UvA ECONOMETRICS

Discussion Paper: 2009/05

Method of moments estimation of GO-GARCH models

H. Peter Boswijk and Roy van der Weide

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

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

UvA UNIVERSITEIT VAN AMSTERDAM

Method of Moments Estimation of GO-GARCH Models

H. Peter Boswijk*
Tinbergen Institute, Netspar &
Amsterdam School of Economics
University of Amsterdam

Roy van der Weide†
Development Research Group (DECRG)
World Bank &
CeNDEF, University of Amsterdam

January 2009

Abstract

We propose a new estimation method for the factor loading matrix in generalized orthogonal GARCH (GO-GARCH) models. The method is based on the eigenvectors of a suitably defined sample autocorrelation matrix of squares and cross-products of the process. The method can therefore be easily applied to high-dimensional systems, where likelihood-based estimation will run into computational problems. We provide conditions for consistency of the estimator, and study its efficiency relative to maximum likelihood estimation using Monte Carlo simulations. The method is applied to European sector returns, and to the correlation between oil and kerosene returns and airline stock returns.

1 Introduction

The GO-GARCH model was proposed by van der Weide (2002), as a generalization of the orthogonal GARCH model of Ding (1994) and Alexander (2001). The starting point of the model is that an observed vector of returns can be expressed as a non-singular linear transformation of latent factors (either independent or conditionally uncorrelated) that have a GARCH-type conditional variance specification.

A restricted version of the model where only a subset of the latent factors has a time-varying conditional variance has been analyzed recently by Lanne and Saikkonen (2007). This shows that a parsimonious version of the factor GARCH model by, e.g., Diebold and Nerlove (1989) and Engle *et al.* (1990) is nested as a special case (where the variance matrix of the idiosyncratic error term will not be of full rank). The closely related model proposed by Vrontos *et al.* (2003) is also nested as a special case by imposing structure on the linear transformation. Recently, Fan *et al.* (2008) studied a general version of the model by relaxing the assumption of independent factors to conditionally uncorrelated factors.

Note that one has considerable flexibility in specifying models for the factors. One could in principle also consider stochastic volatility as an alternative to the GARCH-type models (see e.g. Doz and Renault

*Corresponding author. Address: Department of Quantitative Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. E-mail: H.P.Boswijk@uva.nl.

†E-mail: rvanderweide@worldbank.org.

(2006)). For surveys on multivariate volatility models we refer to e.g. Bauwens *et al.* (2006), and Silvennoinen and Teräsvirta (2008); for an overview on common features, see Urga (2007), and for a glossary to volatility models, see Bollerslev (2008).

The model is designed to balance generality against ease of estimation. Time-varying variances, time-varying correlations and (asymmetric) volatility spillovers are accommodated, which denote the key stylized facts of multivariate financial data. Moreover, the model is closed under linear transformations (i.e. one does not leave the model when rebalancing ones portfolios) and is closed under temporal aggregation, which makes the model analytically convenient. (For temporal aggregation results for multivariate GARCH models, see Hafner (2008).)

van der Weide (2002) proposed a two-step estimation method that requires joint maximum-likelihood (ML) estimation of parameters that feature both in the linear transformation (between factors and observed data) and in the univariate GARCH specifications for the individual factors. (Not all parameters of the linear transformation need to be estimated by ML, more than half can be identified from the spectral decomposition of the unconditional variance matrix.) While the method works well, and numerical optimization of the likelihood function often converges without difficulties for dimensions up to fifteen, maximum-likelihood can become problematic when the dimension is particularly large and/or when the model used to specify the likelihood function is considerably misspecified.

This paper puts forward a three-step estimation method that is easy to implement and is numerically attractive. The first two steps define a method-of-moments (MM) estimator for the linear transformation that does not require any optimization of an objective function, so that the method is free of numerical convergence problems regardless of the dimension. We identify the linear transformation by using the fact that the latent factors are heteroskedastic. All that is assumed is that the factors exhibit persistence in variance and have finite fourth moments. (Note that the idea of identification through heteroskedasticity in simultaneous equation models is not new; see e.g. Sentana and Fiorentini (2001) and Rigobon (2003).) The third and final step involves estimation of the univariate GARCH-type models for each of the factors. (Models may differ between factors.)

Another alternative estimator has recently been proposed by Boswijk and van der Weide (2006). While it also involves a three-step approach where the linear transformation is estimated independently from the univariate GARCH-type models, it is built on a non-linear least-squares (NLS) estimator in the second step and as such requires numerical optimization. Early findings suggest that the NLS estimator falls in between ML and the newly proposed MM estimator, both in terms of ease of estimation and in terms of efficiency (provided that the information from different lags is combined, analgously to the procedure adopted in this paper; this was not yet considered in Boswijk and van der Weide (2006)).

An obvious application of multivariate GARCH models involves forecasting the conditional variance matrix for the purpose of optimal portfolio selection, hedging and risk management, and option pricing. Naturally, the models may also be used to examine patterns in conditional correlations and volatilities over time. Does volatility in one market spill over to other markets? Are correlations increasing or decreasing, i.e., are markets moving closer together over time, are returns on assets becoming more or less sensitive to movements in macro-economic variables such as the interest rate, the exchange rate and the price of oil? Do correlations jump up in periods of extreme volatility (such as in a financial crisis)? In a modest empirical example in this paper we examine the correlation over time between returns on stock prices of airlines and returns on the price of crude oil and kerosene. The correlation may be considered a proxy for an airline's dependence on oil (that varies with the fuel-efficiency of their fleet) and its ability to hedge against movements in the price of oil.

The outline of the paper is as follows. In Section 2, we formulate the GO-GARCH model, and discuss currently available estimation methods. Section 3 introduces our method-of-moments estimator, and discusses how information from autocorrelation matrices at different lags may be efficiently combined. In Section 4 we use Monte Carlo simulations to study the efficiency of our estimator relative to (quasi-) maximum likelihood for different weighting schemes in low-dimensional systems. Section 5 contains two empirical applications to higher-dimensional systems. Section 6 concludes.

2 The GO-GARCH model

2.1 Model and assumptions

Consider an m-vector time series $\{x_t\}_{t\geq 1}$, representing a vector of (daily) returns on m different assets. Letting $\{\mathcal{F}_t\}_{t\geq 0}$ denote the filtration generated by $\{x_t\}_{t\geq 1}$, we assume that any possibly non-zero conditional mean has been subtracted from x_t, so that $E(x_t|\mathcal{F}_{t-1}) = 0$. The GO-GARCH model imposes a structure on the conditional variance matrix $\Sigma_t = \mathrm{var}(x_t|\mathcal{F}_{t-1}) = E(x_t x_t'|\mathcal{F}_{t-1})$, implied by:

Assumption 1 *The process $\{x_t\}_{t\geq 1}$ satisfies the representation*

$$x_t = Zy_t = ZH_t^{1/2}\varepsilon_t, \tag{1}$$

$$H_t = \mathrm{diag}\left(h_{1t}, \ldots, h_{mt}\right), \tag{2}$$

where Z is an $m \times m$ non-singular matrix, where $\{\{h_{it}\}_{t\geq 1}, i = 1, \ldots, m\}$ are positive, $\{\mathcal{F}_{t-1}\}$-adapted processes with $E(h_{it}) = 1$, and where $\{\varepsilon_t\}_{t\geq 1}$ is a vector martingale difference sequence, with $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $\mathrm{var}(\varepsilon_t|\mathcal{F}_{t-1}) = I_m$.

The model implies that the observed vector of returns x_t can be written as a non-singular transformation of a latent vector process y_t (of the same dimension m), the components y_{it} of which satisfy

$$E(y_{it}|\mathcal{F}_{t-1}) = 0, \quad \mathrm{var}(y_{it}|\mathcal{F}_{t-1}) = h_{it}, \quad \mathrm{cov}(y_{it}, y_{jt}|\mathcal{F}_{t-1}) = 0, \qquad i \neq j = 1, \ldots, m,$$

i.e., the components of y_t are conditionally uncorrelated. The original formulation of the GO-GARCH model involved the stronger assumption of independence of the components of y_t, but for the methods presented in the present paper, the conditional uncorrelatedness assumption (proposed by Fan *et al.* (2008)) suffices. The assumptions also imply that y_t is a covariance-stationary process with mean 0 and unconditional variance $E(H_t) = I_m$. This in turn implies that x_t is covariance-stationary with (conditional) mean zero, conditional variance

$$\Sigma_t = \mathrm{var}(x_t|\mathcal{F}_{t-1}) = ZH_tZ', \tag{3}$$

and unconditional variance

$$\Sigma = \mathrm{var}(x_t) = ZZ'. \tag{4}$$

The conditional variances h_{it} are assumed to follow a GARCH-type structure. One possibility, as considered by van der Weide (2002), is to assume separate univariate GARCH(1,1) specifications

$$h_{it} = (1 - \alpha_i - \beta_i) + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1}, \qquad \alpha_i, \beta_i \geq 0, \quad \alpha_i + \beta_i < 1, \tag{5}$$

which, under a suitable starting value assumption on h_{i0}, implies independence of the components y_{it}. Fan *et al.* (2008) propose a more flexible structure, where h_{it} may depend on $y_{j,t-k}, j \neq i, k \geq 1$. A simple extension of (5) is their extended GARCH(1,1) specification:

$$h_{it} = \left(1 - \sum_{j=1}^{m} \alpha_{ij} - \beta_i\right) + \sum_{j=1}^{m} \alpha_{ij} y_{j,t-1}^2 + \beta_i h_{i,t-1}, \qquad \alpha_{ij}, \beta_i \geq 0, \quad \sum_{j=1}^{m} \alpha_{ij} + \beta_i < 1. \tag{6}$$

Intermediate versions, where some of the $\alpha_{ij}, j \neq i$ are restricted to zero, can also be considered. It should be emphasized that Assumption 1, as well as the estimation methods proposed in this paper, also allow for various other specifications of the conditional variance process, including leverage effects and models formulated in terms of log-volatilities. The assumption also allows for stochastic volatility, as long as the projection h_{it} of the latent stochastic volatility process of y_{it} on the observed price history $\mathcal{F}_{t-1}$ is correlated with lagged squares $y_{i,t-k}^2$.

