CORE DISCUSSION PAPER

2006/71

Asymptotic theory for a factor GARCH model

Christian M. Hafner[1] Arie Preminger[2]

Abstract

This paper investigates the asymptotic theory for a factor GARCH model. Sufficient conditions for strict stationarity, existence of certain moments, geometric ergodicity and β- mixing with exponential decay rates are established. These conditions allow for volatility spill-over and integrated GARCH. We then show the strong consistency and asymptotic normality of the quasi-maximum likelihood estimator (QMLE) of the model parameters. The results are obtained under the finiteness of the fourth order moment of the innovations.

Keywords: Multivariate GARCH, factor model, geometric ergodicity, maximum likelihood, consistency, asymptotic normality.
JEL Classification: C14, C22.

[1]Institut de statistique and CORE, Université catholique de Louvain, Voie du Roman Pays 20, B-1348. Louvain-la-Neuve, Belgium, hafner@stat.ucl.ac.be.

[2]CORE, Université catholique de Louvain, Voie du Roman Pays 34, B-1348, Louvain-la-Neuve, Belgium, preminger@core.ucl.ac.be.

1 Introduction

Multivariate GARCH models are becoming more and more popular with the advance of computing power, in particular to model the dynamic volatilities and correlations of financial time series. Many different specifications have been proposed recently; see e.g. the survey by Bauwens et al. (2006). Some of the proposed models such as the Vec or the BEKK (see Bollerslev et al. (1988) and Engle and Kroner (1995)) suffer from the curse of dimensionality in the sense that as the dimension increases, the number of parameters explodes and it becomes unfeasible to estimate these models in high dimensions. To keep the number of parameters under control, many restricted versions of multivariate GARCH models have been proposed, of which factor GARCH models are an example. In this paper we investigate the asymptotic theory for a factor GARCH model which nests some models proposed in the literature. The factors are assumed to be either directly observed or to be linear functions of the underlying variables. Examples for models that are nested in our framework are the orthogonal GARCH model of Alexander (2001), the generalized orthogonal GARCH model of van der Weide (2002), and the full factor GARCH model of Vrontos et al. (2003). The factors are assumed to be conditionally orthogonal, but we allow for Granger causality in variances.

The most commonly employed approach for estimation is based on a quasi-maximum likelihood estimator (QMLE) in which the innovation process is characterized by a multivariate normal distribution. Bollerslev and Wooldridge (1992) show that the QMLE is consistent and asymptotically normal, even if the true DGP is not conditionally normal. For the univariate GARCH(1,1) case, asymptotic theory for a "local" QMLE[1] has been developed by Lee and Hansen (1994) and Lumsdaine (1996). Their results allow for integrated or even mildly explosive GARCH process where the existence of moments of order four or higher of the innovations has been assumed. Under similar assumptions, the case of GARCH(p, q) was considered by Boussama (2000), Berkes et al. (2003) and Francq and Zakoian (2004). Asymptotic theory for multivariate GARCH models is not

[1]Their estimates are "local" in the sense that the likelihood function is maximized in a restricted neighbourhood of true parameter values.

yet established generally. Gourieroux (1997) provides theory for a general formulation using high level assumptions, and Comte and Lieberman (2003) for the BEKK model (see also Ling and McAleer, 2003) under the assumption that the eighth order moments of the observed process exist which rules out the possibility of integrated GARCH and can be inappropriate in some empirical applications.

The main part of this paper investigates the asymptotic theory of the QMLE for the factor GARCH model under mild regularity conditions which allow for integrated processes. Consistency and asymptotic normality are proved under the existence of moments of order two and four of the innovations, respectively. No conditions on the shape of the innovation distribution are required other than the existence of moment conditions.

