Testing the number of factors in GO-GARCH models

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

Roy van der Weide
Development Research Group (DECRG)
World Bank

February 2008; Preliminary and incomplete.

Abstract

We propose two tests for the number of heteroskedastic factors in a generalized orthogonal GARCH (GO-GARCH) model. The first test is the Gaussian likelihood ratio test, the second is a reduced rank test applied to suitably defined autocovariance matrices. We characterize the asymptotic null distributions of the tests, and compare their finite sample size and power properties to an alternative test proposed by Lanne and Saikkonen (2007).

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). Closely related models were proposed by Vrontos *et al.* (2003) and Fan *et al.* (2008). The starting point of these models is that an observed vector of returns can be expressed as a non-singular linear transformation of latent factors that are conditionally uncorrelated, and that have a GARCH-type conditional variance specification.

In a recent paper, Lanne and Saikkonen (2007) consider a special case of the GO-GARCH model, obtained by the restricting the volatility of a subset of the latent factors to be constant, reducing the model to the factor GARCH model of Engle *et al.* (1990) They consider maximum likelihood estimation of the model under the assumption of (mixed) normality of the standardized innovations of the model. In addition, they propose a test for the number of heteroskedastic factors based on a serial correlation test applied to the squares of the estimated factors that are assumed homoskedastic under the null.

In the present paper we consider two alternative tests for the number of heteroskedastic factors. The first is a likelihood ratio test in the model of Lanne and Saikkonen (2007), which one could expect to have higher power. The second is a test on the rank of a suitably defined weighted average of autocovariance matrices of squares and cross-products, inspired by reduced rank regression and canonical

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

correlation tests. The main advantage of this approach is that it does not require a parametric model for the heteroskedastic factors, and as such can be expected to be more robust than the other tests.

The outline of this paper is as follows. In Section 2, we introduce the model and assumptions. Section 3 defines the likelihood ratio test, and discusses its asymptotic distribution, which is affected by the fact that various parameters are not identified under the null hypothesis. Section 4 defines the new reduced rank test, and proposes a (wild) bootstrap procedure to obtain asymptotically valid p-values. Section 5 concludes with a small-scale Monte Carlo experiment to compare the finite-sample size and power properties of the tests.

2 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, such that $E(x_t|\mathcal{F}_{t-1}) = 0$, and

$$\Sigma_t := \mathrm{var}(x_t|\mathcal{F}_{t-1}) = E(x_t x_t'|\mathcal{F}_{t-1}).$$

The GO-GARCH model imposes a structure on the conditional variance matrix Σ_t, 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 an independent and identically distributed (i.i.d.) vector sequence with $E(\varepsilon_t) = 0$, $\mathrm{var}(\varepsilon_t) = I_m$, and with ε_t independent of $\mathcal{F}_{t-1}$.

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

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

such that 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 $E(H_t) = I_m$. These properties of y_t immediately imply that x_t is covariance-stationary with (conditional) mean zero, conditional variance $\Sigma_t = ZH_tZ'$, and unconditional variance $\Sigma = ZZ'$.

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

$$h_{it} = (1-\alpha_i-\beta_i) + \alpha_i y_{i,t-1}^2 + \beta_i h_{i,t-1}, \qquad \alpha_i,\beta_i \geq 0, \quad \alpha_i+\beta_i < 1, \tag{3}$$

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 (3) is their extended GARCH(1,1) specification:

$$h_{it} = \left(1 - \sum_{j=1}^{m} \alpha_{ij} - \beta_i\right) + \sum_{j=1}^{m} \alpha_{ij} y_{j,t-1}^2 + \beta_i h_{i,t-1}, \qquad \alpha_{ij}, \beta_i \geq 0, \quad \sum_{j=1}^{m} \alpha_{ij} + \beta_i < 1. \quad (4)$$

Intermediate versions, where some of the $\alpha_{ij}, j \neq i$ are restricted to zero, can also be considered. Fan *et al.* (2008) propose to use the Bayesian information criterion to select a submodel defined by $\alpha_{ij} = 0$ for some i and $j \neq i$.