Consider the polar decomposition of Z:

$$Z = SU, \tag{7}$$

where S is a positive definite symmetric matrix, and U is an orthogonal matrix. Using (4), it follows that $\Sigma = S^2$, so that S is the symmetric square root of Σ, i.e., $S = PL^{1/2}P'$, where PLP' is the spectral decomposition of Σ. This implies that part of the matrix Z may be identified from the unconditional information $\Sigma = \mathrm{var}(x_t)$, and the problem of estimating Z may be reduced to the problem of identifying the orthogonal matrix U from the conditional information. In other words, defining the (unconditionally) standardized returns

$$s_t = \Sigma^{-1/2} x_t = S^{-1} x_t,$$

we find that s_t follows a GO-GARCH specification $s_t = U y_t$ with an orthogonal link matrix U.

Note that van der Weide (2002) and Boswijk and van der Weide (2006) consider, instead of (7), the singular value decomposition $Z = PL^{1/2}U^*$, where $U^* = P'U$ is another orthogonal matrix. This leads to analyzing the standardized principal components $s_t^* = L^{-1/2}P'x_t$, satisfying $s_t^* = U^* y_t$. Here we follow Lanne and Saikkonen (2007) in using the polar decomposition, which circumvents identification problems that arise when Σ has eigenvalues with a multiplicity. As an extreme example, if $\Sigma = I_m$, then $S = I_m$ and $L = I_m$, but P may be an arbitrary orthogonal matrix; in such cases the principal components s_t^* would form an arbitrary orthogonal transformation of the observation vector x_t, whereas $s_t = x_t$. Note also that the O-GARCH model of Alexander (2001) assumes that the standardized principal components s_t^* are independent GARCH processes, which corresponds to a special case of our model with $U^* = I_m$ (hence $s_t^* = y_t$), which in the parametrization considered here requires $U = P$.

2.2 Reduced factor models

Lanne and Saikkonen (2007) analyze a special case of the GO-GARCH model with independent components, in which only a subset of the components of y_t have a time-varying conditional variance. The motivation for this is that if the number of assets m is large, then it may be reasonable to expect that the conditional variance matrix Σ_t can be described by a number $r < m$ of heteroskedastic factors. Indeed, the model then reduces to a parsimoniously parametrized version of the factor ARCH model of Engle *et al.* (1990) and Diebold and Nerlove (1989).

The variance matrix H_t can in this case be expressed as

$$H_t = \begin{bmatrix} H_{1t} & 0 \\ 0 & I_{m-r} \end{bmatrix}, \qquad H_{1t} = \mathrm{diag}(h_{1t}, \ldots, h_{rt}).$$

Partitioning $Z = [Z_1 : Z_2]$ and $U = [U_1 : U_2]$ conformably with H_t, the model implies

$$\begin{aligned} \Sigma_t &= Z_1 H_{1t} Z_1' + Z_2 Z_2' = \Sigma + Z_1(H_{1t} - I_r)Z_1', \\ \mathrm{var}(s_t|\mathcal{F}_{t-1}) &= U_1 H_{1t} U_1' + U_2 U_2' = I_m + U_1(H_{1t} - I_r)U_1'. \end{aligned}$$

These representations imply that in the reduced-factor model, the matrix U_2 is only identified by the properties $U_1'U_2 = 0$ and $U_2'U_2 = I_{m-r}$. In other words, U_2 and hence $Z_2 = SU_2$ are only identified up to orthogonal transformations of their columns.

In a companion paper (Boswijk and van der Weide (2008)), we propose a testing procedure for the hypothesis of a reduced-factor model based on the same sample autocorrelation matrix that will be used in the next section to estimate U. Unless indicated otherwise, in the present paper we assume a full-factor GO-GARCH model.

2.3 Currently available estimation methods

In this subsection we briefly review the currently available methods for estimating the model implied by Assumption 1, or specific versions thereof. Although the GO-GARCH model can be considerably more parsimonious than alternative multivariate GARCH models, for larger m it will become harder to maximize its likelihood function over the entire parameter space, which has motivated the development of two-step approximations of maximum likelihood, or alternative methods that are easier to apply in larger dimensions.

Gaussian maximum likelihood estimation of the model with independent GARCH factors was analyzed by van der Weide (2002). He considered the standardized returns s_t as observable time series, which leads to a log-likelihood function of the form

$$\ell(\theta) = -\frac{1}{2} \sum_{t=1}^{n} \left\{ m \log(2\pi) + \log |H_t(\theta)| + s_t' U(\theta_1) H_t(\theta)^{-1} U(\theta_1)' s_t \right\}, \tag{8}$$

where $\theta = (\theta_1', \theta_2')$, with θ_1 a vector of dimension $\frac{1}{2}m(m-1)$ characterizing the $m \times m$ orthogonal matrix $U = U(\theta_1)$, and θ_2 a $2m$-dimensional parameter vector of GARCH parameters. A convenient parametrization of $U(\theta_1)$ as the product of $\frac{1}{2}m(m-1)$ rotation matrices, each characterized by one parameter, is discussed in van der Weide (2002). Note that H_t depends on U via $y_{t-1} = U's_t$, so that $H_t = H_t(\theta)$ is characterized by the full parameter vector θ. By applying the general asymptotic results of Comte and Lieberman (2003) for BEKK models (in which the GO-GARCH model is nested), conditions for consistency and asymptotic normality of the maximum likelihood estimator are obtained.

In practice s_t is not observed, and will have to be estimated by $\hat{s}_t = \hat{\Sigma}^{-1/2}x_t$, with $\hat{\Sigma}$ the sample variance matrix $n^{-1}\sum_{t=1}^{n} x_t x_t'$. (If m is large relative to n, one could also consider shrinkage-type estimators of Σ, see e.g. Ledoit and Wolf (2004)). Therefore, in practice the procedure of van der Weide (2002) is a two-step approximation of maximum likelihood. If we let $\theta_0 = \mathrm{vech}(\Sigma)$, then full maximum likelihood would entail maximizing (8), with s_t replaced by $\Sigma(\theta_0)^{-1/2}x_t$, over $(\theta_0', \theta_1', \theta_2')'$. Lanne and Saikkonen (2007) derive asymptotic properties of such a full maximum likelihood procedure for the reduced-factor model considered in Section 2.2.

Boswijk and van der Weide (2006) proposed a non-linear least-squares estimator of U, based on the autocorrelation properties of the matrix-valued process $S_t = s_t s_t' - I_m$. Let $\hat{B}$ be the minimizer, over all symmetric matrices, of the least-squares criterion

$$Q(B) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{tr}\left(S_t - BS_{t-1}B\right)^2.$$

Using the fact that $S_t = UY_tU'$, with $Y_t = y_t y_t' - I_m$, it follows that

$$Q(B) = n^{-1}\sum_{t=1}^{n} \mathrm{tr}\left(Y_t - AY_{t-1}A\right)^2, \qquad A = U'BU.$$

Boswijk and van der Weide (2006) derive conditions under which the probability limit of $\hat{A} = U'\hat{B}U$ is a diagonal matrix, which in turn implies that the eigenvector matrix $\hat{U}$ of $\hat{B}$ will be a consistent estimator of U. This estimator can be embedded in a three-step procedure: first estimate Σ to construct $\hat{s}_t$, then estimate U based on $\hat{s}_t$, and finally estimate the GARCH parameters based on $\hat{y}_t = \hat{U}'\hat{s}_t$.

