The time-variation of volatility and the evolution of expectations

van der Weide, R.

Publication date
2012

Citation for published version (APA):
Chapter 3

Method of moments estimation of GO-GARCH models

3.1 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, 2006). For surveys on multivariate volatility models
we refer to e.g. Bauwens et al. (2006), and Silvennoinen and Teräsvirta (2009). 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., a vector of portfolio returns satisfies the same model, albeit with other parameter values, as the original vector of asset returns. The model is also closed under temporal aggregation, 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 chapter 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 Newton-type optimization of an objective function, but instead only involves iterated matrix rotations, 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.

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 mov-
3.2. THE GO-GARCH MODEL

ing closer together over time, and do correlations jump up in periods of extreme volatility (such as in a financial crisis)? In a modest empirical example in this chapter we examine the correlation over time between returns on European sector indices, and find that the degree of comovement between volatility and correlation depends on the general state of the economy.

The outline of the chapter is as follows. In Section 3.2, we formulate the GO-GARCH model, and discuss currently available estimation methods. Section 3.3 introduces our method-of-moments estimator, and discusses how information from autocorrelation matrices at different lags may be efficiently combined. In Section 3.4 we use Monte Carlo simulations to study the efficiency of our estimator relative to (quasi-) maximum likelihood in low-dimensional systems. Section 3.5 contains an empirical application to a vector of ten European industry returns. Section 3.6 concludes.

3.2 The GO-GARCH model

3.2.1 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 $\{F_t\}_{t \geq 0}$ 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 $E(x_t|F_{t-1}) = 0$. The GO-GARCH model imposes a structure on the conditional variance matrix $\Sigma_t = \text{var}(x_t|F_{t-1}) = E(x_t x_t'|F_{t-1})$, implied by:

Assumption 3.1 The process $\{x_t\}_{t \geq 1}$ satisfies the representation

\[
\begin{align*}
x_t &= Z y_t = Z H_t^{1/2} \varepsilon_t, \\
H_t &= \text{diag}(h_{1t}, \ldots, h_{mt}),
\end{align*}
\]

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

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

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

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 chapter, 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

\[ \Sigma_t = \text{var}(x_t|\mathcal{F}_{t-1}) = ZH_tZ', \quad (3.3) \]

and unconditional variance

\[ \Sigma = \text{var}(x_t) = ZZ'. \quad (3.4) \]

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

\[ h_{it} = (1 - \alpha_i - \beta_i) + \alpha_i y_{it}^2 + \beta_i h_{i,t-1}, \quad \alpha_i, \beta_i \geq 0, \quad \alpha_i + \beta_i < 1, \quad (3.5) \]

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 $y_{jt-k}, j \neq i, k \geq 1$. A simple extension of (3.5) is their extended GARCH(1,1) specification:

\[ h_{it} = \left(1 - \sum_{j=1}^{m} \alpha_{ij} - \beta_i\right) + \sum_{j=1}^{m} \alpha_{ij} y_{jt}^2 + \beta_i h_{i,t-1}, \quad \alpha_{ij}, \beta_i \geq 0, \quad \sum_{j=1}^{m} \alpha_{ij} + \beta_i < 1. \quad (3.6) \]

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 3.1, as well as the estimation methods proposed in this chapter, 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 $y_{it}^2$. 
Consider the polar decomposition of $Z$:

$$Z = SU,$$  \hspace{1cm} (3.7)

where $S$ is a positive definite symmetric matrix, and $U$ is an orthogonal matrix. Using (3.4), it follows that $\Sigma = S^2$, so that $S$ is the symmetric square root of $\Sigma$, i.e., $S = P L^{1/2} P'$, where $P L P'$ is the spectral decomposition of $\Sigma$. This implies that part of the matrix $Z$ may be identified from the unconditional information $\Sigma = \text{var}(x_t)$, 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 (3.7), 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 $\Sigma$ 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 $U = P$.

### 3.2.2 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 $\Sigma_t$ 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).
CHAPTER 3. METHOD OF MOMENTS ESTIMATION OF GO-GARCH

and Diebold and Nerlove (1989).

