}|, \dots, |x_{kt}|)'$$

where x_t follows a multivariate Student-t distribution with zero mean and unit variance and $\lambda(\tau)$ is a $k\times k$ diagonal matrix such that:

$$\begin{aligned}
\lambda(\tau) &= \tau\Xi - (I_k - \tau)\Xi^{-1}\\
\Xi &= diag(\xi)\\
\xi &= (\xi_1, \xi_2, \dots, \xi_k),\ \text{with}\ \xi_i > 0\\
\tau &= diag(\tau_1, \tau_2, \dots, \tau_k),\ \text{with}\ \tau_i \in \{0, 1\}\\
\tau_i &\sim Ber\left(\frac{\xi_i^2}{1+\xi_i^2}\right)
\end{aligned}$$

where $Ber\left(\frac{\xi_i^2}{1+\xi_i^2}\right)$ is a Bernoulli distribution with probability of success $\frac{\xi_i^2}{1+\xi_i^2}$ and the elements of τ are mutually independent. Given that in the GARCH set up, the elements of η_t are zero mean random numbers with unit variance, η_t^* should be standardized such that $\eta_{it} = \frac{\eta_{it}^* - m_i}{s_i}$ where:

$$m_i = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(\xi_i - \frac{1}{\xi_i}\right)$$

$$s_i^2 = \left(\xi_i^2 - \frac{1}{\xi_i^2} - 1\right) - m_i^2.$$

To perform this experiment and the experiment of Section 4.3, we take as parameter values the estimates obtained from real data where we have considered daily returns of three European stock market indices, BEL-20 (Brussels), DAX (Frankfurt) and FTSE-100 (London) for the period January 8, 2002–April 30, 2009. The table below contains some descriptive statistics of the returns series, computed as $100\times\log\left(\frac{p_t}{p_{t-1}}\right)$, of sample size 1774.

	Mean	SD	Skewness	Kurtosis
BEL-20	−0.02	1.45	−0.05	9.12
DAX	−0.01	1.74	0.15	7.80
FTSE-100	−0.01	1.41	−0.03	10.30

Using the first two returns series, we estimate all the models considered, i.e., VAR(1)-CCC, DCC, cDCC and RSDC-GARCH models and their extensions [VAR(1)-ECCC, EDCC, EcDCC, ERSDC-GARCH] with no mean transmissions under the assumption that innovations are distributed as a Student-t. The results are given in Table 2 where series 1 and 2 correspond to BEL-20 and DAX, respectively.

We first generate bivariate series with skewness parameters $\xi_1=\xi_2=\exp(0.4)$ for both series, which implies a skewness of 1.5. Later we take $\xi_1=\exp(0.4)$ and $\xi_2=\exp(-0.7)$ (implying a skewness of –2 for the second series) to see how the results change. Notice that when $\xi_1=\xi_2=1$, we have a symmetric multivariate Student-t distribution.

For each model considered, 1000 bivariate time series vectors of sample sizes T=500 and T=1000 have been generated assuming that innovations follow a skewed Student-t distribution with skewness 1.5 for both series or with skewness {1.5, –2} for the first and second series, respectively and with degrees of freedom 5. Then, the true model is estimated assuming Gaussian or Student-t errors but ignoring skewness.

Figure 5 plots kernel density estimates of the differences between the estimated and the true volatilities and true correlations in five of the models considered when data has been generated using a positively skewed Student-t distribution with the same skewness for both series and estimated assuming Gaussian innovations. In this figure, where T=500, the rows correspond to a different model and the columns represent the kernel densities estimates of the relative deviations of estimated volatilities and correlations from the true ones calculated respectively as:[3]

$$\Delta\hat{h}_i^s = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\hat{h}_{i,t}^s - h_{i,t}}{h_{i,t}}\right\} \tag{15}$$

$$\Delta\hat{p}_{ij}^s = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\hat{p}_{ij,t}^s - p_{ij,t}}{p_{ij,t}}\right\}. \tag{16}$$

As we can see, for the five models in the figure, the kernel densities of the relative deviations of 1-step and multiple steps estimates of volatilities (correlations) from the true ones follow each other closely. It seems that the large positive skewness assumed in the data generating process results in overestimating the conditional correlations while the conditional volatility estimates do not seem to be affected much.

3 Relative deviations are prefered to absolute ones, although conclusions do not change if absolute deviations are plotted.

Table 2 One-step parameter estimates of different models fitted to two real time series assuming Student-t innovations.

	VAR(1)-CCC-GARCH	VAR(1)-DCC-GARCH	VAR(1)-cDCC-GARCH	VAR(1)-RSDC-GARCH
μ_1	0.0936	0.0956	0.0963	0.0920
μ_2	0.1090	0.1143	0.1136	0.1100
β_{11}	0.0087	0.0002	0.0006	0.0030
β_{22}	−0.0503	−0.0488	−0.0469	−0.0485
ω_1	0.0185	0.0156	0.0149	0.0160
ω_2	0.0191	0.0152	0.0141	0.0156
α_{11}	0.0877	0.0963	0.0978	0.0894
α_{22}	0.0732	0.0775	0.0790	0.0717
γ_{11}	0.8987	0.8938	0.8950	0.8998
γ_{22}	0.9195	0.9170	0.9181	0.9218
ρ_{12}	0.7950			
ρ_{12}^L				0.7298
ρ_{12}^H				0.8924
δ_1		0.0376	0.0390	
δ_2		0.9453	0.9465	
π^{LL}				0.9816
π^{HH}				0.9682
	VAR(1)-ECCC-GARCH	**VAR(1)-EDCC-GARCH**	**VAR(1)-EcDCC-GARCH**	**VAR(1)-ERSDC-GARCH**
μ_1	0.0937	0.0958	0.0966	0.0924
μ_2	0.1093	0.1147	0.1140	0.1102
β_{11}	0.0088	−0.0006	−0.0007	0.0025
β_{22}	−0.0498	−0.0494	−0.0476	−0.0486
ω_1	0.0177	0.0137	0.0120	0.0140
ω_2	0.0162	0.0123	0.0107	0.0129
α_{11}	0.0947	0.1038	0.1039	0.0950
α_{21}	0.0034	0.0113	0.0125	0.0075
α_{12}	0.0001	0.0000	0.0000	0.0000
α_{22}	0.0692	0.0685	0.0681	0.0660
γ_{11}	0.8789	0.8831	0.8858	0.8902
γ_{21}	0.0001	0.0000	0.0000	0.0000
γ_{12}	0.0086	0.0042	0.0041	0.0043
γ_{22}	0.9225	0.9192	0.9205	0.9235
ρ_{12}	0.7949			
ρ_{12}^L				0.7280
ρ_{12}^H				0.8921
δ_1		0.0388	0.0411	
δ_2		0.9448	0.9474	
π^{LL}				0.9814
π^{HH}				0.9686

Figure 6 plots the same estimates as Figure 5 but now the estimation has been done assuming a Student-t distribution for the innovations. Similar conclusions can be made about the 1-step correlation estimates for all models. We notice that in the CCC and ECCC-GARCH models, the differences between 1-step and multiple steps estimates of the correlations are very large.

