serie ad

WP-AD 2012-10

Estimating VAR-MGARCH Models in Multiple Steps

M. Angeles Carnero and M. Hakan Eratalay

Ivie

Working papers Working papers Working papers

Versión: marzo 2012 / Version: March 2012

Edita / Published by:
Instituto Valenciano de Investigaciones Económicas, S.A.
C/ Guardia Civil, 22 esc. 2 1º - 46020 Valencia (Spain)

WP-AD 2012-10

Estimating VAR-MGARCH Models in Multiple Steps*

M. Angeles Carnero and M. Hakan Eratalay**

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, this does not seem to be the case if the distribution is a Student-t.

Keywords: Volatility Spillovers, Financial Markets.

JEL classification: C32.

---

* 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.

** M.A. Carnero, Universidad de Alicante, Fundamentos del Análisis Económico. M.H. Eratalay, Carnero, Universidad de Alicante, Fundamentos del Análisis Económico. Corresponding author: acarnero@ua.es.

1 Introduction

Understanding how stock market returns and volatilities move over time has been of interest to researchers into the time series literature. In addition, as the financial crisis has shown, it is also very important to realize that 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 *et al.* (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 *et al.* (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 to be estimated 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, da Veiga and McAleer (2008a, 2008b). 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 one 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 one 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 one-step estimation.

Under the normality assumption, the parameters could also be estimated in two 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 two-step estimators will be consistent and asymptotically normal but not efficient.

The three-steps estimation method is mentioned in Bauwens *et al.* (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 two-steps estimation method for a zero-mean series. Therefore, under normal errors, the three-steps estimators are also consistent and asymptotically normal. Engle and Sheppard (2001) implement the three-steps estimation procedure in the empirical part of their paper.

Evidence gathered over the past decades shows that stock market returns are often far from having a normal distribution. Consequently, we also consider estimating the models assuming a Student-t distribution. The one-step estimator is obtained by maximizing the log-likelihood function based on the multivariate t-distribution; see, for example, Harvey *et al.* (1992) and Fiorentini *et al.* (2003). Although there is no theoretical work studying the properties of multiple steps estimation when assuming a Student-t distribution, we consider two-steps and three-steps estimators. In this line of research, Bauwens and Laurent (2005) and Jondeau and

Rockinger (2005) also analyze two-steps estimators. However, their approach is different in the sense that the first step of their estimation is performed assuming Gaussian errors while we maintain 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 one-step estimator with the two-steps and three-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. 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 one-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 comparison between one-step and two-steps 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 will see, when the errors are assumed to be Gaussian, the small sample behavior of one-step and two-steps estimators is very similar. On the other hand, when the estimation is based on the Student-t distribution, in some cases two-steps estimators deviate from one-step estimators.

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

Some robustness checks have been carried out to study how the results change when the true error distribution is different from the assumed one. Also, we 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 three time series. 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. One-step and multiple steps estimators for the previous 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:

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\beta}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{1}$$

where $Var(\varepsilon_t | \mathbf{Y}_{t-1}, ...\mathbf{Y}_1) = \mathbf{H}_t$, $\mathbf{Y}_t$ is a $k \times 1$ vector of returns, $\boldsymbol{\mu}$ is a $k \times 1$ vector of constants, $\boldsymbol{\beta}$ is a $k \times k$ matrix of autoregressive coefficients and ε_t is a $k \times 1$ vector of error terms as follows.

$$\mathbf{Y}_t = \begin{bmatrix} y_{1t} & y_{2t} & \dots & y_{kt} \end{bmatrix}', \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_k \end{bmatrix}'$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \dots & \beta_{kk} \end{bmatrix}, \qquad \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} & \dots & \varepsilon_{kt} \end{bmatrix}'$$

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

$$|I_k - \boldsymbol{\beta}z| = 0 \tag{2}$$

The number of mean parameters in the coefficient matrices $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ is $k(k+1)$. However, sometimes $\boldsymbol{\beta}$ is assumed to be diagonal. In that case, there will be $2k$ mean parameters to estimate.

The error term ε_t can be written as follows

$$\varepsilon_t = \boldsymbol{H}_t^{1/2}\eta_t$$

where η_t is a $k \times 1$ vector with $E(\eta_t) = 0$ and $Var(\eta_t) = I_k$.

$$\boldsymbol{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t \tag{3}$$

where $\mathbf{D}_t = diag(h_{1t}^{1/2}, h_{2t}^{1/2}, ..., h_{kt}^{1/2})$ and $\mathbf{R}_t$ is the conditional correlation matrix such that

$$\boldsymbol{H}_t = diag(h_{1t}^{1/2}, h_{2t}^{1/2}, ..., h_{kt}^{1/2}) \begin{bmatrix} 1 & \rho_{12t} & \dots & \rho_{1kt} \\ \rho_{12t} & 1 & \dots & \rho_{2kt} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1kt} & \rho_{2kt} & \dots & 1 \end{bmatrix} diag(h_{1t}^{1/2}, h_{2t}^{1/2}, ..., h_{kt}^{1/2})$$

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

From previous equations, given that the conditional correlation matrix, $\mathbf{R}_t$, is always positive definite, it is clear that as long as conditional variances, h_{it}, are positive for any $i = 1, 2, \dots, k$, the conditional variance-covariance matrix, $\mathbf{H}_t$, will be also positive definite. The conditional variances h_{it} are assumed to follow a GARCH(1,1) model. Then,

$$\mathbf{h}_t = \mathbf{W} + \mathbf{A}\boldsymbol{\varepsilon}_{t-1}^{(2)} + \mathbf{G}\mathbf{h}_{t-1} \tag{4}$$

where $\mathbf{h}_t = \begin{bmatrix} h_{1t} & h_{2t} & \dots & h_{kt} \end{bmatrix}'$ and $\boldsymbol{\varepsilon}_t^{(2)} = \begin{bmatrix} \varepsilon_{1t}^2 & \varepsilon_{2t}^2 & \dots & \varepsilon_{kt}^2 \end{bmatrix}'$ are $k \times 1$ vectors of conditional variances and squared errors respectively and $\mathbf{W}$ is a $k \times 1$ and $\mathbf{A}$ and $\mathbf{G}$ are $k \times k$ matrices of coefficients. If $\mathbf{A}$ and $\mathbf{G}$ are restricted to be diagonal; see, for example, Bollerslev (1990) and Engle (2002), then volatility spillovers cannot be captured. Alternatively, if $\mathbf{A}$ and $\mathbf{G}$ are non-diagonal; see, for example, Jeantheau (1998) and Ling and McAleer (2003), then the model allows for volatility spillovers. 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 $\omega_i = [\mathbf{W}]_i$, $\alpha_{ij} = [\mathbf{A}]_{i,j}$ and $\gamma_{ij} = [\mathbf{G}]_{i,j}$. 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 $\mathbf{h}_t$ to have positive elements for all t. They show that off-diagonal elements in $\mathbf{G}$ could be negative while $\mathbf{H}_t$ 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 $|I_k - (\mathbf{A} + \mathbf{G})z| = 0$ 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 $\mathbf{R}_t$ 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, $\mathbf{R}_t = \mathbf{R}$, is equal to the matrix of sample correlations of the standardized residuals, i.e.

$$[\widehat{\mathbf{R}}]_{ij} = \widehat{\rho}_{ij} = \frac{\sum_t \widehat{\nu}_{it}\widehat{\nu}_{jt}}{\sqrt{\left(\sum_t \widehat{\nu}_{it}^2\right)\left(\sum_t \widehat{\nu}_{jt}^2\right)}} \tag{5}$$

where $\nu_t = \mathbf{D}_t^{-1}\varepsilon_t$ 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 $\mathbf{A}$ and $\mathbf{G}$ in (4) are non-diagonal.

The third model we consider is the **DCC-GARCH** in which $\mathbf{R}_t = \mathbf{P}_t\mathbf{Q}_t\mathbf{P}_t$ with $\mathbf{P}_t = diag(\mathbf{Q}_t)^{-1/2}$ and $\mathbf{Q}_t = (1-\delta_1-\delta_2)\overline{\mathbf{Q}} + \delta_1\nu_{t-1}\nu_{t-1}' + \delta_2\mathbf{Q}_{t-1}$ where $\mathbf{Q}_t$ denotes the covariance matrix and $\overline{\mathbf{Q}}$ is the long run covariance (correlation) matrix. The correlation targeting approach suggests replacing $\overline{\mathbf{Q}}$ 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, $\mathbf{R}_t$, 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 (2010), 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 $\overline{\mathbf{Q}}$. As Caporin and McAleer (2009) point out as well, Aielli's conclusion follows from the fact that the unconditional expectations of $\mathbf{Q}_t$ 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 $\mathbf{Q}_t = (1-\delta_1-\delta_2)\overline{\mathbf{Q}} + \delta_1\nu_{t-1}^*\nu_{t-1}^{*\prime} + \delta_2\mathbf{Q}_{t-1}$ where $\boldsymbol{\nu}_t^* = diag(\mathbf{Q}_t)^{1/2}\boldsymbol{\nu}_t$. He argues that in this model a natural estimator for the long run covariance matrix, $\overline{\mathbf{Q}}$, would be the sample covariance matrix of $\boldsymbol{\nu}_t^*$. The number of parameters to be estimated will be then only 2 as in the DCC-GARCH model of Engle (2002).

The last model we will consider in this paper is the **RSDC-GARCH**. 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 $\mathbf{R}_t$, at any time t, could be equal to either $\mathbf{R}^L$ or $\mathbf{R}^H$, 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 $k(k-1)+2$.

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

3.1 Vector Autoregressive CCC, ECCC, DCC and cDCC GARCH models

In this section we analyze three possible procedures to estimate the parameters in equations (1) and (3), denoted by $\Phi = (\theta', \phi', \psi')'$ when $\mathbf{R}_t$ in equation (3) is specified by the CCC-GARCH, ECCC-GARCH, DCC-GARCH or the cDCC-GARCH model.