The variance matrix $H_t$ can in this case be expressed as

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

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

$$\Sigma_t = Z_1 H_{1t} Z_1' + Z_2 Z_2' = \Sigma + Z_1 (H_{1t} - I_r) Z_1',
$$

$$\text{var}(s_t | F_{t-1}) = U_1 H_{1t} U_1' + U_2 U_2' = I_m + U_1 (H_{1t} - I_r) U_1'.$$

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 chapter we assume a full-factor GO-GARCH model.

### 3.2.3 Currently available estimation methods

In this subsection we briefly review the currently available methods for estimating the model implied by Assumption 3.1, 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

$$\ell(\theta) = -\frac{1}{2} \sum_{t=1}^{n} \left\{ m \log (2\pi) + \log |H_t(\theta)| + s_t' U(\theta_1) H_t(\theta)^{-1} U(\theta_1)' s_t \right\}, \quad (3.8)$$

where $\theta = (\theta'_1, \theta'_2)$, with $\theta_1$ a vector of dimension $\frac{1}{2} m(m-1)$ characterizing the $m \times m$ orthogonal matrix $U = U(\theta_1)$, and $\theta_2$ a $2m$-dimensional parameter vector of GARCH pa-
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 $\theta$. 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 $\frac{1}{n-1}\sum_{t=1}^n x_tx_t'$. (If $m$ is large relative to $n$, one could also consider shrinkage-type estimators of $\Sigma$, 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 $\theta_0 = \text{vech}(\Sigma)$, then full maximum likelihood would entail maximizing (3.8), 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 3.2.2.

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

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

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

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

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 $\Sigma$ 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 $E[y_{it}y_{jt}I(B)|B] = 0$ 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
\[ \Psi(U) = \sum_{1 \leq i < j \leq m} \sum_{B \in B} \sum_{k=1}^{p} |u'_i E[s_t s'_t I(s_{t-k} \in B)]u_j| = 0. \] (3.9)

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 \( \Sigma \) 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 common eigenvectors of a sequence of sample moment matrices, and therefore can be applied to arbitrary dimensions \( m \).

### 3.3 Method of moments estimation

#### 3.3.1 The estimator based on a single lag

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

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

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
3.3. METHOD OF MOMENTS ESTIMATION

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 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 $\text{corr}(y_{it}^2, y_{i,t-k}^2) = 0$ for all $k = 1, \ldots, p$). Finally, the zero cross-covariances $\tau_{ijk}$ 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 (3.6), 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

$$\Gamma_k(y) = E(Y_iY_{1-k}), \quad k = 1, 2, \ldots$$

(3.10)

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

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

Therefore, Assumptions 3.1 and 3.2 imply, using $\text{var}(y_{it}^2) = E(y_{it}^4) - E(y_{it}^2)^2 = \kappa_i - 1$,

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

or in other words

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

For the corresponding autocorrelation matrix, we thus find

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

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

$$\Gamma_k(s) = E(S_iS_{1-k}) = E(UY_iU'Y_{1-k}S') = U\Gamma_k(y)U',$$

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 3.1 and 3.2, $U$ may be identified by the eigenvectors of either $\Gamma_k(s)$ or $\Phi_k(s)$.

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

\[
\hat{\Gamma}_k(s) = \frac{1}{n} \sum_{t=k+1}^{n} S_t S_{t-k} = \frac{1}{n} \sum_{t=k+1}^{n} (s_t s'_t - I_m)(s_{t-k} s'_{t-k} - I_m),
\]
\[
\hat{\Phi}_k(s) = \hat{\Gamma}_0(s)^{-1/2} \tilde{\Phi}_k(s) \hat{\Gamma}_0(s)^{-1/2},
\]

(3.11) (3.12)

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 $\tilde{\Phi}_k(s) = \frac{1}{2}(\hat{\Phi}_k(s) + \hat{\Phi}_k(s)')$ of $\Phi_k(s)$. 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 $\tilde{\Phi}_k(s)$ leads to a more efficient estimator.