On the other hand, when the series have different skewness and the estimation is performed assuming Gaussian errors, volatilities and correlations seem to be underestimated in all these five models. The figures corresponding to different skewness are not included in the paper to save space. One-step correlation estimates seem to be slightly less affected by the skewness than the multiple step estimates. As well when the estimation is based on Student-t errors, the 1-step estimators underestimate the volatilities and correlations. In general, 1-step estimators are less affected by the skewness than multiple steps estimators, except for the volatility estimates of ECCC-GARCH model. In the case of DCC and cDCC-GARCH models, the multiple steps estimates deviate slightly from the 1-step estimates. It should be noted that one of the series have higher

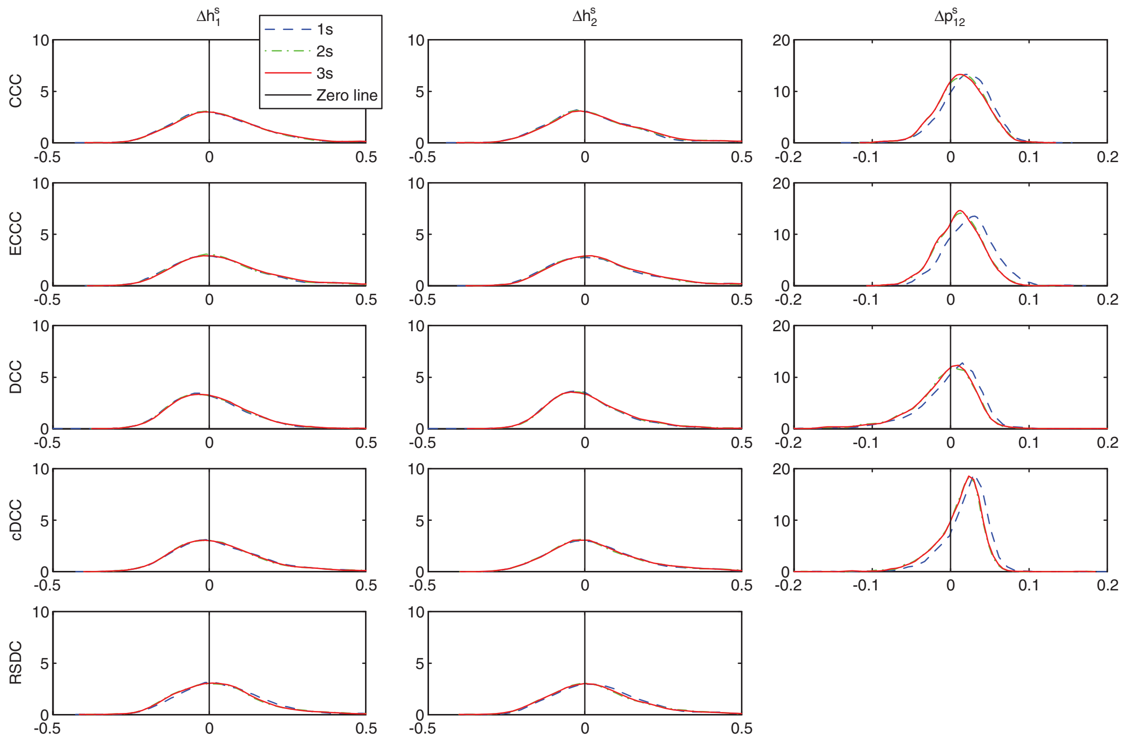

Figure 5 Kernel density estimates of deviations from estimated to true volatility and true correlation for some of the models considered. Series have been generated assuming Student-t innovations with same skewness parameter and models have been estimated assuming Gaussian innovations. T=500.

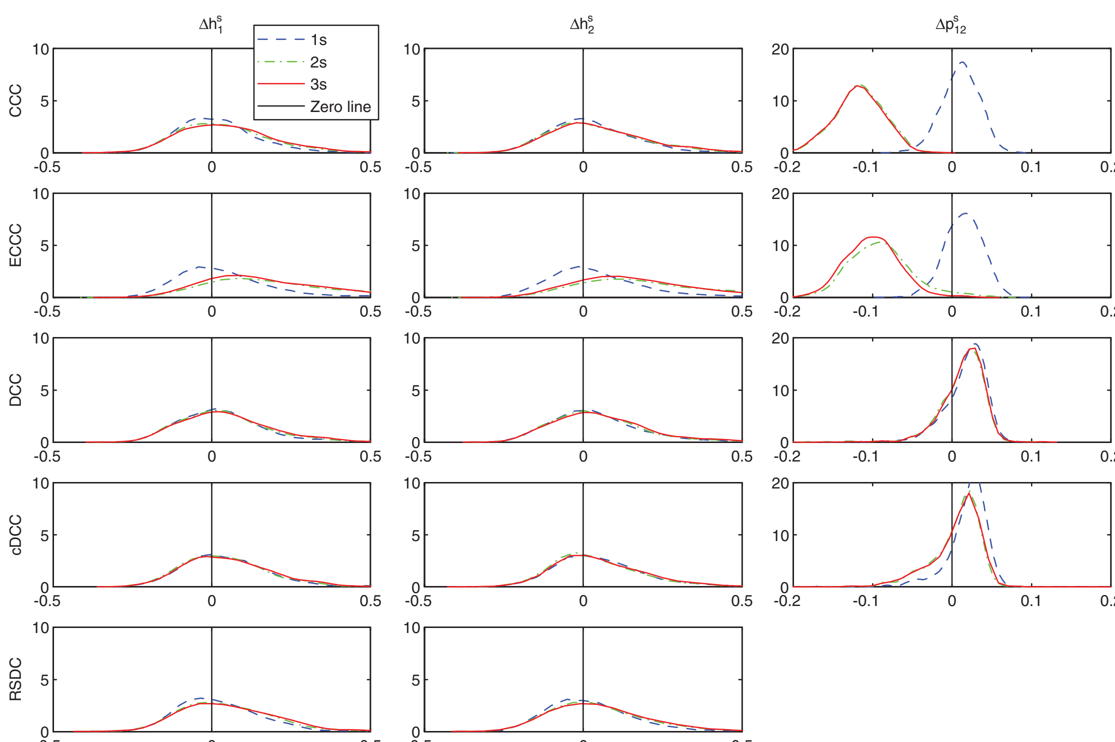

Figure 6 Kernel density estimates of deviations from estimated to true volatility and true correlation for some of the models considered. Series have been generated assuming Student-t innovations with same skewness parameter and models have been estimated assuming Student-t innovations. T=500.