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-varing 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). Therefore, in this paper we consider tests for the hypothesis

$$\mathcal{H}_r : h_{it} = 1, \quad i = r+1, \ldots, m.$$

Note that the ordering of the factors is arbitrary, so the restriction that the *final* $m - r$ components of y_t are homoskedastic is a notational convention. Also, the constant variances are restricted to 1 because we have imposed that the unconditional variance of y_t is given by I_m, which in turn is without loss of generality because the link matrix Z in (1) is unrestricted. The null hypothesis $\mathcal{H}_r$ is tested against the unrestricted GO-GARCH alternative $\mathcal{H}_m$.

Under the parametric assumptions (3) or (4), the null hypothesis corresponds to the parametric restrictions $\alpha_i = \beta_i = 0$ or $\alpha_{ij} = \beta_i = 0$, for $r < i \leq m$ and all j. For the independent GARCH case, Lanne and Saikkonen (2007) test $\mathcal{H}_r$ by testing for multivariate heteroskedasticity in the estimated components $(\hat{y}_{r+1,t}, \ldots, \hat{y}_{mt})$, obtained from estimating the model under $\mathcal{H}_r$. In the next section we consider a likelihood ratio test for the hypothesis, based on either (3) or (4), and the assumption $\varepsilon_t \sim$ i.i.d. $N(0, I_m)$. Next, we will propose a reduced rank test which is not based on such parametric restrictions. Instead, this test will require the following:

Assumption 2 *The process* $\{\mathrm{vech}(y_t y_t')\}_{t \geq 1}$ *is covariance-stationary, with* $\gamma_{i0} := \mathrm{var}(y_{it}^2) < \infty$*, and with autocorrelations* $\rho_{ik} := \mathrm{corr}(y_{it}^2, y_{i,t-k}^2)$ *and cross-covariances* $\tau_{ijk} = \mathrm{cov}(y_{it}^2, y_{i,t-k} y_{j,t-k})$*, satisfying for some integer* p*,*

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

The covariance-stationarity assumption, as well as the assumptions on the (cross-) autocorrelations, 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 such restrictions, this assumption is not without loss of generality. The

non-zero autocorrelation assumption allows us to identify conditional heteroskedasticity in the first r components of y_{it} from the first p autocorrelation coefficients of y_{it}^2, where p will appear as a lag length parameter in the construction of the reduced rank test. 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 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 (4), possibly augmented with $y_{i,t-1}$ and $y_{j,t-1}$ (but not their product) to allow for leverage effects.

We conclude this section with some remarks on parametrization of the model, and an assumption needed for identification. Let $\Sigma = PLP'$ be the spectral decomposition of the Σ, such that $L = \text{diag}(l_1, \ldots, l_m)$ contains its eigenvalues and P contains the corresponding eigenvectors. Since $\Sigma = ZZ'$, the singular value decomposition of Z is given by $Z = PL^{1/2}U$, where U is an orthogonal matrix (containing the eigenvectors of $Z'Z$). It will be convenient to define $s_t = L^{-1/2}P'x_t$, the (unconditionally) standardized returns, such that $s_t = Uy_t$, and hence $\Omega_t := \text{var}(s_t|F_{t-1}) = UH_tU'$ and $\text{var}(s_t) = I_m$. Since information on P and L is directly available from the unconditional variance matrix Σ of the returns, the main task for inference on the loading matrix Z is to identify the orthogonal matrix U. Note that s_t in fact contains the (standardized) principal components of x_t; the O-GARCH model of Alexander (2001) assumes that these are independent GARCH processes, and therefore is obtained as a special case with $U = I_m$ and h_{it} given by (3).

Under $\mathcal{H}_r$, partition $Z = [Z_1 : Z_2]$ and $U = [U_1 : U_2]$, where the $m \times r$ matrices Z_1 and U_1 correspond to the heteroskedastic components in y_t, and the $m \times (m-r)$ matrices Z_2 and U_2 contain the loadings of the homoskedastic factors (with $Z_i = P\Lambda^{1/2}U_i, i = 1, 2$). The matrices P and L are identified from the unconditional variance matrix Σ as before. Using the result

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

and similarly $\Omega_t = U_1 H_{1t} U_1' + U_2 U_2' = I_m + U_1(H_{1t} - I_r)U_1'$, we observe that the matrices Z_2 and U_2 no longer determine the conditional variance of x_t or s_t, so they are not identified (other than by the properties $U_2'U_1 = 0$ and $Z_2 Z_2' = \Sigma - Z_1 Z_1'$, i.e., as functions of the other parameters). For identification of U and hence Z in the unrestricted model, and of U_1 and hence Z_1 in the restricted model, we require the time-varing variances to be distinct:

Assumption 3 *In the model $\mathcal{H}_r$, with $r \leq m$, it is the case that $h_{it} \neq h_{jt}$ for all $1 \leq i < j \leq r$.*

If this assumption is violated, i.e., if $h_{it} = h_{jt}$ (a.s.) for some $i \neq j$, then the corresponding submatrices of Z and U (or Z_1 and U_1) may be multiplied by any orthogonal matrix, and therefore is not identified. This occurs because in that case $\text{var}\,[(y_{it}, y_{jt})'|\mathcal{F}_{t-1}] = h_{it} I_2$, such that $Q(y_{it}, y_{jt})'$ will have the same conditional variance matrix for any orthogonal matrix Q. One example of a violation of Assumption 3 is when $h_{it} = h_{jt} = 1$ for $i \neq j \leq r$, i.e., when the hypothesis $\mathcal{H}_q$ with $q \leq r - 2$ is satisfied, but the model $\mathcal{H}_r$ is analyzed.

Under Assumption 3, the number of identified parameters is determined as follows. In the unrestricted model $\mathcal{H}_m$, Z is an unrestricted $m \times m$ matrix, therefore containing m^2 unknown parameters. These may be separated into $\frac{1}{2}m(m-1)$ and m parameters characterizing the orthogonal matrix P and the diagonal matrix L, respectively, (which together determine the $\frac{1}{2}m(m+1)$ parameters in Σ) and the remaining $\frac{1}{2}m(m-1)$ parameters in U. Under Assumption 3, all m^2 parameters are locally identified, in the sense that Z is observationally equivalent only to matrices Z^* with a permutation of the columns of Z, possibly multiplied by -1 (corresponding to a reordering and sign change of the components of y_t). Exactly the same local identification property applies to the matrix U. In the restricted model $\mathcal{H}_r$, the semi-orthogonal matrix U_1 is characterized by $mr - \frac{1}{2}r(r+1)$ parameters, which are again locally identified under Assumption 3. Therefore, the difference in identified parameters (in the loading matrix Z) in the unrestricted and restricted models is given by $\frac{1}{2}m(m-1) - mr + \frac{1}{2}r(r+1) = \frac{1}{2}(m-r)(m-r-1)$.

3 Likelihood ratio test

In this section we discuss the likelihood ratio test for $\mathcal{H}_r$ against $\mathcal{H}_m$, in a Gaussian GO-GARCH model, i.e., with $\{\varepsilon_t\} \sim$ i.i.d. $N(0, I_m)$. The log-likelihood function therefore is given by

$$\ell(\theta, \Sigma) = -\frac{nm}{2}\log 2\pi - n\log|Z| - \frac{1}{2}\sum_{t=1}^{n}\left(\sum_{i=1}^{m}\log h_{it} + x_t' Z^{-1\prime} H_t^{-1} Z^{-1} x_t\right) \tag{5}$$

$$= -\frac{nm}{2}\log 2\pi - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{t=1}^{n}\sum_{i=1}^{m}\left(\log h_{it} + s_t' U H_t^{-1} U' s_t\right). \tag{6}$$

Here θ contains the GARCH parameters characterizing H_t, together with $\frac{1}{2}m(m-1)$ parameters characterizing U. The first expression is most convenient for full maximization of the log-likelihood over all parameters. However, we may also follow the suggestion of van der Weide (2002) to replace Σ and s_t in (6) by $\hat{\Sigma}$ and $\hat{s}_t$, respectively, where $\hat{\Sigma} = n^{-1}\sum_{t=1}^{n} x_t x_t'$ and $\hat{s}_t = \hat{L}^{-1/2}\hat{P}' x_t$, with $\hat{P}$ and $\hat{L}$ are the matrices of eigenvectors and eigenvalues of $\hat{\Sigma}$, and then maximize the resulting estimated loglikelihood over the remaining parameters θ. This considerably reduces the dimension of the parameter vector to be estimated by maximum likelihood. A parametrization of the orthogonal matrix U in terms of the product of $\frac{1}{2}m(m-1)$ rotation matrices, each parametrized by an angle ϕ_j, is discussed by van der Weide (2002).