3.3.2 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 $\hat{\Phi}_k = \hat{\Phi}_k(s)$. The property that each of the population matrices $\Phi_k = \Phi_k(s)$ have the same matrix of eigenvectors is shared with the so-called common principal components (CPC) model, see Flury (1984), where $\{\Phi_k\}_{k=1}^{p}$ represent covariance matrices of $p$ different random vectors. Under the assumption that these are Gaussian, and that we have $p$ independent i.i.d. samples, Flury (1984) derives the maximum likelihood estimator of the common eigenvector matrix $U$. Because these assumptions are clearly violated, we consider a closely related least-squares estimator,
3.3. METHOD OF MOMENTS ESTIMATION

discussed by Beaghen (1997). This estimator minimizes the criterion function

\[ S(U) = \sum_{k=1}^{p} \text{tr}(U'\tilde{\Phi}_k U - \text{diag}(U'\tilde{\Phi}_k U))(U'\tilde{\Phi}_k U - \text{diag}(U'\tilde{\Phi}_k U)) \]  

(3.13)

over all orthogonal matrices \( U \). Although this minimization problem does not have a closed-form solution, Beaghen (1997) shows that the so-called F-G algorithm of Flury and Gautschi (1986) can be easily adapted to this least-squares estimator. This algorithm involves an iteration of rotations until the first-order conditions are satisfied, which can be rewritten as

\[ u_i' \left( \sum_{k=1}^{p} (\lambda_{ki} - \lambda_{kj}) \tilde{\Phi}_k \right) u_j = 0, \quad i \neq j = 1, \ldots, m, \]  

(3.14)

\[ \lambda_{ki} = u_i' \tilde{\Phi}_k u_i, \quad i = 1, \ldots, m; \quad k = 1, \ldots, p, \]  

(3.15)

with \( u_i \) the columns of \( U \). This shows that the solution diagonalizes a weighted average of the matrices \( \tilde{\Phi}_k \), with weights proportional to the difference in eigenvalues. Matrices \( \tilde{\Phi}_k \) that are less informative about the matrix \( U \), because of the near multiplicity of their eigenvalues, are therefore given less weight.

The estimator we propose can be summarized as follows.

**Summary 3.1** 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 (3.12);

3. using the symmetrized autocorrelation matrices \( \tilde{\Phi}_k = \frac{1}{2}(\hat{\Phi}_k(s) + \hat{\Phi}_k(s)') \), estimate \( U \) by minimizing the criterion function \( S(U) \) given in (3.13), based on the adapted F-G algorithm;
CHAPTER 3. METHOD OF MOMENTS ESTIMATION OF GO-GARCH

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 \( y_t \) by quasi-maximum likelihood.

3.3.3 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):

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

The motivation for \( D(\cdot, \cdot) \) is that in the model \( s_t = U y_t \), the columns of the matrix \( U \) may be reordered and multiplied by \(-1\), by changing rows of \( y_t \) in the same way. In other words, \( U \) is invariant under permutation and sign change of its columns. 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:

**Assumption 3.3** In the model defined by Assumptions 3.1 and 3.2, 

\[
\max_{1 \leq k \leq p} \min_{1 \leq i < j \leq m} |\rho_{ik} - \rho_{jk}| > 0.
\]

The assumption excludes the possibility that two squared components \( y^2_{it} \) and \( y^2_{jt} \) have the same autocorrelation function for \( k = 1, \ldots, p \). 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 \( \hat{U} \).

**Theorem 3.1** Consider the MM-CPC estimator \( \hat{U} \), minimizing (3.13). Then, under As-
3.4 Monte Carlo simulations

In this section we study the finite-sample performance of the estimator proposed in this chapter, in comparison with maximum likelihood. We focus on a trivariate system \((m = 3)\), and we consider four different data-generating processes (DGPs) for the conditionally uncorrelated process \(\{y_t\}_{t \geq 1}\). The (relative) efficiency of the two estimators is evaluated using 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 both the MM-CPC estimator and the ML estimator,
the distance $d(U, \hat{U})$ is invariant to $U$, in the sense that if $\hat{U}_1$ and $\hat{U}_2$ are estimates based on data generated using $U_1$ and $U_2$, respectively, then $d(U_1, \hat{U}_1) = d(U_2, \hat{U}_2)$. Therefore, the choice of $U$ in the DGP is irrelevant. Three sample sizes, with $n \in \{500, 1000, 2000\}$, are considered.

We first consider two DGPs in which the components of $y_t$ are independent Gaussian GARCH(1,1) processes, so the maximum likelihood estimator is based on a well-specified model:

- **DGP A**: $(\alpha_1, \beta_1) = (0.055, 0.94); (\alpha_2, \beta_2) = (0.16, 0.8); (\alpha_3, \beta_3) = (0.25, 0.65);
- **DGP B**: $(\alpha_1, \beta_1) = (0.095, 0.9); (\alpha_2, \beta_2) = (0.26, 0.7); (\alpha_3, \beta_3) = (0.25, 0.65).

