UvA ECONOMETRICS

Discussion Paper: 2006/03

Wake me up before you GO-GARCH

H. Peter Boswijk and Roy van der Weide

www.fee.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Wake me up before you GO-GARCH*

H. Peter Boswijk
Tinbergen Institute &
Department of Quantitative Economics
Universiteit van Amsterdam

Roy van der Weide†
Poverty Research Group
The World Bank
Washington DC

September 2006

Abstract

In this paper we present a new three-step approach to the estimation of Generalized Orthogonal GARCH (GO-GARCH) models, as proposed by van der Weide (2002). The approach only requires (non-linear) least-squares methods in combination with univariate GARCH estimation, and as such is computationally attractive, especially in larger-dimensional systems, where a full likelihood optimization is often infeasible. The effectiveness of the method is investigated using Monte Carlo simulations as well as a number of empirical applications.

1 Introduction

Daily time series of financial returns are well known to display volatility clustering; i.e., highly volatile periods alternate with periods of low volatility. The class of GARCH (generalized autoregressive-conditional heteroskedasticity) models have proved to be a suitable approach to describe this property, and to exploit it for purposes of risk management and derivative pricing. When such applications depend on the joint evolution of a vector of returns, for example for risk management of a portfolio (with time-varying weights), then multivariate generalizations of the GARCH model are required. Until fairly recently, the application of such multivariate GARCH models was confined to low-dimensional systems, because the dimension of the parameter space increases very rapidly with the dimension of the system.

Balancing generality and feasibility is largely what defines the challenge in multivariate GARCH modelling. One recent model is the Generalized Orthogonal GARCH model (GO-GARCH; see van der Weide, 2002). The GO-GARCH model essentially consists of two parts: (i) a set of conditionally uncorrelated univariate GARCH processes; and (ii) a linear map that

*This research has been supported by the Netherlands Organization for Scientific Research (NWO) under a NWO-MaG Pionier grant.

†Corresponding author: Roy van der Weide, DECRG, The World Bank, Washington DC, USA. E-mail: rvanderweide@worldbank.org.

relates these components to the observed data. The model was obtained as a generalization of the O-GARCH model (Alexander, 2001), which is nested as a special case where the linear map is orthogonal. The practical power of the O-GARCH model lies in its two-step method. In particular, all GARCH parameters can be conveniently estimated by appealing to univariate GARCH models for the principal components of the original series. The first step, where the orthogonal map is estimated, involves only solving an eigenvalue problem.

With the extension from an orthogonal link matrix to an arbitrary invertible link matrix, the GO-GARCH model has not kept this practical two-step structure intact. The method originally proposed by van der Weide (2002) is in fact a two-step approach, but not one where only univariate GARCH estimation is needed in the second step. In the first step, only part of the invertible matrix is identified. The remaining parameters of this link matrix, together with the univariate GARCH parameters for the factors, are then estimated by maximizing a multivariate likelihood function. The latter can be troublesome for high dimensions.

Therefore, this paper puts forward a new method for estimating the GO-GARCH model. This alternative method is meant to make the model more attractive purely from a practical point of view. In comparison to the two-step method originally proposed, our alternative is more practical in terms of implementation, as well as more feasible in terms of estimation. With improved feasibility we mean that the new approach is less likely to experience convergence difficulties when compared with the original method. This advantage will be most noticeable, and most welcomed, in higher dimensions. The price we pay for these practical advances is a loss of efficiency.

These practical improvements have been realized by adding a third step to the original two-step method. With this extra step, the estimation of the second part of the link matrix is separated from estimation of the univariate GARCH models. Apart from the practical aspects, keeping estimation of the link matrix and the GARCH parameters separate also has another notable advantage. The link matrix is estimated without imposing any knowledge of GARCH structure (for the uncorrelated components). Although this is part of where the loss in efficiency stems from, at the same time our estimator of the link matrix will also be robust with respect to any misspecification in the GARCH structure.

Another multivariate GARCH class is the Dynamic Conditional Correlation (DCC) model by Engle (2002), which was obtained as a generalization of the Constant Conditional Correlation (CCC) model by Bollerslev (1990). The CCC model has a convenient structure, where estimation of the parameters may (also) be done by means of univariate GARCH estimation, after which the constant correlation may be easily estimated. By extending the model to allow for time-varying correlation, DCC managed to keep this practical two-step approach intact. For an overview of the recent advances on multivariate GARCH modelling we refer to Bauwens *et al.* (2006).

The generality of any multivariate GARCH model can be measured by the ability to account for the key stylized facts of multivariate data: (i) persistence in volatility and covariation; (ii)

time-varying correlation; and (iii) spill-over effects in volatility. Additionally it is desired that the model is closed under linear transformations. The DCC model incorporates the first two items, but does not allow for volatility spill-overs, nor is it closed under linear transformations. GO-GARCH meets all criteria, but gives in on DCC in terms of practicability. With our alternative three-step estimation approach we managed to improve on feasibility of the GO-GARCH model, while preserving its level of generality.

The remainder of the paper is structured as follows. Section 2 discusses the GO-GARCH model, and the originally proposed two-step estimation procedure. The three-step estimator is derived in Section 3. Section 4 provides a brief examination of the loss function. Section 5 studies some of the finite-sample properties of the new estimator. Some empirical applications to a number of widely traded stock indices of the United States and Europe are presented in Section 6 and 7, and Section 8 concludes.

2 The model

Consider an m-vector process $\{x_t, t = 1, 2, \ldots\}$ of financial returns. We are interested in modelling the conditional variance matrix $V_t = \mathrm{Var}(x_t|\mathcal{F}_{t-1})$, where $\mathcal{F}_t = \{x_t, x_{t-1}, \ldots\}$. The starting point of the GO-GARCH model is the assumption that there exists an $m \times m$ non-singular matrix Z, and an m-vector process $\{y_t = (y_{1t}, \ldots, y_{mt})'\}$ of *independent*[1] component processes (i.e., $\{y_{it}\}$ is independent of $\{y_{jt}\}$, $i \neq j$), such that

$$x_t = Zy_t. \tag{1}$$

Each of the component processes $\{y_{it}\}$ is modelled as a univariate GARCH process, i.e.,

$$y_{it}|\mathcal{F}_{t-1} \sim N(0, h_{it}), \tag{2}$$

$$h_{it} = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1}. \tag{3}$$

This implies that

$$x_t|\mathcal{F}_{t-1} \sim N(0, V_t), \tag{4}$$

$$V_t = ZH_tZ', \tag{5}$$

with $H_t = \mathrm{diag}(h_{1t}, \ldots, h_{mt})$. The specification (2)–(3) and hence (4)–(5) may be extended in various directions, to allow for a non-zero conditional mean of x_t, a non-Gaussian conditional distribution, or conditional variance specifications other than GARCH(1,1). The essential assumption is (1), together with independence of the processes $\{y_{it}\}$.

The model was proposed by van der Weide (2002), as a generalization of the orthogonal GARCH model pioneered by Ding (1994) and popularized by Alexander (2001), where the

[1] Although the discussion is facilitated by assuming the component processes to be independent, all that is needed is that they are conditionally uncorrelated, with conditional variances that depend only on the past of the separate processes.

matrix Z is restricted to be orthogonal. It may also be viewed as a factor GARCH model where the number of factors equals the dimension of the system (and with no idiosyncratic noise). As such, it was analyzed by Vrontos *et al.* (2003), although they restrict Z to be lower triangular, which is not without loss of generality.

Throughout this paper we will assume that y_t and hence x_t is covariance stationary, such that the unconditional variances $H := \mathrm{Var}(y_t)$ and $V := \mathrm{Var}(x_t) = ZHZ'$ exist. In the O-GARCH model of Alexander (2001), the time series $\{x_t\}$ has been standardized such that $V := \mathrm{Var}(x_t)$ is in fact the correlation matrix of x_t. Let P be the orthogonal matrix of eigenvectors of V, and Λ the diagonal matrix containing the corresponding eigenvalues, such that $V = P\Lambda P'$. The principal component vector $y_t = P'x_t$ then satisfies $H = \mathrm{Var}(y_t) = P'VP = \Lambda$, such that the components of y_t are unconditionally uncorrelated. This property is then strengthened to the assumption that the components of y_t are independent, or at least conditionally uncorrelated, with no dynamic links between their respective volatility processes h_{it}. Therefore this is a special case[2] of the GO-GARCH model, with $Z = P$, an orthogonal matrix, and y_t the principal components of x_t.

In the GO-GARCH model, we may impose that each of the components y_{it} has an unconditional variance of unity, which in the GARCH(1,1) specification (3) leads to the restriction

$$\omega_i = 1 - \alpha_i - \beta_i. \tag{6}$$

This is an identifying restriction, since the variance of x_t is determined by Z and hence not restricted by (6). Indeed, this restriction implies $\mathrm{Var}(y_t) = E(H_t) = I_m$, and hence

$$V = \mathrm{Var}(x_t) = ZZ'. \tag{7}$$

Consider the singular value decomposition of Z:

$$Z = P\Lambda^{1/2}U', \tag{8}$$

where P contains the orthonormal eigenvectors of $ZZ' = V$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ contains the corresponding eigenvalues (with $\lambda_i > 0$ for all i), and where U is the orthogonal matrix of eigenvectors of $Z'Z$. Note that P and Λ are identified from the unconditional variance matrix V. The matrix U, on the other hand, is only identified from the structure of the *conditional* variance matrix V_t. The O-GARCH model (with a slight redefinition of $y_t \mapsto \Lambda^{1/2}y_t$, loosening the restriction $\mathrm{Var}(y_t) = I_m$) corresponds to the restriction $U = I_m$. Note that in the GO-GARCH model the components of x_t do not have to be standardized as in the O-GARCH

[2]For larger-dimensional systems, Alexander (2001) proposes to include only the first $k < m$ principal components in y_t, which together explain a sufficient percentage of the total variation in x_t. Then Z only contains the first k columns from P. Note however, that the implied conditional variance matrix of x_t is singular if $k < m$. Therefore, in the GO-GARCH model x_t and y_t always have the same dimension. Recently, Lanne and Saikkonen (2005) introduced a test procedure that can be used to identify the number of factors needed to model the conditional variance of the observed vector x_t, in a GO-GARCH setting (they consider a non-singular Z).

model, which would lead to a different matrix P; however, for comparison purposes between the two models we will use the standardization in the empirical analysis of this paper.