A key ingredient in developing asymptotic theory for QMLE is the existence of a stationary and ergodic solution for the underlying GARCH process. Bougerol and Picard (1992b) give necessary and sufficient conditions for strict stationarity and ergodicity of a univariate GARCH(p, q) model in terms of the top Lyapounov exponent. Their results generalize Nelson's (1990) results for the univariate GARCH(1,1) model. Unfortunately, the approach of Bougerol and Picard (1992b) cannot be extended to the multivariate case in general. Boussama (1998) gave a counterexample to this extent. For the multivariate GARCH model, sufficient conditions for geometric ergodicity were given by Boussama (1998) and Kristensen (2005). Ling and McAleer (2003) provide sufficient conditions under which the CCC model of Bollerslev (1990), has a strictly stationary solution. All of these results were established under conditions for covariance stationarity for the multivariate GARCH model (see e.g. Engle and Kroner, 1995).

We provide sufficient conditions for strict stationarity, geometric ergodicity, regular mixing with exponential decay rates and existence of higher order moments for the observations. The existence of moments and the mixing property allow verification of how the model can fit stylized facts such as the fat tails and the temporal persistence observed in financial data. Our conditions allow for integrated or even mildly explosive factor GARCH processes. For the univariate GARCH case, similar results were obtained by Carrasco and Chen (2002), Meitz and Saikkonen (2004) and Francq and Zakoian (2005). These results are useful in establishing limit theorems for the processes and for the proofs of our asymptotic theory.

The plan of this paper is as follows. In the next section the factor GARCH model is introduced. In Section 3 we investigate its properties. In Section 4 we establish the consistency and asymptotic normality of the QMLE. Concluding remarks are offered in Section 6. All proofs are given in the Appendix.

Throughout the paper we use $\|\cdot\|$ as a matrix operator norm induced by some vector norm. As we use the Euclidean vector norm, $\|\cdot\|$ is the spectral norm, i.e. $\|A\|$ =max $\{\sqrt{\lambda}$:λ is eigenvalue of $AA'\}$. $O(1)$ (or $o(1)$) denotes a series of nonstochastic variables that are bounded (or converge to zero); $O_{\mathrm{P}}(1)$ (or $o_{\mathrm{P}}(1)$) denotes a series of random variables that are bounded (or converge to zero) in probability. The symbol$\rightarrow_{a.s.}$ ($\rightarrow_{\mathrm{D}}$) denotes convergence almost surely (or in distribution).

2 The model

Let $\{y_t\}, y_t \in \mathbb{R}^N, t \in \mathbb{Z}$ be a stochastic process. We consider the following factor model.

$$y_t = W f_t \tag{1}$$

$$f_t = \Sigma_t^{1/2} \xi_t \tag{2}$$

where $\xi_t \sim i.i.d(0, I_N)$. The factors f_t are conditionally heteroskedastic with conditional covariance matrix Σ_t. The loading matrix W is of dimension $N \times N$ and of full rank. The model implies that the conditional covariance matrix of y_t is given by

$$H_t = W \Sigma_t W'. \tag{3}$$

This model can be considered as a full factor model in the terminology of Vrontos et al. (2003), since the factor space has the same dimension as the process space, and there is no idiosyncratic noise. It includes as special cases e.g. the popular Orthogonal GARCH model of Alexander (2001), the GO-GARCH model of van der Weide (2002) and the full factor model of Vrontos et al. (2003). We assume that the factors are conditionally orthogonal, i.e., Σ_t is diagonal, but allow for Granger causality in the variances. The conditional factor variances are modeled as

$$\sigma_{it}^2 = 1 + \sum_{j=1}^{N} \alpha_{ij} f_{j,t-1}^2 + \sum_{j=1}^{N} \beta_{ij} \sigma_{j,t-1}^2, \quad i = 1, \ldots, N \tag{4}$$

or in vector form

$$\sigma_t^2 = \iota + A f_{t-1}^2 + B \sigma_{t-1}^2 \tag{5}$$

with $\iota = (1, \ldots, 1)'$ an $N \times 1$ vector and $A = (\alpha_{ij})$, $B = (\beta_{ij})$ being $N \times N$ coefficient matrices such that $\sigma_t^2 > 0$ a.s. The restriction that the constant term in (4) equals 1 is an identification restriction. Defining $c_t = \iota + A f_{t-1}^2$, we can write