Let Θ_m denote the unrestricted parameter space of θ, and let Θ_r be the parameter space implied by the null hypothesis $\mathcal{H}_r$, i.e., with $\alpha_i = \beta_i = 0$ for $i = r+1, \ldots, m$. The likelihood ratio statistic is given by

$$LR_r = -2\left(\max_{\theta\in\Theta_r, \Sigma>0}\ell(\theta,\Sigma) - \max_{\theta\in\Theta_m, \Sigma>0}\ell(\theta,\Sigma)\right).$$

As discussed by van der Weide (2002) and Lanne and Saikkonen (2007), the GO-GARCH model is a special case of the BEKK model of Engle and Kroner (1995), and as such the general results of Comte and Lieberman (2003) concerning consistency and asymptotic normality of maximum likelihood estimators can be directly applied. Asymptotic normality of the restricted and unrestricted maximum likelihood estimators of (θ, Σ) requires the model to be a smooth submodel of the BEKK model, with

identified parameters. Under the null hypothesis $\mathcal{H}_r$ and Assumption 3, the restricted model does indeed have identified parameters. However, in the unrestricted model some parameters are not identified under $\mathcal{H}_r$: the matrix U_2 and hence Z_2 is only identified up to an orthogonal rotation, and furthermore the parameters β_i, $i = r+1, \ldots, m$ in (3) are not identified because $\alpha_i = 0$ implies $h_{it} = 1$ for all values of $0 \leq \beta_i < 1$. The latter identification problem is similar to the lack of identification in ARMA models with common AR and MA roots.

These proporties put this testing problem in the framework of likelihood ratio testing with nuisance parameters that are identified only under the alternative, as analyzed, e.g., by Andrews and Ploberger (1994). Their asymptotic analysis may be used to characterize the asymptotic null distribution of LR_r, and in practice simulation methods may be used to obtain approximate p-values. In the specific case of $r = m - 1$, where the test biols down to an LR test for $\alpha_m = \beta_m = 0$ in the GARCH(1,1) model for y_{mt}, the limiting distribution is as given by Andrews and Ploberger (1996), and the likelihood ratio statistic can be compared with the critical values as given in their Table 1. We will consider this special case in the Monte Carlo simulation study in Section 5.

4 Reduced rank test

Consider the autocovariance matrices

$$\Gamma_k(y) := E\left[(y_t y_t' - I_m)(y_{t-k} y_{t-k}' - I_m)\right], \qquad k = 1, \ldots, p. \tag{7}$$

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 $y_t y_t'$), 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

$$\Gamma_k(y)_{ij} = \mathrm{cov}(y_{it}^2, y_{i,t-k} y_{j,t-k}) = \begin{cases} \gamma_{i0}\rho_{ik} & , \qquad j = i, \\ \tau_{ijk} = & 0, \qquad j \neq i, \end{cases}$$

or in other words, $\Gamma_k(y)_{ij} = \mathrm{diag}(\gamma_{10}\rho_{1k}, \ldots, \gamma_{m0}\rho_{mk})$.

The corresponding autocovariance matrices for the process $s_t = L^{-1/2} P' x_t = U y_t$ satisfy, using the orthogonality of U,

$$\Gamma_k(s) = E\left[(s_t s_t' - I_m)(s_{t-k} s_{t-k}' - I_m)\right] = U \Gamma_k(y) U', \qquad k = 1, \ldots, p. \tag{8}$$

Since $\Gamma_k(y)$ is a diagonal matrix and U is orthogonal, it follows that under Assumptions 1 and 2 $\Gamma_k(y)$ contains the eigenvalues and U contains the eigenvectors of $\Gamma_k(s)$. Because these eigenvectors do not vary with k, the same result applies to a weighted average of the $\Gamma_k(s)$ matrices, with non-negative weights $\{w_1, \ldots, w_p\}$ summing to one. Defining

$$\mu_i = \gamma_{i0} \sum_{k=1}^p w_k \rho_{ik}, \qquad i = 1, \ldots, m, \tag{9}$$

we find

$$\Gamma_w(s) := \sum_{k=1}^p w_k \Gamma_k(s) = U \left(\sum_{k=1}^p w_k \Gamma_k(y) \right) U' = U \, \mathrm{diag}(\mu_1, \ldots, \mu_m) U, \tag{10}$$

Under the null hypothesis $\mathcal{H}_r$, we find $\rho_{ik} = 0$ for $i = r+1, \ldots, m$ and $k = 1, \ldots, p$, such that $\mu_{r+1} = \ldots = \mu_m = 0$, i.e., the final $m - r$ eigenvalues of $\Gamma_w(s)$ are zero, and

$$\Gamma_w(s) = U_1 \operatorname{diag}(\mu_1, \ldots, \mu_r) U_1'.$$

In other words, we may interpret $\mathcal{H}_r$ as a *reduced rank* hypothesis on $\Gamma_w(s)$.