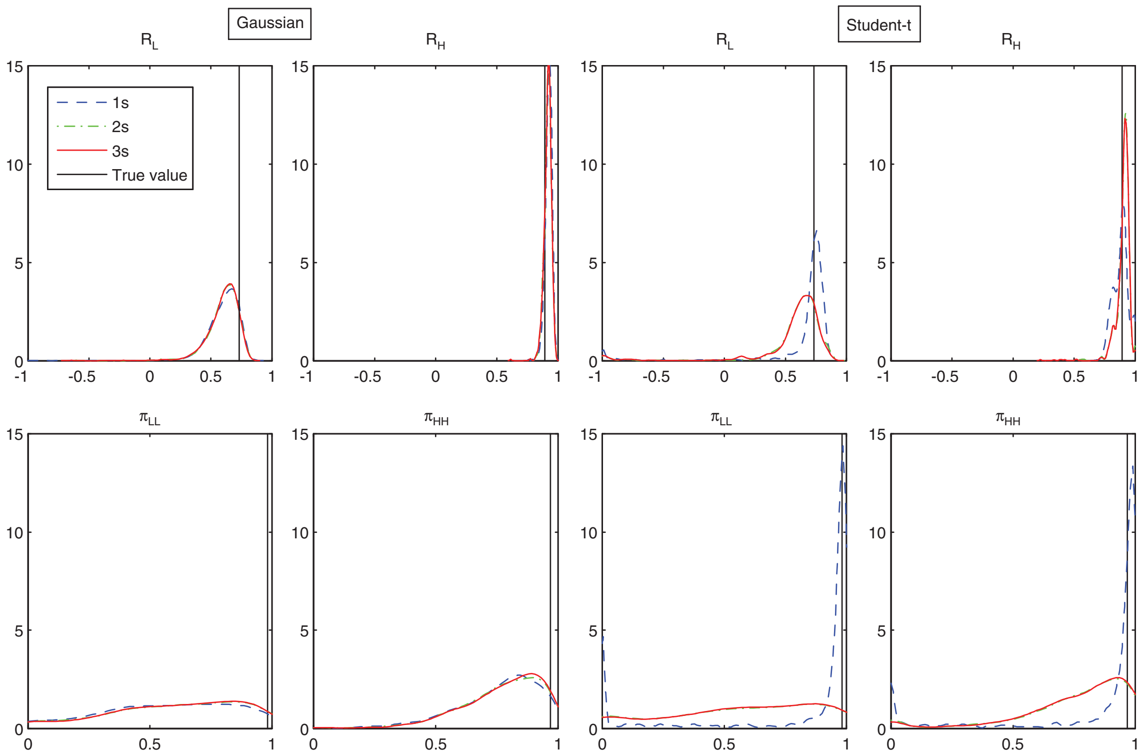

Figure 7 Kernel density estimates of estimated correlation parameters for the RSDC-GARCH model. Series have been generated assuming Student-t innovations with same skewness, and models have been estimated assuming Gaussian and Student-t errors, respectively. T=500.

skewness when λ={exp(0.4), exp(−0.7)} compared to the case when λ={exp(0.4), exp(0.4)} and this could be the reason behind the underestimation of volatilities and correlations with both Gaussian and Student-t errors.

For the RSDC-GARCH model, multiple steps estimators of conditional volatilities behave similar to the 1-step estimators as illustrated in Figure 4 and this does not seem to depend on the skewness. Figure 7 plots kernel density estimates of estimated correlation parameters when the series have the same skewness and errors are assumed to follow a Gaussian or Student-t distribution. As we can see, when the estimations are based on Gaussian errors, the 1-step and multiple steps estimators of the correlation parameters are behaving similarly. Although the corresponding figure is not included in the paper, when the skewnesses of the two series are different, the multiple steps estimates of R_L and π_{LL} deviate slightly from the 1-step estimates. When Student-t errors are used in the estimation, the differences between the behavior of 1-step and multiple steps estimators become more apparent.

For the extended models, results are very similar and also results do not change when T=1000. To save space, the corresponding figures are not included in the paper, but they are available from authors upon request.

Newey and Steigerwald (1997) argue that when the true distribution is not symmetric, the 1-step estimator based on Student-t innovations is not consistent in general. This is confirmed in our experiments since when T=500, we can see in Figure 6 that 1-step QML estimators based on Student-t errors are overestimating the correlations, and this overestimation does not disappear as the sample size increases.

Finally, when the data generating process is symmetric and the estimation is based on Gaussian errors, the kernel density estimates of relative differences between 1-step and multiple steps estimates of the volatilities and correlations from the true values are very close to each other for all the models and their extensions as mentioned before. When the estimation is based on Student-t errors, the multiple steps volatility estimates of all extended models and the multiple steps correlation estimates of CCC and ECCC-GARCH models are far

Table 3 Parameter estimates of different models fitted to three real time series assuming Gaussian innovations.

	VAR(1)-CCC-GARCH			VAR(1)-DCC-GARCH			VAR(1)-cDCC-GARCH			VAR(1)-RSDC-GARCH		
	1-Step	2-Steps	3-Steps	1-Step	2-Steps	3-Steps	1-Step	2-Steps	3-Steps	1-Step	2-Steps	3-Steps
μ_1	0.1017	0.0937	−0.0167	0.0982	0.0937	−0.0167	0.1008	0.0936	−0.0166	0.0851	0.0936	−0.0166
μ_2	0.1110	0.0843	−0.0061	0.1041	0.0843	−0.0061	0.1065	0.0843	−0.0062	0.0927	0.0843	−0.0061
μ_3	0.0690	0.0465	−0.0128	0.0646	0.0463	−0.0128	0.0669	0.0465	−0.0128	0.0611	0.0465	−0.0128
β_{11}	−0.0297	−0.0055	0.0641	−0.0277	−0.0055	0.0640	−0.0278	−0.0055	0.0641	−0.0245	−0.0055	0.0640
β_{22}	−0.0934	−0.0549	−0.0465	−0.0748	−0.0549	−0.0466	−0.0738	−0.0549	−0.0465	−0.0743	−0.0549	−0.0466
β_{33}	−0.0991	−0.1033	−0.0821	−0.0938	−0.1027	−0.0821	−0.0935	−0.1034	−0.0820	−0.0888	−0.1034	−0.0820
ω_1	0.0288	0.0218	0.0210	0.0248	0.0218	0.0209	0.0240	0.0218	0.0210	0.0233	0.0218	0.0210
ω_2	0.0259	0.0210	0.0201	0.0226	0.0210	0.0201	0.0213	0.0210	0.0201	0.0212	0.0210	0.0201
ω_3	0.0171	0.0102	0.0097	0.0143	0.0094	0.0098	0.0131	0.0102	0.0097	0.0173	0.0102	0.0098
α_{11}	0.0940	0.1331	0.1231	0.1118	0.1331	0.1231	0.1183	0.1331	0.1231	0.0874	0.1331	0.1231
α_{22}	0.0774	0.0955	0.0927	0.0887	0.0955	0.0927	0.0933	0.0955	0.0927	0.0687	0.0955	0.0927
α_{33}	0.0777	0.1045	0.1041	0.0935	0.0938	0.1041	0.0985	0.1045	0.1041	0.0779	0.1045	0.1041
γ_{11}	0.8819	0.8573	0.8673	0.8705	0.8573	0.8673	0.8695	0.8573	0.8673	0.8956	0.8573	0.8673
γ_{22}	0.9095	0.8977	0.9011	0.9005	0.8977	0.9011	0.9004	0.8977	0.9011	0.9217	0.8977	0.9011
γ_{33}	0.9063	0.8921	0.8936	0.8948	0.9015	0.8936	0.8948	0.8921	0.8936	0.9088	0.8921	0.8936
ρ_{12}	0.7911	0.7865	0.7866									
ρ_{12}^L										0.6571	0.6674	0.6455
ρ_{12}^H										0.8782	0.8802	0.8773
ρ_{13}	0.7751	0.7642	0.7644									
ρ_{13}^L										0.6286	0.6286	0.6060
ρ_{13}^H										0.8695	0.8702	0.8658
ρ_{23}	0.8050	0.8013	0.8016									
ρ_{23}^L										0.6477	0.6633	0.6371
ρ_{23}^H										0.9054	0.9087	0.9063
δ_1				0.0411	0.0459	0.0494	0.0405	0.0439	0.0449			
δ_2				0.9215	0.9172	0.9116	0.9272	0.9226	0.9217			
π^{LL}										0.8718	0.8934	0.8681
π^{HH}										0.9272	0.9213	0.9207

from the true ones. The multiple steps volatility and correlation estimates of DCC and cDCC-GARCH models follow closely the 1-step estimates and are not far from the true values as in Figure 2. All the figures with the results are not reported in the paper, but are available from the authors upon request.