An alternative estimator of U was proposed by Fan *et al.* (2008). The starting point of their analysis is that the conditionally uncorrelated restriction $E(y_{it}y_{jt}|\mathcal{F}_{t-1}) = 0$ is equivalent to $E[y_{it}y_{jt}I(B)] = 0$ for all $B \in \mathcal{F}_{t-1}$, where $I(.)$ is the indicator function. Let u_i denote the i-th vector of U, so that $y_{it} = u_i' s_t$, let $\mathcal{B}$ be a collection of subsets of $\mathbb{R}^m$, and let p be an arbitrary integer. Then the columns of U should satisfy the (population) criterion

$$\Psi(U) = \sum_{1\leq i<j\leq m}\sum_{B\in\mathcal{B}}\sum_{k=1}^{p}\left|u_i' E[s_t s_t' I(s_{t-k}\in B)]u_j\right| = 0. \tag{9}$$

Fan *et al.* (2008) propose to estimate U by minimizing a sample analog of $\Psi(U)$, and provide a bootstrap inference procedure for this estimator, and for a test of the conditionally uncorrelatedness hypothesis. Again, this estimator of U should be preceded by the estimation of Σ and s_t, and followed by the estimation of the (extended) GARCH models for $\hat{y}_t = \hat{U}'\hat{s}_t$.

All methods considered in this subsection require numerical maximization of a criterion function over a high-dimensional parameter space. Therefore, as m increases, each of these methods is likely to run into numerical problems, such as failure of a Newton-type optimization procedure to converge, or the possibility of ending up in a local maximum. The estimator proposed in the next section, on the other hand, only requires the calculation of eigenvalues and eigenvectors of a particular sample moment matrix, and therefore can be applied to arbitrary dimensions m.

3 Method of moments estimation

3.1 The estimator

The starting point of our method-of-moments estimator is the same as in Boswijk and van der Weide (2006), i.e., the autocorrelation properties of the (mean-zero) matrix-valued processes $S_t = s_t s_t' - I_m$ and $Y_t = y_t y_t' - I_m$. For the autocorrelation matrices of these processes to be well-defined (and consistently estimated by their sample analogs) and to be able to identify U from these, we make the following assumption.

Assumption 2 *The process $\{y_t\}_{t\geq 1}$ is strictly stationary and ergodic, and has finite fourth moments $\kappa_i = E(y_{it}^4) < \infty, i = 1, \ldots, m$. Furthermore, the autocorrelations $\rho_{ik} = \mathrm{corr}(y_{it}^2, y_{i,t-k}^2)$ and cross-covariances $\tau_{ijk} = \mathrm{cov}(y_{it}^2, y_{i,t-k}y_{j,t-k})$ satisfy, for some integer p,*

$$\min_{1\leq i\leq m} \max_{1\leq k\leq p} |\rho_{ik}| > 0, \qquad \max_{1\leq k\leq p, 1\leq i<j\leq m} |\tau_{ijk}| = 0.$$

The stationarity assumption, as well as the assumptions on the moments, would be implied by independent GARCH processes for y_{it}, under suitable parameter restrictions to guarantee a finite kurtosis, see He and Teräsvirta (1999). Because estimated GARCH parameters in practice do not always satisfy the finite kurtosis restrictions, this assumption is not without loss of generality. In the next section, we investigate the sensitivity of our method to deviations from this assumption through Monte Carlo simulations. The non-zero autocorrelation assumption allows us to identify U from the the first p autocorrelation coefficients of y_{it}^2. It would be hard to think of processes that do display volatility clustering but violate this assumption (i.e., with $\mathrm{corr}(y_{it}^2, y_{i,t-k}^2) = 0$ for all $k = 1, \ldots, p$). Finally, the zero cross-covariances τ_{ijk} exclude dependence in h_{it} on whether $y_{i,t-k}$ and $y_{j,t-k}$ have the same sign. Although this may exclude particular asymmetries in volatility, note that the assumption does allow for the extended GARCH model (6), possibly augmented with $y_{i,t-1}$ and $y_{j,t-1}$ (but not their product) to allow for leverage effects.

Define the autocovariance matrices

$$\Gamma_k(y) = E(Y_t Y_{t-k}), \qquad k = 1, 2, \ldots \tag{10}$$

Note that $\Gamma_k(y)$ does not contain all separate k-th order (cross-) autocovariances of squares and cross-products of y_t (which would require vectorizing Y_t), but is an $m \times m$ matrix with elements

$$\Gamma_k(y)_{ij} = \sum_{\ell=1}^{m} \mathrm{cov}(y_{it}y_{\ell t}, y_{\ell,t-k}y_{j,t-k}).$$

Therefore, Assumptions 1 and 2 imply, using $\mathrm{var}(y_{it}^2) = E(y_{it}^4) - E(y_{it}^2)^2 = \kappa_i - 1$,

$$\Gamma_k(y)_{ij} = \mathrm{cov}(y_{it}^2, y_{i,t-k}y_{j,t-k}) = \begin{cases} (\kappa_i - 1)\rho_{ik}, & j = i, \\ \tau_{ijk} = 0, & j \neq i, \end{cases}$$

or in other words

$$\Gamma_k(y) = \mathrm{diag}((\kappa_1 - 1)\rho_{1k}, \ldots, (\kappa_m - 1)\rho_{mk}).$$

For the corresponding autocorrelation matrix, we thus find

$$\Phi_k(y) = \Gamma_0(y)^{-1/2}\Gamma_k(y)\Gamma_0(y)^{-1/2} = \mathrm{diag}(\rho_{1k}, \ldots, \rho_{mk}).$$

For the process $s_t = Uy_t$, the corresponding autocovariance and autocorrelation matrices satisfy

$$\Gamma_k(s) = E(S_t S_{t-k}) = E(UY_tU'UY_{t-k}S') = U\Gamma_k(y)U',$$

and hence

$$\Phi_k(s) = \Gamma_0(s)^{-1/2}\Gamma_k(s)\Gamma_0(s)^{-1/2} = U\Phi_k(y)U'.$$

Because $\Gamma_k(y)$ and $\Phi_k(y)$ are diagonal matrices and U is an orthogonal matrix, we find that under Assumptions 1 and 2, U may be identified by the eigenvectors of either $\Gamma_k(s)$ or $\Phi_k(s)$.

Consider the sample analogs of $\Gamma_k(s)$ or $\Phi_k(s)$:

$$\hat{\Gamma}_k(s) = \frac{1}{n}\sum_{t=k+1}^{n} S_t S_{t-k} = \frac{1}{n}\sum_{t=k+1}^{n} (s_t s_t' - I_m)(s_{t-k}s_{t-k}' - I_m), \tag{11}$$

$$\hat{\Phi}_k(s) = \hat{\Gamma}_0(s)^{-1/2}\hat{\Gamma}_k(s)\hat{\Gamma}_0(s)^{-1/2}, \tag{12}$$

where $\hat{\Gamma}_0(y)^{-1/2}$ is the symmetric square root of $\hat{\Gamma}_0(y)^{-1}$. We define our estimator $\hat{U}_k$ as the matrix of eigenvectors of the symmetrized version $\frac{1}{2}(\hat{\Phi}_k(s) + \hat{\Phi}_k(s)')$ of $\hat{\Phi}_k(s)$. Although in principle one could also take the eigenvectors of the corresponding symmetric version of $\hat{\Gamma}_k(s)$ as estimator of U, preliminary Monte Carlo experiments have indicated that the standardization used to construct $\hat{\Phi}_k(s)$ leads to a more efficient estimator.

3.2 Combining information from different lags

Although one could in principle use the estimator $\hat{U}_k$ proposed in the previous subsection for one particular choice of the lag length k, we may obtain a more efficient estimator by combining information from different lags. This is relevant in particular for daily financial data, where the autocorrelation function of the squares typically is small but slowly decaying. This implies that the eigenvalues $\{\rho_{ik}\}_{i=1}^m$ of $\Phi_k(s)$ will be close to zero (and and hence close to each other), yielding weakly identified eigenvectors for fixed k. Provided that the autocorrelation functions $\{\rho_{ik}\}_{k=1}^\infty$ are sufficiently different, pooling the information from different $\Phi_k(s)$ matrices will then increase the efficiency of the estimator.