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 one-step estimator will be the maximum likelihood estimator of Φ and it can be found by maximizing the multivariate Gaussian log-likelihood function:

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

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

From equation (3) we have that

$$L(\Phi) = -\frac{Tk}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\log|\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t| - \frac{1}{2}\sum_{t=2}^{T}\varepsilon_t'(\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t)^{-1}\varepsilon_t =$$

$$= -\frac{Tk}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\log|\mathbf{R}_t| - \sum_{t=2}^{T}\log|\mathbf{D}_t| - \frac{1}{2}\sum_{t=2}^{T}\nu_t'\mathbf{R}_t^{-1}\nu_t \tag{6}$$

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 *et al.* (2003):

$$L(\Phi,\eta) = T\log\left[\Gamma\left(\frac{\eta k+1}{2\eta}\right)\right] - T\log\left[\Gamma\left(\frac{1}{2\eta}\right)\right] - \frac{Tk}{2}\log\left(\frac{1-2\eta}{\eta}\right) - \frac{Tk}{2}\log(\pi)$$

$$-\sum_{t=2}^{T}\left[\frac{1}{2}\log|\mathbf{H}_t| + \left(\frac{\eta k+1}{2\eta}\right)\log\left(1+\frac{\eta}{1-2\eta}\nu_t'\mathbf{R}_t^{-1}\nu_t\right)\right] \tag{7}$$

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

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

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, one-step estimators of the parameters, Φ, computed by maximizing the likelihood function in (7), are consistent and asymptotically normal; see Fiorentini *et al.* (2003). It is important to note that if the true distribution of the data is Student-t, Maximum Likelihood (ML) estimators (in this case, one-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 two steps following Engle (2002) and Engle and Sheppard (2001). They proposed to use two-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 $\mathbf{R}_t$ is replaced by the identity matrix. Therefore, in the first step, the function to be maximized is the following:

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

If volatility spillovers are not allowed, i.e. $\mathbf{A}$ and $\mathbf{G}$ 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 the second step, given the estimates from the first step, $\widehat{\theta}$ and $\widehat{\phi}$, the correlation coefficients are estimated by maximizing the following function

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

where $\widehat{\nu}_t$ 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, two-steps estimators are also consistent. Furthermore, Engle and Sheppard (2001) give conditions for the DCC-GARCH model under which two-steps estimators are also asymptotically normal.

Next, we also consider two-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 $\mathbf{R}_t$ has been replaced by I_k. That is

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

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

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

where $\widehat{\nu}_t$ 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 $\mathbf{R}_t$ is equal to the identity matrix for all t. Therefore, the function to be maximized is the following

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

where $\mathbf{D} = diag(h_1^{1/2}, h_2^{1/2}, ..., h_k^{1/2})$ 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, $\widehat{\theta}$, and substituting the correlation matrix $\mathbf{R}_t$ by I_k. This leads to the maximization of the following function:

$$L_2\left(\phi|\widehat{\theta}\right) = -\frac{Tk}{2}\log(2\pi) - \sum_{t=2}^{T}\log|\mathbf{D}_t| - \frac{1}{2}\sum_{t=2}^{T}\tilde{\nu}_t'\tilde{\nu}_t$$

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

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

where $\widehat{\nu}_t$ 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, three-step estimators are also consistent and their asymptotic distribution is very similar to that of the two-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

$$L_1(\theta, \eta, h_i) = T \log\left[\Gamma\left(\frac{\eta k + 1}{2\eta}\right)\right] - T \log\left[\Gamma\left(\frac{1}{2\eta}\right)\right] - \frac{Tk}{2}\log\left(\frac{1-2\eta}{\eta}\right) - \frac{Tk}{2}\log(\pi)$$

$$-\sum_{t=2}^{T}\left[\log|\mathbf{D}| + \left(\frac{\eta k + 1}{2\eta}\right)\log\left(1 + \frac{\eta}{1-2\eta}\nu_t'\nu_t\right)\right]$$

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, $\widehat{\theta}$. The function to be maximized is the following

$$L_2\left(\phi, \eta|\widehat{\theta}\right) = T \log\left[\Gamma\left(\frac{\eta k + 1}{2\eta}\right)\right] - T \log\left[\Gamma\left(\frac{1}{2\eta}\right)\right] - \frac{Tk}{2}\log\left(\frac{1-2\eta}{\eta}\right) - \frac{Tk}{2}\log(\pi)$$

$$-\sum_{t=2}^{T}\left[\log|\mathbf{D}_t| + \left(\frac{\eta k + 1}{2\eta}\right)\log\left(1 + \frac{\eta}{1-2\eta}\tilde{\nu}_t'\tilde{\nu}_t\right)\right]$$

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

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

where $\widehat{\nu}_t$ 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 $\mathbf{R}_t$ 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 one-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 $\mathbf{R}_t$ 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 $\mathbf{R}_t$ 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 three-steps estimation procedure without paying much attention to the mean parameters or a two-steps estimation method for a zero mean series.

Finally, the asymptotic properties of the one-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**	**One-step**	**Two-steps**	**Three-steps**
Gaussian	Gaussian	Consistent	Consistent	Consistent
Student-t	Student-t	Consistent	.	.
Student-t	Gaussian	Consistent	Consistent	Consistent
Gaussian	Student-t	Consistent	.	.

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 one-step, two-steps and three-steps estimators when they are used to estimate the parameters of first order Vector Autoregressive CCC, ECCC, DCC, cDCC and RSDC-GARCH models. 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 one-step, two-steps and three-steps 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 four different experiments we have carried out. In the first one, we simulate time series vectors following the five 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 one-step and multiple steps estimators are analyzed. In a second experiment we study how robust the results obtained in the previous experiment are to the error distribution. With this objective, we simulate data from the five models considered assuming a Gaussian distribution for the innovations and estimate the true model under the assumption that errors follow a Student-t distribution. In adition, time series vectors are generated using a Student-t distribution for the errors and then, true models are estimated under the Gaussianity assumption. In a third experiment we analyze how good or bad volatilities and correlations generated from a given model can be estimated using a different model. Finally, in the fourth and last experiment 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.

4.1 Innovations distributed as a Gaussian or Student-t

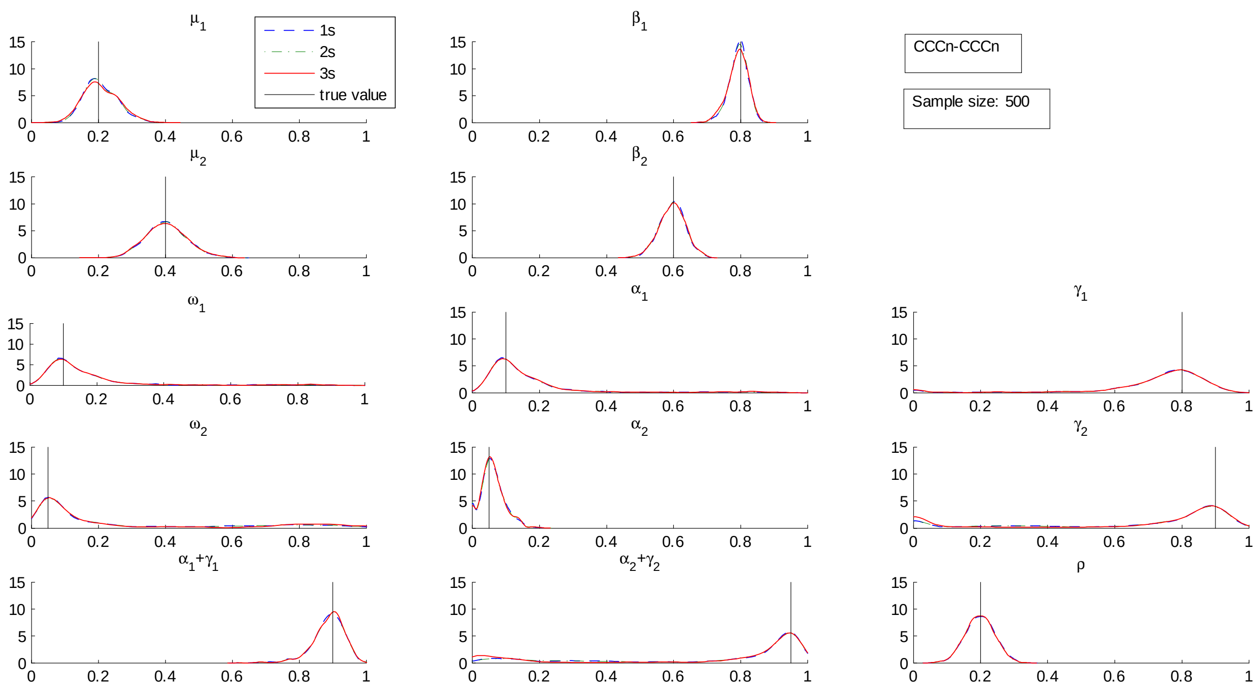
We start by considering the case in which data is generated and estimated assuming a normal distribution. Let us consider a bivariate model given by equations (1) to (3) with a diagonal matrix β and $\mathbf{R}_t = \mathbf{R}$ as given by the CCC-GARCH model. The unconditional mean and variance are fixed 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 one-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 one-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 one-step, two-steps and three-steps 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.