To sum up, our results suggest that even though the data generating process is skewed, when the estimation is based on Gaussian errors, multiple-steps estimators could still be preferred to 1-step estimators given that their performances are very similar. On the other hand, as noted by Newey and Steigerwald (1997), if the data generating process is skewed, the 1-step QML estimator based on Student-t errors is not consistent. Hence, when the true distribution is skewed, one should be cautious in using 1-step or multiple-steps estimators based on Student-t errors.

4.3 Robustness to model

The next question we address is how bad (or well) volatilities and correlations can be estimated when the model is misspecified. We analyze the differences between true conditional volatilities and correlations and the estimated ones when the model used to generate the data is different from the estimated model. To perform this experiment, parameter values are taken from the estimates obtained from daily returns of the three European stock market indices considered before. Using the returns series, we estimate all the models considered, i.e., VAR(1)-CCC, DCC, cDCC and RSDC-GARCH models and their extensions [VAR(1)-ECCC, EDCC, EcDCC, ERSDC-GARCH] with no mean transmissions assuming Gaussian errors. The results are given in Tables 3 and 4 in which series 1, 2 and 3 correspond to BEL-20, DAX and FTSE-100, respectively.

Table 4 Parameter estimates of different models fitted to three real time series assuming Gaussian innovations (continued).

	VAR(1)-ECCC-GARCH			VAR(1)-EDCC-GARCH			VAR(1)-EcDCC-GARCH			VAR(1)-ERSDC-GARCH		
	1-Step	2-Steps	3-Steps	1-Step	2-Steps	3-Steps	1-Step	2-Steps	3-Steps	1-Step	2-Steps	3-Steps
μ_1	0.0997	0.1021	−0.0167	0.1119	0.1021	−0.0167	0.1232	0.1021	−0.0167	0.0817	0.1021	−0.0167
μ_2	0.1097	0.0705	−0.0061	0.1184	0.0706	−0.0061	0.1333	0.0706	−0.0061	0.0892	0.0706	−0.0061
μ_3	0.0683	0.0472	−0.0128	0.0806	0.0471	−0.0128	0.0821	0.0471	−0.0129	0.0595	0.0471	−0.0128
β_{11}	−0.0267	−0.0160	0.0641	−0.0283	−0.0160	0.0641	−0.0359	−0.0160	0.0641	−0.0237	−0.0160	0.0641
β_{22}	−0.0903	−0.0453	−0.0466	−0.0686	−0.0453	−0.0466	−0.0805	−0.0453	−0.0465	−0.0739	−0.0453	−0.0466
β_{33}	−0.0956	−0.1023	−0.0820	−0.0861	−0.1022	−0.0820	−0.0983	−0.1023	−0.0821	−0.0871	−0.1022	−0.0820
ω_1	0.0322	0.0210	0.0196	0.0309	0.0210	0.0196	0.0305	0.0210	0.0195	0.0244	0.0210	0.0195
ω_2	0.0245	0.0183	0.0172	0.0292	0.0183	0.0172	0.0266	0.0183	0.0172	0.0189	0.0183	0.0172
ω_3	0.0168	0.0105	0.0097	0.0161	0.0105	0.0096	0.0168	0.0105	0.0096	0.0150	0.0105	0.0096
α_{11}	0.0772	0.1081	0.0767	0.1023	0.1082	0.0767	0.1140	0.1081	0.0766	0.0690	0.1082	0.0766
α_{21}	0.0032	0.0331	0.0249	0.0177	0.0331	0.0249	0.0183	0.0331	0.0248	0.0066	0.0331	0.0248
α_{31}	0.0224	0.0337	0.0228	0.0377	0.0338	0.0228	0.0404	0.0338	0.0228	0.0128	0.0338	0.0228
α_{12}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
α_{22}	0.0794	0.0676	0.0676	0.0941	0.0676	0.0676	0.1001	0.0677	0.0676	0.0689	0.0676	0.0676
α_{32}	0.0038	0.0000	0.0000	0.0036	0.0000	0.0000	0.0039	0.0000	0.0000	0.0000	0.0000	0.0000
α_{13}	0.0476	0.0488	0.0742	0.0672	0.0487	0.0742	0.0657	0.0487	0.0742	0.0502	0.0487	0.0742
α_{23}	0.0000	0.0000	0.0052	0.0000	0.0000	0.0052	0.0000	0.0000	0.0053	0.0000	0.0000	0.0053
α_{33}	0.0599	0.0748	0.0832	0.0847	0.0748	0.0832	0.0882	0.0748	0.0832	0.0717	0.0747	0.0832
γ_{11}	0.8399	0.7986	0.8360	0.8055	0.7985	0.8360	0.8012	0.7985	0.8364	0.8507	0.7985	0.8364
γ_{21}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
γ_{31}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
γ_{12}	0.0083	0.0000	0.0000	0.0079	0.0000	0.0000	0.0078	0.0000	0.0000	0.0000	0.0000	0.0000
γ_{22}	0.9065	0.9023	0.9054	0.8777	0.9022	0.9054	0.8749	0.9022	0.9054	0.9181	0.9022	0.9054
γ_{32}	0.0029	0.0000	0.0000	0.0027	0.0000	0.0000	0.0027	0.0000	0.0000	0.0000	0.0000	0.0000
γ_{13}	0.0000	0.0453	0.0137	0.0000	0.0454	0.0137	0.0008	0.0454	0.0134	0.0198	0.0454	0.0134
γ_{23}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
γ_{33}	0.8883	0.8838	0.8885	0.8532	0.8837	0.8885	0.8463	0.8837	0.8885	0.9036	0.8837	0.8885
ρ_{12}	0.7921	0.7853	0.7870									
ρ_{12}^L										0.6634	0.6721	0.6667
ρ_{12}^H										0.8818	0.8823	0.8819
ρ_{13}	0.7770	0.7677	0.7670									
ρ_{13}^L										0.6365	0.6416	0.6295
ρ_{13}^H										0.8714	0.8743	0.8734
ρ_{23}	0.8054	0.7990	0.8013									
ρ_{23}^L										0.6570	0.6688	0.6611
ρ_{23}^H										0.9074	0.9099	0.9098
δ_1				0.0592	0.0483	0.0465	0.0570	0.0455	0.0432			
δ_2				0.8985	0.9249	0.9278	0.9043	0.9296	0.9342			
π^{LL}										0.8836	0.9089	0.8941
π^{HH}										0.9257	0.9258	0.9214

As we can see in both Tables 3 and 4, 3-steps estimates of the mean parameters are the same, as expected, since the mean equation is the same for all the models. Correlation estimates for the CCC and ECCC models are also very similar. The correlation parameter estimates of the dynamic correlation models are significantly different from zero, suggesting that correlations are not constant during this period. When looking at the other parameters, as expected, the differences between 1-step, 2-step and 3-step estimates are not very large. Figures 8 and 9 plot the volatilities and correlations estimates, respectively using different models. Similar plots were obtained for the EDCC, EcDCC and ERSDC models and are available from the authors upon request. We can see that the three different estimators provide very similar estimates. The graphs containing the correlation estimates obtained from DCC and cDCC models suggest that the correlation between the returns of these markets in the period analyzed has been changing over time.