Let p denote the maximal lag length, and let $\{w_k\}_{k=1}^p$ be a set of weights, satisfying $w_k \geq 0$ and $\sum_{i=1}^p w_k = 1$. The information from different lags can be combined in various alternative ways. A simple approach would be to average the moment matrices $\hat{\Gamma}_k(s)$ or $\hat{\Phi}_k(s)$, and calculate the eigenvectors from these averages. Alternatively, one could take the weighted geometric average of $\hat{U}_k$, or the orthogonal part of the polar decomposition of the weighted arithmetic average. In preliminary Monte Carlo experiments, it has appeared that each of these possibilities is dominated by the following simple approach. Consider the inverse Cayley transforms (see Liebeck and Osborne (1991))

$$\hat{S}_k = (I_m - \hat{U}_k)(I_m + \hat{U}_k)^{-1},$$

which results in skew-symmetric matrices $\{\hat{S}_k\}_{k=1}^p$. The average $\hat{S} = \sum_{k=1}^p w_k \hat{S}_k$ of such matrices is also a skew-symmetric matrix. Our pooled estimator is then defined as the Cayley transform of $\hat{S}$, i.e.,

$$\begin{aligned} \hat{U} &= (I_m - \hat{S})(I_m + \hat{S})^{-1} \\ &= \left(I_m - \sum_{k=1}^p w_k (I_m - \hat{U}_k)(I_m + \hat{U}_k)^{-1}\right)\left(I_m + \sum_{k=1}^p w_k (I_m - \hat{U}_k)(I_m + \hat{U}_k)^{-1}\right)^{-1}. \end{aligned} \tag{13}$$

When averaging (Cayley-transformed) orthogonal matrices, care has to be taken of the ordering and sign of the columns of $\hat{U}_k$. Because the model can be expressed as $s_t = U y_t = \sum_{i=1}^m u_i y_{it}$, with u_i the ith column of U, it follows that U is equivalent to a matrix containing $\{u_1, \dots, u_m\}$, in arbitrary order, and possibly multiplied by -1. We use the following procedure to match an orthogonal matrix U as closely as possible to another orthogonal matrix V, with columns $\{v_i\}_{i=1}^m$:

- the first column of the matched matrix $\tilde{U}$ is that column u_j such that $|v_1' u_j| = \max_{1 \le i \le m} |v_1' u_i|$;
- for $\ell = 2, \dots, m$, the ℓth column of $\tilde{U}$ is that column u_j such that $|v_\ell' u_j| = \max_{i \in I_\ell} |v_\ell' u_i|$, where $\{u_i, i \in I_\ell\}$ is the set of remaining columns of U, after the first $\ell - 1$ columns of $\tilde{U}$ have been determined;
- the sign of each of the columns of $\tilde{U}$ is such that the diagonal elements of this matrix are positive;
- if the resulting matrix has determinant -1, then the column of $\tilde{U}$ that has the smallest absolute inner product with the corresponding column of V is multiplied by -1.

Using this procedure, we first match $\hat{U}_1$ to the identity matrix I_m (which means that columns are rearranged such that the largest elements end up along the diagonal), and then we match $\hat{U}_k, k = 1, \dots, p$, to $\hat{U}_1$.

We consider two choices for the weight functions: equal weights $w_k = 1/p$, and weights depending on the eigenvalues $\{\hat{\lambda}_{ik}\}_{i=1}^m$ of $\frac{1}{2}(\hat{\Phi}_k(s) + \hat{\Phi}_k(s)')$:

$$w_k = \frac{\min_{1 \le i < j \le m} (\hat{\lambda}_{ik} - \hat{\lambda}_{jk})^2}{\sum_{k=1}^p \min_{1 \le i < j \le m} (\hat{\lambda}_{ik} - \hat{\lambda}_{jk})^2}. \tag{14}$$

This weight function reflects the fact that the estimator precision of eigenvectors of a matrix estimator depends on the degree to which the eigenvalues are distinct. For example, when the estimator $\hat{\Sigma}$ of a positive definite matrix Σ follows a Wishart distribution, then the asymptotic variance of the eigenvectors (corresponding to non-zero eigenvalues) is proportional to $\sum_{i\neq j}\lambda_i\lambda_j/(\lambda_i-\lambda_j)^2$ (Anderson, 1984, Section 13.5).

The estimator we propose can be summarized as follows.

Summary 1 *Starting from an m-vector of daily returns $\{x_t\}_{t=1}^n$, possibly corrected (by least-squares) for a constant mean and serial correlation, the model is estimated in the following steps:*

1. *estimate the unconditional variance matrix $\hat{\Sigma}=n^{-1}\sum_{t=1}^n x_t x_t'$, its spectral decomposition $\hat{\Sigma}=PLP'$, and hence its symmetric square root $S=PL^{1/2}P'$ and the standardized returns $s_t=S^{-1}x_t=PL^{-1/2}P'x_t$;*

2. *calculate the matrix-valued series $S_t=s_t s_t'-I_m$, its sample autocovariance matrices $\hat{\Gamma}_k(s)=n^{-1}\sum_{t=1}^n S_t S_{t-k}$, $k=0,\ldots,p$, and its sample autocorrelation matrices $\hat{\Phi}_k(s), k=1,\ldots,p$, from (12);*

3. *calculate the spectral decomposition of the symmetrized autocorrelation matrices $\frac{1}{2}(\hat{\Phi}_k(s)+\hat{\Phi}_k(s)')=\hat{U}_k\hat{\Lambda}_k\hat{U}_k'$; from the diagonal elements $\{\hat{\lambda}_{ik}\}_{i=1}^m$ of $\hat{\Lambda}_k$, calculate the weights w_k from (14), match the orthogonal matrices $\{\hat{U}_k\}_{k=1}^p$ to each other using the procedure described in the previous subsection, and use these to calculate the pooled estimate $\hat{U}$ as given in (13);*

4. *estimate the conditionally uncorrelated components y_t by $\hat{y}_t=\hat{U}'s_t$, and estimate separate GARCH-type models for the components of y_{it} by quasi-maximum likelihood.*

3.3 Consistency

In this subsection we prove consistency of the estimator $\hat{U}$ defined in the previous section. We use the square root $d(\cdot,\cdot)$ of a symmetric version of the distance measure $D(\cdot,\cdot)$ for orthogonal matrices introduced by Fan *et al.* (2008):

$$
\begin{aligned}
d(U,\hat{U}) &= \sqrt{\tfrac{1}{2}\left[D(U,\hat{U})+D(\hat{U},U)\right]},\\
D(\hat{U},U) &= 1-\frac{1}{m}\sum_{i=1}^m \max_{1\le j\le m}\left|u_i'\hat{u}_j\right|.
\end{aligned}
$$

The motivation for $D(\cdot,\cdot)$ is the same equivalence of U under permutation and sign change of its columns as discussed above. The modification $d(\cdot,\cdot)$ is a distance function that satisfies the properties of a metric (symmetry, triangle inequality[1]), provided that an orthogonal matrix is identified by its equivalence class.

An identification assumption needed for consistency of $\hat{U}$ is the following:

[1] Although we have not been able to prove the triangle inequality, numerically it appears that $d(\cdot,\cdot)$ satisfies this property; the original distance $D(\cdot,\cdot)$ violates this property.

Assumption 3 *In the model defined by Assumptions 1 and 2,*

$$\max_{1\le k\le p}\min_{1\le i<j\le m}\left|\rho_{ik}-\rho_{jk}\right|>0.$$

The assumption excludes the possibility that two squared components y_{it}^2 and y_{jt}^2 have the same autocorrelation function for $k = 1,\dots,p$. The reason for this assumption is that the autocorrelations are the eigenvalues of the matrix $\Phi_k(s)$, and if this matrix has eigenvalues with a multiplicity, then the corresponding submatrix of eigenvectors is only identified up to orthogonal transformations. Because such transformations will typically destroy the property of the true matrix U, that $U's_t = y_t$ is a vector of conditionally uncorrelated components, this would result in an inconsistent estimator $\hat{U}_k$. For the weighted estimator based on the eigenvalue-based weight function (14), such inconsistent estimates will automatically get a zero weight, so that Assumption 3 is sufficient for consistency. For the equally-weighted estimator, however, the assumption needs to be strengthened to

Assumption 4 *In the model defined by Assumptions 1 and 2,*

$$\min_{1\le k\le p, 1\le i<j\le m}\left|\rho_{ik}-\rho_{jk}\right|>0.$$

Theorem 1 *Consider the estimator (13), with weight function (14). Then, under Assumptions 1–3, and as $n\to\infty$,*

$$d(U,\hat{U})\xrightarrow{P}0.$$

Under Assumptions 1–4, the same result applies to the estimator with weight function $w_k = 1/p$.

Proof. From the law of large numbers for stationary ergodic Markov chains, see Jensen and Rahbek (2007), it follows that under Assumptions 1 and 2, and as $n\to\infty$,

$$\hat{\Gamma}_k(s)\xrightarrow{P}\Gamma_k(s),\qquad \hat{\Phi}_k(s)\xrightarrow{P}\Phi_k(s).$$

If $\Phi_k(s)$ has distinct eigenvalues, this implies $d(U,\hat{U}_k)\xrightarrow{P}0$. By the continuous mapping theorem, for these values of k we have

$$w_k=\frac{\min_{1\le i<j\le m}(\hat{\lambda}_{ik}-\hat{\lambda}_{jk})^2}{\sum_{k=1}^{p}\min_{1\le i<j\le m}(\hat{\lambda}_{ik}-\hat{\lambda}_{jk})^2}\xrightarrow{P}\frac{\min_{1\le i<j\le m}(\rho_{ik}-\rho_{jk})^2}{\sum_{k=1}^{p}\min_{1\le i<j\le m}(\rho_{ik}-\rho_{jk})^2}>0, \tag{15}$$

whereas those $\hat{U}_k$ matrices for which $\Phi_k(s)$ has eigenvalues with a multiplicity will get zero weight. Under Assumption 3, there is at least one k such that (15) holds, which implies $d(U,\hat{U})\xrightarrow{P}0$. For the equally-weighted estimator, any inconsistent $\hat{U}_k$ will get a non-zero weight, so that they need to be consistent for all $k = 1,\dots,p$, which requires Assumption 4. $\square$

A next step could be to derive the asymptotic distribution of the estimator, as well as a rate of convergence. Both would require assuming finite eighth moments of y_{it}, which is likely to be violated in practical applications, and is therefore not consider here. Alternatively, bootstrap procedures could be devised to conduct asymptotic inference on $\hat{U}$. We leave this for future work.