The distinction between 'unconditional information' (which identifies P and Λ) and 'conditional information' (which identifies U) is the basis of the two-step estimation procedure proposed by van der Weide (2002). Define $s_t := \Lambda^{-1/2}P'x_t$, such that the GO-GARCH model implies $s_t = U'y_t$. Since V is estimated consistently by $\hat{V} = n^{-1}\sum_{t=1}^{n} x_t x_t'$, this yields consistent estimates $\hat{P}$ and $\hat{\Lambda}$ of P and Λ, respectively, and hence $\hat{s}_t := \hat{\Lambda}^{-1/2}\hat{P}'x_t$. The remaining parameters U and $\{\alpha_i, \beta_i\}_{i=1}^{m}$ may then be estimated by maximum likelihood, replacing the unobserved s_t by $\hat{s}_t$. Note that orthogonality of U implies that this matrix is characterized by $m(m-1)/2$ free parameters.

Although this procedure is feasible for moderately large portfolios, for larger values of m the full maximization of the likelihood function over the free parameters in U becomes problematic. Therefore, this paper proposes an alternative procedure to estimate the matrix U and hence Z. Note that given a consistent estimator $\hat{Z}$ of Z, we may construct $\hat{y}_t := \hat{Z}^{-1}x_t$, and estimate the parameters $\{\alpha_i, \beta_i\}_{i=1}^{m}$ in univariate GARCH models for $\{\hat{y}_{it}\}$.

3 Estimation

The idea of our approach is to identify U from the (cross-) autocorrelation structure of $s_t s_t'$, where $s_t = \Lambda^{-1/2}P'x_t$ is the standardized and orthogonalized version of x_t. In order to motivate the estimator proposed below, we start by analyzing some moment conditions on squares and cross-products of the components of y_t, and then investigate what they imply for $s_t s_t'$.

Let ρ_i denote the first-order autocorrelation of y_{it}^2. If h_{it} follows a GARCH(1,1) specification, as assumed above, then an explicit expression for ρ_i is available, implying $\rho_i > 0$. However, the procedure proposed here is intended to be applicable for a wider range of GARCH-type specifications, as long as $\rho_i := \mathrm{Corr}(y_{it}^2, y_{i,t-1}^2) > 0$.

Because y_{it}^2 is independent of $y_{j,t-1}^2$ and uncorrelated with $y_{i,t-1}y_{j,t-1}$, $j \neq i$, the linear projection of $y_{it}^2 - 1$ on $w_{t-1} := \mathrm{vech}(y_{t-1}y_{t-1}')$ is given by

$$y_{it}^2 - 1 = \rho_i(y_{i,t-1}^2 - 1) + \upsilon_{ii,t}, \tag{9}$$

where $E(\upsilon_{ii,t}) = 0$ and $\mathrm{Cov}(\upsilon_{ii,t}, w_{t-1}) = 0$. Next, because $E(y_{it}y_{jt}|\mathcal{F}_{t-1}) = 0$, it follows that the linear projection of $y_{it}y_{jt}$ on w_{t-1} is 0, such that $y_{it}y_{jt} = \upsilon_{ij,t}$, where $\upsilon_{ij,t}$ again satisfies the properties $E(\upsilon_{ij,t}) = 0$ and $\mathrm{Cov}(\upsilon_{ij,t}, w_{t-1}) = 0$. These results may be summarized as

$$\begin{aligned} y_t y_t' - I_m &= D_\rho(y_{t-1}y_{t-1}' \odot I_m - I_m) + \Upsilon_t \\ &= D_\rho^{1/2}\left(y_{t-1}y_{t-1}' \odot I_m - I_m\right) D_\rho^{1/2} + \Upsilon_t, \end{aligned} \tag{10}$$

where $D_\rho = \mathrm{diag}(\rho_1, \ldots \rho_m)$, where "$\odot$" denotes the Hadamard (element-by-element) product, and where $\Upsilon_t = (\upsilon_{ij,t})$ has mean zero and is uncorrelated to w_{t-1}.

Suppose now that we estimate the following regression model

$$y_t y_t' - I_m = A\left(y_{t-1}y_{t-1}' - I_m\right)A + E_t, \tag{11}$$

with A a symmetric matrix, by non-linear least-squares. That is, suppose that we minimize

$$S(A) = \frac{1}{n}\sum_{t=1}^{n} \operatorname{tr}\left(E_t^2\right) = \frac{1}{n}\sum_{t=1}^{n} \operatorname{tr}\left(\left[y_t y_t' - I_m - A\left(y_{t-1}y_{t-1}' - I_m\right)A\right]^2\right) \tag{12}$$

over symmetric matrices A. The difference between (10) and (11) is that the diagonal matrix $y_{t-1}y_{t-1}' \odot I_m$ has been replaced by $y_{t-1}y_{t-1}'$. This means, in particular, that the off-diagonal elements of the matrix $A\left(y_{t-1}y_{t-1} - I_m\right)A$ are no longer zero, even if A is diagonal. In this sense, (11) is a misspecified regression model. However, we might expect that the pseudo-true value of A, i.e., the solution to the equations $E\left[\partial S(A)/\partial A\right] = 0$, is still a diagonal matrix. This question is analysed in Theorem 1 below, but first we demonstrate how this result is used to obtain an estimate of U.

Using $y_t = Us_t$, and the properties $\operatorname{tr} AB = \operatorname{tr} BA$ and $U'U = UU' = I_n$, we may rewrite the criterion function as

$$\begin{aligned} S(A) &= \frac{1}{n}\sum_{t=1}^{n} \operatorname{tr}\left(\left[Us_t s_t' U' - I_m - AU\left(s_{t-1}s_{t-1} - I_m\right)U'A\right]^2\right) \\ &= \frac{1}{n}\sum_{t=1}^{n} \operatorname{tr}\left(\left[s_t s_t' - I_m - U'AU\left(s_{t-1}s_{t-1} - I_m\right)U'AU\right]^2\right) \\ &= \frac{1}{n}\sum_{t=1}^{n} \operatorname{tr}\left(\left[s_t s_t' - I_m - B\left(s_{t-1}s_{t-1} - I_m\right)B\right]^2\right) \\ &= S^*(B). \end{aligned} \tag{13}$$

It is immediately clear from the first-order conditions that if $\hat{A}$ and $\hat{B}$ minimize $S(A)$ and $S^*(B)$ over symmetric matrices, respectively, then $\hat{B} = U'\hat{A}U$. Therefore, if $\operatorname{plim}_{n\to\infty}\hat{A} = D_a = \operatorname{diag}(a_1, \ldots, a_m)$, then $\operatorname{plim}_{n\to\infty}\hat{B} = U'D_aU$. This in turn means that the eigenvectors of $\hat{B}$ provide a consistent estimator of U', provided that the corresponding eigenvalues $a_1, \ldots, a_m$ are distinct. Let us provide a summary.

Summary 1 *The observed data is described by $s_t = U'y_t$, where y_t represents a vector of independent processes and where U denotes an orthogonal matrix. The assumptions imply $Es_t s_t' = Ey_t y_t' = I_m$. Our objective is to estimate U. The proposed method considers the following matrix regression:*

$$s_t s_t' - I_m = B\left(s_{t-1}s_{t-1}' - I_m\right)B + \Gamma_t, \quad E(\Gamma_t) = 0, \tag{14}$$

which is equivalent to equation (11). The relation $y_t = Us_t$ implies $B = U'AU$, which we will utilize to estimate U. Given that A is diagonal, U may be obtained as the eigenvector matrix of B, provided that the elements of $A = D_a$ are distinct. Consistent estimates of $B = U'D_aU$ will then provide consistent estimates of U.

The proposed method of estimation is built on the assumption that A is diagonal, at least asymptotically. Theorem 1 proves that this assumption is valid. An explicit expression of the pseudo-true value of A is derived, which confirms that $A = D_a$ is diagonal. It can be seen that D_a does not necessarily equal $D_\rho^{\frac{1}{2}}$. In general, it is a function of both the first-order autocorrelations ρ_i of y_{it}^2 and the kurtosis κ_i of y_{it}, $i = 1, \ldots, m$.

Theorem 1 *Assume that the m-vector process $\{y_t, t \geq 1\}$ consists of independent stationary processes $\{y_{it}, t \geq 1\}$ with mean zero and unit unconditional variance. Next, assume that $\rho_i = \mathrm{Corr}(y_{it}^2, y_{i,t-1}^2) > 0$ and define $\kappa_i = E(y_{it}^4)$ (possibly infinite). Then the equation $E\left[\partial S(A)/\partial A\right] = 0$, with $S(A)$ defined in (12), has a diagonal solution $A = D_a = \mathrm{diag}(a_1, \ldots, a_m)$, where a_i satisfies*

$$a_i \left(\rho_i(\kappa_i - 1) - a_i^2(\kappa_i - 1) - \sum_{j \neq i} a_j^2 \right) = 0, \qquad i = 1, \ldots, m. \tag{15}$$

We distinguish three cases:

1. *The equations (15) have a solution with $a_i^2 > 0$ for all i, given by*

$$\begin{pmatrix} a_1^2 \\ \vdots \\ a_m^2 \end{pmatrix} = \begin{pmatrix} \rho_1(1+\theta_1) \\ \vdots \\ \rho_m(1+\theta_m) \end{pmatrix} - \left(\frac{\rho_1(1+\theta_1) + \ldots + \rho_m(1+\theta_m)}{1 + \theta_1 + \ldots + \theta_m} \right) \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_m \end{pmatrix}, \tag{16}$$

   where $\theta_i = \frac{1}{\kappa_i - 2}$.

2. *The equations (15) have $a_i = 0$, $\forall i$ as the only solution.*

3. *The equations (15) have a partly positive solution, with some of the diagonal elements a_i equal to zero. Letting P denote the set of indices corresponding to the nonzero positive solutions, we find*

$$a_i^2 = \rho_i(1+\theta_i) - \left(\frac{\sum_{j \in P} \rho_j(1+\theta_j)}{1 + \sum_{j \in P} \theta_j} \right) \theta_i. \tag{17}$$

   If only one positive solution remains, $D_a = \mathrm{diag}\left(0, \ldots, \sqrt{\rho_i}, \ldots, 0\right)$.