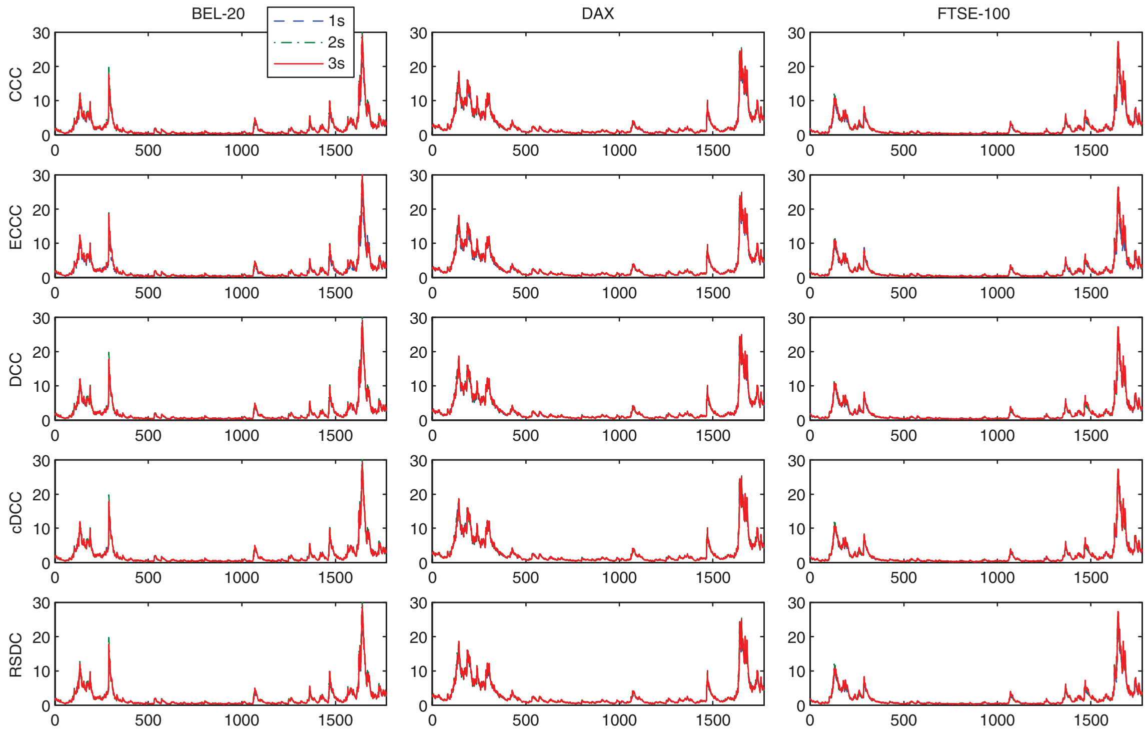

Figure 8 1-Step, 2-step and 3-step estimates of the volatilities of BEL-20, DAX and FTSE-100 observed from January 8, 2002 to April 30, 2009, assuming Gaussian innovations.

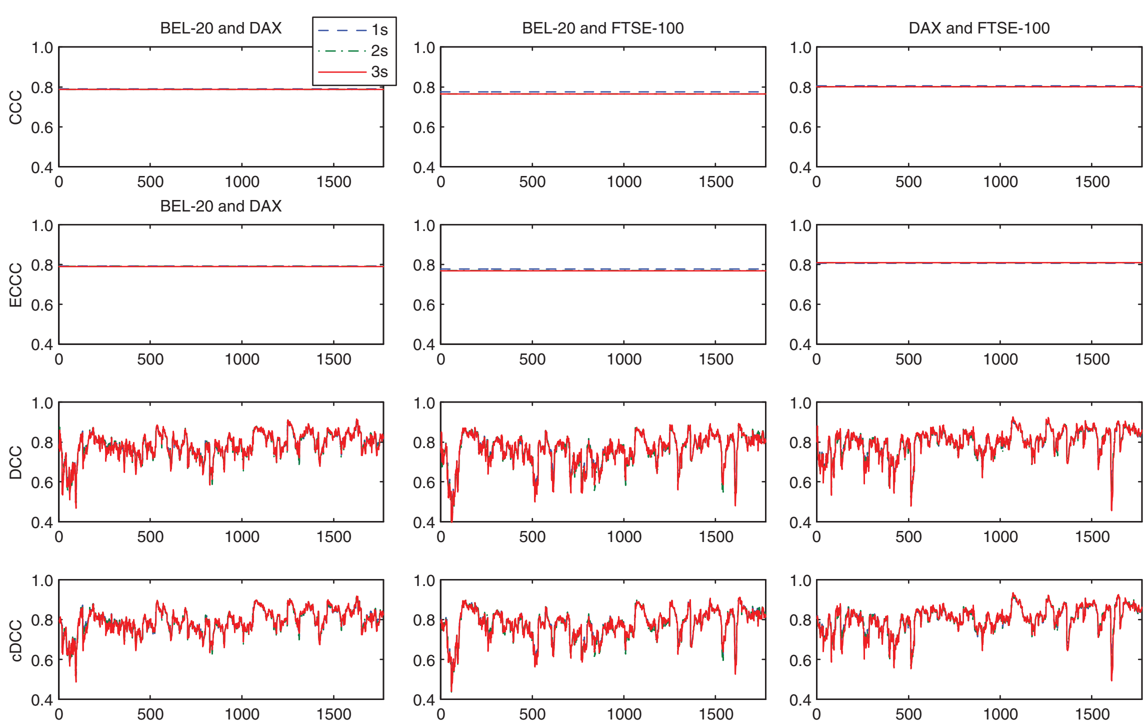

Figure 9 1-Step, 2-step and 3-step estimates of the correlations between the returns on BEL-20, DAX and FTSE-100 indices observed from January 8, 2002 to April 30, 2009, assuming Gaussian innovations.

For the Monte Carlo experiments, we take the 1-step estimates obtained in this empirical exercise as the true parameter values to generate the data sets. The first set of experiments considers VAR(1)-CCC, ECCC, DCC, cDCC and RSDC-GARCH models. For each model, we generate 1000 trivariate time series vectors of sample size 1000 and given each of the time series vectors, we estimate the five models considered. We perform these 25 experiments assuming a Gaussian error distribution for generating the data and also for estimating the parameters. In a second set of experiments the extended models, VAR(1)-ECCC, EDCC, EcDCC and ERSDC-GARCH models are considered in a similar way counting up to 15 more experiments.