These results suggest to use eigenvalues of the sample analog of $\Gamma_k(s)$ or its weighted average $\Gamma_w(s)$ (or the eigenvalues of suitably defined symmetric versions of these matrices) as a basis for a test for $\mathcal{H}_r$ against $\mathcal{H}_m$. Therefore, consider

$$\begin{aligned}
\hat{\Gamma}_k(\hat{s}) &= \frac{1}{n} \sum_{t=1}^{n} (\hat{s}_t \hat{s}_t' - I_m)(\hat{s}_{t-k} \hat{s}_{t-k}' - I_m), \\
\hat{\Gamma}_w(\hat{s}) &= \sum_{k=1}^{p} w_k \hat{\Gamma}_k(\hat{s}) = \frac{1}{n} \sum_{t=1}^{n} (\hat{s}_t \hat{s}_t' - I_m) \sum_{k=1}^{p} w_k (\hat{s}_{t-k} \hat{s}_{t-k}' - I_m),
\end{aligned}$$

where the pre-sample values $\{\hat{s}_t \hat{s}_t' - I_m\}_{t \leq 0}$ are set to zero.

The main argument for considering the weighted average $\hat{\Gamma}_w(\hat{s})$ is that it allows us to pool the information in the separate $\hat{\Gamma}_k(\hat{s})$ matrices, which should yield more powerful tests. The power of the test is essentially its ability to distinguish the non-zero eigenvalues $(\mu_1, \ldots, \mu_{r_0})$ from the zero eigenvalues $(\mu_{r_0+1}, \ldots, \mu_m)$, where r_0 is the true rank of $\Gamma_w(s)$. We conjecture that for this purpose, p and $\{w_k\}_{k=1}^p$ should maximize $\operatorname{corr}\left(y_{it}^2, \sum_{k=1}^p w_k y_{i,t-k}^2\right)$, or its minimum over $i = 1, \ldots, r_0$. Given the autocorrelation structure in empirically observed squared returns, a fairly robust choice would be an exponentially weighted moving average $w_k = w_1 \nu^{k-1}$ with ν close to one (a common choice for daily financial returns is $\nu = 0.94$) and p chosen such that $\sum_{k=p+1}^{\infty} \nu^k = \nu^{p+1}/(1-\nu)$ is small relative to $\sum_{k=1}^{\infty} \nu^k = \nu/(1-\nu)$.

For any choice of weights, define $Y_t = (s_t s_t' - I_m)$, $X_t = \sum_{k=1}^p w_k (s_{t-k} s_{t-k}' - I_m)$, and let $\hat{Y}_t$ and $\hat{X}_t$ denote the corresponding quantities with s_t replaced by $\hat{s}_t$. This implies that we now wish to develop a test for the rank of the matrix $\hat{\Gamma}_w(\hat{s}) = n^{-1} \sum_{t=1}^n \hat{Y}_t \hat{X}_t =: S_{\hat{Y}\hat{X}}$. Although this is not a standard reduced-rank regression problem, because $\hat{Y}_t$ and $\hat{X}_t$ are (estimated) matrices instead of (observed) vectors, a natural test statistic to consider is the reduced-rank regression (or canonical correlation) test statistic (see Anderson (1951))

$$Q_r = -n \sum_{i=r+1}^{m} \log(1 - \hat{\lambda}_i), \tag{11}$$

where $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_m \geq 0$ are the descending eigenvalues of the generalized eigenvalue problem

$$\left| \lambda S_{\hat{Y}\hat{Y}} - S_{\hat{Y}\hat{X}} S_{\hat{X}\hat{X}}^{-1} S_{\hat{X}\hat{Y}} \right| = 0, \tag{12}$$

with $S_{\hat{Y}\hat{Y}} = n^{-1} \sum_{t=1}^n \hat{Y}_t^2$, $S_{\hat{X}\hat{X}} = n^{-1} \sum_{t=1}^n \hat{X}_t^2$, and $S_{\hat{X}\hat{Y}} = S_{\hat{Y}\hat{X}}' = n^{-1} \sum_{t=1}^n \hat{X}_t \hat{Y}_t$.