Proof. It will be convenient to introduce $Q_t = y_t y_t' - I_m$. Using symmetry of $E_t = Q_t - AQ_{t-1}A$ for all symmetric matrices A, the partial derivative $\partial S(A)/\partial A$ is obtained from

$$\begin{aligned} d \operatorname{tr} E_t^2 &= \operatorname{tr} E_t dE_t + \operatorname{tr}(dE_t) E_t \\ &= 2 \operatorname{tr} E_t dE_t \\ &= -2 \operatorname{tr} E_t A Q_{t-1} dA - 2 \operatorname{tr} E_t (dA) Q_{t-1} A \\ &= -4 \operatorname{tr} E_t A Q_{t-1} dA \\ &= -4 (\operatorname{vec} E_t A Q_{t-1})' D (d \operatorname{vech} A), \end{aligned}$$

where D is the duplication matrix, such that $\operatorname{vec} A = D \operatorname{vech} A$ for any symmetric matrix A. Therefore,

$$\frac{\partial S(A)}{\partial \operatorname{vech} A} = \frac{-4}{n} D' \sum_{t=1}^{n} \operatorname{vec} E_t A Q_{t-1}.$$

The pseudo-true value A^* is obtained from setting the expected partial derivative to zero. Therefore, it solves the equations

$$E\left[\left(Q_t - A^* Q_{t-1} A^*\right) A^* Q_{t-1}\right] = 0. \tag{18}$$

The assumptions imply that

$$Q_t = D_\rho^{1/2}(Q_{t-1} \odot I_m) D_\rho^{1/2} + \Upsilon_t, \tag{19}$$

see (10). Substitution of (19) into (18) leads to

$$\begin{aligned} & E\left[\left(D_\rho^{1/2}(Q_{t-1} \odot I_m) D_\rho^{1/2} + \Upsilon_t - A^* Q_{t-1} A^*\right) A^* Q_{t-1}\right] \\ = \; & E\left[D_\rho^{1/2}(Q_{t-1} \odot I_m) D_\rho^{1/2} A^* Q_{t-1}\right] - E\left[A^* Q_{t-1} A^{*2} Q_{t-1}\right] = 0. \end{aligned}$$

We now investigate if this equation has a diagonal solution $A^* = D_a$. For convenience, replace Q_{t-1} by $Q = (q_{ij})$, and note that

$$\begin{aligned} E(q_{ii}) &= E(q_{ij}) = 0, \\ E(q_{ii}^2) &= \kappa_i - 1, \qquad E(q_{ij}^2) = 1, \\ E(q_{ii} q_{jj}) &= E(q_{ii} q_{ij}) = E(q_{ii} q_{jk}) = E(q_{ij} q_{kl}) = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} E\left[D_\rho^{1/2}(Q \odot I_m) D_\rho^{1/2} D_a Q\right] &= E\left(\begin{bmatrix} \rho_1 a_1 q_{11}^2 & \cdots & \rho_1 a_1 q_{11} q_{1m} \\ \vdots & \cdots & \vdots \\ \rho_m a_m q_{mm} q_{m1} & \cdots & \rho_m a_m q_{mm}^2 \end{bmatrix}\right) \\ &= \begin{bmatrix} \rho_1 a_1 (\kappa_1 - 1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_m a_m (\kappa_m - 1) \end{bmatrix}, \end{aligned}$$

and

$$\begin{aligned} E\left[D_a Q D_a^2 Q\right] &= E\left(\begin{bmatrix} a_1^2 q_{11} & \cdots & a_1 a_m q_{1m} \\ \vdots & \ddots & \vdots \\ a_m a_1 q_{m1} & \cdots & a_m^2 q_{mm} \end{bmatrix} \begin{bmatrix} a_1 q_{11} & \cdots & a_1 q_{1m} \\ \vdots & \ddots & \vdots \\ a_m q_{m1} & \cdots & a_m q_{mm} \end{bmatrix}\right) \\ &= \begin{bmatrix} a_1\left(a_1^2(\kappa_1 - 1) + \sum_{i \neq 1} a_i^2\right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_m\left(a_m^2(\kappa_m - 1) + \sum_{i \neq m} a_i^2\right) \end{bmatrix}. \end{aligned}$$

The relevant equations therefore become

$$a_i \left(\rho_i(\kappa_i - 1) - a_i^2(\kappa_i - 1) - \sum_{j \neq i} a_j^2 \right) = 0, \qquad i = 1, \ldots, m,$$

which has the trivial solution $a_i = 0 \; \forall i$, a partly positive-, and a strictly positive solution. The positive elements can be expressed as a solution to following matrix equation:

$$\left(D_\kappa + \mathbf{1}\mathbf{1}' \right) \begin{pmatrix} a_1^2 \\ \vdots \\ a_m^2 \end{pmatrix} = \begin{pmatrix} (\kappa_1 - 1)\rho_1 \\ \vdots \\ (\kappa_m - 1)\rho_m \end{pmatrix}, \tag{20}$$

where $D_\kappa = \operatorname{diag}(\kappa_1 - 2, \ldots, \kappa_m - 2)$, and $\mathbf{1}$ is a vector of ones. Isolating the elements of D_a by inverting the matrix on the left we obtain

$$\begin{pmatrix} a_1^2 \\ \vdots \\ a_m^2 \end{pmatrix} = \left(D_\kappa^{-1} - \frac{\theta\theta'}{1 + tr\left(D_\kappa^{-1}\right)} \right) \begin{pmatrix} (\kappa_1 - 1)\rho_1 \\ \vdots \\ (\kappa_m - 1)\rho_m \end{pmatrix}, \tag{21}$$

where $\theta = (\theta_1, \ldots, \theta_m)'$ with $\theta_i = \frac{1}{\kappa_i - 2}$. By rearranging terms we find

$$\begin{pmatrix} a_1^2 \\ \vdots \\ a_m^2 \end{pmatrix} = \begin{pmatrix} \rho_1(1 + \theta_1) \\ \vdots \\ \rho_m(1 + \theta_m) \end{pmatrix} - \left(\frac{\sum_i \rho_i(1 + \theta_i)}{1 + \sum_i \theta_i} \right) \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_m \end{pmatrix}, \tag{22}$$

which corresponds to the expression presented in the theorem. The partly positive solution is obtained by first deciding on the zero elements, those who can not satisfy (17), and then solve the positive elements by considering the linear equation for these elements only. □

Note that the conditions in Theorem 1 place no restrictions on the fourth moments of the independent processes $\{y_{it}\}$. If the kurtosis of the ith process, κ_i, tends to infinity, then $\theta_i \to 0$, and hence $a_i \to \sqrt{\rho_i}$. However, the consistency proof of the estimator $\hat{A}$, defined by the first-order condition $\partial S(\hat{A})/\partial A = 0$, requires the fourth moments to be finite. This is because $\operatorname{plim}_{n\to\infty} \hat{A} = D_a$ follows from uniform convergence of $\partial S(A)/\partial A$ to a non-stochastic function $G(A) = E[\partial S(A)/\partial A]$, satisfying $G(D_a) = 0$. When $\kappa_i \to \infty$, then $E[\partial S(A)/\partial A]$ is not defined for any other value than $A = D_a$, so that $\partial S(A)/\partial A$ does not converge uniformly, and hence consistency cannot be proved. In Section 5, we investigate by Monte Carlo simulation how the procedure performs in the infinite-kurtosis case.

Let $\hat{B}$ be the least-squares estimator in (14), i.e., $\hat{B} = \arg\min_B S^*(B)$, where $S^*(B)$ is defined in (13), and the minimization is over symmetric matrices B. Theorem 1 implies that if $\kappa_i < \infty$ for all i, then $\operatorname{plim}_{n\to\infty} \hat{B} = U' D_a U$. To conclude from this that the matrix of eigenvalues $\hat{U}$ of $\hat{B}$ is a consistent estimator of U, the diagonal elements of D_a need to be distinct. This leads to the next corollary.

Corollary 1 *The estimator $\hat{U}$ defined in Summary 1 is consistent if:*
(i) $\kappa_i < \infty \; \forall i$, and
(ii) the solutions $a_i, i = 1, \ldots, m$ described in Theorem 1 are distinct.
Lack of identification, and hence inconsistency of $\hat{U}$, may occur if $a_i = a_j = 0$, or if $a_i = a_j > 0$ for some $i \neq j$. Let P denote the set of indices corresponding to the positive elements. Then $a_i = a_j > 0$ occurs if

$$\rho_i(1+\theta_i)\left(1+\sum_{l\in P}\theta_l\right) - \theta_i \sum_{l\in P}\theta_l = \rho_j(1+\theta_j)\left(1+\sum_{l\in P}\theta_l\right) - \theta_j \sum_{l\in P}\theta_l. \tag{23}$$

The trivial solution here is given by $\rho_i = \rho_j$ and $\theta_i = \theta_j$ (implying $\kappa_i = \kappa_j$).

Henceforth we will refer to the region in parameter space described by Corollary 1 as the 'non-identification area'.

4 The objective function and its maxima

In this section we will derive the asymptotic objective function and investigate the nature of its maxima as we move through the parameter space. This also allows us to observe how zeros enter the diagonal of the pseudo-true value of $A = D_a$. The example we use here will also be used in the Monte Carlo section. As such the knowledge accumulated in this section will directly carry over to the next.

For the sake of clarity we consider the bivariate case. Each of the two independent components is described by a stationary ARCH(1) process:

$$y_{1,t} = \sqrt{h_{1,t}}\varepsilon_{1,t}, \qquad y_{2,t} = \sqrt{h_{2,t}}\varepsilon_{2,t}, \tag{24}$$

where $E\varepsilon_{i,t} = E\varepsilon_{j,t} = E\varepsilon_{i,t}\varepsilon_{j,t} = 0$, and where

$$h_{1,t} = (1-\alpha_1) + \alpha_1 y_{1,t-1}^2 \tag{25}$$

$$h_{2,t} = (1-\alpha_2) + \alpha_2 y_{2,t-1}^2 \tag{26}$$

This means $Ey_t y_t' = I_2$ while $E_{t-1} y_t y_t' = H_t = \mathrm{diag}\,(h_{1,t}, h_{2,t})$.

Next we investigate how these model parameters determine the pseudo-true value of $A = D_a$. From Theorem 1 we know that the model parameters enter the objective function via the implied autocorrelations ρ_i and the kurtosis κ_i, if finite. For the ARCH(1) process these are given by

$$\rho_i = \alpha_i \tag{27}$$

$$\kappa_i = \frac{3\left(1-\alpha_i^2\right)}{1-3\alpha_i^2}, \tag{28}$$

We will use these in the proof of the next lemma.

Lemma 1 *The curves in the (α_1, α_2)-plane along which a zero enters the diagonal of D_a are described by*

$$C_1: \qquad 2\alpha_1 = \alpha_2\left(1 - 3\alpha_1^2\right) \tag{29}$$

$$C_2: \qquad 2\alpha_2 = \alpha_1\left(1 - 3\alpha_2^2\right). \tag{30}$$

The partly positive solutions for the diagonal elements of D_a are given by $\left(\sqrt{\alpha_1}, 0\right)$ and $\left(0, \sqrt{\alpha_2}\right)$.