In order to check the robustness of the results, we consider different scenarios by changing

Table 1: Monte Carlo mean and standard deviations of one-step, two-steps and three-steps estimators of a bivariate Gaussian VAR(1)-CCC-GARCH model

		One-step			Two-steps			Three-steps		
Parameter	Value	$T = 500$	$T = 1000$	$T = 5000$	$T = 500$	$T = 1000$	$T = 5000$	$T = 500$	$T = 1000$	$T = 5000$
μ_1	0.20	0.207 (0.050)	0.204 (0.036)	0.201 (0.016)	0.207 (0.050)	0.204 (0.037)	0.201 (0.017)	0.208 (0.053)	0.204 (0.039)	0.201 (0.017)
μ_2	0.40	0.403 (0.060)	0.403 (0.043)	0.400 (0.017)	0.403 (0.061)	0.403 (0.044)	0.400 (0.018)	0.403 (0.062)	0.404 (0.044)	0.400 (0.018)
β_1	0.80	0.793 (0.028)	0.796 (0.020)	0.799 (0.009)	0.793 (0.029)	0.796 (0.020)	0.799 (0.009)	0.792 (0.030)	0.796 (0.022)	0.799 (0.010)
β_2	0.60	0.596 (0.038)	0.598 (0.026)	0.600 (0.011)	0.597 (0.039)	0.598 (0.027)	0.600 (0.012)	0.596 (0.039)	0.597 (0.027)	0.600 (0.012)
ω_1	0.10	0.180 (0.179)	0.124 (0.079)	0.103 (0.019)	0.182 (0.183)	0.123 (0.072)	0.103 (0.019)	0.183 (0.184)	0.124 (0.075)	0.103 (0.019)
ω_2	0.05	0.270 (0.308)	0.120 (0.177)	0.053 (0.015)	0.273 (0.311)	0.132 (0.198)	0.053 (0.015)	0.290 (0.339)	0.146 (0.231)	0.054 (0.031)
α_1	0.10	0.108 (0.044)	0.103 (0.030)	0.099 (0.012)	0.109 (0.044)	0.103 (0.030)	0.099 (0.013)	0.106 (0.043)	0.102 (0.030)	0.099 (0.013)
α_2	0.05	0.061 (0.036)	0.054 (0.023)	0.050 (0.009)	0.061 (0.037)	0.054 (0.024)	0.050 (0.009)	0.061 (0.035)	0.054 (0.023)	0.050 (0.009)
γ_1	0.80	0.706 (0.203)	0.772 (0.096)	0.796 (0.027)	0.705 (0.206)	0.773 (0.089)	0.796 (0.027)	0.705 (0.208)	0.772 (0.093)	0.797 (0.027)
γ_2	0.90	0.660 (0.322)	0.822 (0.192)	0.897 (0.021)	0.656 (0.325)	0.810 (0.212)	0.897 (0.021)	0.637 (0.355)	0.796 (0.243)	0.896 (0.035)
ρ	0.20	0.199 (0.044)	0.201 (0.031)	0.200 (0.014)	0.198 (0.043)	0.199 (0.031)	0.200 (0.014)	0.198 (0.043)	0.199 (0.031)	0.200 (0.014)

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



the parameter values in Table 1 and repeat the Monte Carlo experiment. Table 2 contains the new parameter values and experiments considered. First, we consider the case in which the unconditional variance of one of the series is more than six times the other (Experiment 2). In addition, we repeat the experiment with the unconditional mean of one series being larger than the other (Experiment 3). We also consider the case in which the coefficients of the first variance equation are changed (Experiment 4). The other case we analyze is when interactions among the series are allowed (Experiment 5). Finally, we consider a trivariate model (Experiment 6). The results obtained from all these experiments can be summarized in tables and graphs similar to Table 1 and Figure 1. All the results are similar to the ones discussed before and summarized in Table 1 and they are not included in the paper to save space but they are available from the authors upon request.

Since, as mentioned before, 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 have also calculated the estimated volatilities and correlations obtained from one-step, two-steps and three-steps estimators. For a sample size T, let us denote by $\widehat{h}_{i,t}^{s}$ the estimated volatilities of series i at time t obtained from estimator s (one-step, two-steps or three-steps) 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

$$\Delta\widehat{h}_{i}^{s} = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{h}_{i,t}^{s} - h_{i,t}\right) \tag{13}$$

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

$$\Delta\widehat{p}_{ij}^{s} = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{p}_{ij,t}^{s} - p_{ij,t}\right) \tag{14}$$

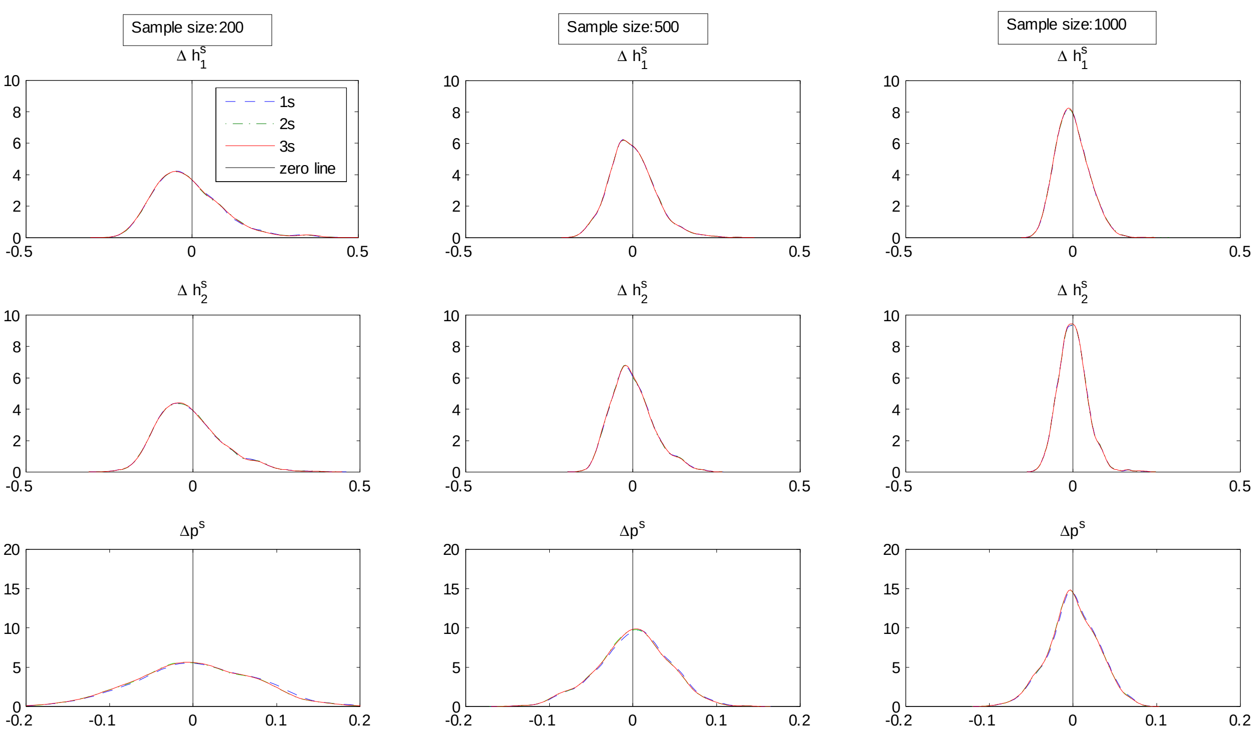
Figure 2 plots kernel density estimates of the differences between the estimated and the true volatilities and correlations measured as in (13) and (14) for a VAR(1)-CCC-GARCH(1,1) model with parameter values as in Experiment 1 (see Table 1) and sample sizes $T = 200$, $T = 500$ and $T = 1000$. As the graph illustrates, one-step, two-steps and three-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. $\frac{\widehat{h}_{i,t}^{s}-h_{i,t}}{h_{i,t}}$; $\frac{\widehat{p}_{ij,t}^{s}-p_{ij,t}}{p_{ij,t}}$ 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

Table 2: Parameter values of the VAR(1)-CCC-GARCH model for different Monte Carlo experiments

Parameter	Basic (Table 1)	Experiment 2	Experiment 3	Experiment 4	Experiment 5	Experiment 6
μ_1	0.20	0.20	0.30	0.20	0.10	0.20
μ_2	0.40	0.40	0.40	0.40	0.40	0.40
μ_3	-	-	-	-	-	0.30
β_{11}	0.80	0.80	0.80	0.80	0.80	0.80
β_{12}	0.00	0.00	0.00	0.00	0.10	0.00
β_{21}	0.00	0.00	0.00	0.00	0.10	0.00
β_{22}	0.60	0.60	0.60	0.60	0.60	0.60
β_{33}	-	-	-	-	-	0.70
ω_1	0.10	0.10	0.10	0.10	0.10	0.10
ω_2	0.05	1.00	0.05	0.05	0.05	0.05
ω_3	-	-	-	-	-	0.05
α_1	0.10	0.10	0.10	0.35	0.10	0.10
α_2	0.05	0.15	0.05	0.05	0.05	0.05
α_3	-	-	-	-	-	0.15
γ_1	0.80	0.80	0.80	0.55	0.80	0.80
γ_2	0.90	0.70	0.90	0.90	0.90	0.90
γ_3	-	-	-	-	-	0.80
ρ_{12}	0.20	0.20	0.20	0.20	0.20	0.10
ρ_{13}	-	-	-	-	-	0.20
ρ_{23}	-	-	-	-	-	0.30

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



correlations in VAR(1)-DCC, cDCC, ECCC and RSDC-GARCH(1,1) models were computed. The parameter values used in this case for the mean equation (1), i.e. $\boldsymbol{\mu}, \boldsymbol{\beta}$ are the same as the ones in Table 1. The variance parameters in equation (4) are also the same as the ones in Table 1 for VAR(1)-DCC, cDCC and RSDC-GARCH(1,1) models. For the VAR(1)-ECCC-GARCH(1,1) model they are $\omega_1 = 0.2$, $\omega_2 = 0.3$, $\alpha_{11} = 0.25$, $\alpha_{12} = 0.05$, $\alpha_{21} = 0.10$, $\alpha_{22} = 0.20$, $\gamma_{11} = 0.50$, $\gamma_{12} = 0.10$, $\gamma_{21} = 0.05$ and $\gamma_{22} = 0.40$. Finally, the correlation parameter is the same as the one in Table 1 for the VAR(1)-ECCC-GARCH(1,1) model. Other correlation parameters are $\bar{\mathbf{Q}}_{12} = 0.20$, $\delta_1 = 0.04$ and $\delta_2 = 0.94$ for the VAR(1)-DCC and cDCC-GARCH(1,1) models, and $\pi_{LL} = 0.80$, $\pi_{HH} = 0.90$, $R_{12}^L = 0.20$ and $R_{12}^H = 0.80$ for the VAR(1)-RSDC-GARCH(1,1) model. Since the graphs are very similar to Figure 2 they are not included in the paper. Consequently, our results suggest that under normal innovations, using multiple step estimators is 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 and cDCC-GARCH models the results are similar to the ones obtained under the normal assumption. Figure 3 contains, as an example, kernel density estimates of the differences between the estimated and the true volatilities and correlations in a VAR(1)-DCC-GARCH(1,1) model. As we can see, one-step, two-steps and three-steps estimators provide volatilities and correlations estimates which are very close to each other. 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 one-step estimation. In the case of the cDCC-GARCH model, results are very similar and the graphs are not included to save space.

However, for three of the models considered, namely the VAR(1)-CCC-GARCH(1,1), VAR(1)-ECCC-GARCH(1,1) and VAR(1)-RSDC-GARCH(1,1) models, important differences appear when estimating the correlations (or correlation parameters and transition probabilities for the RSDC-GARCH model) with different estimators. In this case, one-step estimator provides the best estimates. Figure 4 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 4 and it is not included in the paper. For the RSDC-GARCH model, Figure 5 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.