$$\sigma_t^2 = c_t + B \sigma_{t-1}^2 = \sum_{j=0}^{\infty} B^j c_{t-j} \tag{6}$$

which will be used later. If A and B are diagonal, a standard univariate GARCH(1,1) model results for each of the factor variances. Note that under diagonal Σ_t, (2) is equivalent to $f_{it} = \sigma_{it} \xi_{it}, \;\; i = 1, \ldots, N$. Thus, we can write (5) as

$$\sigma_t^2 = \iota + (A \Xi_t + B) \sigma_{t-1}^2 \tag{7}$$

where $\Xi_t = diag(\xi_{1t}^2, \ldots, \xi_{Nt}^2)$, or

$$\sigma_t^2 = \iota + C_t \sigma_{t-1}^2 = \iota + \sum_{j=0}^{\infty} C_t C_{t-1} \cdots C_{t-j} \; \iota \tag{8}$$

with $C_t = A \Xi_t + B$ being a random i.i.d. coefficient matrix.

Note that model (1) is a special case of the Vec model of Bollerslev et al. (1988). To see this, we can write $h_t = \text{vec}(H_t)$ as

$$h_t = \omega + \tilde{A} \text{vec}(y_{t-1} y_{t-1}') + \tilde{B} h_{t-1}$$

where[2] $\omega = (W \otimes W) P_N \iota$, $\tilde{A} = (W \otimes W) P_N A P_N' (W \otimes W)^{-1}$ and $\tilde{B} = (W \otimes W) P_N B P_N' (W \otimes W)^{-1}$. Furthermore, if there is no spill-over, that is, $A = \text{diag}(\alpha_1, \ldots, \alpha_N)$ and $B =$

[2]The $N^2 \times N$ matrix P_N is defined by the property $P_N' \text{vec}(A) = dg(A)$ for any $N \times N$ matrix A, where the operator $dg(\cdot)$ transforms the diagonal of a matrix into a column vector.

$\mathrm{diag}(\beta_1, \ldots, \beta_N)$, then the model is a special case of the BEKK$(1, 1, N)$ model, since it can be written as

$$H_t = WW' + \sum_{i=1}^{N} A_i y_{t-1} y_{t-1}' A_i' + B_i H_{t-1} B_i',$$

where $A_i = \sqrt{\alpha_i} w_i \gamma_i'$, $B_i = \sqrt{\beta_i} w_i \gamma_i'$, γ_i' is the i-th row of W^{-1} and w_i is the i-th column of W. Note that in the BEKK model the matrix W is restricted to be triangular, see Engle and Kroner (1995).

3 Geometric ergodicity

The factor GARCH model can be expressed as a Markov chain. In what follows, we provide sufficient conditions which allow us to establish not only that a stationary solution exists for $\{y_t\}$, but also that the Markov chain, irrespective of its initialisation will converge in the total variation norm towards it. The key concept to establish this type of asymptotic stability is geometric ergodicity. These conditions involve the notion of the top Lyapounov exponent for a sequence $\{C_t\}$ of i.i.d. $N \times N$ matrices which is defined as

$$\gamma = \inf_{n \in \mathrm{N}} \left\{ \tfrac{1}{n} \mathrm{E} \log \|C_n \cdots C_1\| \right\}.$$

If $\mathrm{E} \log^+ \|C_1\| < \infty$, where $\log^+ x = \max(0, \log(x))$, then an application of the subadditive ergodic theorem (see Kingman (1973)) yields that

$$\gamma = \lim_{n \to \infty} \tfrac{1}{n} \mathrm{E} \log \|C_n \cdots C_1\| \qquad a.s.$$

The definition of γ does not depend on the choice of the norm that is used and its value cannot be calculated explicitly for the model under study. However, the top Lyapounov exponent can be determined via Monte Carlo simulations of the random matrices C_n. The work by Goldsheid (1991) even allows us to give asymptotic confidence bands through a central limit theorem (CLT). Next, we give sufficient conditions for the process to be geometrically ergodic.