Proof. Working with the expressions for positive a_i from Theorem 1, and setting them equal to zero, we obtain

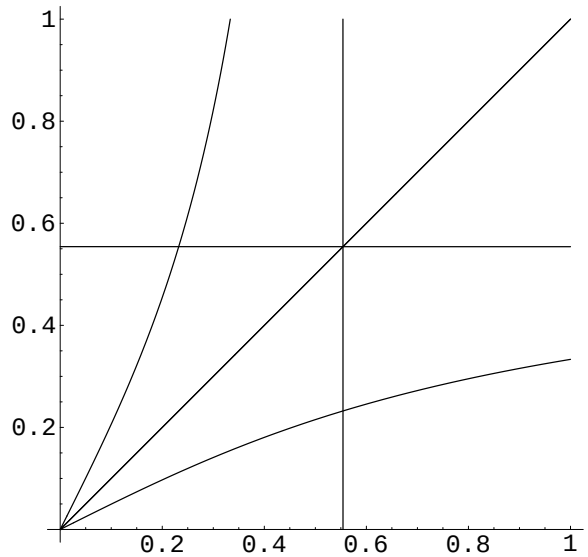
$$(\kappa_1 - 1)\rho_1 = \rho_2 \tag{31}$$

$$(\kappa_2 - 1)\rho_2 = \rho_1. \tag{32}$$

Substituting $\rho_i = \alpha_i$ and $\kappa_i = \frac{3\left(1-\alpha_i^2\right)}{1-3\alpha_i^2}$ from equations (27) and (28) yields the result presented in the lemma. □

A plot of the curves in the (α_1, α_2)-plane is provided by Figure 1. The two straight lines, parallel to the axis, describe the limit values for α_1 and α_2 above which the fourth moments do not exist. According to equation (28) the kurtosis tends to infinity when $\alpha = \sqrt{\frac{1}{3}} \approx 0.57$.

Figure 1: The different regions in the (α_1, α_2)-plane corresponding to the different type of solutions for the pseudo-true value of $A = D_a$. The upper and the lower curve are respectively denoted by C_1 and C_2.



Let us briefly summarize the figure. In the upper right corner, where both fourth moments do not exist, the pseudo-true value of A is given by $D_a = D_\rho^{\frac{1}{2}} = \mathrm{diag}(\sqrt{\alpha_1}, \sqrt{\alpha_2})$. Moving down until α_2 falls below $\alpha_2^* = \sqrt{\frac{1}{3}}$, but remaining above the lower curve, the second kurtosis κ_2 comes into existence. As a result a_2 declines further but remains positive, $a_2 = (\rho_2 - \frac{\rho_1}{\kappa_2 - 1})^{\frac{1}{2}} < \sqrt{\rho_2}$.

The first diagonal element remains $a_1 = \sqrt{\rho_1}$. If we stay above the lower curve and move to the left until the first kurtosis κ_1 also comes into existence, then a_1 too will show additional decline. In that case, $D_a = \text{diag}(a_1, a_2) < \text{diag}(\sqrt{\rho_1}, \sqrt{\rho_2})$. A zero enters the diagonal if we cross either C_1 or C_2 anywhere in the (α_1, α_2)-plane; At any point below C_2 (the lower curve) we have $D_a = \text{diag}(\sqrt{\rho_1}, 0) = \text{diag}(\sqrt{\alpha_1}, 0)$, while anywhere left from C_1 (the upper curve) we obtain $D_a = \text{diag}(0, \sqrt{\rho_2}) = \text{diag}(0, \sqrt{\alpha_2})$. Indeed, note the symmetry with respect to the diagonal $\alpha_1 = \alpha_2$. At the origin all elements are zero, i.e. $D_a = 0$.

By inspecting the (asymptotic) objective function we will see how the pseudo-true value of A attains the different types of solutions sketched above. The following lemma derives an explicit expression for minus the expected objective function[3] $f(A) = -E[S(A)]$ in the bivariate case where we assume an infinite number of observations.

Lemma 2 *Let $A = (a_{ij})$ be a symmetric 2×2 matrix. For any bivariate example characterized by its autocorrelations ρ_i and kurtosis κ_i, the asymptotic objective function $f(A) = -E[S(A)]$ is given by*

$$\begin{aligned} f_{\rho,\kappa}(A) &= 2\left(\rho_1(\kappa_1 - 1)a_{11}^2 + \rho_2(\kappa_2 - 1)a_{22}^2\right) - (\kappa_1 + \kappa_2) \\ &- \left((\kappa_1 - 1)(a_{11}^2 + a_{12}^2)^2 + (\kappa_2 - 1)(a_{12}^2 + a_{22}^2)^2\right) - tr\left(LA^2LA^2\right), \end{aligned}$$

where $L = \mathbf{1}\mathbf{1}' - I_2$. In case of the bivariate ARCH(1) process described by eq. (24-26), for parameter values $\alpha_i < \sqrt{\frac{1}{3}}$, this translates to

$$\begin{aligned} f_\alpha(A) &= 4\left(\frac{\alpha_1 a_{11}^2}{1 - 3\alpha_1^2} + \frac{\alpha_2 a_{22}^2}{1 - 3\alpha_2^2}\right) - 3\left(\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} + \frac{1 - \alpha_2^2}{1 - 3\alpha_2^2}\right) \\ &- 2\left(\frac{(a_{11}^2 + a_{12}^2)^2}{1 - 3\alpha_1^2} + \frac{(a_{12}^2 + a_{22}^2)^2}{1 - 3\alpha_2^2}\right) - tr\left(LA^2LA^2\right) \end{aligned}$$

Proof. The asymptotic objective function f is defined by

$$f(A) = -E\left[\text{tr}(E_t^2)\right] = -E\left[\text{tr}(Q_t - AQ_{t-1}A)^2\right], \tag{33}$$

where $Q_t = y_t y_t' - I_m$. Let us first focus on $\text{tr}\, E_t^2$. We will take expectations and add the minus later. Developing the matrix product yields

$$\begin{aligned} \text{tr}(E_t^2) &= \text{tr}(Q_t^2) - 2\,\text{tr}\left(Q_t A Q_{t-1} A\right) + \text{tr}\left(Q_{t-1} A^2 Q_{t-1} A^2\right) && (34) \\ &= \text{tr}\left(M_1\right) - 2\,\text{tr}\left(M_2\right) + \text{tr}\left(M_3\right). && (35) \end{aligned}$$

Note that the second term contains both Q_t and Q_{t-1}, which are not independent. The model assumptions imply (see equation (10))

$$Q_t = D_\rho\left(Q_{t-1} \odot I_m\right) + \Psi_t, \tag{36}$$

[3]The minus sign implies that we focus on the *maximum* of $f(A)$ in the figures in this section instead of the minimum.

where $E[\Psi_t] = E[Q_{t-1}\Psi_t] = 0$. Substituting this into M_2, and taking expectations, gives

$$E[M_2] = E[D_\rho (Q_{t-1} \odot I_m) A Q_{t-1} A]. \tag{37}$$

At this point we are safe to drop the time subscript, i.e. Q_t and Q_{t-1} both become Q. It will be convenient to decompose Q into

$$Q = Q_1 + Q_2, \tag{38}$$

where $Q_1 = \text{diag}(q_{11}, q_{22})$ and where $Q_2 = q_{12}L$, with $L = \mathbf{1}\mathbf{1}' - I_m$. Using this decomposition, the relevant elements of $E[M_2]$ are seen to be

$$\begin{aligned} E[M_2] &= E[D_\rho Q_1 A Q_1 A] + E[D_\rho Q_1 A Q_2 A] \\ &= E\begin{bmatrix} q_{11}(a_{11}^2 q_{11} + a_{12}^2 q_{22})\rho_1 & \\ & q_{22}(a_{12}^2 q_{11} + a_{22}^2 q_{22})\rho_2 \end{bmatrix} + E[q_{12} D_\rho Q_1 A L A]. \end{aligned}$$

The only stochastic variables in the equation are provided by Q. The moments of its elements are given by

$$E[q_{ii}] = E[q_{ij}] = 0 \tag{39}$$

$$E[q_{ii}q_{jj}] = E[q_{ii}q_{ij}] = 0 \tag{40}$$

$$E\left[q_{ii}^2\right] = \kappa_i - 1 \tag{41}$$

$$E\left[q_{ij}^2\right] = 1. \tag{42}$$

For M_2 this implies

$$E[M_2] = \begin{bmatrix} (\kappa_1 - 1)\rho_1 a_{11}^2 & \\ & (\kappa_2 - 1)\rho_2 a_{22}^2 \end{bmatrix}. \tag{43}$$

Let us now center attention on the third term, M_3. Substituting the decomposition for Q, and rearranging terms, we find

$$M_3 = QA^2QA^2 = Q_1A^2Q_1A^2 + Q_1A^2Q_2A^2 + Q_2A^2Q_1A^2 + Q_2A^2Q_2A^2. \tag{44}$$

Here the relevant elements (ignoring off-diagonal elements) are seen to be

$$\begin{aligned} M_3 &= \text{diag}\begin{bmatrix} q_{11}(a_{11}^4 q_{11} + a_{12}^4 q_{11} + 2a_{11}a_{12}^2 a_{22} q_{22} + a_{12}^2 a_{22}^2 q_{22} + a_{11}^2 a_{12}^2 (2q_{11} + q_{22})) \\ q_{22}(a_{22}^4 q_{22} + a_{12}^4 q_{22} + 2a_{11}a_{12}^2 a_{22} q_{11} + a_{11}^2 a_{12}^2 q_{11} + a_{12}^2 a_{22}^2 (q_{11} + 2q_{22})) \end{bmatrix} \\ &+ 2q_{12}Q_1A^2LA^2 + q_{12}^2 LA^2LA^2. \end{aligned}$$

Taking expectations, and simplifying the expression, leaves us with

$$E[M_3] = \begin{bmatrix} (\kappa_1 - 1)(a_{11}^2 + a_{12}^2)^2 & \\ & (\kappa_2 - 1)(a_{12}^2 + a_{22}^2)^2 \end{bmatrix} + LA^2LA^2. \tag{45}$$

Last, we turn to the first term, M_1, which is the most compact of the three

$$M_1 = Q^2 = \begin{bmatrix} q_{11}^2 + q_{12}^2 & q_{11}q_{12} + q_{12}q_{22} \\ q_{11}q_{12} + q_{12}q_{22} & q_{12}^2 + q_{22}^2 \end{bmatrix}. \tag{46}$$

In expectation this reduces to

$$E\left[M_1\right] = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix}. \tag{47}$$

By putting the three terms together, summing their diagonal elements and adding the minus sign, we finally obtain the expression for the asymptotic objective function as presented in the lemma. □

Figures 2 to 4 provide plots of the asymptotic objective function $f(A)$ for different values of (α_1, α_2), having set $a_{12} = 0$. The figures on the right display the contour plots in the (a_1, a_2)-plane while the figures on the left plot the values of $f(A)$ in 3D.