4 Monte Carlo simulations

In this section we study the finite-sample performance of the estimator proposed in this paper for various choices of lag length and weight function, and we compare its performance to the maximum likelihood estimator for various sample sizes. We consider a trivariate system ($m = 3$), with true value of U given by the product of rotation matrices:

$$U = \begin{bmatrix} \cos\phi_1 & -\sin\phi_1 & 0 \\ \sin\phi_1 & \cos\phi_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi_2 & 0 & -\sin\phi_2 \\ 0 & 1 & 0 \\ \sin\phi_2 & 0 & \cos\phi_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_3 & -\sin\phi_3 \\ 0 & \sin\phi_3 & \cos\phi_3 \end{bmatrix},$$

with $\phi_1 = \pi/3$, $\phi_2 = \pi/5$, and $\phi_3 = \pi/7$. The components of y_t are assumed to be independent Gaussian GARCH(1,1) processes, with two alternative parameter combinations:

- DGP A: $(\alpha_1, \beta_1) = (0.03, 0.96)$; $(\alpha_2, \beta_2) = (0.09, 0.90)$; $(\alpha_3, \beta_3) = (0.17, 0.78)$;
- DGP B: $(\alpha_1, \beta_1) = (0.05, 0.94)$; $(\alpha_2, \beta_2) = (0.15, 0.84)$; $(\alpha_3, \beta_3) = (0.25, 0.74)$.

Under DGP A, all components of y_t have finite kurtosis, so that Assumption 2 is satisfied. The first and second component have the same autocorrelation decay rate $\alpha + \beta = 0.99$, but a different initial autocorrelation ρ_1, so that the autocorrelation functions do not cross. The parameters of the third component have been chosen such that the cumulative autocorrelation $\sum_{k=1}^{\infty} \rho_k$ is (almost) the same as for y_{1t}; the autocorrelation decay is faster for the third component ($\alpha + \beta = 0.95$), but ρ_1 is larger, so that the autocorrelation functions of y_{1t}^2 and y_{3t}^2 cross (around $k = 40$). We would expect that in such cases, the eigenvalue-weighted estimator will perform much better than the equally-weighted estimator.

Under DGP B, the second and third component have infinite kurtosis; this DGP is included to investigate how sensitive the estimator is to deviations from the finite-fourth-moment assumption.

Figure 1 depicts the root mean square distance (RMSD), i.e., the square root of the average of $d(U, \hat{U})^2$ over 5000 Monte Carlo replications, for the two DGPs discussed above, with $T \in \{800, 1600, 3200, 6400\}$ and $p \in \{1, 5, 10, 25, 50, 100, 200\}$. The left panels refer to DGP A, the right panels to DGP B.

Figure 1 about here

The top panels of Figure 1 depict the RMSD as a function of the lag length p, for $n = 1600$ and $n = 6400$. We find that the eigenvalue-weighted estimator always outperforms the equally-weighted estimator, although the difference is not larger in DGP A, as we would have expected. In all cases there is a clear gain from combining lags: the minimum achievable average RMSD is about half the distance for $p = 1$. For DGP A, the optimal lag length is about $p = 50$, whereas for DGP A, $p = 100$ gives the best results (for $n = 1600$; for the larger sample size, higher lag orders are preferred). For the eigenvalue-weighted estimator, we find that over-specifying p does not lead to a loss of information; on the other hand, the RMSD of the equally-weighted estimator does increase as p increases beyond the optimal choice.

The bottom panels show the RMSD, both of the ML estimator and the two versions of the MM estimator, as function of the sample size. Here p has been chosen to minimize the distance. We observe that the MM estimators have a much larger RMSD than the ML estimator. We also observe that the RMSD of the MM estimator is approximately proportional to $n^{-1/2}$. (In a bivariate model where U is characterized by a single angle ϕ, it can be shown that the average $d(U, \hat{U})^2$ is approximately equal to half the mean squared error of $\hat{\phi}$.) Most striking is that the estimator in DGP B, which does not satisfy Assumption 2 because some of its components have infinite kurtosis, has the same qualitative behaviour as the estimator in DGP A, and in fact has a smaller RMSD. Therefore, although the definition of the moment matrices $\Gamma_k(s)$ and $\Phi_k(s)$ require the existence of fourth moments, we observe that the estimator is not negatively affected by a departure from this assumption.

The clear superiority of the ML estimator leads to the conclusion that this is the preferred method of estimation when the dimension of the problem is not too large. One might expect that the ML estimator will be less robust against departures from the assumption of independent GARCH(1,1) components, for example in case of extended GARCH or non-Gaussian innovations. Unreported simulation evidence suggests that although the ML estimator may indeed be negatively affected by such departures, the available alternative estimators will also perform worse in these cases. While the misspecification reduces the gap in efficiency between the two estimators, the same ranking of methods applies. What is evident is that in case of severe misspecification, likelihood maximization will run into convergence problems, and the same clearly applies when the dimension m of the system increases. Therefore, we recommend our method-of-moments estimator in cases where numerical problem are experienced.

5 Empirical applications

5.1 European sector indices

In this subsection we analyse empirical GO-GARCH models for Dow Jones STOXX 600 European stock market (super-)sector indices. From `www.stoxx.com`, we downloaded daily data, January 1987 through December 2007, on the 15 super-sector indices that were available for this period[2], yielding $n = 5420$ daily log-returns. These data displayed virtually no autocorrelation, so that they have only been corrected for a constant mean.

We start by estimating a trivariate model for the sectors Automobiles & Parts (A), Banks (B), and Oil & Gas (O). The unconditional standard deviations and correlations of these three returns are given by

$$\hat{\sigma}_A = 0.0141, \qquad \hat{\sigma}_B = 0.0115, \qquad \hat{\sigma}_O = 0.0120,$$
$$\hat{\rho}_{A,B} = 0.765, \qquad \hat{\rho}_{A,O} = 0.531, \qquad \hat{\rho}_{B,O} = 0.604.$$

The method-of-moments estimator with $p = 100$ lags (denoted $\hat{U}_{MM}$), and the maximum likelihood estimator based on independent Gaussian GARCH(1,1) factors (denoted $\hat{U}_{ML}$) are given by

[2] Automobiles & Parts; Banks; Basic Recources; Chemicals; Construction and Materials; Financial Services; Food & Beverage; Health Care; Industrial Goods & Services; Insurance; Media; Oil & Gas; Technology; Telecommunications; Utilities.

$$\hat{U}_{MM} = \begin{bmatrix} 0.973 & -0.157 & 0.172 \\ 0.039 & 0.839 & 0.543 \\ -0.229 & -0.522 & 0.822 \end{bmatrix}, \qquad \hat{U}_{ML} = \begin{bmatrix} 0.775 & -0.631 & 0.012 \\ 0.563 & 0.683 & -0.465 \\ 0.285 & 0.367 & 0.885 \end{bmatrix},$$

with a distance of $d(\hat{U}_{MM}, \hat{U}_{ML}) = 0.504$. The estimated GARCH parameters for the three factors from both methods are

$$\begin{aligned} (\hat{\alpha}_1, \hat{\beta}_1)_{MM} &= (0.060, 0.926), & (\hat{\alpha}_1, \hat{\beta}_1)_{ML} &= (0.095, 0.881), \\ (\hat{\alpha}_2, \hat{\beta}_2)_{MM} &= (0.042, 0.954), & (\hat{\alpha}_2, \hat{\beta}_2)_{ML} &= (0.054, 0.937), \\ (\hat{\alpha}_3, \hat{\beta}_3)_{MM} &= (0.072, 0.907), & (\hat{\alpha}_3, \hat{\beta}_3)_{ML} &= (0.033, 0.964). \end{aligned}$$

We observe rather different estimates of both the link matrix U and the GARCH parameters. To investigate the consequences of this, Figure 2 compares the estimated conditional volatilities (on the diagonal), conditional correlations (above the diagonal) and conditional covariances (below the diagonal) of the three series, based on both methods.

Figure 2 about here

We observe that, although the parameter estimates based on both methods are rather different, they seem to agree to a large part on the estimated volatilities and covariances. The estimated correlations differ more. On average, the series based on ML estimation seem to display more variation than the corresponding series based on MM estimation (most notably in the correlation between the first two series).