Assumption 3.1 *The sequence* $\{\xi_t\}$ *is i.i.d. on* $\mathbb{R}^N$. *The distribution of* ξ_t *has a positive density that is lower semicontinuous w.r.t. the Lebesgue measure*[3]. *The initial condition* σ_0^2 *is independent of* $\{\xi_t\}$.

Assumption 3.2 *The top Lyapounov exponent associated with* $\{C_t\}$ *is strictly negative.*

Assumption 3.3 $\mathrm{E}|\xi_t|^s < \infty$ *for some* $s > 0$.

Assumption 3.4 *For some* $r \geq 1$, $\mathrm{E}|\xi_t|^r < \infty$ *and* $\mathrm{E}\,\|C_t\|^r < 1$.

The first assumption is satisfied for a wide range of density functions for the innovation process such as the multivariate Gaussian and Student densities. The i.i.d. assumption on the error terms makes $\{y_t\}$ a Markov chain. The second assumption is similar to the top Lyapunov exponent condition imposed by Bougerol and Picard (1992b) in order to establish the existence of a stationary solution for the univariate GARCH model. However, in our set-up this condition not only induces stationarity but also geometric ergodicity. Along with this result we are also able to verify the existence of the r-th order moment. Based on Feigin and Tweedie (1985) the next theorem is obtained.

Theorem 1 *Assume that Assumption 3.1 holds.*

a) Under Assumptions 3.2 and 3.3, the process $\{y_t\}$ *is geometrically ergodic and* $\mathrm{E}|y_t|^r < \infty$ *for some* $r \in (0, s)$.

b) Under Assumption 3.4, the result of 1.a) holds and $\mathrm{E}|y_t|^r < \infty$ *for* $r \geq 1$.

Theorem 1.a) implies that if the process starts from its stationary distribution or in the infinite past $\{y_t\}$ is strictly stationary and ergodic. This result will be used in the next section to show that the QMLE of the factor GARCH model is strongly consistent

[3] A function $f : \mathrm{R}^N \to \mathrm{R}$ is lower semicontinuous if $\liminf_{x \to y} f(x) \geq f(y)$ for all y. The Vitali-Caratheodory Theorem implies that any function in L^1 can be approximated in that space by lower semicontinuous functions, and hence any distribution that possesses a density may be approximated in total variation norm by a distribution satisfying Assumption 3.1.

and asymptotically normal. Next, note that a sufficient condition for strict stationarity of $\{y_t\}$ is given by

$$\mathrm{E}\log\|C_t\| < 0 \tag{9}$$

To show that condition (9) allows for integrated factor GARCH, consider the simple case $N = 2$, $B = 0$ and $A = I_2$. By simulation we find that $\mathrm{E}\log\|C_t\| = \mathrm{E}\log(\max(\xi_{1t}^2, \xi_{2t}^2)) = -0.1045 < 0$.

A particular consequence of geometric ergodicity is that the Markov chain, for any initial distribution, converges in total variation to its stationary measure with a uniform geometric rate. This implies that the process is β-mixing with geometric rate (see e.g. Meyn and Tweedie (1993), Ch.16). Note that besides asymptotic stability of the model, existence of higher order moments is established in Theorem 1.b). This can be useful in providing the law of large numbers or the CLT for the process, regardless of the initial distribution, see e.g. Jones (2004). Furthermore, these limit theorems, and the existence of analytical expressions for certain moments of data such as the second and fourth moments can be exploited for a generalized method of moments (GMM) type estimation of the model parameters.