Figure 2: The asymptotic objection function $f(A)$ for $\alpha = (0.10; 0.15)$, with $a_{12} = 0$

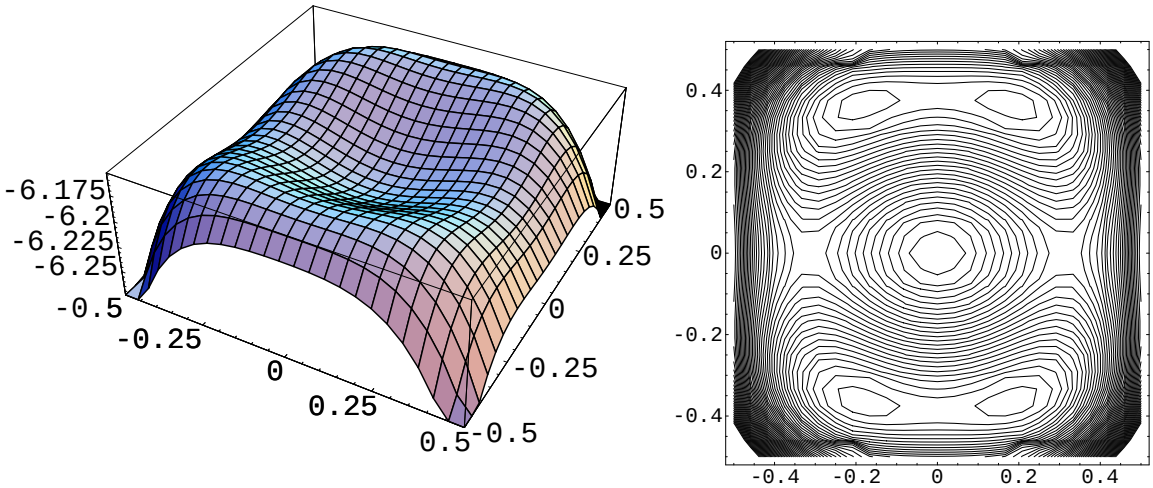


Figure 3: The asymptotic objection function $f(A)$ for $\alpha = (0.10; 0.30)$, with $a_{12} = 0$

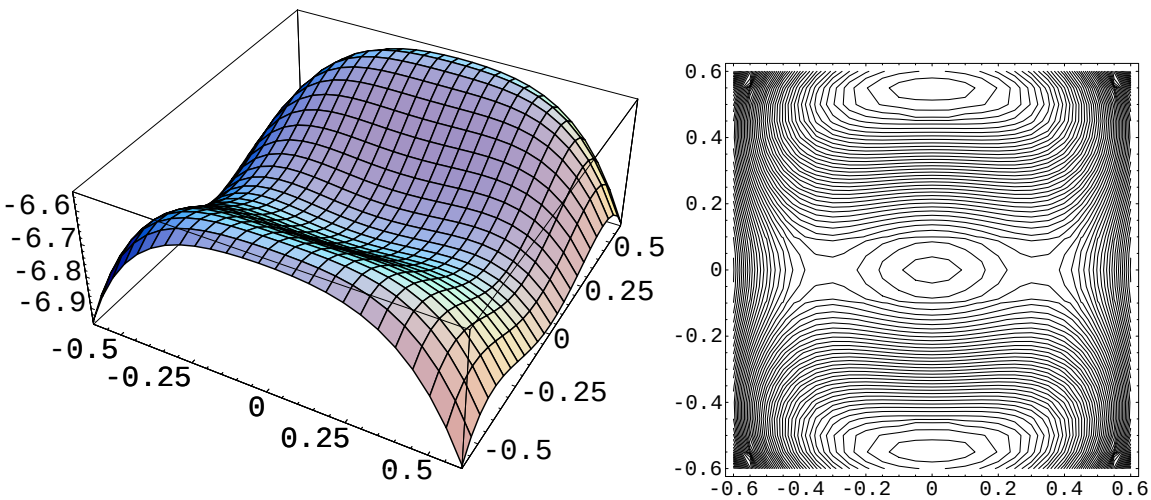
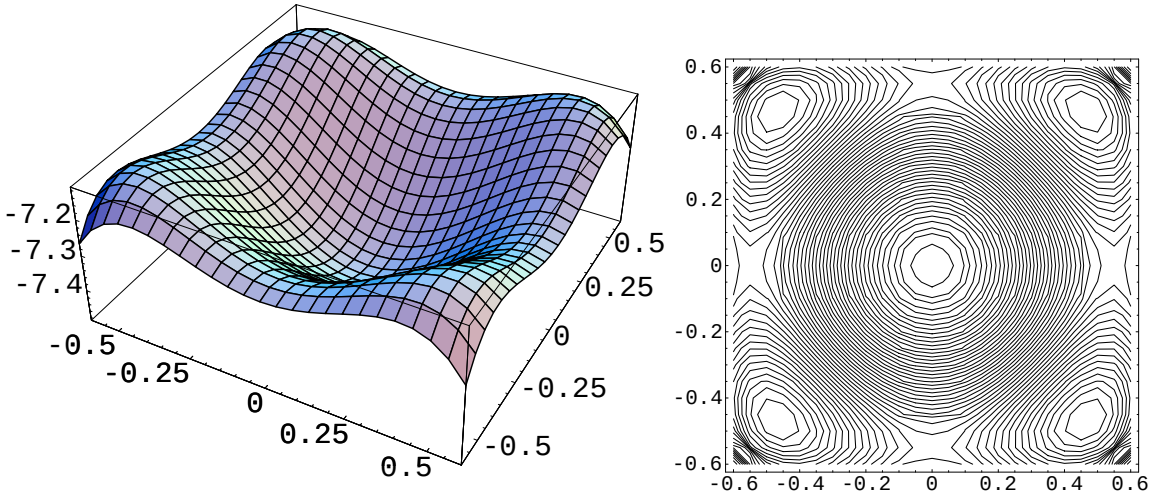


Figure 4: The asymptotic objection function $f(A)$ for $\alpha = (0.30; 0.30)$, with $a_{12} = 0$



In Figure 2, it can be verified, the asymptotic objective function has four local maxima, each of which correspond to the same value of $(a_1, a_2) \approx (0.187, 0.366)$ following from (17). In Figure 3 two maxima remain, with $(a_1, a_2) \approx (0, 0.548)$; note that $(\alpha_1, \alpha_2) = (0.1, 0.3)$ indeed meets the condition of Lemma 1 for a partly positive solution. Finally, Figure 4 corresponds to the special case where the diagonal elements are no longer distinct; It can be verified that $\alpha_1 = \alpha_2 = 0.3$ implies $a_1 = a_2 \approx 0.466$.

5 Some Monte Carlo evidence

To assess the finite-sample behavior of the estimator of U proposed in the previous sections, labelled $\hat{U}_{NLS}$, and its loss of efficiency relative to the ML estimator $\hat{U}_{ML}$, this section considers a number of small-scale Monte Carlo experiments. The section is divided into two subsections. The first subsection explores how performance of the proposed estimator varies with the underlying model parameters. The second evaluates performance for different lengths of the sample size.

To compare the performance of the estimators, we require a single descriptive statistic that measures the accuracy with which we are able to estimate U. Given the true value U_0 and an estimate $\hat{U}$, any matrix norm $\|\hat{U} - U_0\|$ could be used. An advantage of the 2-dimensional case is that the orthogonal matrix can be parameterized by a single parameter, which we will denote φ, which defines the rotation angle[4] with respect to I_2. In all experiments we set $\varphi_0 = \pi/6$ to characterize the orthogonal matrix U_0. The root mean squared error (RMSE) will be our

[4]Note that this parameterization may also be extended to the m-dimensional case. Then $\frac{1}{2}m(m-1)$ rotation angles are required to identify U (see van der Weide, 2002).

choice of summary statistic:

$$RMSE(n) = \left(\frac{1}{N} \sum_{i=1}^{N} (\hat{\varphi}_i - \varphi_0)^2 \right)^{\frac{1}{2}}, \tag{48}$$

where N denotes the number of Monte Carlo replications.

Note that without further constraints different rotation angles may correspond to orthogonal matrices that are in effect identical. To avoid such identification problems, we decide on an explicit structure for U:

$$U(\varphi) = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}, \tag{49}$$

where $\varphi \in [0, \frac{\pi}{2})$. In words, estimates of U will often show variations of (49) where rows are exchanged and/or rows have been multiplied by -1. We deal with that by first imposing (49) before we determine the rotation angle as $\varphi = \arccos(U_{11})$.

One reason we do not center attention on the ARCH parameters in this Monte Carlo exercise is that they are in essence a derivative. The proposed estimator for U requires no model specification for the volatility dynamics. It merely assumes that volatility varies over time. Robustness to any misspecification in the GARCH-type model marks one of the advantages of the proposed estimator. Given an estimate for U, one subsequently transforms the data using $\hat{U}$ and estimates the GARCH parameters in a separate stage.

Finally note that the next subsection will use the same example that was introduced, and examined, in the previous section. Accordingly we benefit from the theoretical results accumulated as we conduct and interpret the simulation experiments.

5.1 For different values of the model parameters

As stated in the outset of this section, we consider the bivariate ARCH(1) example introduced in the previous section (see equations (24)–(26)). From the results presented in the estimation section we know that the effectiveness of the proposed estimator is not independent of the model parameters that describe the persistence in volatility. In this example the persistence in volatility is entirely described by the ARCH parameters α_1 and α_2. We will evaluate performance as we vary α in the range $[0, 1]$. An application of Corollary 1 for this bivariate ARCH(1) example leads us to the following lemma.

Lemma 3 *The non-identification area in the (α_1, α_2)-plane, along which the proposed estimator $\hat{U}$ breaks down asymptotically, is given by*

$$\alpha_1 = \alpha_2. \tag{50}$$

Proof. Applying Corollary 1 to the bivariate case yields

$$(\kappa_1 - 1)\kappa_2\rho_1 = \kappa_1(\kappa_2 - 1)\rho_2. \tag{51}$$

By directly substituting the expressions for the autocorrelations ρ and kurtosis κ given the ARCH(1) specification (see equations (27) and (28)), we obtain

$$3\left(\frac{2\alpha_1}{1-3\alpha_1^2}\right)\left(\frac{1-\alpha_2^2}{1-3\alpha_2^2}\right) = 3\left(\frac{2\alpha_2}{1-3\alpha_2^2}\right)\left(\frac{1-\alpha_1^2}{1-3\alpha_1^2}\right), \tag{52}$$

which is seen to simplify to

$$\alpha_1\left(1-\alpha_2^2\right) = \alpha_2\left(1-\alpha_1^2\right). \tag{53}$$

The latter equality holds if and only if $\alpha_1 = \alpha_2$, which is what is stated in the lemma. □

The non-identification area marks one region in parameter space where the proposed estimator breaks down asymptotically. In this example the problematic region in the (α_1, α_2)-plane is described by a line, the diagonal $\alpha_1 = \alpha_2$. For finite samples however the bounds of the region will be less clear-cut. Even if estimates remain consistent away from the diagonal $\alpha_1 = \alpha_2$, it seems plausible that efficiency of the proposed estimator declines as the model parameters approach the non-identification area.