Next, we estimate a 15-variate GO-GARCH model for all sector indices. Maximum likelihood estimation, which involves maximization over a parameter space of dimension $\frac{1}{2}m(m-1)+2m = 135$, takes a long time to converge, but eventually a (possibly local) maximum of the likelihood function is reached. The method of moments estimator (still with $p = 100$), on the other hand, still can be computed within a few seconds. The distance between the estimates now is given by $d(\hat{U}_{ML}, U_{MM}) = 0.591$. Figure 3 depicts the estimated volatilities, covariances and correlations for the same three series as in Figure 2, but now based on the 15-variate model.

Figure 3 about here

Again we observe that the estimated volatilies and covariances from both estimation methods are rather similar, and that the largest differences occur in the conditional correlations. Moreover, comparing the two figures, it appears that moving from the trivariate model to the 15-variate model leads to smoother behaviour in all estimated volatilities, covariances and correlations, regardless of which estimation method is used. In fact the conditional correlations in the large model seem to display very little variation around their unconditional mean.

The results so far were based on estimated univariate GARCH(1,1) models for the factors $\{\hat{y}_{it}\}$. It should be noted, however, that the misspecification test of Ling and Li (1997), based on the sample autocorrelation function of $\hat{y}_t'\hat{H}_t^{-1}\hat{y}_t$, rejects the null hypothesis of correct specification, for both estimation methods, and both dimensions ($m = 3$ and $m = 15$). Therefore, we have also estimated extended GARCH specifications for the factors (based on method-of-moments estimation of U). Although the Ling and Li (1997) test statistics become substantially smaller, they still lead to a rejection of the null hypothesis at any reasonable significance level. The corresponding graphs of volatilities, covariances and correlations are virtually indistinguishable from those given in Figures 2 and 3.

5.2 Fuel prices and transportation stock prices

The empirical application in this subsection will examine the conditional correlations between daily returns on stock prices from the transportation sector and daily returns on the price of crude oil and kerosene (jet fuel) in the United States. The stocks included are American Airlines, South-West Airlines, Boeing and FedEx, which we downloaded from Yahoo Finance: `http://finance.yahoo.com`. The price data for crude oil and kerosene denote U.S. Energy Information Administration (EIA) daily energy spot prices that can be downloaded from EconStats: `http://www.econstats.com`. The six-variate sample covers the period between July 19, 2003, and August 12, 2008, yielding $n = 3790$ daily returns. We will work with the residuals from a VAR(1) model that corrects for the little autocorrelation present in the data.

A motivation for looking at these correlations is that they may be revealing of fuel efficiency of the transportation companies and how well these companies are able to hedge fuel price risk. A reason why we may expect to see persistent correlation between price movements on a daily scale is that they reflect simultaneous revision of market expectations concerning the future price of fuel and its impact on the value of the transportation company. In other words, the level of correlation in a way measures the market's assessment of the significance of fuel prices for the present value of the airline, given the market's assessment of the (future) fuel efficiency of the airline, the uncertainty in future oil prices, and the airline's ability to hedge uncertain fuel prices. The question is, how do airlines compare over time?

The six variables combined yield a total of fifteen correlations between all possible pairs. In what follows we will focus our attention on the correlations between the return on jet fuel and the returns on the four stock prices. The unconditional volatilities and correlations are given by

$$\hat{\sigma}_K = 0.025, \qquad \hat{\sigma}_A = 0.039, \qquad \hat{\sigma}_S = 0.024, \qquad \hat{\sigma}_B = 0.020, \qquad \hat{\sigma}_F = 0.020,$$
$$\hat{\rho}_{K,A} = -0.172, \qquad \hat{\rho}_{K,S} = -0.088, \qquad \hat{\rho}_{K,B} = -0.027, \qquad \hat{\rho}_{K,F} = -0.062,$$

where K refers to kerosene, A to American Airlines, S to South-West Airlines, B to Boeing and F to FedEx. We find that the two airlines exhibit a stronger (negative) correlation with the price of jet fuel than Boeing and FedEx. American Airlines shows the highest correlation in absolute value (it also has the highest volatility). Boeing reports the lowest correlation (in absolute value) which may hint to the fact that it is the only non-transportation company in the list, such that its operational costs and profitability depends less crucially on the price of kerosene.

We continue by estimating the six-variate model. The maximum-likelihood estimator is based on independent Gaussian GARCH(1,1) factors. For the method-of-moments estimator we set the number of lags at $p = 50$. Figure 4 plots the four conditional correlations over time. For ease of exposition we plotted the ML estimates only. (A comparison between MM and ML is shown in the next figure.)

Figure 4 about here

It can be seen that all four correlations show the same pattern over time. Most of the differences between variables concern the levels. The figure confirms that the two airlines appear to be more correlated with changes in the price of jet fuel than FedEx and Boeing (where Boeing is least correlated). (The separation between the airlines and the non-airlines is most noticeable since 2004.) What stands out is that American Airlines shows the strongest correlation over the entire period.

When comparing airlines, South-West seems to be less sensitive to movements in the price of jet fuel. This finding is consistent with the fact that South-West is known to have a long-term and active program to hedge fuel prices. In short, they buy fuel options when the price of oil is believed to be low to hedge against potentially high fuel prices years later. These hedges have for example helped South-West through the Iraq war and hurricane Katrina when oil prices indeed increased significantly. In the third quarter of 2008 (the end of our sample), South-West recorded its first loss in many years, partly because the then noticeable drop in oil prices had rendered the fuel hedging strategy of lesser value. (Also, South-West has a somewhat younger fleet compared to American Airlines, which may benefit fuel efficiency.)[3]

The fact that time patterns in correlations are largely the same for the series considered may have various explanations. Where a variation in fuel efficiency and/or different strategies to hedge fuel prices may be able to explain the differences in levels, it is thinkable that the co-variation over time may be attributed to isolated events that apply to all. It is also possible however, that whenever hedging the price of fuel is proving to be more difficult for one airline, it has also become more difficult for others, which too may explain why correlations largely move together over time.

Figure 5 compares MM estimates of the conditional correlations to those obtained with ML. Note that in contrast to the previous example, see Figure 2, the MM-based correlations now display more variation than their ML-based counterparts.

Figure 5 about here

The two different estimators largely agree on the correlations for the two airlines, but show clear differences for both Boeing and FedEx. It is possible that MM has not been able to identify selected factors due to multiplicity in sample eigenvalues; these factor may have been more important for Boeing and FedEx than for the two airlines. Indeed, inspection of the MM estimates of the link matrix Z and the GARCH parameters reveals that three factors with very similar GARCH parameters (and hence similar eigenvalues) have the strongest effect on Boeing and FedEx, and much less on the other returns.

Let us examine more closely the correlations over time for the two airlines. Most noticeable are the two large drops in correlation between end of 2002 and end of 2003 from -0.1 to around -0.4. After

[3] See `http://en.wikipedia.org/wiki/Southwest_Airlines#Hedging_fuel` for references.

the first drop late 2002 correlations quickly climbed to previous levels of around -0.1, only to drop again shortly after, creating a 'W' pattern. The most important event that occurred around the time of these volatile movements in correlation was the Iraq war (also known as the second Gulf war) which began on March 20, 2003. The invasion of Iraq was indeed a significant event for oil markets because of Iraq's large oil reserves, which probably led to an increase in levels of uncertainty concerning future oil prices. While the Iraq war, as an isolated incident, probably does not change the long-term price of oil, it will have short-term effects via the updating of expectations by the market.

Next we observe a small spike in correlation towards the end of 2005. If it is not due to noise in our estimates, this may well reflect the effects of hurricane Katrina that was effective between August 23 and August 30, 2005. Katrina damaged or destroyed a number of oil platforms, caused closure of some refineries, and caused serious damage to important infrastructure in the Gulf Coast which is an important oil producing region in/for the United States.

Towards the end of our sample, from 2007 onwards, the correlation steadily declined from around -0.15 to below -0.50, well below the unconditional mean. The important event that coincides with this trend in correlation is the financial crisis that started around July, 2007.

Inspired by these findings we conjecture that the conditional correlation, between changes in the price of jet fuel and changes in the price of airline stocks, are linked to levels of uncertainty concerning future price levels of oil (and jet fuel). Correlations are high (in absolute value) when uncertainty is high. A possible rationale for this relationship is that when the market is uncertain about future price levels of oil, and thus about the operational costs the airline may face, the price of oil will get more weight when assessing the present value of the airline, compared to factors that exhibit lesser uncertainty. Also, in times of high uncertainty, the market will more frequently be revising their expectations both about the price of oil and its effect on the present value of the airline – which too may strengthen correlations between the two.

Let us also explain how this conjecture is consistent with our empirical findings (the time-variation in estimated conditional correlations).