The results are reported in Tables 5 and 6, in which the models used to generate the data appear in the first column and the estimated models are in the second row. For each series, each replication and at each time t, the relative and absolute relative differences between estimated volatility (correlation) and true[4] volatility (correlation) are calculated and then the average is computed across the number of series k, replications R and sample size T. For example, for the volatilities, the relative and the absolute relative difference between the estimated and the true ones is given by

$$ratio_{h,est}^{true}=\frac{1}{TRk}\sum_{t=1}^{T}\sum_{r=1}^{R}\sum_{i=1}^{k}\left(\frac{\hat{h}_{i,t}^{r}-h_{i,t}}{h_{i,t}}\right) \text{ and } |ratio|_{h,est}^{true}=\frac{1}{TRk}\sum_{t=1}^{T}\sum_{r=1}^{R}\sum_{i=1}^{k}\left|\frac{\hat{h}_{i,t}^{r}-h_{i,t}}{h_{i,t}}\right| \quad (17)$$

where in our case, k=3, R=1000 and T=1000. We also calculate, in a similar manner, the ratio for the covariance estimates $\hat{h}_{ij,t}$, where $i\neq j$. The ratios corresponding to the 1-step estimation of a model that is correctly specified is, by construction, equal to 0. Therefore, the ratios reported in Tables 5 and 6 are relative ratios and they should be read as a measure of the performance of the corresponding estimator in a certain model when estimating the volatility (correlation or covariance), relative to the 1-step estimator in the correctly specified model. The results are reported in three parts: volatilities, correlations and covariances. In general, we can see that the ratios and absolute ratios are close to zero, indicating that, on average, volatilities and correlations are relatively well estimated even when using a misspecified model.

When we look at the volatility ratios in Table 5, we can see that the largest ratio is 0.0155 and it appears when the true volatilities are generated by the VAR(1)-ECCC-GARCH model and estimated by the VAR(1)-DCC-GARCH in 3-step. Other large ratios correspond to the 3-steps estimators of all the models considered when the data have been generated by the VAR(1)-ECCC-GARCH model. On the other hand, the largest absolute ratio for the volatility is 0.0291 and it appears when the true volatilities are generated by the VAR(1)-RSDC-GARCH model and estimated by VAR(1)-ECCC-GARCH model in three steps; followed by 0.0259 and 0.0245 which correspond to the 3 steps VAR(1)-ECCC-GARCH estimates of the volatilities generated by the VAR(1)-DCC-GARCH and VAR(1)-cDCC-GARCH models. These results are not surprising and suggest that it seems hard for the constant correlation models to capture the volatilities generated by the dynamic correlation models and also for the dynamic correlation models to capture the volatilities generated by the VAR(1)-ECCC-GARCH model. When looking at Table 6, where the results are presented for the extended models, similar conclusions arise. Although, now all models allow for volatility spillovers, we can see that the largest ratio and absolute ratio appear when the true volatilities are generated by the VAR(1)-EDCC-GARCH model and estimated by the VAR(1)-ECCC-GARCH.

When we look at the correlation ratios in Table 5, results are again, the expected ones. We can see that the largest ratio (in modulus) is −0.0064 and it appears when the true correlations are generated by the VAR(1)-ECCC-GARCH model and estimated by the VAR(1)-DCC-GARCH in 3 steps. The largest absolute ratios are obtained when the true correlations are dynamic and they are estimated by a constant correlation model. On the other hand, the absolute ratios for the correlations estimated by the VAR(1)-DCC-GARCH and VAR(1)-cDCC-GARCH models are relatively much closer to zero; implying that they can capture well the correlations generated by the constant correlation models and also the correlations generated by each other. A similar comment can be made for the VAR(1)-CCC-GARCH and VAR(1)-ECCC-GARCH models. Table 6 presents the results for the extended models and similar conclusions can be reached.

4 In order to better interpret the numbers, true volatility is computed substituting the true parameter values by the 1-step estimates of the correct model.

Table 5 Volatility, covariance and correlation ratios and absolute ratios.

	Estimated model														
	VAR(1)-CCC-GARCH			VAR(1)-ECCC-GARCH			VAR(1)-DCC-GARCH			VAR(1)-cDCC-GARCH			VAR(1)-RSDC-GARCH		
Simulated model	**1-Step**	**2-Step**	**3-Step**	**1-Step**	**2-Step**	**3-Step**	**1-Step**	**2-Step**	**3-Step**	**1-Step**	**2-Step**	**3-Step**	**1-Step**	**2-Step**	**3-Step**
Ratios															
Volatility															
VAR(1)-CCC-GARCH	0.0000	−0.0034	0.0041	−0.0038	−0.0055	−0.0035	−0.0032	−0.0030	0.0049	−0.0023	−0.0033	0.0039	−0.0023	−0.0035	0.0046
VAR(1)-ECCC-GARCH	0.0097	0.0066	0.0139	0.0000	−0.0003	0.0021	0.0080	0.0067	0.0155	0.0082	0.0066	0.0143	0.0091	0.0066	0.0138
VAR(1)-DCC-GARCH	0.0083	0.0035	0.0120	0.0056	0.0011	0.0028	0.0000	0.0035	0.0104	0.0044	0.0037	0.0099	0.0050	0.0032	0.0107
VAR(1)-cDCC-GARCH	0.0014	0.0005	0.0092	−0.0011	−0.0052	−0.0024	−0.0031	0.0010	0.0080	0.0000	0.0005	0.0102	0.0024	−0.0002	0.0086
VAR(1)-RSDC-GARCH	−0.0010	−0.0008	0.0080	−0.0008	−0.0029	−0.0022	−0.0027	−0.0009	0.0068	−0.0027	−0.0009	0.0068	0.0000	−0.0007	0.0080
Correlation															
VAR(1)-CCC-GARCH	0.0000	−0.0022	−0.0039	−0.0002	−0.0032	−0.0038	−0.0003	−0.0023	−0.0042	−0.0004	−0.0022	−0.0040			
VAR(1)-ECCC-GARCH	−0.0024	−0.0045	−0.0062	0.0000	−0.0031	−0.0034	−0.0026	−0.0045	−0.0064	−0.0025	−0.0044	−0.0062			
VAR(1)-DCC-GARCH	0.0056	0.0022	0.0007	0.0059	0.0011	0.0008	0.0000	−0.0021	−0.0036	0.0019	−0.0004	−0.0021			
VAR(1)-cDCC-GARCH	0.0033	0.0005	−0.0012	0.0035	−0.0009	−0.0015	−0.0015	−0.0039	−0.0052	0.0000	−0.0023	−0.0045			
VAR(1)-RSDC-GARCH															
Covariance															
VAR(1)-CCC-GARCH	0.0000	−0.0061	−0.0010	−0.0044	−0.0101	−0.0094	−0.0036	−0.0058	−0.0005	−0.0028	−0.0059	−0.0012			
VAR(1)-ECCC-GARCH	0.0067	0.0009	0.0059	0.0000	−0.0040	−0.0024	0.0047	0.0011	0.0071	0.0049	0.0011	0.0063			
VAR(1)-DCC-GARCH	0.0148	0.0051	0.0113	0.0116	0.0009	0.0021	0.0000	0.0009	0.0058	0.0063	0.0028	0.0069			
VAR(1)-cDCC-GARCH	0.0053	0.0004	0.0066	0.0029	−0.0074	−0.0058	−0.0045	−0.0034	0.0017	0.0000	−0.0023	0.0039			
VAR(1)-RSDC-GARCH															
Absolute ratios															
Volatility															
VAR(1)-CCC-GARCH	0.0000	0.0050	0.0122	0.0050	0.0186	0.0239	−0.0047	0.0051	0.0132	−0.0034	0.0049	0.0122	−0.0048	0.0049	0.0130
VAR(1)-ECCC-GARCH	0.0128	0.0128	0.0195	0.0000	0.0162	0.0203	0.0099	0.0124	0.0214	0.0100	0.0126	0.0199	0.0110	0.0126	0.0190
VAR(1)-DCC-GARCH	0.0170	0.0104	0.0197	0.0217	0.0237	0.0259	0.0000	0.0100	0.0165	0.0003	0.0106	0.0174	0.0112	0.0098	0.0170
VAR(1)-cDCC-GARCH	0.0163	0.0102	0.0197	0.0182	0.0201	0.0245	−0.0005	0.0103	0.0171	0.0000	0.0101	0.0217	0.0091	0.0089	0.0185
VAR(1)-RSDC-GARCH	0.0109	0.0119	0.0220	0.0211	0.0265	0.0291	0.0058	0.0116	0.0196	0.0068	0.0113	0.0202	0.0000	0.0119	0.0221
Correlation															
VAR(1)-CCC-GARCH	0.0000	−0.0007	−0.0004	−0.0004	−0.0008	−0.0005	0.0002	0.0003	0.0006	0.0001	0.0002	0.0005			
VAR(1)-ECCC-GARCH	0.0003	0.0001	0.0006	0.0000	−0.0001	0.0001	0.0003	0.0008	0.0014	0.0003	0.0008	0.0014			
VAR(1)-DCC-GARCH	0.0257	0.0255	0.0256	0.0260	0.0258	0.0257	0.0000	0.0007	0.0015	0.0003	0.0006	0.0013			
VAR(1)-cDCC-GARCH	0.0247	0.0243	0.0244	0.0245	0.0243	0.0244	0.0002	0.0009	0.0017	0.0000	−0.0003	0.0010			
VAR(1)-RSDC-GARCH															
Covariance															
VAR(1)-CCC-GARCH	0.0000	0.0016	0.0070	0.0008	0.0096	0.0130	−0.0050	0.0022	0.0081	−0.0041	0.0020	0.0074			
VAR(1)-ECCC-GARCH	0.0078	0.0037	0.0092	0.0000	0.0133	0.0161	0.0040	0.0037	0.0110	0.0040	0.0039	0.0097			
VAR(1)-DCC-GARCH	0.0355	0.0187	0.0256	0.0356	0.0243	0.0264	0.0000	0.0075	0.0130	0.0004	0.0079	0.0136			
VAR(1)-cDCC-GARCH	0.0333	0.0182	0.0251	0.0314	0.0210	0.0238	0.0001	0.0082	0.0137	0.0000	0.0074	0.0169			
VAR(1)-RSDC-GARCH															