In addition to consistency problems due to lack of identification, it is of interest to investigate the effect of infinite kurtosis ($\alpha_i \geq \sqrt{\frac{1}{3}}$) or even infinite variance ($\alpha_i = 1$) on the behavior of the estimator. Furthermore, we can investigate whether the behavior of the estimator is substantially different between the cases where both a_i's are positive, and the case where one of them is zero (i.e., whether it matters if we are inside our outside the parameter region bounded by the curves C_1 and C_2 in Figure 1).

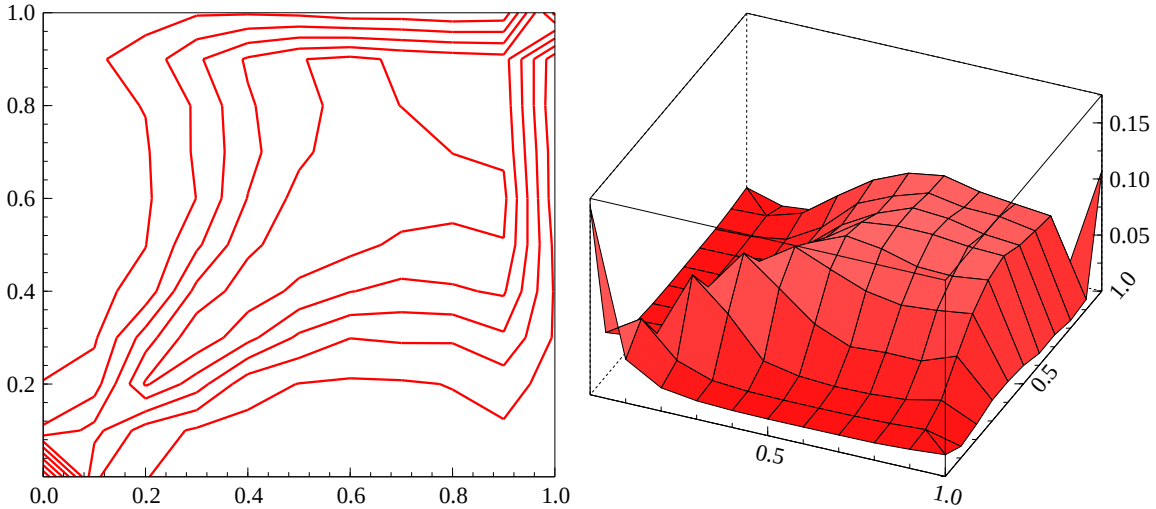
In this Monte Carlo experiment all parameters other than $\alpha = (\alpha_1, \alpha_2)$ will be kept constant. Their values are listed in Table 1.

parameter	value
φ_0	$\frac{\pi}{6}$
sample size n	3000
replications N	2500

Table 1: Parameters kept fixed during Monte Carlo experiment

Figure 5 summarizes our first results, a contour and surface plot of the RMSE for $\alpha \in [0,1] \times [0,1]$. We observe the following properties. First, the RMSE is high on the diagonal, i.e., in the non-identification area characterized by Lemma 3. Second, the RMSE is high when both processes have an infinite kurtosis. Both properties are as expected from Corollary 1. However, we also note that the estimator performs quite well if only one of the processes has infinite kurtosis (for example, when $\alpha_1 = 0.1$ and $\alpha_2 = 0.8$). Although we do not have a formal explanation of this, it is clear that the estimator benefits from the fact that the two independent processes have a very different autocorrelation structure in the squares in this case; the existence of moments is then relatively unimportant.

Figure 5: Values of the RMSE for $\alpha \in [0,1] \times [0,1]$



Perhaps even more striking is the fact that the RMSE is very low if one of the processes has $\alpha_i = 1$ (i.e., an integrated ARCH process, with infinite variance), as long as the other process has $\alpha_j < 1$ (whether the other process has finite higher moments appears to be irrelevant). Again, we see that the different autocorrelations in squares between the processes is very helpful for $\hat{U}_{NLS}$.

5.2 For different values of the sample size

This subsection investigates how efficiency of the proposed estimator increases with the sample size. The results will be placed against the background of the ML estimator, which allows us to judge the loss of efficiency. Figure 6 displays the results, for $n \in \{800, 1600, 3200, 6400\}$, and for two bivariate systems where the independent factors y_{it} follow a GARCH(1,1) process, with parameter values inspired by the empirical results in the next section. In the left panel of Figure 6, the GARCH parameters are set equal to

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 0.09 \\ 0.90 \end{pmatrix}, \qquad \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0.04 \\ 0.95 \end{pmatrix}, \tag{54}$$

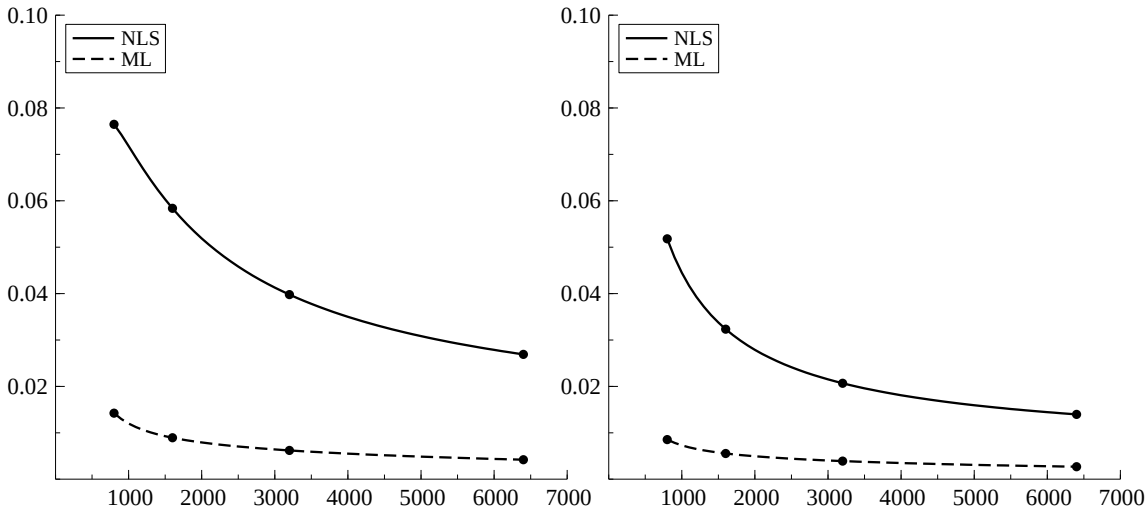
whereas the right panel, they are

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 0.16 \\ 0.83 \end{pmatrix}, \qquad \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0.03 \\ 0.96 \end{pmatrix}. \tag{55}$$

Note that in all cases, $\alpha_i + \beta_i = 0.99$, which corresponds to the values typically found in practice. In the second set of parameter values, the difference in α_1 and α_2 (and hence in the first-order autocorrelations ρ_1 and ρ_2) is larger, such that we expect our estimator to perform better. Note that in all cases, the kurtosis of y_{it} is finite. Recall that for a mean-zero Gaussian GARCH(1,1) process y_t, we have (see Bollerslev, 1986, 1988):

$$\rho = \mathrm{Corr}(y_t^2, y_{t-1}^2) = \frac{\alpha(1-\alpha\beta-\beta^2)}{1-2\alpha\beta-\beta^2}, \qquad \kappa = \frac{E(y_t^4)}{E(y_t^2)^2} = 3\frac{1-\alpha^2-\beta^2-2\alpha\beta}{1-3\alpha^2-2\alpha\beta-\beta^2}.$$

Figure 6: Values of the RMSE for different lengths of the sample size, for parameter values given in (54) (left panel) and (55) (right panel)



From Figure 6, we observe that the RMSE of both estimators decreases as the sample size increases, as expected. We also see that our estimator has a higher RMSE than the MLE in both cases, and that both estimators perform better in the second case, where the difference between the two independent GARCH processes is larger. Although the absolute difference in RMSE between the two estimators decreases as $n \to \infty$ and as the difference between the two α's gets larger, the relative efficiency does not change much; in all cases the RMSE of $\hat{U}_{NLS}$ is about five times larger than that of $\hat{U}_{ML}$. Clearly this relative inefficiency is the price we have to pay for the relative simplicity of our proposed method, but also for its robustness: the maximum likelihood estimator can be expected to be much more sensitive to misspecification in the GARCH processes for the independent components.

6 Empirical application: Dow Jones and Nasdaq in the 1990s

This section considers an empirical application of the method proposed in the previous sections. Following Engle (2002), we consider 10 years of daily returns (March 26, 1990 through March 23, 2000) on the Dow Jones Industrial index and the Nasdaq Composite index ($n = 2609$).

The data are displayed in Figure 7. Both series clearly display volatility clustering. We observe that the volatility of Nasdaq is higher, on average, than the Dow Jones volatility. Furthermore, both volatilities clearly decrease in the middle of the sample, and increase substantially towards the end of the 1990s.

We compare estimated O-GARCH, GO-GARCH and DCC models for the time-varying correlation between the Dow Jones and Nasdaq, where we apply our proposed NLS-based three-step method to the GO-GARCH model. Later, we will compare the NLS model parameter estimates with those obtained from the two-step ML method.