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

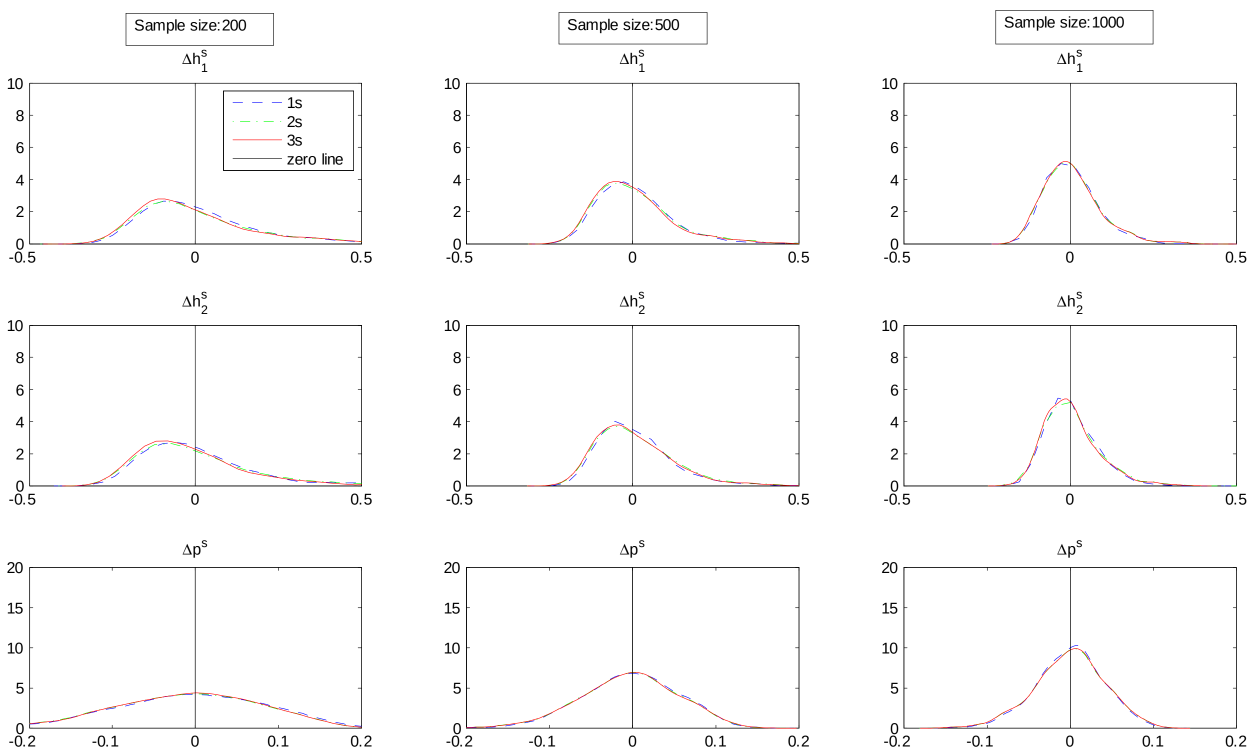


Figure 4: Kernel density estimates of deviations from estimated to true volatility in a VAR(1)-CCC-GARCH(1,1) model with Student-t innovations

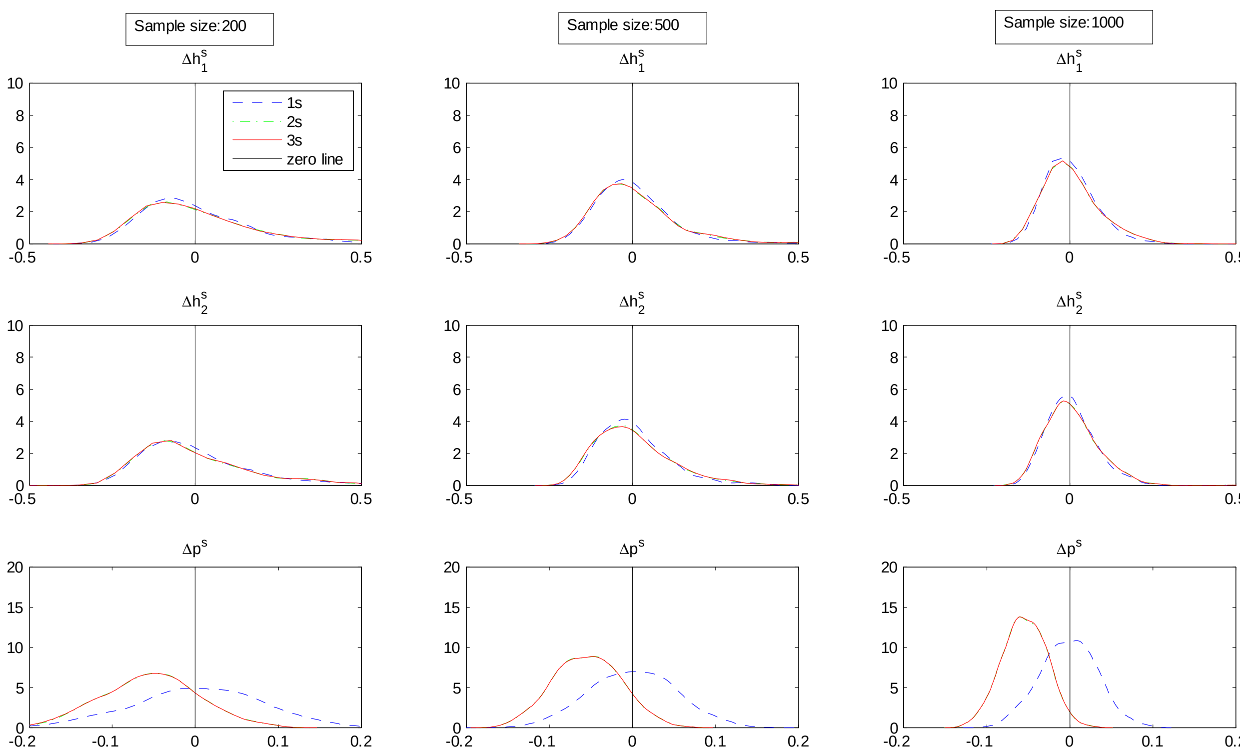
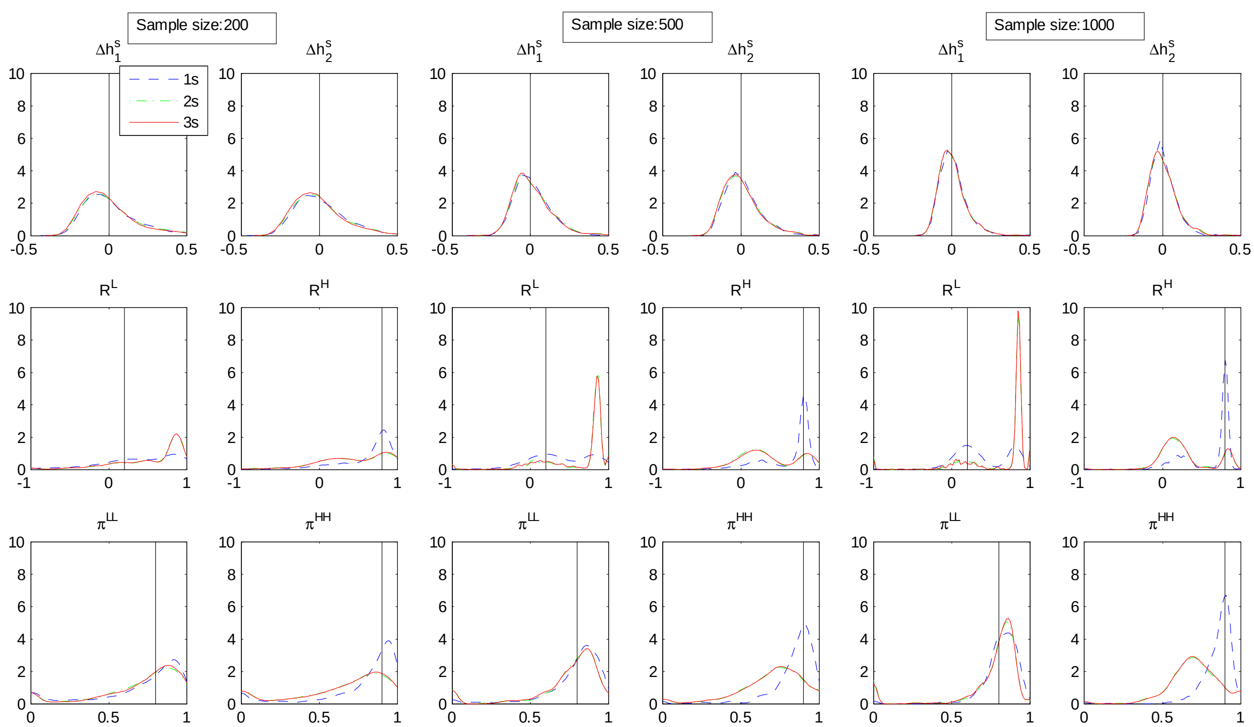


Figure 5: Kernel density estimates of deviations from estimated to true volatility, of estimated correlation parameters and of estimated transition probabilities in a VAR(1)-RSDC-GARCH(1,1) model with Student-t innovations



As we can see, estimates obtained with multiple steps estimators seem to be far from the ones obtained with the one-step estimator. Therefore, our results suggest that when innovations are distributed as a Student-t, using multiple steps estimators under the correct error distribution might not be a good idea.

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. In another experiment, we simulate data in which innovations follow a Student-t distribution and estimate the true model assuming Gaussian errors.

Figure 6 contains kernel density estimates of the differences between the estimated and the true volatilities and correlations in a VAR(1)-ECCC-GARCH(1,1) model when the data is generated using a Student-t distribution for the errors and estimated assuming Gaussian errors. Differences between one-step and multiple steps seem to be, again, negligible. Compared to the case in which the true and assumed error distributions are both normal, the estimated densities in Figure 6 have fatter tails. Finally, the results illustrate how the density estimates of the differences between the estimated and the true volatilities and correlations tend to zero as the sample size increases. Similar figures are obtained for the other four models.