Table 6 Volatility, covariance and correlation ratios and absolute ratios for the extended models.

	Estimated model											
	VAR(1)-ECCC-GARCH			VAR(1)-EDCC-GARCH			VAR(1)-EcDCC-GARCH			VAR(1)-ERSDC-GARCH		
Simulated model	**1-Step**	**2-Step**	**3-Step**	**1-Step**	**2-Step**	**3-Step**	**1-Step**	**2-Step**	**3-Step**	**1-Step**	**2-Step**	**3-Step**
Ratios												
Volatility												
VAR(1)-ECCC-GARCH	0.0000	−0.0003	0.0021	0.0008	0.0007	0.0015	0.0015	0.0007	0.0013	0.0187	0.0008	0.0016
VAR(1)-EDCC-GARCH	0.0211	0.0080	0.0099	0.0000	0.0078	0.0098	0.0078	0.0081	0.0102	0.0168	0.0080	0.0104
VAR(1)-EcDCC-GARCH	0.0147	0.0003	0.0033	−0.0083	0.0004	0.0029	0.0000	0.0006	0.0027	0.0090	0.0005	0.0030
VAR(1)-ERSDC-GARCH	0.0022	−0.0008	0.0011	0.0017	−0.0003	0.0005	−0.0002	−0.0008	0.0011	0.0000	−0.0007	0.0002
Correlation												
VAR(1)-ECCC-GARCH	0.0000	−0.0031	−0.0034	0.0001	−0.0030	−0.0031	0.0001	−0.0029	−0.0031			
VAR(1)-EDCC-GARCH	0.0240	0.0152	0.0152	0.0000	−0.0035	−0.0034	0.0043	0.0011	0.0011			
VAR(1)-EcDCC-GARCH	0.0187	0.0095	0.0093	−0.0052	−0.0087	−0.0086	0.0000	−0.0041	−0.0042			
VAR(1)-ERSDC-GARCH												
Covariance												
VAR(1)-ECCC-GARCH	0.0000	−0.0040	−0.0024	0.0012	−0.0024	−0.0019	0.0018	−0.0023	−0.0020			
VAR(1)-EDCC-GARCH	0.0518	0.0224	0.0244	0.0000	0.0037	0.0057	0.0117	0.0086	0.0107			
VAR(1)-EcDCC-GARCH	0.0409	0.0096	0.0122	−0.0130	−0.0084	−0.0059	0.0000	−0.0036	−0.0017			
VAR(1)-ERSDC-GARCH												
Absolute ratios												
Volatility												
VAR(1)-ECCC-GARCH	0.0000	0.0162	0.0203	−0.0088	0.0063	0.0075	−0.0087	0.0062	0.0073	0.0072	0.0063	0.0074
VAR(1)-EDCC-GARCH	0.0441	0.0163	0.0182	0.0000	0.0163	0.0186	0.0027	0.0164	0.0183	0.0302	0.0165	0.0184
VAR(1)-EcDCC-GARCH	0.0431	0.0137	0.0170	−0.0021	0.0138	0.0164	0.0000	0.0138	0.0165	0.0279	0.0139	0.0160
VAR(1)-ERSDC-GARCH	0.0060	0.0126	0.0155	0.0116	0.0125	0.0146	0.0115	0.0130	0.0152	0.0000	0.0134	0.0142
Correlation												
VAR(1)-ECCC-GARCH	0.0000	−0.0001	0.0001	0.0011	0.0009	0.0010	0.0010	0.0010	0.0011			
VAR(1)-EDCC-GARCH	0.0551	0.0555	0.0555	0.0000	−0.0096	−0.0094	0.0042	−0.0070	−0.0068			
VAR(1)-EcDCC-GARCH	0.0529	0.0534	0.0535	−0.0029	−0.0117	−0.0115	0.0000	−0.0093	−0.0090			
VAR(1)-ERSDC-GARCH												
Covariance												
VAR(1)-ECCC-GARCH	0.0000	0.0133	0.0161	−0.0079	0.0052	0.0059	−0.0079	0.0049	0.0057			
VAR(1)-EDCC-GARCH	0.0991	0.0458	0.0479	0.0000	0.0081	0.0100	0.0045	0.0091	0.0109			
VAR(1)-EcDCC-GARCH	0.0983	0.0433	0.0460	−0.0028	0.0057	0.0080	0.0000	0.0067	0.0090			
VAR(1)-ERSDC-GARCH												

In general terms, when volatilities and correlations that have been generated by a particular model are estimated by another model, their estimates seem to be worse as the number of steps used in the estimation increase. Nevertheless, volatilities and correlations are relatively well estimated even when using a misspecified model. The largest average ratio and absolute ratio are 5.18% and 9.91%, respectively, that could be interpreted as follows: on average, multiple steps estimates of volatilities (correlations) deviate from the corresponding true volatilities (correlations) at most about 5.2% (and about 10% in absolute terms) more than the amount that 1-step estimates of the correctly specified model deviates.