Figure 7: Daily Dow Jones and Nasdaq returns, March 1990 – March 2000

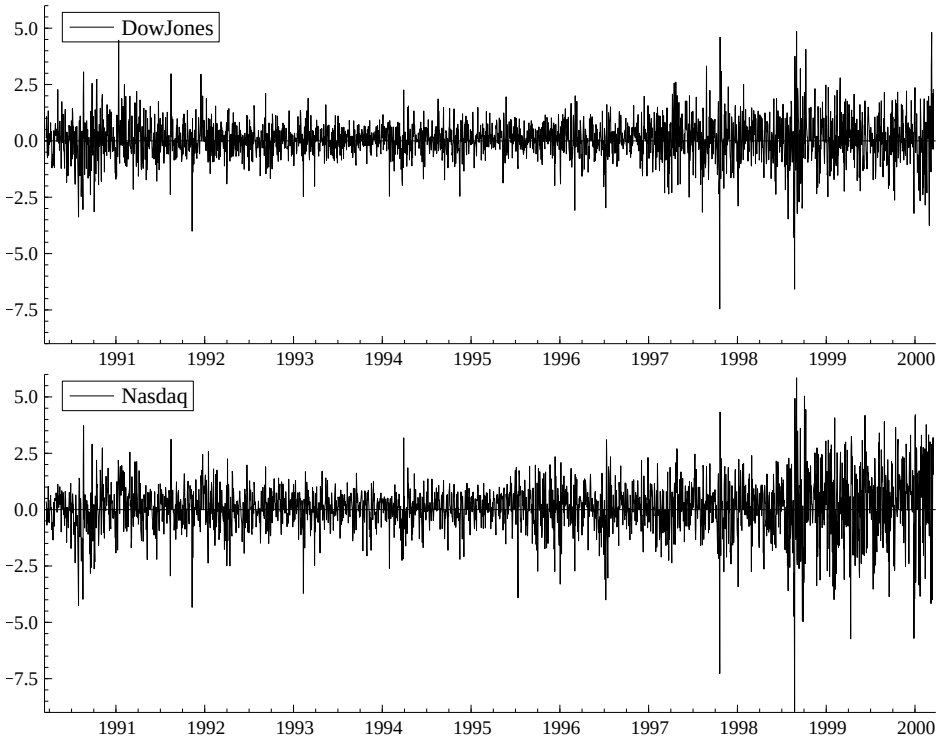
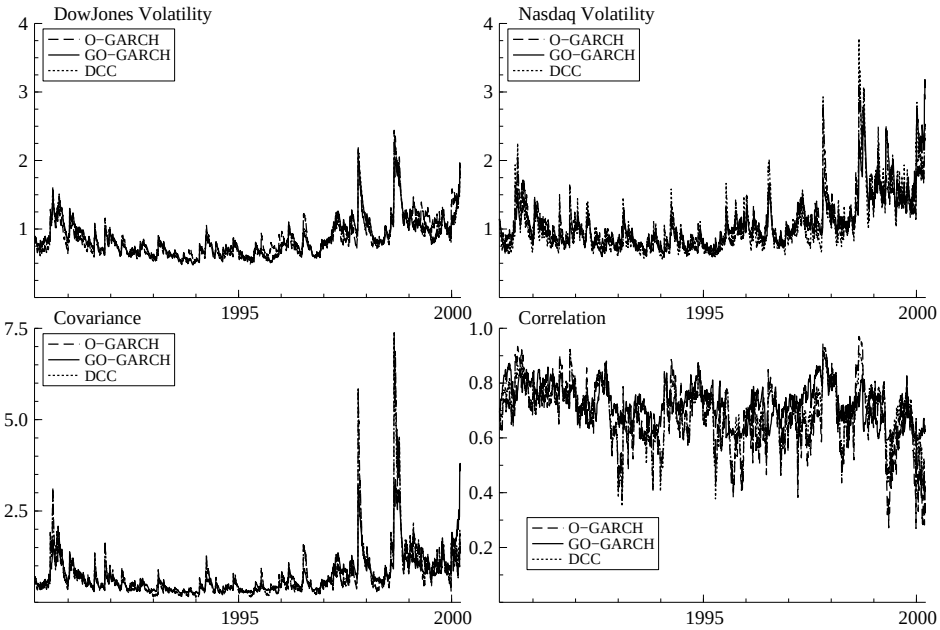


Figure 8 displays the estimated volatilities, covariances and correlations based on the three models. At first sight, all three methods seem to imply very similar volatilities, in particular for the Dow Jones.

Figure 8: Estimated volatilities, covariances and correlations

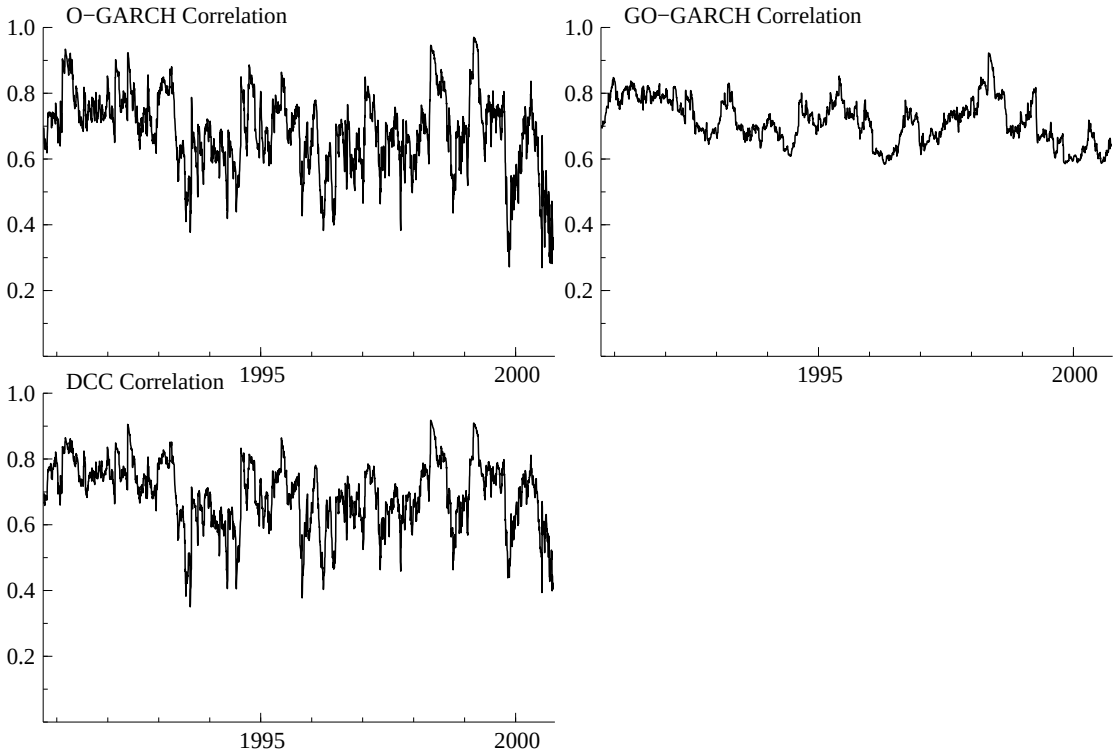


The larger differences occur in the volatility peaks, which is not surprising, since the main

difference between empirical GARCH specifications for the same data-set usually concerns the height of the volatility during the peaks (as well as the subsequent speed of mean reversion). A similar picture emerges from the covariances: the methods agree about the general pattern of covariances over time, but differ in the height of the covariances during the peak periods.

For the time-varying correlations, the differences between the three methods are more pronounced. This is more obvious in Figure 9, where we have depicted the three correlation series in separate graphs.

Figure 9: Estimated correlations



The most obvious difference between the GO-GARCH correlations and the other two specifications is the range in which they vary: the GO-GARCH correlations always lie between 0.6 and 0.9, whereas the other two can become as low as 0.3–0.4. We also see that the O-GARCH and DCC correlation patterns are very similar, and that the GO-GARCH correlation series behaves like a smoothed version of the other two.

The fact that GO-GARCH correlation series sometimes display upper and lower bounds was also noted by van der Weide (2002). Whether this property should be seen as beneficial or not is up to the user. It is quite debatable whether the short periods of very low correlation implied by the O-GARCH and DCC models are genuine; they may be fully driven by the volatility patterns in those periods, and in that case the less volatile behavior of the GO-GARCH correlations may provide a better indication of the actual correlation between these two series[5].

[5]Recall that O-GARCH is nested as a special case in the GO-GARCH model. The data thus has the option to pick the former. Yet the empirical observations decide otherwise as indicated by a likelihood-ratio test, which

The estimates of the matrix U from the NLS-based and ML methods are given by

$$\hat{U}_{NLS} = \begin{bmatrix} 0.523 & -0.852 \\ 0.852 & 0.523 \end{bmatrix}, \qquad \hat{U}_{ML} = \begin{bmatrix} 0.856 & 0.517 \\ -0.517 & 0.856 \end{bmatrix}.$$

These two estimates are very similar, once we take into account that the rows of U may be interchanged (corresponding to the components of y_t being interchanged) or multiplied by -1 (corresponding to replacing y_{it} by $-y_{it}$). Parametrizing the orthogonal matrix as in (49), with $\varphi \in [0, \frac{1}{2}\pi)$, we find that $\hat{\varphi}_{NLS} = 0.175\pi$ and $\hat{\varphi}_{ML} = 0.173\pi$.

The two corresponding estimates of the matrix Z are given by

$$\hat{Z}_{NLS} = \begin{bmatrix} 0.149 & 0.989 \\ 0.814 & 0.581 \end{bmatrix}, \qquad \hat{Z}_{ML} = \begin{bmatrix} 0.990 & -0.142 \\ 0.587 & -0.810 \end{bmatrix}.$$

The cosine of the angle between the two columns of Z is estimated as 0.654 (NLS) or -0.651 (ML). Note that in the O-GARCH model, the columns of Z are orthogonal. Therefore, this cosine, or the estimates of φ given above, may be interpreted as the deviation of the estimated GO-GARCH model from an O-GARCH specification. We have not tried to construct standard errors for these estimates, such that their significance cannot be established. However, a likelihood ratio test statistic for the hypothesis $U = I_2$ is equal to 166, suggesting that this restriction may be safely rejected. (Unreported Monte Carlo simulations indicate that the asymptotic $\chi^2(1)$ null distribution of the likelihood ratio statistic provides a good approximation to the actual null distribution.)

The estimated GARCH(1,1) parameters for the two independent components (based on $\hat{Z}_{NLS}$) are given by

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 0.088 \\ 0.905 \end{pmatrix}, \qquad \begin{pmatrix} \hat{\alpha}_2 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0.044 \\ 0.952 \end{pmatrix}. \tag{56}$$

Note that the volatility persistence $\alpha + \beta$ is very similar for both factors, but the estimates of α are different, indicating that the first factor will display more short-run variation in the volatility. The estimates based on $\hat{Z}_{ML}$ instead of $\hat{Z}_{NLS}$ are very similar.

We conclude this section with some diagnostics for the GO-GARCH model. To this end, we construct standardized residuals $\hat{z}_t = \hat{V}_t^{-1/2} x_t$, where $\hat{V}_t^{1/2}$ denotes the symmetric square root of $\hat{V}_t = \hat{Z}\hat{H}_t\hat{Z}'$, based on either the ML or the NLS estimator. Next, we estimate a first-order vector autoregression for the vector $r_t = (\hat{z}_{1t}^2, \hat{z}_{2t}^2, \hat{z}_{1t}\hat{z}_{2t})'$, and test whether all nine coefficients in this VAR model are equal to zero. When using the NLS estimators the resulting Wald statistic is given by 14.12, with a p-value of 0.12; for the ML estimator the Wald statistic is 9.66, with a p-value of 0.38. Hence we see that the GO-GARCH model, estimated using either one of the two available methods, captures the volatility and correlation dynamics and volatility spill-overs quite well.

rejects O-GARCH in favour of GO-GARCH.