When we simulate the data with Gaussian errors and estimate the model under the Student-t distribution assumption, for the models considered[2], i.e. VAR(1)-DCC and cDCC-GARCH(1,1) models, figures look very similar to the case when the true and the assumed distribution are both normal. Figure 7 shows the results for the VAR(1)-DCC-GARCH(1,1) model in this case. This similarity makes sense since Student-t distribution has an extra parameter, namely the degrees of freedom, such that this distribution could approximate Gaussian distribution when this parameter is sufficiently large. In fact, in the experiments for this last case, we obtained very large estimates for the degrees of freedom of the Student-t distribution.

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

[2] Models VAR(1)-CCC, ECCC and RSDC-GARCH(1,1) have been excluded since, as we have previously seen, multiple steps estimators do not perform well when the estimation is done under the assumption of Student-t innovations.

Figure 6: Kernel density estimates of deviations from estimated to true volatility in a VAR(1)-ECCC-GARCH(1,1) model generated with Student-t innovations and estimated assuming Gaussian errors

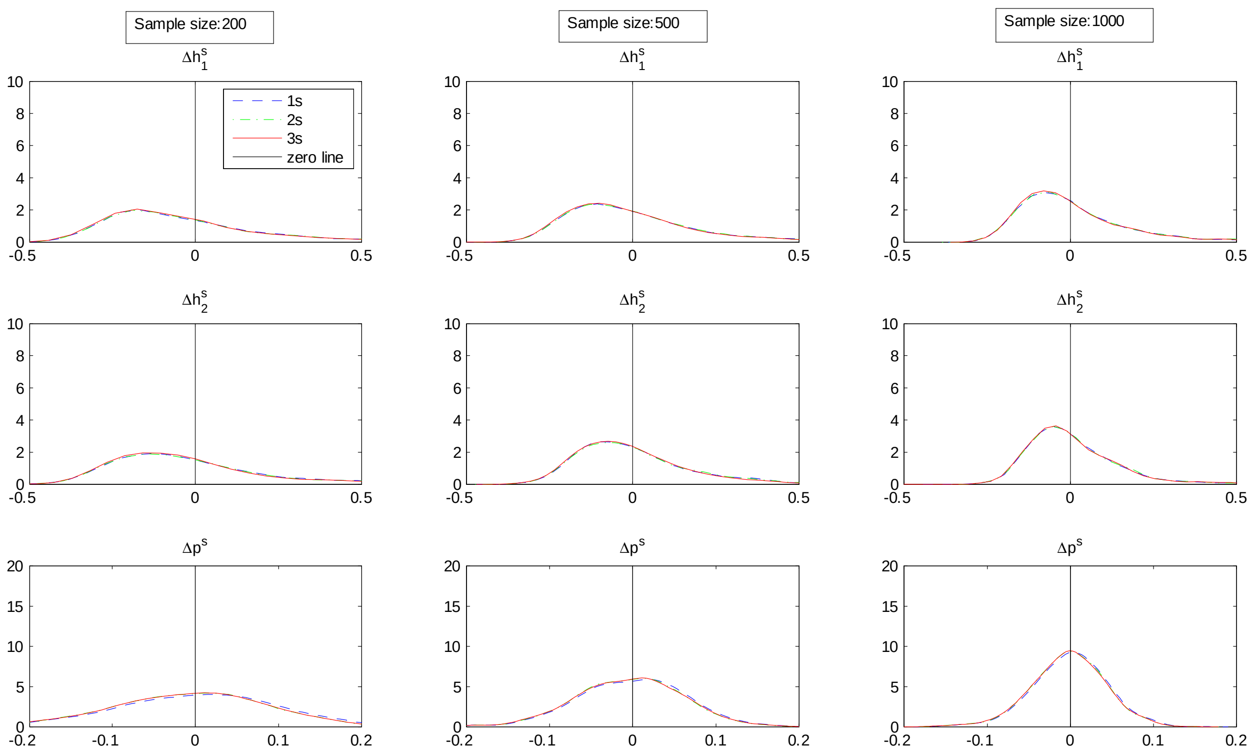
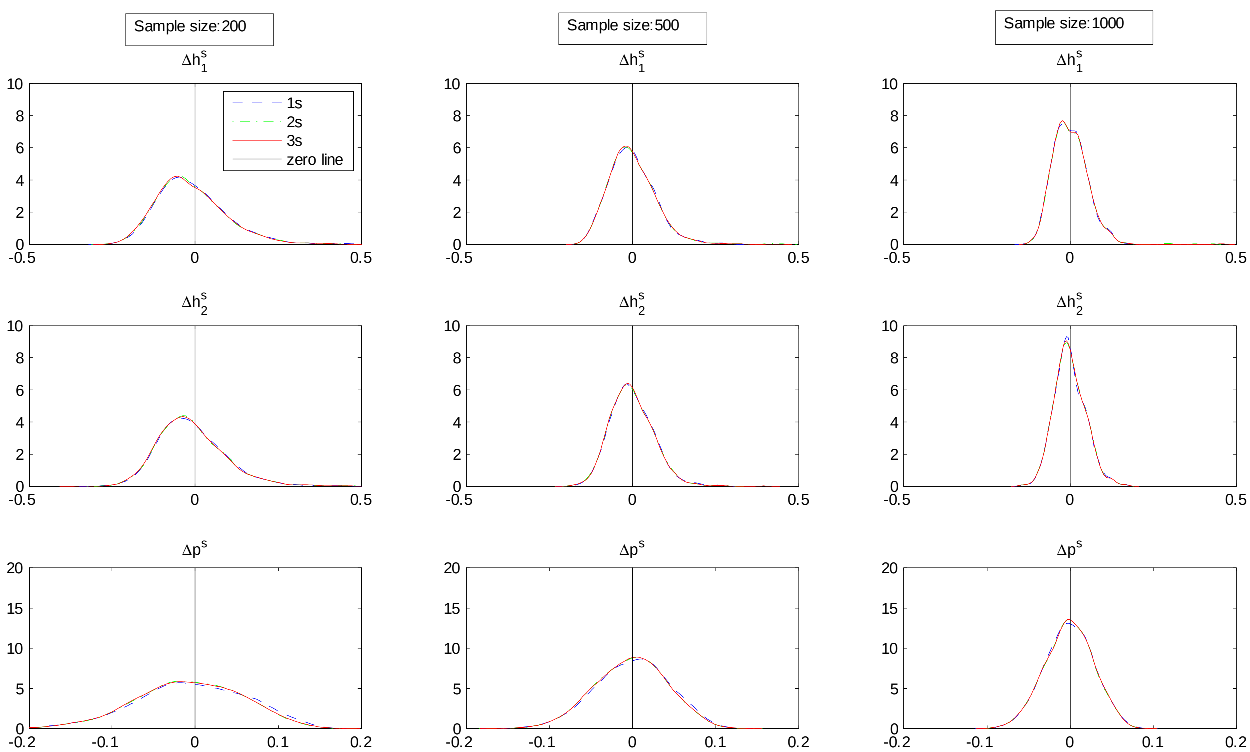


Figure 7: Kernel density estimates of deviations from estimated to true volatility in a VAR(1)-DCC-GARCH(1,1) model generated with Gaussian innovations and estimated assuming Student-t errors



from the estimated model. To perform this experiment, we take the parameter values from real data. 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 returns series, we estimate all the five models considered, i.e. VAR(1)- CCC, ECCC, DCC, cDCC and RSDC-GARCH models with no mean transmissions assuming Gaussian errors. The results are given in Table 3 in which series 1,2 and 3 correspond to BEL-20, DAX and FTSE-100, respectively.

As we can see in the table, three-steps estimates of the mean parameters are the same, as expected, since the mean equation is the same for all the models. Two-steps mean parameter estimates are also very similar, with the exception of the ECCC-GARCH model, since the variance equation is the same across the other 4 models. Correlation estimates for the CCC and ECCC-GARCH 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 one-step, two-steps and three-steps estimates are not very large. Figures 8 and 9 plot the volatilities and correlations estimates respectively. We can see that estimates obtained using different estimators are very similar. The graphs containing the correlation estimates obtained from DCC and cDCC models show that the correlation between the returns of these markets in the period analyzed has been changing over time.

For the Monte Carlo experiments, we take the one-step estimates obtained in this empirical exercise as the true parameter values to generate the data sets. Given that there are five models, it adds up to 25 different experiments. 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 the experiments assuming a Gaussian error distribution for generating the data and also for estimating the parameters.

The results are reported in Table 4, 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 difference between estimated volatility (correlation) and true volatility (correlation) is 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

Table 3: Parameter estimates for three real time series under Gaussian innovations

	VAR(1)-CCC-GARCH			VAR(1)-ECCC-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	1-step	2-steps	3-steps
μ_1	0.1017	0.0937	−0.0167	0.0997	0.0969	−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	0.1097	0.0707	−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	0.0683	0.0503	−0.0128	0.0646	0.0463	−0.0128	0.0669	0.0465	−0.0128	0.0611	0.0465	−0.0128
β_1	−0.0297	−0.0055	0.0641	−0.0267	−0.0112	0.0641	−0.0277	−0.0055	0.0640	−0.0278	−0.0055	0.0641	−0.0245	−0.0055	0.0640
β_2	−0.0934	−0.0549	−0.0465	−0.0903	−0.0455	−0.0465	−0.0748	−0.0549	−0.0466	−0.0738	−0.0549	−0.0465	−0.0743	−0.0549	−0.0466
β_3	−0.0991	−0.1033	−0.0821	−0.0956	−0.1020	−0.0820	−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	0.0322	0.0101	0.0170	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	0.0245	0.0183	0.0172	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	0.0168	0.0162	0.0000	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	0.0772	0.0402	0.0489	0.1118	0.1331	0.1231	0.1183	0.1331	0.1231	0.0874	0.1331	0.1231
α_{21}				0.0032	0.0328	0.0249									
α_{31}				0.0224	0.0684	0.0373									
α_{12}				0.0000	0.0000	0.0000									
α_{22}	0.0774	0.0955	0.0927	0.0794	0.0678	0.0676	0.0887	0.0955	0.0927	0.0933	0.0955	0.0927	0.0687	0.0955	0.0927
α_{32}				0.0038	0.0000	0.0000									
α_{13}				0.0476	0.0907	0.0973									
α_{23}				0.0000	0.0000	0.0051									
α_{33}	0.0777	0.1045	0.1041	0.0599	0.0605	0.0343	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	0.8399	0.0000	0.8577	0.8705	0.8573	0.8673	0.8695	0.8573	0.8673	0.8956	0.8573	0.8673
γ_{21}				0.0000	0.0000	0.0000									
γ_{31}				0.0000	0.2938	0.6074									
γ_{12}				0.0083	0.0000	0.0000									
γ_{22}	0.9095	0.8977	0.9011	0.9065	0.9023	0.9054	0.9005	0.8977	0.9011	0.9004	0.8977	0.9011	0.9217	0.8977	0.9011
γ_{32}				0.0029	0.0000	0.0000									
γ_{13}				0.0000	0.9680	0.0000									
γ_{23}				0.0000	0.0000	0.0000									
γ_{33}	0.9063	0.8921	0.8936	0.8883	0.5322	0.2512	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	0.7921	0.7853	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	0.7770	0.7662	0.7654									
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	0.8054	0.7983	0.8000									
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