For example, under Assumptions 3.1 and 3.4 for $r \geq 2$ (which implies that all eigenvalues of the matrix $A + B$ have a modulus smaller than one), the unconditional covariance matrix is given by $W\mathrm{E}(\Sigma_t)W'$, where $\mathrm{E}(\Sigma_t) = diag(\sigma_1^2, \ldots, \sigma_N^2)$ and the vector $\sigma^2 = (\sigma_1^2, \ldots, \sigma_N^2)'$ is given by $\sigma^2 = (I_N - A - B)^{-1}\iota$, see also Engle and Kroner (1995). For the existence of fourth moments of y_t and the auto- and cross-correlation of $\eta_t = f_t \odot f_t$ we can use Assumption 3.4 with $r \geq 4$ (which implies that all eigenvalues of the matrix $Z = (A \otimes A)M_\xi + A \otimes B + B \otimes A + B \otimes B$ have a modulus smaller than one). Let $M_\xi = \mathrm{E}[\Xi_t \otimes \Xi_t]$, $\Sigma_\eta = \mathrm{E}(\eta_t\eta_t')$, $\Sigma_\sigma = \mathrm{E}(\sigma_t^2\sigma_t^{2\prime})$. Similar to Hafner (2003), based on the VARMA representation we can show that the fourth moments of f_t and the τ-autocorrelation of η_t, $\Gamma(\tau) = \mathrm{E}[(\eta_t - \sigma^2)(\eta_{t-\tau} - \sigma^2)]$, are given by

$$\mathrm{E}[vec(\Sigma_\eta)] = M_\xi(I_{N^2} - Z)^{-1}vec\{\iota\iota' + (A + B)\sigma^2\iota' + \iota\sigma^{2\prime}(A + B)'\}$$

and

$$\Gamma(\tau) = (A + B)^{\tau-1}(A\Sigma_\eta + B\Sigma_\sigma) - (A + B)^\tau\sigma^2\sigma^{2\prime}, \quad \tau \geq 1,$$

respectively. Expressions for the fourth moments of y_t and the crosscorrelation of $y_t \odot y_t$ can also be derived. For the univariate GARCH model, GMM type estimators based on the ARMA representation of the squared GARCH process also have been proposed, see e.g. Baillie and Chung (2001), Kristensen and Linton (2006) and Storti (2006). The development of such an estimate in the multivariate context is however beyond the scope of this paper. Note that these estimators are based on the existence of higher order moments of the observed process, which might be too restrictive in financial applications. Further, similar to Bougerol and Picard (1992a) and Berkes et al. (2003) we can also cover the case that (ξ_t) is a strictly stationary martingale difference sequence, showing the existence of a stationary solution and r-th order moment for the process. However, this result is weaker than the geometric ergodicity as discussed above and for simplicity, in what follows, the i.i.d. assumption is used.

4 Quasi maximum likelihood estimation

We now consider estimation by maximum likelihood assuming that the innovation distributions are Gaussian. This need not be true, but under quite general conditions we know that it provides consistent estimates, see e.g. Bollerslev and Wooldridge (1992), and is commonly called quasi maximum likelihood estimation. The log likelihood, up to an additive constant, for a sample of n observations takes the form

$$L_n(\theta) = -\tfrac{1}{2}\sum_{t=1}^{n} (\log\det(H_t(\theta) + y_t' H_t^{-1} y_t) = \sum_{t=1}^{n} l_t(\theta). \tag{10}$$

where the starting value H_1 is a fixed matrix. Define the QMLE as $\hat{\theta}_n = \arg\max_{\theta\in\Theta} L_n(\theta)$. Let $\tilde{H}_t$ denote the covariance process where the starting value is drawn from its stationary distribution, and let $\tilde{L}_n$, $\tilde{l}_t$, $\tilde{\Sigma}_t$ and $\tilde{\sigma}_t^2$ be accordingly. These terms will be used in the proofs. Note that in practice the use of these values is not possible.

Let us decompose the parameter vector as $\theta = (\theta_1', \theta_2', \theta_3')$, where $\theta_1 = vec(W)$, $\theta_2 = vec(A)$, $\theta_3 = vec(B)$ and assume that $\theta \in \Theta \subset \mathrm{R}^p$. Furthermore, denote the true parameter vector by θ_0. To simplify the proofs and in order to ensure that $\sigma_t^2 > 0$ a.s., we assume that all elements of A and B are nonnegative. This assumption is also used by

Ling and McAleer (2003)[4]. It is worth noting that we allow for the possibility that the process is a pure ARCH process, i.e. $B = 0$. To identify the model in case of conditional homoskedasticity (i.e. $A = 0$), we restrict $B = 0$ and impose appropriate restrictions such as triangularity on W. To show strong consistency the following assumptions are made.