The effect of the Iraq war starting March 2003: Towards the start of the war, as plans for the U.S. led invasion were unfolding, uncertainty concerning the future price of oil was probably increasing, together with the correlation (in absolute value) between oil and the airlines' stock prices. During a short period after the invasion, the perception was that the war would quickly come to an end ('mission accomplished'), which may explain why correlations promptly returned to previous levels (shortly after the first drop). But then the market realized that the end was nowhere in sight, re-introducing increased levels of uncertainty concerning the price of oil, and correlations dropped again to below -0.4.

The effect of the financial crisis starting July 2007: At the outset of the financial crisis, oil prices were still climbing. (They reached their peak mid 2008 when the crisis was well under way.) Thus when the recession in the United States was already looming, the price of oil was still high. Importantly, there was considerable uncertainty at the time whether the recession would bring down oil prices. Much of the (new) demand for oil has come from emerging economies such as China, as well as from Europe. It was uncertain if and how the recession in the United States would carry over to the emerging markets and Europe. In case of a "decoupling" the demand for oil would stay strong despite the recession originating in the United States. In this time of uncertainty, correlations increased in absolute value. When it then became clear that the recession would be global also oil prices came down, together with stock prices. We can already see correlations increasing again at the end of our sample. (Our conjecture is that correlations will have continued climbing in the period that followed our sample.)

6 Concluding remarks

We have put forward a method-of-moments estimator for the factor loading matrix in GO-GARCH models. The method is based on the eigenvectors of a suitably defined sample autocorrelation matrix of squares and cross-products of the observed data. This means that estimation does not require any optimization of an objective function, so that it is free of numerical convergence problems regardless of the dimension. The parameters from the univariate GARCH-type models can be estimated separately for each individual factor, given our estimate for the factor loading matrix, which makes the method particularly easy to implement.

Our method of estimation provides an alternative to the estimator originally proposed by van der Weide (2002), which jointly estimates parameters that feature both in the factor loading matrix and the univariate GARCH-type specifications by means of maximum likelihood. ML estimates of the factor loading matrix thus depend on the choice of GARCH-type models used to specify the likelihood function.

The ML estimator is found to be considerably more efficient than the MM estimator, particularly when the likelihood function is consistent with the data generating process. The loss in efficiency is the price the MM estimator pays for its numerical convenience (no optimization required), and the fact that no assumptions need to be made concerning the model specification for the individual factors. Interestingly, Monte Carlo simulations suggest that ML is fairly robust to misspecifications in the underlying model. While model misspecification reduces the gap between MM and ML, the latter tends to remain the more efficient estimator.

The MM estimator will be a welcome alternative however, whenever ML experiences convergence difficulties. ML estimation can become problematic when the dimension is particularly large and/or when the model used to specify the likelihood function is considerably misspecified, while the MM estimator always works. One could also consider taking the MM estimate as an initial value for ML, in an effort to resolve any convergence problems[4].

References

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

[4] In that case, it is more convenient to parametrize the U matrix in the likelihood function in terms of a skew-symmetric matrix using the Cayley transform.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis* (2nd ed.). New York: John Wiley and Sons.

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

Bollerslev, T. (2008), "Glossary to ARCH (GARCH)", CREATES Research Paper 2008-49, `ftp://ftp.econ.au.dk/creates/rp/08/rp08_49.pdf`.

Boswijk, H. P. and R. van der Weide (2006), "Wake me up before you GO-GARCH", UvA-Econometrics Discussion Paper 2006/03, `http://www.ase.uva.nl/pp/bin/381fulltext.pdf`.

Boswijk, H. P. and R. van der Weide (2008), "Testing the number of factors in GO-GARCH models", Working paper, University of Amsterdam.

Comte, F. and O. Lieberman (2003), "Asymptotic theory for multivariate GARCH processes", *Journal of Multivariate Analysis*, 84, 61–84.

Diebold, F. X. and M. Nerlove (1989), "The dynamics of exchange rate volatility: A multivariate latent factor ARCH model", *Journal of Applied Econometrics*, 4, 1–21.

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

Doz, C. and E. Renault (2006), "Factor stochastic volatility in mean models: a GMM approach", *Econometric Reviews*, 25, 275–309.

Engle, R. F. and K. F. Kroner (1995), "Multivariate simultaneous generalized GARCH", *Econometric Theory*, 11, 122–150.

Engle, R. F., V. K. Ng and M. Rotschild (1990), "Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills", *Journal of Econometrics*, 45, 213–237.

Fan, J., M. Wang and Q. Yao (2008), "Modelling multivariate volatilities via conditionally uncorrelated components", *Journal of the Royal Statistical Society Series B*, 70, 679–702.

Jensen, S. T. and A. Rahbek (2007), "On the law of large numbers for (geometrically) ergodic Markov chains", *Econometric Theory*, 23, 761–766.

Hafner, C. (2008), "Temporal aggregation of multivariate GARCH processes", *Journal of Econometrics*, 142, 467–483.

He, C. and T. Teräsvirta (1999), "Properties of moments of a family of GARCH processes", *Journal of Econometrics*, 92, 173–192.

Lanne, M. and P. Saikkonen (2007), "A multivariate generalized orthogonal factor GARCH model" *Journal of Business & Economic Statistics*, 25, 61–75.

Ledoit, O. and M. Wolf (2004), "A well-conditioned estimator for large-dimensional covariance matrices", *Journal of Multivariate Analysis*, 88, 365–411.

Liebeck, H. and A. Orborne (1991), "The generation of all rational orthogonal matrices", *The American Mathematical Monthly*, 98, 131–133.

Ling, S. and W. K. Li (1997), "Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors", *Journal of Time Series Analysis*, 18, 447–464.

Rigobon, R. (2003), "Identification through heteroskedasticity", *Review of Economics and Statistics*, 85, 777–792.

Sentana, E. and G. Fiorentini (2001), "Identification, estimation and testing of conditionally heteroskedastic factor models", *Journal of Econometrics*, 102, 143–164.

Silvennoinen, A. and T. Teräsvirta (2008), "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.

Urga, G. (2007), "Common features in economics and finance: An overview of recent developments", *Journal of Business & Economic Statistics*, 25, 2–11.

van der Weide, R. (2002), "GO-GARCH: A multivariate generalized orthogonal GARCH model", *Journal of Applied Econometrics*, 17, 549–564.

Vrontos, I. D., P. Dellaportas and D. N. Politis (2003), "A full-factor multivariate GARCH model", *Econometrics Journal*, 6, 312–334.

Figure 1: Root mean square distance of MM and ML estimators for various choices of p and n.

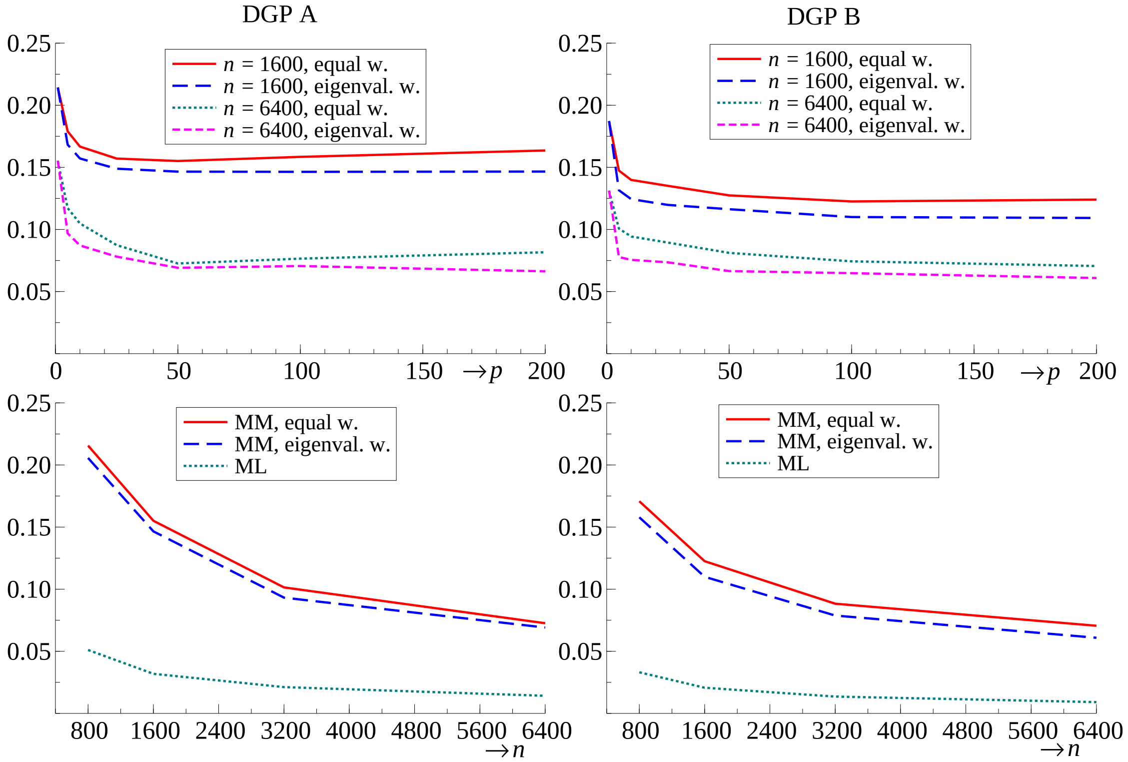

Figure 2: Estimated conditional volatilities, correlations and covariances based on MM and ML estimation, trivariate model. The first row and column refer to Automobiles & Parts, the second row and column to Banks, and the final row and column to Oil & Gas.