Table 4: Volatility, covariance and correlation ratios

Simulated model		Estimated model														
		VAR(1)-CCC-GARCH			VAR(1)-ECCC-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	1-step	2-steps	3-steps
VAR(1)-CCC-GARCH	Volatility	0.0000	−0.0034	0.0041	0.0021	−0.0025	0.0029	−0.0032	−0.0030	0.0049	−0.0023	−0.0033	0.0039	−0.0023	−0.0035	0.0046
VAR(1)-ECCC-GARCH		0.0095	0.0064	0.0137	0.0000	−0.0001	0.0023	0.0078	0.0065	0.0154	0.0080	0.0065	0.0141	0.0089	0.0064	0.0136
VAR(1)-DCC-GARCH		0.0083	0.0035	0.0120	0.0143	0.0042	0.0059	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	0.0165	0.0015	0.0068	−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	0.0067	0.0005	0.0013	−0.0027	−0.0009	0.0068	−0.0027	−0.0009	0.0068	0.0000	−0.0007	0.0080
VAR(1)-CCC-GARCH	Correlation	0.0000	−0.0022	−0.0039	0.0007	−0.0027	−0.0034	−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	0.0071	0.0013	0.0010	0.0000	−0.0021	−0.0036	0.0019	−0.0004	−0.0021			
VAR(1)-cDCC-GARCH		0.0033	0.0005	−0.0012	0.0059	−0.0003	−0.0012	−0.0015	−0.0039	−0.0052	0.0000	−0.0023	−0.0045			
VAR(1)-RSDC-GARCH																
VAR(1)-CCC-GARCH	Covariance	0.0000	−0.0035	−0.0025	−0.0003	−0.0034	−0.0037	−0.0026	−0.0039	−0.0022	−0.0019	−0.0036	−0.0017			
VAR(1)-ECCC-GARCH		0.0135	0.0025	0.0035	0.0000	−0.0051	−0.0048	0.0108	0.0025	0.0044	0.0124	0.0029	0.0038			
VAR(1)-DCC-GARCH		0.0255	−0.0003	−0.0034	0.0232	−0.0024	−0.0015	0.0000	−0.0057	−0.0051	0.0066	0.0010	0.0000			
VAR(1)-cDCC-GARCH		0.0198	−0.0050	−0.0070	0.0210	−0.0063	−0.0086	−0.0048	−0.0155	−0.0132	0.0000	−0.0069	−0.0130			
VAR(1)-RSDC-GARCH																

Figure 8: One-step, two-steps and three-steps estimates of the volatilities of BEL-20, DAX and FTSE-100 observed from January 8, 2002 to April 30, 2009, asuming Gaussian innovations.

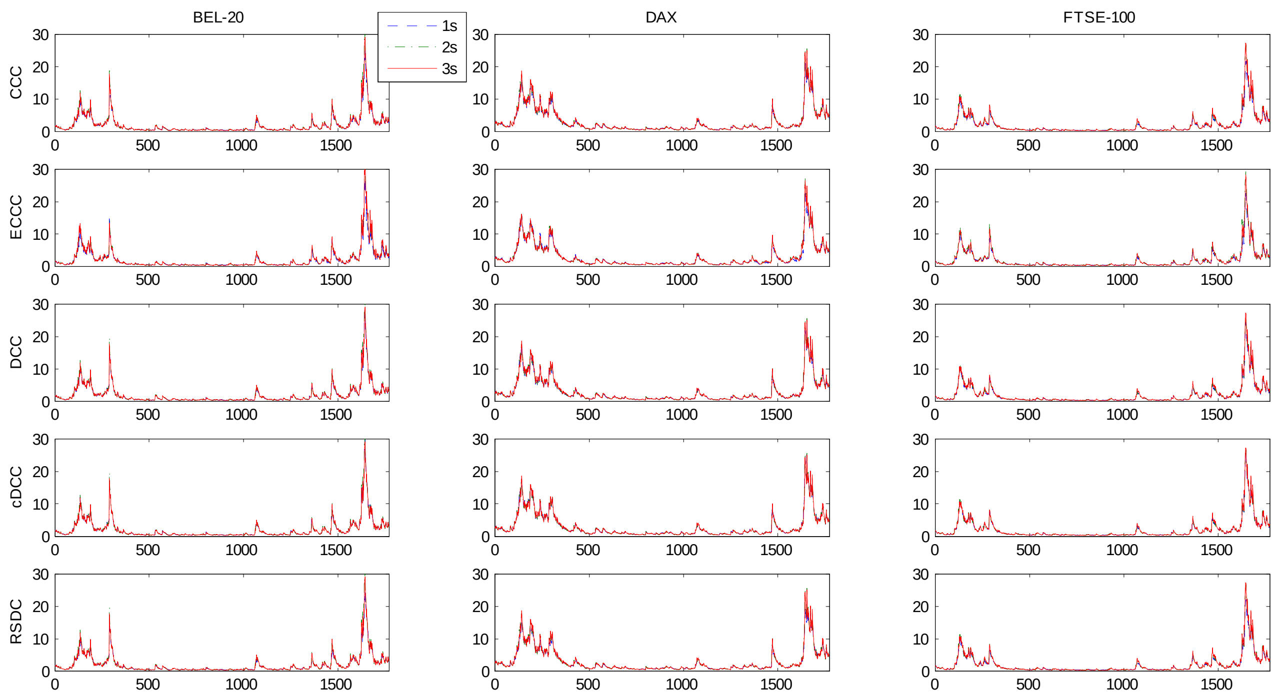
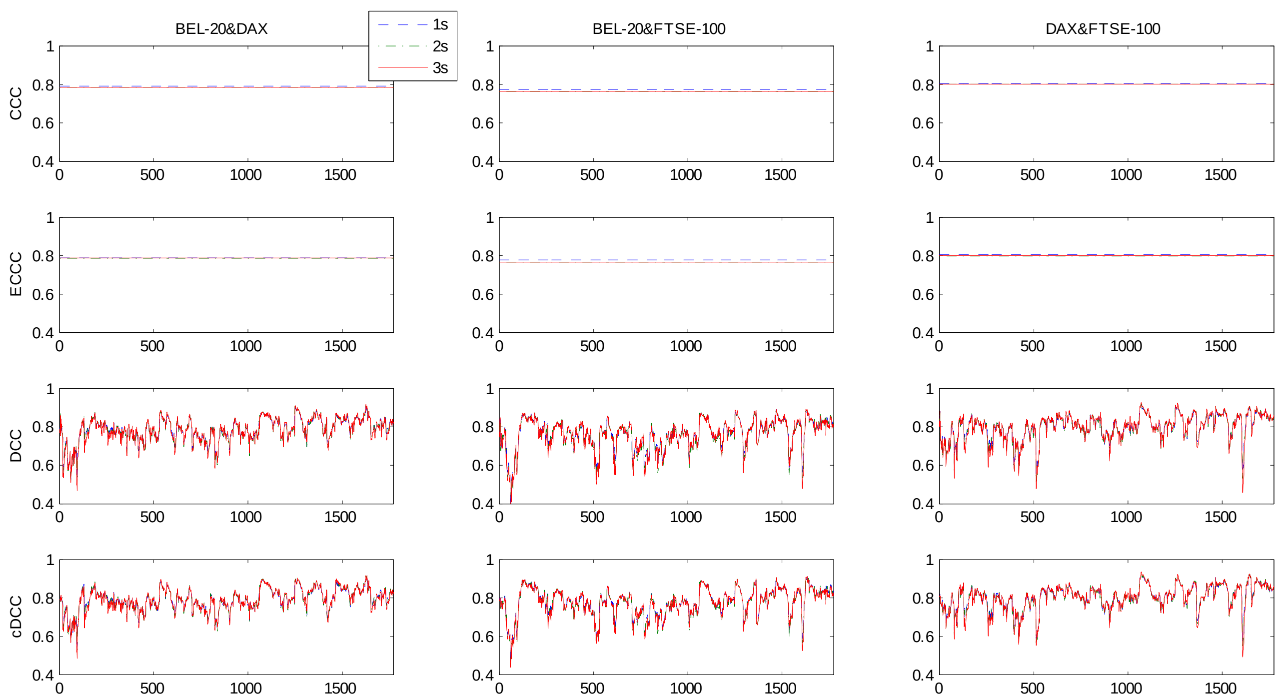


Figure 9: One-step, two-steps and three-steps estimates of the correlations between BEL-20, DAX and FTSE-100 indices observed from January 8, 2002 to April 30, 2009, asuming Gaussian innovations.



relative difference between the estimated and the true ones is given by the ratio

$$ratio_{h,est}^{true} = \frac{1}{TRk} \sum_{t=1}^{T} \sum_{r=1}^{R} \sum_{i=1}^{k} \left(\frac{\widehat{h}_{i,t}^{r} - h_{i,t}}{h_{i,t}} \right) \tag{15}$$

where in our case, $k = 3$, $R = 1000$ and $T = 1000$. The ratios corresponding to the one-step estimation of a model that is correctly specified is set to be equal to 0. Therefore, the ratios reported in Table 4 are relative ratios and they should be read as the performance of the corresponding estimator in a certain model when estimating the volatility (correlation), relative to the one-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 are all very close to zero, indicating that, on average, volatilities and correlations are relatively well estimated even when using a misspecified model.