Assumption 4.1 *The parameter space* Θ *is compact and* $\|B\| < 1$.

Assumption 4.2 *Assumption 3.2 holds for* $\theta = \theta_0$.

Assumption 4.3 *Assumption 3.1 holds,* $\mathrm{E}|\xi_t|^2 < \infty$ *(*$var(\xi_t) = I_N$*).*

Assumption 4.4 *The model is identifiable, thus* $\forall \theta, \theta_0 \in \Theta$, $H_{t,\theta} = H_{t,\theta_0}$ *a.s. then* $\theta = \theta_0$.

Under Assumptions 4.2-4.3 and Theorem 1 we have that the factor GARCH process converges to a strictly stationary and ergodic solution with some fractional moment. As was shown in the previous section, these assumptions allow for integrated or even mildly explosive ($||A + B|| > 1$) GARCH process for the factors. Assumption 4.3 strengthens Assumption 3.3, by requiring the existence of second order moment for the innovations. The requirement that the variance of ξ_t equals the identity matrix is made to ensure identifiability and it is not restrictive since $\mathrm{E}|\xi_t|^2 < \infty$. Assumption 4.4 is a high level identification condition. For this assumption we can apply the identification results in Engle and Kroner (1995) in the special case of no spill-over. Other primitive conditions can be found in Jeantheau (1998).

The method underlying the proofs basically consists of two main stages. In the first stage it is assumed that the process is initiated from its stationary distribution and we establish the finiteness of various moments of the score and higher order derivatives of the likelihood function. This part is justified by the second stage in which we show that the choice of the initial values does not matter for the asymptotic properties of the estimator. Our first result is given as follows.

[4]We can also allow that some rows of A are zero, in which case our model reduces to a model with K conditionally heteroskedastic factors, and $N - K$ conditionally homoskedastic factors, see Lanne and Saikkonen (2005).

Theorem 2 *Under Assumptions 4.1-4.4,* $\hat{\theta}_n \to_{a.s.} \theta_0$.

Next we state results for the asymptotic distribution.

Assumption 4.5 θ_0 *is an interior point.*

Assumption 4.6 $\mathrm{E}|\xi_t|^4 < \infty$.

Assumption 4.5 is needed to establish asymptotic normality. Otherwise, if the parameters are on the boundary, other methods and assumptions should be used and the asymptotic distribution of the parameters is not standard. Andrews (1999) studied in detail the distribution of the QMLE for the univariate GARCH (1,q) model, see also Francq and Zakoian (2006). This issue is beyond the scope of this paper. Assumption 4.6 strengthens Assumption 4.3 to the existence of the fourth order moment of the error term. This assumption is needed to establish that the variance of the score function exists. Note that for our asymptotic theory, no conditions on the shape of the innovation distribution are required. The next theorem establishes the asymptotic normality.

Theorem 3 *Under Assumptions 4.1-4.6,* $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_{\mathrm{D}} N(0, J^{-1}VJ^{-1})$, *where*

$$V = \mathrm{E}\left(\frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta}\frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta'}\right) \quad \textit{and} \quad J = -\mathrm{E}\left(\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta \partial \theta'}\right)$$

Using Lemma 1 in Comte and Lieberman (2003), we have that if $\varepsilon_t \sim i.i.N(0, I_N)$, then $\sqrt{n}(\hat{\theta}_n - \theta_0) \to N(0, J^{-1})$. Let $\hat{J}_n$ and $\hat{V}_n$ be the sample counterpart of J and V where $\hat{\theta}_n$ is used. Under our assumptions and as a by-product of the proof of Theorem 3, it is straightforward to show that $J_n^{-1}V_nJ_n^{-1}$ is a strongly consistent estimate of $J^{-1}VJ^{-1}$. Note that the QMLE here is the global maximum over the whole parameter set. In Theorem 3 we have shown that the score function obeys the standard CLT. However, as in Lee and Hansen (1996) and Preminger and Storti (2006), it is possible to establish the functional central limit theorem and law of iterated logarithm, respectively, for the score function. These additional results come at no cost and could be useful in other applications.