Under DGP A, all components of $y_t$ have finite kurtosis, so that Assumption 3.2 is satisfied. The autocorrelation decay rate $\alpha + \beta$ varies from 0.995 to 0.9, reflecting empirically relevant persistence in the autocorrelation functions of $y^2_{it}$. The initial autocorrelations are such that the three autocorrelation functions cross each other, implying that the average autocorrelations (over $p$ lags) for the three series can be close to each other (depending on $p$). Under DGP B, the first and second component have infinite kurtosis; this DGP is included to investigate how sensitive the estimator is to deviations from the finite-fourth-moment assumption.

The number of lags used in the MM-CPC estimator is fixed at $p = 100$. Preliminary simulations have indicated that for the type of persistence in autocorrelations considered here, the efficiency does not improve much beyond $p = 50$, so that $p = 100$ may be considered a conservative choice. It should be emphasized that the improvements over $p = 1$ are substantial: the RMSD of the MM estimator based on only the first lag is typically about 3 to 5 times larger than that of the MM-CPC estimator with $p = 100$. The results, for the first two DGPs are given in Figure 3.4, top panels.

We observe that for the first two DGPs, the MM-CPC estimator has a RMSD which is about two and a half times larger than that of the ML estimator. This should be seen as the price that we pay for using a more robust and computationally less intensive estimator than ML. The RMSD of both estimators appears to decrease with the sample size at approximately the rate $n^{-1/2}$. (In a bivariate model where $U$ is characterized by a single angle $\phi$, 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 MM estimator in DGP B, which does not satisfy Assumption 3.2 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 $\Phi_k(s)$ requires the existence of fourth moments, we observe that the estimator is not negatively affected by a departure from this assumption.

To investigate the effect of misspecification on the relative performance of the two methods, we now consider the following data-generating processes. DGP C is an extended GARCH specification, of the form

$$
\begin{align*}
h_{1t} &= 0.005 + 0.035 y_{1,t-1}^2 + 0.02 y_{3,t-1}^2 + 0.94 h_{1,t-1}, \\
h_{2t} &= 0.04 + 0.08 y_{1,t-1}^2 + 0.08 y_{2,t-1}^2 + 0.8 h_{2,t-1}, \\
h_{3t} &= 0.1 + 0.25 y_{3,t-1}^2 + 0.65 h_{3,t-1},
\end{align*}
$$

where the standardized innovations $\varepsilon_{it}$ now follow a standardized Student’s $t(5)$ distribution (with $\varepsilon_{it}$ independent of $\varepsilon_{jt}$, $i \neq j$). Extending the results of Hafner (2003), it can be shown that the components of $y_t$ have finite kurtosis under this DGP. Clearly, the ML estimator is now based on a misspecified likelihood (assuming independent GARCH processes), whereas the MM estimator is still valid. Finally, DGP D has the same GARCH parameters as DGP A, but now the standardized innovations $\varepsilon_{it}$ follow a standardized Student’s $t(3)$ distribution (implying infinite kurtosis for all components of $y_t$). In this case
the misspecification of the likelihood function only concerns the shape of the innovation density, not the volatility dynamics.

The results in the bottom panels of Figure 3.4 show that for these two DGPs and the sample sizes considered here, the MM estimator is in fact more efficient than the ML estimator. Most striking about the results for DGP C is that the RMSD appears to decrease slower as a function of the sample size than in the other cases. Comparison of the results for DGP D and DGP A (which differ only in the innovation distribution) shows that the performance of the MM estimator has improved by having fat-tailed innovations (despite the lack of finite fourth moments), whereas the ML estimator performs much worse.

The superiority of the ML estimator for the first two DGPs indicates that this is the preferred method of estimation when the dimension of the problem is not too large, and the latent processes are known to be independent Gaussian GARCH processes. On the other hand, the results for DGP C and D illustrate that maximum likelihood estimation can be sensitive to both volatility misspecification and innovation distribution misspecification, and that the MM-CPC method can be more robust in such cases. Furthermore, for larger dimensions, likelihood maximization will run into convergence problems, especially for misspecified models, which is not the case for the MM-CPC method.