More specifically, looking at the results for the volatilities, we can see that the largest ratio is 0.0165 and it appears when the true volatilities are generated by the VAR(1)-cDCC-GARCH model and estimated by the VAR(1)-ECCC-GARCH in one step. Other large ratios correspond to three-steps estimators of all the models considered when the data have been generated by the VAR(1)-ECCC-GARCH model. For example, when the true volatilities are generated by the VAR(1)-ECCC-GARCH model and estimated by the VAR(1)-CCC-GARCH in three steps, the ratio is 0.0137, when they are estimated by the VAR(1)-DCC-GARCH, the ratio is 0.0154 and when using the VAR(1)-cDCC-GARCH and the VAR(1)-RSDC-GARCH models to estimate the volatilities, the ratio is 0.0141 and 0.0136 respectively. The reason could be that with the exception of the correlation structure, all the models considered are nested in the VAR(1)-ECCC-GARCH model, being this one more general and therefore, the rest of the models have problems in explaining the volatilities generated by the VAR(1)-ECCC-GARCH model. On the other hand, the true volatilities generated by the VAR(1)-CCC-GARCH model can be well estimated by the other models since CCC-GARCH is nested within all of them.

When looking at the results for the correlations, we can see that the largest ratio is 0.0071 and it appears when the true correlations are generated by the VAR(1)-DCC-GARCH model and estimated by the VAR(1)-ECCC-GARCH in one step. As expected, when the correlations are generated by a dynamic correlation model, their estimation is better when assuming another dynamic correlation model than when a constant correlation model is used. Also expected is the fact that the VAR(1)-ECCC-GARCH model produces better estimates of the correlations generated by the VAR(1)-CCC-GARCH model than the estimates produced by the VAR(1)-CCC-GARCH model when estimating the correlations generated by the VAR(1)-ECCC-GARCH model. The reason could be that the VAR(1)-CCC-GARCH model can not capture the volatility spillovers which indirectly can affect correlations.

In general terms, when volatilities and correlations which have been generated by a particular model are estimated by another model, their estimation seem to get worse as the number of steps used in the estimation increase. On the other hand, the average ratios do not deviate from zero more than 2 % in most of the cases. An interpretation of this result could be that, on average, multiple steps estimates of volatilities (correlations) deviate from the corresponding true volatilities (correlations) at most 2 % more than the amount that one-step estimates of the correctly specified model do.

4.4 Innovations distributed as a Skewed Student-t

In this section, we analyze the case in which innovations follow a skewed Student-t 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:

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

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:

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

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:

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

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 the Monte Carlo experiments we use the estimates reported in Table 5[3] as the true parameter values. We generate 1000 bivariate time series vectors of sample size equal to 1000. Innovations are generated from 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 we estimate the true model assuming Gaussian or Student-t errors but ignoring skewness.

Figure 10 plots the results of the Monte Carlo experiment 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, 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:[4]

$$\Delta\widehat{h}_i^s = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\widehat{h}_{i,t}^s - h_{i,t}}{h_{i,t}}\right\} \tag{16}$$

$$\Delta\widehat{p}_i^s = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\widehat{p}_{i,t}^s - p_{i,t}}{p_{i,t}}\right\} \tag{17}$$

As we can see, for all the models, the kernel densities of the relative deviations of one-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 overestimation of the conditional correlations by around 2-3 % in each model while the conditional volatility estimates don't seem to be affected much.

Figure 11 plots the same estimates as Figure 10 but now the estimation has been done assuming a Student-t distribution for the innovations. Similar conclusions can be made about the one step correlation estimates for all models. We notice that in the CCC and ECCC-GARCH models, the differences between one 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 the five models considered. The figures corresponding to different skewness are not included in the

[3]Table 5 contains parameter estimates for the 5 models considered using daily returns of the BEL-20 and DAX stock market indices under the assumption that innovations are distributed as a Student-t.

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

Table 5: One-step parameter estimates for two real time series under Student-t innovations

	VAR(1)-CCC-GARCH	VAR(1)-ECCC-GARCH	VAR(1)-DCC-GARCH	VAR(1)-cDCC-GARCH	VAR(1)-RSDC-GARCH
μ_1	0.0936	0.0937	0.0956	0.0963	0.0920
μ_2	0.1090	0.1093	0.1143	0.1136	0.1100
β_1	0.0087	0.0088	0.0002	0.0006	0.0030
β_2	-0.0503	-0.0498	-0.0488	-0.0469	-0.0485
ω_1	0.0185	0.0177	0.0156	0.0149	0.0160
ω_2	0.0191	0.0162	0.0152	0.0141	0.0156
α_{11}	0.0877	0.0947	0.0963	0.0978	0.0894
α_{21}		0.0001			
α_{12}		0.0034			
α_{22}	0.0732	0.0692	0.0775	0.0790	0.0717
γ_{11}	0.8987	0.8789	0.8938	0.8950	0.8998
γ_{21}		0.0086			
γ_{12}		0.0001			
γ_{22}	0.9195	0.9225	0.9170	0.9181	0.9218
ρ_{12}	0.7950	0.7949			
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

Figure 10: Kernel density estimates of deviations from estimated to true volatility and correlation for all the models considered. The series in the data are generated using Student-t innovations with same skewness parameter and estimated assuming Gaussian innovations.

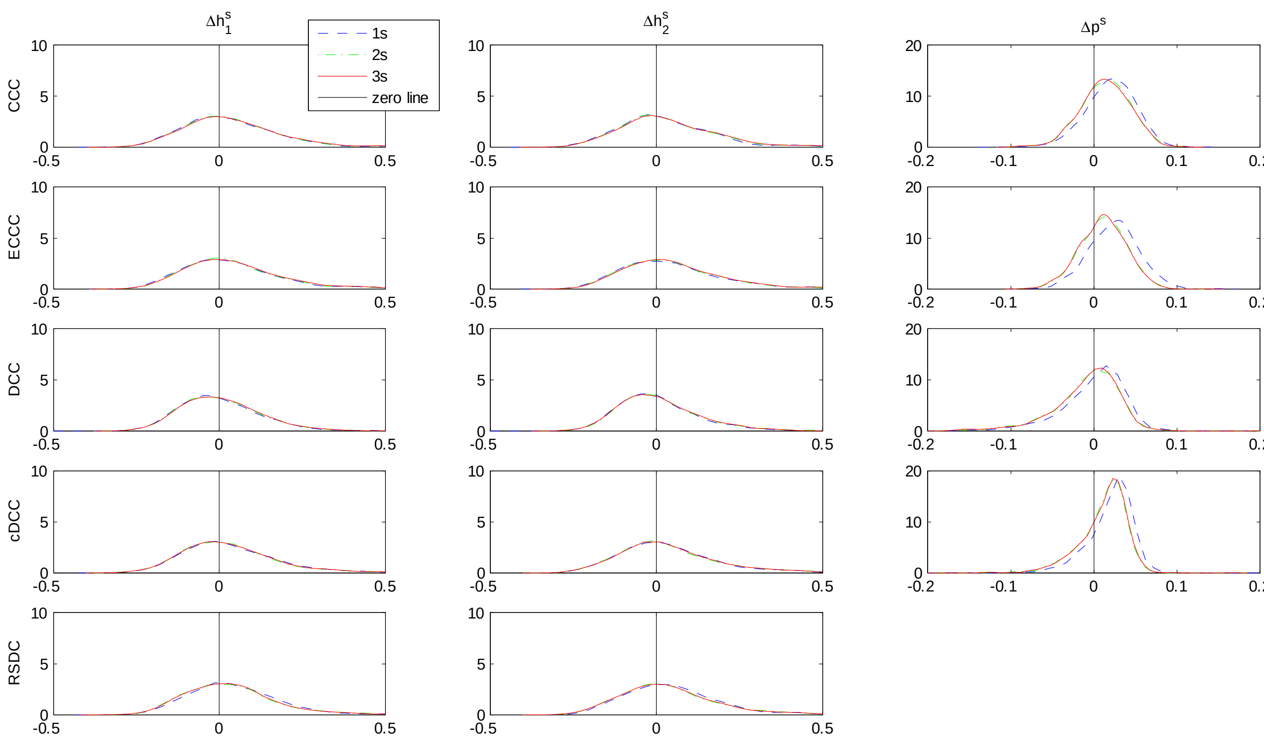
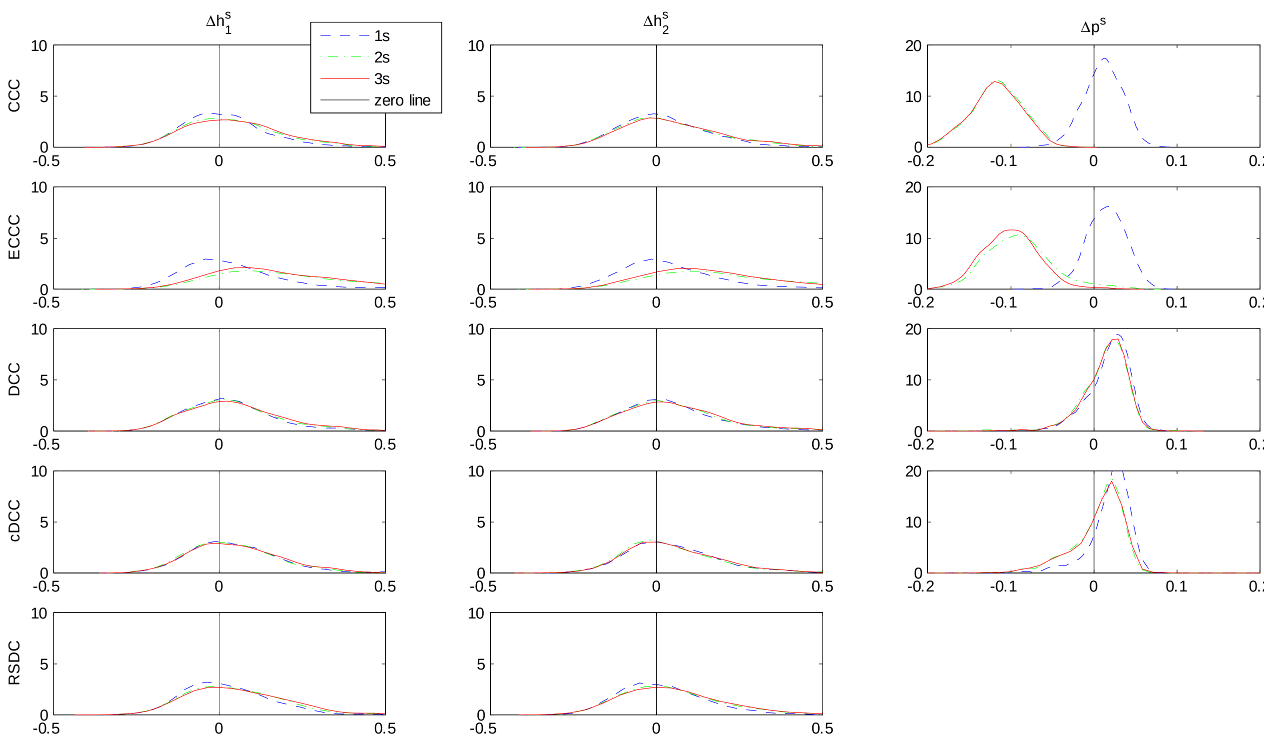


Figure 11: Kernel density estimates of deviations from estimated to true volatility and correlation for all the models considered. The series in the data are generated using Student-t innovations with same skewness parameter and estimated assuming Student-t innovations.