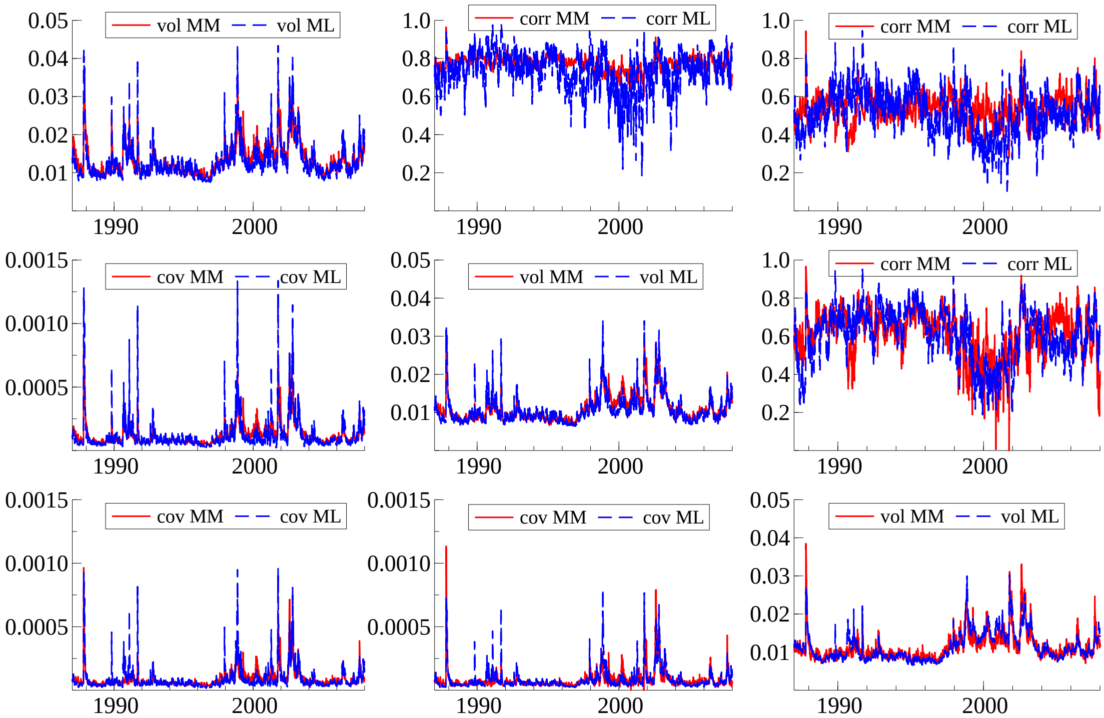

Figure 3: Estimated conditional volatilities, correlations and covariances based on MM and ML estimation, 15-variate model. The first row and column refer to Automobiles & Parts, the second row and column to Banks, and the final row and column to Oil & Gas.

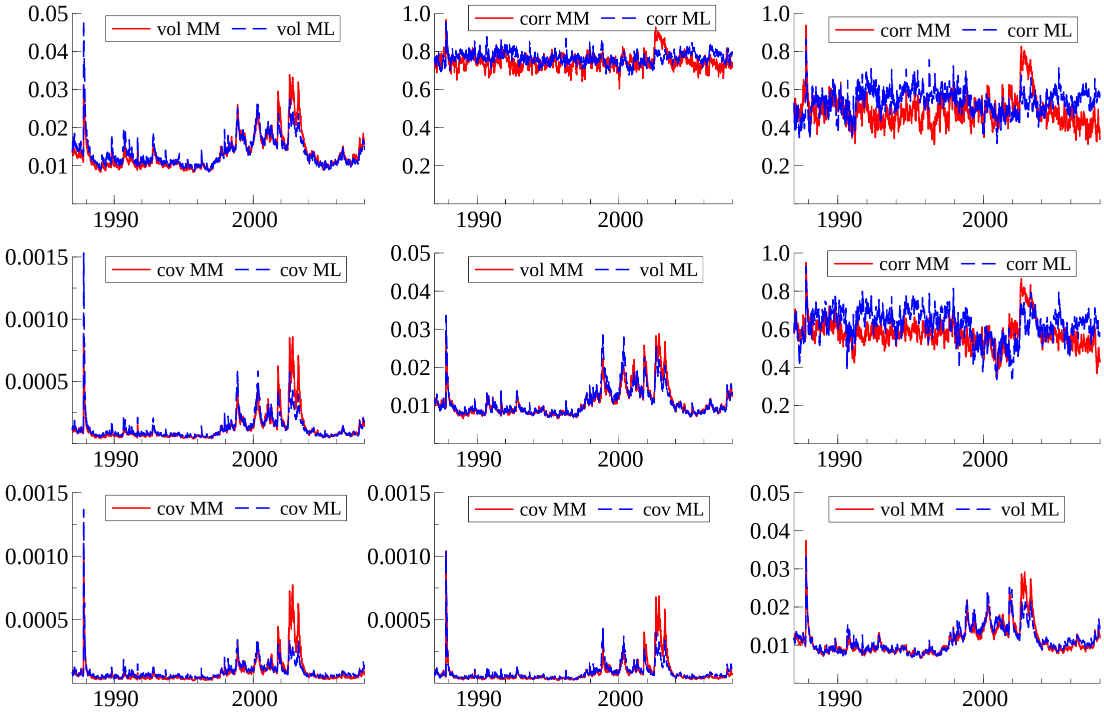

Figure 4: Estimated conditional correlations between Kerosene returns, and stocks returns of American Airlines, South-West Airlines, Boeing and FedEx; ML estimation.

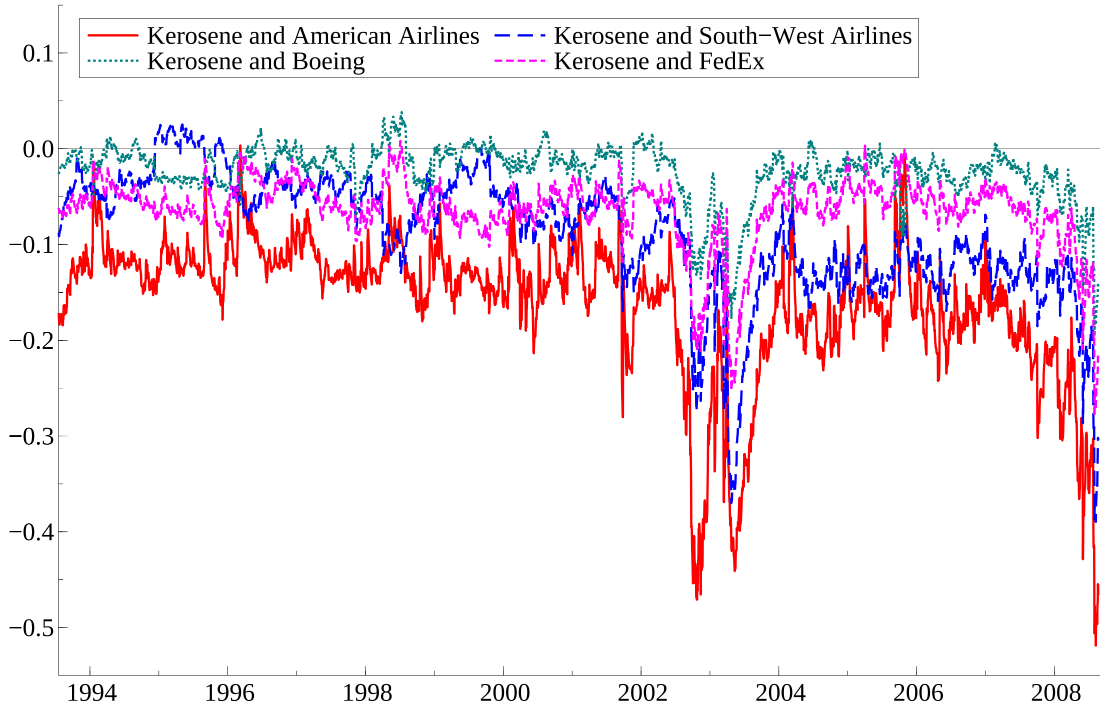

Figure 5: MM and ML estimated conditional correlations between Kerosene returns, and stocks returns of American Airlines, South-West Airlines, Boeing and FedEx.