The standard proof that the reduced rank test statistic Q_r has an asymptotic $\chi^2((m-r)^2)$ distribution under the null hypothesis is obtained as follows. First, it is shown that

$$Q_r = n \operatorname{tr}\left\{ (U_2' S_{\hat{Y}\hat{Y}} U_2)^{-1} U_2' S_{\hat{Y}\hat{X}} U_2 (U_2' S_{\hat{X}\hat{X}} U_2)^{-1} U_2' S_{\hat{X}\hat{Y}} U_2) \right\} + o_p(1).$$

Next, it is shown that $\sqrt{n}U_2'S_{\hat{X}\hat{Y}}U_2$ has a limiting normal distribution with mean zero, and with a variance matrix that is consistently estimated by $(U_2'S_{\hat{Y}\hat{Y}}U_2)\otimes(U_2'S_{\hat{X}\hat{X}}U_2)$. In the present case, the second step cannot be proved, essentially because even under the null hypothesis, $U_2'Y_t^2U_2$ is correlated with $U_2'X_t^2U_2$. The same problem would arise in reduced rank regressions with conditional heteroskedasticity. The consequence is that $\sqrt{n}U_2'S_{\hat{X}\hat{Y}}U_2$ will still have a limiting normal distribution with mean zero, but its conditional variance matrix cannot be written as a suitably defined kronecker product. This in turn implies that the limiting distribution of Q_r, based on an inconsistent estimate of the variance matrix of $\sqrt{n}U_2'S_{\hat{X}\hat{Y}}U_2$, will not be $\chi^2((m-r)^2)$ but the distribution of a weighted sum of independent $\chi^2(1)$ random variables.

As a solution to this problem, we propose a bootstrapping approach, obtained by combining a wild bootstrap resampling scheme for the heteroskedastic factors (see Liu (1988) and Mammen (1993)) with an ordinary resampling scheme for the homoskedastic factors. Specifically, let $\hat{U}$ denote the matrix of eigenvectors of the eigenvalue problem (12), and let $\hat{y}_t = \hat{U}'\hat{s}_t = (\hat{y}_{1t}', \hat{y}_{2t}')'$, where $\hat{y}_{1t}$ contains the first r components of $\hat{y}_t$, and $\hat{y}_{2t}$ the remaining $m-r$ components. The proposed procedure is to obtain B bootstrap replications $\{Q_r^{(i)}, i = 1,\ldots,B\}$ of the test statistic, based on resampled observations $\{y_t^{(i)} = (y_{1t}^{(i)\prime}, y_{2t}^{(i)\prime})', t = 1,\ldots,n, i = 1,\ldots,B\}$, where the components of $y_{1t}^{(i)}$ are given by the corresponding components of $\hat{y}_{1t}$, multiplied by a random sign change, whereas $\{y_{2t}^{(i)}, t = 1,\ldots,n\}$ is obtained by taking n random draws (with replacement) from $\{\hat{y}_{21},\ldots,\hat{y}_{2n}\}$. The wild bootstrap preserves the volatility clustering in y_{1t}, whereas the ordinary (non-parametric) bootstrap preserves the marginal distribution of y_{2t}, while forcing these to be i.i.d. The bootstrap p-value of Q_r is now simply defined as $1 - F_B(Q_r)$, where $F_B(x)$ is the bootstrap null distribution $F_B(x) = \#(Q_r^{(i)} < x)/B$. This is an estimate of the asymptotic p-value $1 - F(Q_r)$, where $F(x)$ is the asymptotic null distribution $F(x) = \lim_{n\to\infty} P(Q_r < x|\mathcal{H}_r)$.