### 3.5 Empirical application to European sector indices

In this section we analyse an empirical GO-GARCH model for Dow Jones STOXX 600 European stock market industry indices. From [www.stoxx.com](http://www.stoxx.com), we downloaded daily data, from the beginning of January 1992 through the end of January 2010, on the 10 industry indices, yielding \( n = 4650 \) daily log-returns.

Table 3.5 provides some descriptive statistics. The sample means are relatively small, which is due to the recent financial crisis: by the end of 2009 many stock prices were back at their 2003 levels. At the high end we find Basic Materials, at the low end we have Consumer Services. The latter is also among the least volatile industries, whereas the most volatile is the Technology industry. Note that while the first-order autocorrelation is statistically significant for some of the industries, the coefficients are all small in size. For the analysis all first-order autocorrelation is removed from the data.

Estimates not reported in the chapter show that all unconditional correlations range between 0.5 and 0.9, which suggests that the industries exhibit strong linkages. The in-
### Table 3.1: Annualized means and standard deviations, and first-order autocorrelations of returns.

<table>
<thead>
<tr>
<th>Industry</th>
<th>mean</th>
<th>std. dev.</th>
<th>ac(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Materials</td>
<td>0.085</td>
<td>0.235</td>
<td>0.027*</td>
</tr>
<tr>
<td>Consumer Goods</td>
<td>0.060</td>
<td>0.169</td>
<td>−0.004</td>
</tr>
<tr>
<td>Consumer Services</td>
<td>0.028</td>
<td>0.187</td>
<td>0.013</td>
</tr>
<tr>
<td>Financials</td>
<td>0.036</td>
<td>0.239</td>
<td>0.057**</td>
</tr>
<tr>
<td>Industrials</td>
<td>0.049</td>
<td>0.196</td>
<td>0.058**</td>
</tr>
<tr>
<td>Health Care</td>
<td>0.072</td>
<td>0.184</td>
<td>0.024*</td>
</tr>
<tr>
<td>Oil and Gas</td>
<td>0.065</td>
<td>0.231</td>
<td>−0.003</td>
</tr>
<tr>
<td>Technology</td>
<td>0.037</td>
<td>0.314</td>
<td>0.040**</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.050</td>
<td>0.251</td>
<td>0.026*</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.065</td>
<td>0.176</td>
<td>−0.020</td>
</tr>
</tbody>
</table>

*, ** significant at the 10% and 5% level.

Industries that show some of the highest correlations include: Industrials, Financials and Consumer Services. Industries that exhibit comparatively low correlations include: Oil and Gas, Technology and Health Care.

For our method-of-moments estimator, denoted by $\hat{U}_{MM}$, we used $p = 250$ lags. For the maximum likelihood estimator, denoted by $\hat{U}_{ML}$, we assume independent Gaussian GARCH(1,1) factors. The distance between the two estimates, $d(\hat{U}_{MM}, \hat{U}_{ML}) = 0.385$, indicates that the two methods provide different estimates for the link matrix $U$. How large these differences really are, and the consequences of this, can best be seen by comparing the estimates of the GARCH parameters, and ultimately the estimates of the volatilities and correlations.

The estimates for the GARCH parameters are shown in Table 3.5. For most factors, we observe that the MM and ML estimates are not that different from each other.

Figure 3.5 presents the estimated time-varying annualized volatilities of the 10 industry returns, obtained from the GO-GARCH model estimated by maximum likelihood or the method of moments. These are compared with the volatilities obtained from univariate GARCH(1,1) models for the returns, to check if the GO-GARCH model produces sensible estimates. We observe that for all returns, the three volatility estimates follow similar patterns, which are also remarkably similar across industries. After a relatively stable period with low volatilities in the 1990s, volatilities tend to increase and display more variation around the time of the burst of the internet bubble in 2000. After a few years this again is followed by a stable period, which ends with the peak in the credit crisis in
CHAPTER 3. METHOD OF MOMENTS ESTIMATION OF GO-GARCH