7 Higher-variate examples: US versus European stock indices

The empirical application in this section involves two higher-variate examples: Five widely traded stock indices in the US versus the national indices of the five largest markets in Europe. For the US, we have the Dow Jones Composite (DJ), the Nasdaq 100 (NAS), the New York Stock Exchange Composite (NYA), the Standard and Poor 500 index (SP), and the Major Market Index (XMI). The European markets we consider for our example are located in Amsterdam (AEX), Paris (CAC), Frankfurt (DAX), London (FTSE), and in Zurich (SMI). The samples are of a different size, for we have about 19 years for the US (December 1, 1987 through August 1, 2006), compared to roughly 14 years for Europe (November 11, 1992 through August 1, 2006). This amounts to respectively 4708 and 3502 daily returns[6]. Note that we started our sample of US data well after black Monday, the major stock market crash of 1987, to prevent this outlier from possibly influencing the results[7].

Where the bivariate example from the first empirical application provided a suitable point of departure to study the estimator in more detail, the two five-variate examples in this section will be more of a test-case. Estimation of the multi-variate GARCH model is particularly challenging for larger dimensional systems, and as such the higher-variate examples make an interesting comparison between the different estimators. Naturally, the higher-variate examples also provide a useful test for the model's ability to fit the data. For comparison we included the more parsimonious O-GARCH model.

Descriptive statistics of the data are displayed in Tables 2 and 3. We observe that the standard deviation of the daily returns shows little variation across the indices of both the US and Europe. The Nasdaq 100 appears an exception, as an index for the more volatile high-tech stocks; its standard deviation is almost double the size of the other indices. If we exclude the Nasdaq, we find that the European indices are perhaps slightly more volatile than the US indices. The London FTSE index is least volatile, at a level comparable to the US. More apparent differences between the US and Europe concern the skewness and kurtosis. The empirical densities for the US are noticeably more skewed, where the European appear almost symmetric, and exhibit heavier tails.

[6]Note that there will be days where not all stock exchanges are open for trade. We observed a number of such cases for the different European markets, while it appeared not much of an issue for the US indices from our example. Overall, we find that the London exchange has a slightly larger number of trading days compared to the other exchanges. Some apparent dates include the 26th of December (whenever it is not part of the weekend), and September 11, 2001 (the terorist attack in New York), where the Paris exchange appeared the only market in Europe open for trade. For our European sample there were 20 days where we had missing price data for four out of the five indices. These observations were dropped from our sample. There are also days where only one or two indices report missing values, for only their markets would be closed for trade. In these cases, we decided to impute a zero return. After data-cleaning, our sample of European indices consists out of 3502 daily returns. No observations needed to be dropped from the US sample.

[7]Results not reported here show that including the stock market crash in our sample noticeably affects the estimates from the two-step ML method. The three-step NLS estimator appears largely robust to this outlier.

	DJ	NAS	NYA	SP	XMI	AEX	CAC	DAX	FTSE	SMI
SD	0.009	0.018	0.009	0.010	0.010	0.013	0.013	0.014	0.010	0.011
SK	−0.35	−0.10	−0.34	−0.22	−0.32	−0.13	−0.07	−0.17	−0.16	−0.14
K	7.9	8.2	7.7	7.3	7.8	7.8	5.5	6.0	6.0	7.2

Table 2: Standard deviation (SD), skewness (SK) and kurtosis (K) for raw returns of both US and European indices

These are interesting differences, for the GARCH models used here neither accommodate skewness, nor allow for particularly heavy tails. As such, these differences will plausibly challenge the robustness of the estimators with respect to such misspecifications in the model. Finally, Table 3 shows some differences in the (unconditional) correlations: levels ranging from 0.59 to 0.98 for US indices compared to 0.71 to 0.83 for the European indices. Overall, it appears the US indices are slightly more correlated.

DJ	NAS	NYA	SP	XMI	AEX	CAC	DAX	FTSE	SMI
1.00	0.65	0.94	0.92	0.91	1.00	0.83	0.79	0.78	0.78
	1.00	0.72	0.81	0.59		1.00	0.78	0.78	0.73
		1.00	0.98	0.92			1.00	0.70	0.71
			1.00	0.90				1.00	0.72

Table 3: Unconditional correlations between raw returns, for both US and European indices

As stated, the key objective will be to compare our three-step NLS estimator with the originally proposed two-step ML estimator. In terms of convergence of the numerical optimization procedure, we experienced no difficulties with the NLS estimator; with a first pick of initial values we obtained convergence for both data sets. We also obtained convergence for the ML estimator, although we had to play a little with the initial values for the US data.

To assess how well each estimates the GO-GARCH model, i.e. fit the model to the data, we will appeal to a simple test for misspecification. Let $\hat{z}_t = \hat{V}_t^{-\frac{1}{2}} x_t$ be the standardized residuals, where the matrix $\hat{V}_t$ denotes the estimated (conditional) covariance of x_t, based on either the NLS or the ML estimator. The matrix $\hat{V}_t^{-\frac{1}{2}}$ is obtained as the inverted symmetric square root of $\hat{V}_t$. Let $r_t = (\hat{z}_{1t}^2, \ldots, \hat{z}_{5t}^2, \hat{z}_{1t}\hat{z}_{2t}, \ldots, \hat{z}_{4t}\hat{z}_{5t})'$, where r_t contains all multiples $\hat{z}_{it}\hat{z}_{jt}$ for $j > i$ and $i, j = 1, \ldots, 5$. Then we estimate a first-order vector autoregression for the vector r_t, and use the adjusted-R^2's as our diagnostics. If the model is correctly specified, and the estimators provide accurate estimates, then the VAR model should not pick up much structure in the standardized residuals. The results are presented in Tables 4 and 5. To save space we only report on the adjusted-R^2's for $\hat{z}_{1t}^2$ to $\hat{z}_{5t}^2$ as our dependent variables.

variable	O	GO-ML	GO-NLS	RAW
$\hat{z}_{1t}^2$	0.20	0.11	0.07	0.04
$\hat{z}_{2t}^2$	0.02	0.01	0.01	0.09
$\hat{z}_{3t}^2$	0.02	0.05	0.03	0.05
$\hat{z}_{4t}^2$	0.04	0.04	0.04	0.06
$\hat{z}_{5t}^2$	0.06	0.03	0.01	0.05
likelihood	$-30,851$	$-29,769$	$-30,470$	

Table 4: US example: adjusted-R^2's for regressions against squared residuals, as well as the value of the likelihood function for each model

variable	O	GO-ML	GO-NLS	RAW
$\hat{z}_{1t}^2$	0.01	0.02	0.02	0.13
$\hat{z}_{2t}^2$	0.01	0.00	0.02	0.10
$\hat{z}_{3t}^2$	0.06	0.01	0.01	0.07
$\hat{z}_{4t}^2$	0.00	0.00	0.00	0.11
$\hat{z}_{5t}^2$	0.02	0.01	0.02	0.13
likelihood	$-22,726$	$-22,394$	$-22,916$	

Table 5: EU example: adjusted-R^2's for regressions against squared residuals, as well as the value of the likelihood function for each model

We observe a number of apparent results. Firstly, our NLS estimator performs either as good as the ML estimator (when both perform well), or it seems to perform better. Secondly, the results for the US and Europe are notably different; judging by our diagnostics, the GO-GARCH model fits the European data better than the data from the US. Note that for the European indices, O-GARCH too provides decent results.

It is an interesting detail that GO-GARCH fits the European data better than the US data. It is even more interesting that, for the latter data, the NLS estimator outperforms the ML estimator. Where the ML estimator performs poorly, yet better than the O-GARCH model, the NLS estimator provides a decent fit. Let us put forward a possible explanation for why the NLS estimator, which should be less efficient when the model is correctly specified, is able to do better than the ML estimator[8]. It is anticipated that our three-step NLS estimator is more robust against misspecification. From the descriptive statistics, we read that the empirical densities

[8]Note that the value of the likelihood function is highest for the ML estimate. In addition to the value of the link matrix, however, the likelihood function also depends on the exact model specification for the independent components, in this case a GARCH(1,1) model with normal disturbances. The three-step NLS estimator does not require this model to be specified; its estimate of the link matrix is considered optimal regardless of which model is assumed for the components. Therefore, if our model is misspecified, the value of the likelihood function

associated with the US indices exhibit substantial skewness as well as particularly heavy tails. The model is not tailored to account for these stylized facts, and as such is misspecified, which will affect the ML estimates. Consistent with this reasoning is that the ML estimator performs well for the European data, judging by our diagnostics slightly better than the NLS estimator. Indeed, misspecification is less of an issue here, for the data do not share the same stylized facts that would particularly hamper the ML estimates; the skewness is fairly low, while the kurtosis consistent with normal disturbances that feature GARCH structure.

Something worth trying is to replace the normal distribution for the uncorrelated components with a multivariate skewed t-distribution, see for example Bauwens and Laurent (2005). As such the model is better equipped to handle the stylized facts observed for the US data. If it is indeed the robustness to misspecification of the NLS estimator that sets it apart from the ML estimator, this exercise is anticipated to reduce this discrepancy.

8 Concluding remarks

This paper puts forward an alternative method of estimation for the GO-GARCH model, by appealing to a NLS-based three-step estimator, which is meant to make the model more attractive purely from a practical point of view. The method is better suited for higher-dimensional systems, where the two-step ML procedure originally proposed by van der Weide (2002) is more likely to experience convergence difficulties. The preliminary empirical and Monte Carlo results indicate that the new estimation procedure is promising, although it is noticeably less efficient than the ML procedure. This loss of efficiency however, is accompanied by an increase in robustness: possible misspecification of the GARCH models of the independent components will have no effect on the estimation of the link matrix. Our empirical examples seem to suggest that our NLS estimator indeed benefits from this robustness.

References

Alexander, C. (2001), "Orthogonal GARCH", Chapter 2 (pp. 21–28) in C. Alexander (Ed.), *Mastering Risk*. London: Financial Times-Prentice Hall.

Bauwens, L. and S. Laurent (2005), "A New Class of Multivariate Skew Densities, with Application to GARCH Models", *Journal of Business and Economic Statistics*, 23, 346–354.

Bauwens, L., Laurent, S., and J. Rombouts (2006), "Multivariate GARCH models: a survey", *Journal of Applied Econometrics*, 21, 79–109.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307–327.

need not be entirely informative when comparing the different models or the outcomes of the different estimation methods. It does confirm that our implementation of the two-step ML method did manage to find the maximum of the likelihood function.

Bollerslev, T. (1988), "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process", *Journal of Time Series Analysis*, 9, 121–131.

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

Ding, Z. (1994), *Time Series Analysis of Speculative Returns*. PhD Dissertation, University of California at San Diego.

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

Lanne, M. and P. Saikkonen (2005), "A Multivariate Generalized Orthogonal Factor GARCH Model", *HECER Discussion Paper*, No. 63.

Vrontos, I. D., P. Dellaportas and D. N. Politis (2003), "A Full-Factor Multivariate GARCH Model", *The Econometrics Journal*, 6, 312–334.

Weide, R. van der (2002), "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model", *Journal of Applied Econometrics*, 17, 549–564.