Consistency of such bootstrap p-values builds on consistency of $\hat{U}_1$, containing the first r rows of $\hat{U}$. For this purpose, we need a slightly stronger version of Assumption 3:

Assumption 4 *In the model $\mathcal{H}_r$, with $r \leq m$, the r largest eigenvalues $(\mu_1,\ldots,\mu_r)$ of $\Gamma_w(s)$ are distinct and non-zero.*

Theorem 1 *Under Assumptions 1–2 and 4, and as $n \to \infty$,*

$$F_B(\cdot) \Rightarrow F(\cdot),$$

where "$\Rightarrow$" denotes weak convergence. Under the alternative hypothesis $\mathcal{H}_m\backslash\mathcal{H}_r$, the test statistic Q_r diverges, such that the test is consistent.

5 Monte Carlo simulations

In this section we study the finite-sample size and power performance of the tests proposed in this paper, as well as a version of the test proposed by Lanne and Saikkonen (2007), in a small-scale Monte Carlo

experiment. We consider a data-generating process with $m = 2$,

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \qquad U = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix},$$

with $\rho = 0.4$ and $\phi = \pi/6$. The components y_{1t} and y_{2t} are generated by independent Gaussian GARCH(1,1) processes, with $(\alpha_1, \beta_1) = (0.1, 0.8)$, and with $(\alpha_2, \beta_2) = (0, 0)$ under the null hypothesis $r = 1$, whereas $(\alpha_2, \beta_2) = (0.1, 0.4)$ under the alternative $r = 2$. The sample size is set to $n = 2000$.

The likelihood ratio test statistic is compared with the 5% asymptotic critical value 6.01 obtained from Table 1 of Andrews and Ploberger (1996). The reduced rank test is constructed based on $p = 10$ lags and exponential weights $w_k = w_0 0.94^k$, $k = 1, \ldots, 10$, where $w_0 = (1 - 0.94)/(1 - 0.94^{10})$. The number of bootstrap replications is set to $B = 99$. The implementation of the Lanne and Saikkonen (2007) test we consider is a Q-statistic with 10 lags applied to $\tilde{y}_{2t}^2$, where $\tilde{y}_{2t} = \tilde{U}_2' \hat{s}_t$, with $\tilde{U}$ the maximum likelihood estimator under the null hypothesis. The results in Table 1 are based on 5000 replications.

Table 1: Size ($r = 1$) and size-corrected power ($r = 2$) of the likelihood ratio (LR), reduced rank (Q) and Lanne-Saikkonen (LS) tests for $r = 1$.

	LR	Q	LS
$r = 1$	0.075	0.053	0.054
$r = 2$	0.959	0.667	0.899

We observe that in the present DGP all tests have reasonable size, although the LR test has displays some over-rejection. Possibly better results can be obtained by bootstrapping its null distribution, or by modelling y_{2t} as a higher-order ARCH instead of a GARCH(1,1) process (as this would avoid the idenfication problems with the GARCH(1,1) model with $\alpha_2 = \beta_2 = 0$).

Concerning the power, we observe that the LR and LS have a very high and comparable power against the alternative considered here. The power of the reduced-rank test is clearly less but still substantial. This power loss may be seen as a price to be paid for the fact that the test is based on less parametric restrictions. In case of misspecification of the GARCH components, we would expect the LR and LS statistics to display bigger size distortions, and our Q_r test to gain power relatively to the other two. We intend to investigate this in a more extensive Monte Carlo study. Furthermore, the distinct advantage of the Q_r test is that it can be used to test for the number of heteroskedastic factors before a parametric model for these factors needs to be specified and estimated, and as such avoids the need to estimate partially unidentified models.

Appendix

Proof of Theorem 1 to be completed.

References

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

Anderson, T. W. (1951), "Estimating linear restrictions on regression coefficients for multivariate normal distributions", *Annals of Mathematical Statistics*, 22, 327–351.

Andrews, D. W. K. and W. Ploberger (1996), "Optimal tests when a nuisance parameter is present only under the alternative", *Econometrica*, 62, 1383–1414.

Andrews, D. W. K. and W. Ploberger (1996), "Testing for serial correlation against an ARMA(1,1) process", *Journal of the American Statistical Association*, 91, 1331–1342.

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

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

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

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 Tor 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*, forthcoming, `http://stats.lse.ac.uk/q.yao/qyao.links/paper/fwy05.pdf`.

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.

Liu, R. Y. (1988), "Bootstrap procedures under some non-I.I.D. models", *Annals of Statistics*, 16, 1696–1708.

Mammen, E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models", *Annals of Statistics*, 21, 255–285.

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.