<table>
<thead>
<tr>
<th>Factor</th>
<th>$(\hat{\alpha}, \hat{\beta})_{ML}$</th>
<th>$(\hat{\alpha}, \hat{\beta})_{MM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0.055, 0.943)</td>
<td>(0.051, 0.948)</td>
</tr>
<tr>
<td>2</td>
<td>(0.074, 0.920)</td>
<td>(0.083, 0.910)</td>
</tr>
<tr>
<td>3</td>
<td>(0.088, 0.894)</td>
<td>(0.082, 0.901)</td>
</tr>
<tr>
<td>4</td>
<td>(0.105, 0.892)</td>
<td>(0.089, 0.908)</td>
</tr>
<tr>
<td>5</td>
<td>(0.088, 0.903)</td>
<td>(0.064, 0.931)</td>
</tr>
<tr>
<td>6</td>
<td>(0.027, 0.968)</td>
<td>(0.028, 0.967)</td>
</tr>
<tr>
<td>7</td>
<td>(0.033, 0.965)</td>
<td>(0.032, 0.966)</td>
</tr>
<tr>
<td>8</td>
<td>(0.075, 0.923)</td>
<td>(0.060, 0.938)</td>
</tr>
<tr>
<td>9</td>
<td>(0.024, 0.973)</td>
<td>(0.026, 0.971)</td>
</tr>
<tr>
<td>10</td>
<td>(0.053, 0.937)</td>
<td>(0.035, 0.961)</td>
</tr>
</tbody>
</table>

Table 3.2: Estimated GARCH parameters for the factors.

September–October 2008, leading to a sharp increase in volatility, followed by a gradual decline. By late January 2010 (the end of our sample), volatilities are back to their 2007 levels.

Industries mainly differ in the impact of the internet bubble crash, which, as expected, clearly led to higher volatilities in the Technology and Telecom industries than in other industries. The main difference between the outcomes of the different estimation methods is the height of the volatility peaks. For example, for October 30, 2008, the GO-GARCH-based estimated volatilities of the Technology returns are 1.31 (ML) and 1.25 (MM), respectively, whereas the univariate GARCH-based volatility estimate is 0.67. Closer inspection reveals that on average, the ML-based volatilities are closer to the univariate GARCH estimates than the MM-based estimates, although the differences are very small; this is most evident in the results for the Technology and Telecom industries.

Note that the 10 industries yield a total of 45 different pairs. There would be little value added to report all 45 estimates of the conditional correlation processes. Pairs whose conditional correlation exhibits considerable variation over time would be most interesting, and would provide a test whether the different estimators identify the same trends and patterns in conditional correlation. The Oil and Gas industry is an example of such an industry. Correlations between Industrials and Oil and Gas for example are seen to fall as low as 0.25 in 2002, and then climb to levels as high as 0.85 in 2009. For many other pairs, especially those with high unconditional correlation, we observe relatively little variation in conditional correlation over time (which leaves little space for the two estimators to potentially disagree on). Figure 3.5 presents estimates for the conditional correlations...
between the Oil and Gas industry and a choice of six other industries. We observe that, although the estimates for $U$ show some difference, MM and ML largely agree on the conditional correlations.

What is apparent is that correlations are relatively low during the first period of higher volatility (between 1999 and 2003), while they climb from around 0.50 to 0.80 during the second period of high volatility in our sample period (between 2008 and 2009). This shows that correlations need not necessarily be higher in periods of higher volatility. What sets these two periods apart is that between 1999 and 2003 the rise in volatility is more gradual in a time when stock prices too are rising (which was the case for all industries except Technology and Telecom; not reported here). The rise in volatility in 2008 is more abrupt, in a time when stock prices started coming down as the financial crisis unfolded. This is

---

**Figure 3.2:** Estimated annualized volatilities: MM, ML, and univariate GARCH.
consistent with earlier findings that correlations tend to increase in times of crisis, which are characterized by periods of extreme volatility combined with falling stock prices (see e.g. Longin and Solnik, 2001; Ang and Bekaert, 2002; Ang and Chen, 2002; Das and Uppal, 2004; and Bekaert et al., 2005).

### 3.6 Conclusion

We have put forward a method-of-moments estimator for the factor loading matrix in GO-GARCH models. The method is based on the common eigenvectors of suitably defined sample autocorrelation matrices of squares and cross-products of the observed data. This means that estimation does not require any Newton-type 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.

In a Monte Carlo experiment, the ML estimator is found to be more efficient than the MM estimator 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. On the other hand, Monte Carlo simulations show that the MM estimator can be more efficient than ML when the latter is based on a misspecified likelihood function. Moreover, the MM estimator is a welcome alternative 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 does not suffer from such problems.