5 Conclusions

In this paper we have carried out several Monte Carlo experiments to study the performance in finite samples of 1-step and multiple steps estimators of vector autoregressive multivariate conditional correlation GARCH models. Although 1-step estimators are preferable because of their theoretical properties, they are not always feasible and therefore, estimating the parameters of a model in multiple steps could be a reasonable alternative. Our results indicate that, when the distribution of the errors is Gaussian, multiple steps estimators have a very good performance even in small samples. However, when the estimation is based on Student-t errors, we find that multiple steps estimators do not always perform well even when the data follows a Student-t distribution.

Our results also show that if the true error distribution is Student-t but estimation is based on the Gaussian distribution, kernel density estimates of the estimates of volatility and correlation obtained from 1-step and multiple steps estimators are quite similar. Analogously, if the true error distribution is Gaussian but estimation is based on the Student-t distribution, we obtain similar results as when the true and assumed distribution is a Student-t. When errors are distributed as a skewed Student-t but the estimation is performed assuming symmetric innovations, we find that kernel density estimates of the difference between 1-step and multiple steps estimates of volatilities and correlations from their true values are very similar when the estimation is based on a Gaussian distribution. However, this is not true when the estimation is based on Student-t errors. In any case, when the true distribution is skewed, one should be cautious in using 1-step or multiple-steps estimators based on Student-t errors since both are inconsistent estimators.

Finally, we also analyze the robustness of our results to the misspecification of the model when the estimation is based on Gaussian errors. We find that, on average, volatilities and correlations are relatively well estimated even when using a misspecified model. The multiple-steps estimates of volatilities (correlations) deviate from the true values at most by 10% more than what 1-step estimates of the correctly specified model do.

Acknowledgement: We are very grateful to an anonymous referee for helpful comments. Financial support from IVIE (Instituto Valenciano de Investigaciones Económicas) to the project “Estimating Multivariate GARCH Models in Multiple Steps with an Application to Stock Markets” is gratefully acknowledged. We also acknowledge the Spanish Government for grant ECO2011-29751.

References

Aielli, G. P. 2008. “Consistent Estimation of Large Scale Dynamic Conditional Correlations.” Unpublished paper: Department of Economics, Statistics, Mathematics and Sociology, University of Messina. Working paper n. 47.

Aielli, G. P. 2011. “Dynamic Conditional Correlations: on Properties and Estimation.” Available at SSRN: http://ssrn.com/abstract=1507743 or http://dx.doi.org/10.2139/ssrn.1507743.

Bauwens, L., and S. Laurent. 2005. “A New Class of Multivariate Skew Densities, With Application to Generalized Autoregressive Conditional Heteroscedasticity Models.” *Journal of Business and Economics Statistics* 23(3): 346–354.

Bauwens, L., S. Laurent, and J. V. K. Rombouts. 2006. “Multivariate GARCH Models: A Survey.” *Journal of Applied Econometrics* 21: 79–109.

Bollerslev, T. 1990. "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model." *The Review of Economics and Statistics* 72: 498–505.

Bollerslev, T., and J. M. Wooldridge. 1992. "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances." *Econometric Reviews* 11: 143–172.

Caporin, M., and M. McAleer. 2009. "Do We Really Need Both BEKK and DCC? A Tale of Two Covariance Models." Department of Economics, University of Padova.

Conrad, C., and M. Karanasos. 2010. "Negative Volatility Spillovers in the Unrestricted ECCC-GARCH Model." *Econometric Theory* 26: 838–862.

Engle, R. 2002. "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models." *Journal of Business and Economic Statistics* 20: 339–350.

Engle, R. 2009. *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press, Princeton, New Jersey.

Engle, R., and K. Sheppard. 2001. "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH." NBER Working Paper, No: W8554.

Engle, R., C. W. J. Granger, and D. Kraft. 1984. "Combining Competing Forecasts of Inflation using a Bivariate ARCH Model." *Journal of Economics Dynamics and Control* 8: 151–165.

Engle, R., N. Shephard, and K. Sheppard. 2008. "Fitting and Testing Vast Dimensional Time-Varying Covariance Models." NYU Working Paper No. FIN-07-046.

Fiorentini, G., E. Sentana, and G. Calzolari. 2003. "Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroskedastic Dynamic Regression Models with Student-T Innovations." *Journal of Business and Economic Statistics* 21: 532–546.

Hafner, C. M., and P. H. Franses. 2009. "A Generalized Dynamic Conditional Correlation Model: Simulation and Application to Many Assets." *Econometrics Reviews* 28: 6.

Hafner, C. M., and O. Reznikova. 2012. "On the Estimatiion of Dynamic Conditional Correlation Models." *Computational Statistics and Data Analysis* 56: 3533–3545.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Harvey A., E. Ruiz, and N. Shephard. 1992. "Unobservable Component Time Series Models with ARCH Disturbances." *Journal of Econometrics* 52: 129–158.

Jeantheau, T. 1998. "Strong Consistency of Estimators for Multivariate ARCH Models." *Econometric Theory* 14: 70–86.

Jondeau, E., and M. Rockinger. 2005. "Conditional Asset Allocation Under Non-Normality: How Costly is the Mean-Variance Criterion?" EFA 2005 Moscow Meeting. Discussion Paper.

Ling, S., and M. McAleer. 2003. "Asymptotic Theory for a Vector ARMA-GARCH Model." *Econometric Theory* 19: 280–310.

Longin, F., and B. Solnik. 1995. "Is the Correlation in International Equity Returns Constant: 1960–1990?" *Journal of International Money and Finance* 14(1): 3–26.

McAleer, M., and B. da Veiga. 2008a. "Forecasting Value-at-Risk with a Parsimonious Portfolio Spillover GARCH (PS-GARCH) Model." *Journal of Forecasting* 27(1): 1–19

McAleer, M., and B. da Veiga. 2008b. "Single Index and Portfolio Models for Forecasting Value-at-Risk Thresholds." *Journal of Forecasting* 27(3): 217–235.

Nakatani, T., and T. Teräsvirta. 2008. "Positivity Constraints on the Conditional Variances in the Family of Conditional Correlation GARCH Model." *Finance Research Letters* 5: 88–95.

Newey, W. K., and D. McFadden. 1994. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics*. Vol. 4. Elsevier, North Holland.

Newey, W. K., and D. G. Steigerwald. 1997. "Asymptotic Bias for Quasi Maximum Likelihood Estimators in Conditional Heteroskedasticity Models." *Econometrica* 3: 587–599.

Pelletier, D. 2006. "Regime Switching for Dynamic Correlations." *Journal of Econometrics* 131: 445–473.

Silvennoinen, A., and T. Teräsvirta. 2009. "Multivariate GARCH Models." In *Handbook of Financial Time Series*, edited by T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch. New York: Springer.

Tse, Y. K., and A. K. C. Tsui. 2002. "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations." *Journal of Business and Economic Statistics* 20: 351–362.