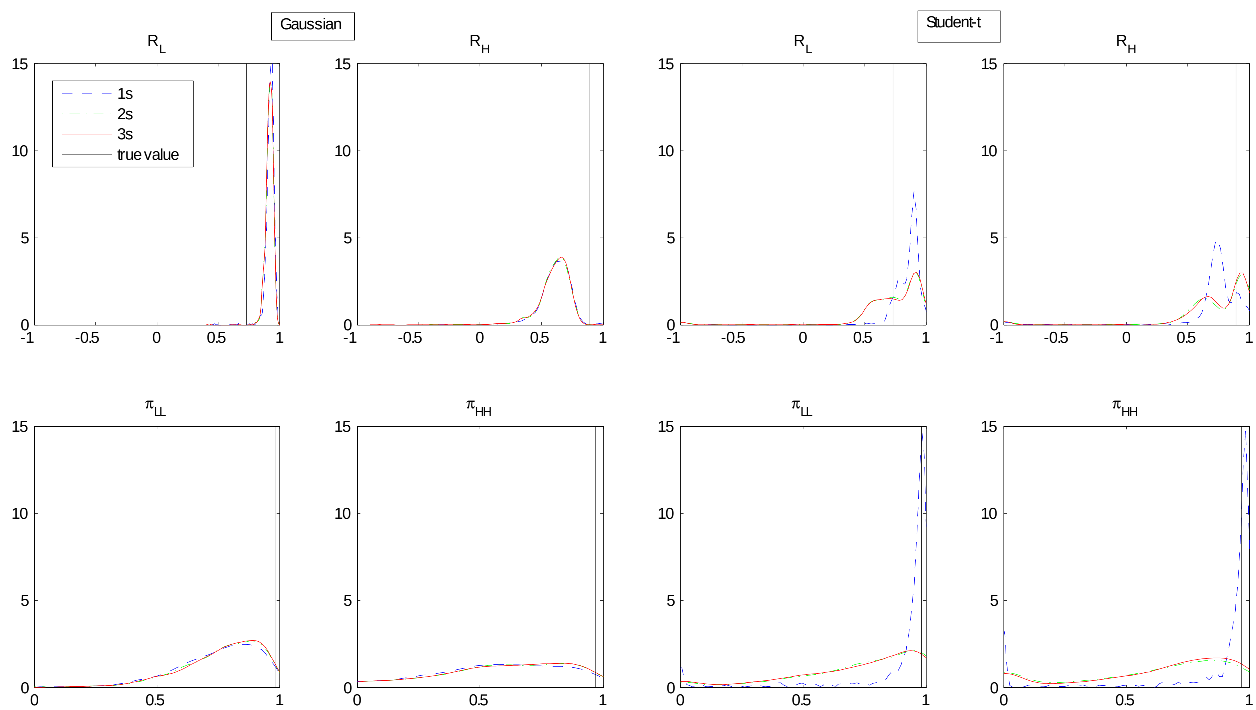
paper to save space. One-step correlation estimates seem to be slightly less affected by the skewness than the multiple step estimates. As well when the estimation is based on Student-t errors, the one-step estimators underestimate the volatilities and correlations. In general, one-step estimators are 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 one step estimates. It should be noted that one of the series have higher skewness when $\lambda = \{\exp(0.4), \exp(-0.7)\}$ compared to the case when $\lambda = \{\exp(0.4), \exp(0.4)\}$ and this could be the reason behind the underestimation of volatilities and correlations with both Gaussian and Student-t errors.

Newey and Steigerwald (1997) suggest that when the data is not symmetrically distributed, the one-step QML method based on Student-t errors do not produce consistent estimators in general. Therefore in this case what is expected is that even though the estimation is performed in one-step, the estimates could be far from the true values and the differences might not disappear in larger samples. In our experiments with a data of length $T = 1000$, we see that one-step QML estimators based on Student-t errors are over/underestimating the volatilities and correlations. We would expect that this result holds for larger datasets produced with the same parameter values and skewness.

For the RSDC-GARCH model, multiple steps estimators of conditional volatilities behave similar to the one-step estimators as illustrated in Figure 5 and this does not seem to depend on the skewness. In this model, the conditional correlations follow an unobserved Markov Chain, therefore instead of reporting correlation estimates, we report the correlation parameter estimates, R_L, R_H, π_{LL}, π_{HH} together with their true values. Figure 12 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 one-step and multiple steps estimators of the correlation parameters are behaving similarly when the series have the same skewness. Although the corresponding figure is not included in the paper, when the skewness of both series is different, the multiple steps estimates of R_L and π_{LL} deviate slightly from the one step estimates. When Student-t errors are used in the estimation, the differences between the behavior of one-step and multiple steps estimators become more apparent.

Finally, when the data generating process is symmetric and the estimation is based on Gaussian errors, the kernel density estimates of relative differences between one-step and multiple steps estimates of the volatilities and correlations from the true values are very close to each other for all the models as was illustrated in Figure 6. Also when the estimation is based on Student-t errors, the multiple steps correlation estimates of CCC and ECCC-GARCH models are far from the true ones as was shown in Figure 4. The multiple steps estimates of DCC and cDCC-GARCH models of volatilities and correlations follow closely the one-step estimates

Figure 12: Kernel density estimates of estimated correlation parameters for the RSDC-GARCH model. The series in the data are generated with Student-t innovations and with same skewness, and estimated assuming Gaussian and Student-t errors, respectively.



and are not far from the true values as in Figure 3. These results are also not reported in the paper, but are available from the authors upon request.

To sum up, we have seen 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 one-step estimators given that their performances are very similar. Given that the estimation based on Gaussian errors is a Quasi-maximum Likelihood estimation, as Bollerslev and Wooldridge (1992) show, it produces consistent estimators. Therefore our results from Section 4.1 and 4.2 still prevail in the existence of skewness. On the other hand, as noted by Newey and Steigerwald (1997), if the data generating process is skewed, the one-step QML estimator based on Student-t errors do not produce consistent estimators. In conformance with this, we have found in this section that the correlations are over-estimated in all models with one-step and and also with multiple steps estimators. Hence, when the true distribution is skewed, one should be cautious in using one-step or multiple-steps estimators based on Student-t errors.

5 Conclusions

In this paper we have carried out several Monte Carlo experiments to study the performance in finite samples of one-step and multiple steps estimators of Vector Autoregressive Multivariate Conditional Correlation GARCH models. Although one-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 one-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 the same results as when the true and assumed distribution is a Student-t.

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 2 % more than what one-step estimates of the correctly specified model do.

Finally, when errors are distributed as a skewed Student-t but the estimation is performed assuming non-skewed Gaussian or Student-t errors, we find that kernel density estimates of the difference between one-steps 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 one-step or multiple-steps estimators based on Student-t errors since both are inconsistent estimators.

References

[1] 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.

[2] Bauwens, L., Laurent, S. (2005), A New Class of Multivariate Skew Densities, With Application to Generalized Autoregressive Conditional Heteroscedasticity Models, *Journal of Business and Economics Statistics,* 23(3), pp 346-354.

[3] Bauwens, L., Laurent, S. and J.V.K. Rombouts (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics*, 21, 79-109.

[4] 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.

[5] 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.

[6] 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.

[7] Conrad, C. and M. Karanasos (2010). Negative volatility spillovers in the unrestricted ECCC-GARCH model. *Econometric Theory*, 26, pp 838-862

[8] 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.

[9] Engle,R. (2009) Anticipating Correlations: A New Paradigm for Risk Management. Princeton University Press

[10] 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.

[11] Engle, R. and K. Sheppard (2001). Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. NBER Working Paper, No: W8554.

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

[13] 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.

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

[15] Hafner C.M. and Reznikova, O. (2010) On the estimatiion of dynamic conditional correlation models, *Computational Statistics and Data Analysis,* forthcoming.

[16] Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press.

[17] Harvey A., E. Ruiz and N. Shephard (1992). Unobservable component time series models with ARCH disturbances. *Journal of Econometrics*, 52: 129–158.

[18] Jeantheau, T. (1998). Strong Consistency of Estimators For Multivariate ARCH Models. *Econometric Theory*, 14, 70-86.

[19] Jondeau, E., Rockinger, M. (2005). Conditional asset allocation under non-normality: how costly is the mean-variance criterion? EFA 2005 Moscow Meetings, Discussion Paper.

[20] Ling, S. and M. McAleer (2003). Asymptotic Theory for a Vector ARMA-GARCH Model. *Econometric Theory*, 19, 280-310.

[21] 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.

[22] 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

[23] 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.

[24] 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.

[25] Newey W.K. and D.G. Steigerwald (1997). Asymptotic bias for quasi maximum likelihood estimators in conditional heteroskedasticity models. *Econometrica*, 3, 587–599.

[26] Pelletier D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics*, 131, 445-473.

[27] Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH Models. In T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch, eds. *Handbook of Financial Time Series.* New York: Springer.

[28] 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.

Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Department of Economics
University of Alicante

Campus San Vicente del Raspeig
03071 Alicante - Spain
Phone: +34 965 903 563
Fax: +34 965 903 898

Website: www.ivie.es
E-mail: publicaciones@ivie.es