Under the results above the Gaussian QMLE is $\sqrt{n}$ consistent for the true parameter values. However, in the presence of non-Gaussian innovations, this estimator can fail to produce asymptotically efficient estimates. Assuming a non-normal distribution for the likelihood function entails the risk of inconsistent parameter estimation if the distribution is misspecified. Given the results of Theorems 2-3 and mild regularity conditions on the innovation terms which appear in Hafner and Rombouts (2006), we can construct semi-parametric estimators which are asymptotically more efficient than the QMLE. Hafner and Rombouts (2006) show that for the CCC model of Bollerslev (1990), one can even attain the parametric lower bound for the parameters describing the volatility dynamics. Whether or not this so-called adaptive estimation is possible in the current model framework is left for future research. It seems that a direct application of Corollary 3.1 of Drost et al. (1997) does not yield the result in our factor model, so other methods should be investigated.

5 Conclusions

In this paper, we investigate the asymptotic theory for a factor GARCH model. We start by providing sufficient conditions for strict stationarity, geometric ergodicity and existence of certain moments. All of our conditions are expressed in terms of the model parameters and can be easily checked. Using these results and other mild regularity conditions, we proceed to show consistency and asymptotic normality of the QMLE. Our conditions do not involve the existence of second order moments for the process, which is particularly appealing in financial applications. The improvement of the efficiency of the QMLE via a semi-parametric approach is discussed. Our method could also be applied to establish the asymptotic theory for the CCC model of Bollerslev (1990). However, in general we can not apply our results to other classes of multivariate GARCH models and this issue should be investigated in future work. In addition, the small sample properties of the QMLE should be studied via Monte Carlo simulation. In our setting we have assumed that the innovations are independent and identically distributed, but this can probably be weakened by assuming that they are strictly stationary, ergodic and a martingale difference sequence. This is also left for future research.

Acknowledgments

The authors thank participants of the *Journées Modelisation Aléatoire et Statistique de la Société de Mathématiques Appliquées et Industrielles*, Lille 2006, Farid Boussama, Christian Francq, Lajos Horvath, Robert de Jong, Sharon Rubin and Jean-Michel Zakoian for their helpful comments. Hafner acknowledges financial support by the *Fonds Spéciaux de Recherche* (FSR 05) of the Université catholique de Louvain, Louvain-la-Neuve, Belgium. Preminger acknowledges financial support provided through the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, MICFINMA and the Ernst foundation.

Appendix

Proof of Theorem 1.

a) The process (8) forms a homogenous Markov chain with state space $(\mathbb{R}_+^N, \mathcal{B})$, where $\mathbb{R}_+ = (0, \infty)$, $\mathcal{B}$ is a Borel σ-algebra on $\mathbb{R}_+^N$. By an application of the dominated convergence theorem we can show that the Markov chain is Feller. Given Assumptions 3.1 and 3.2, we can show similarly to Bougerol and Picard (1992a) that the infinite series (8) converges a.s. to a stationary solution. Let μ be the stationary measure of the process on $(\mathbb{R}_+^N, \mathcal{B})$. It follows that for all $x \in \mathbb{R}_+^N$ there exists an $n \geq 1$ such that $\mathrm{P}^n(x, \mathcal{A}) > 0$ whenever $\mu(\mathcal{A}) > 0$. Hence the chain is μ-irreducible (here $\mathrm{P}^n(x, \mathcal{A})$ denotes the n-step transition probability from $\sigma_0^2 = x$ to $\sigma_n^2 \in \mathcal{A}$). Assumption 3.2 implies that for n sufficiently large, $S = \mathrm{E} \log \|\prod_{l=0}^n C_{n-l}\| < 0$. Let $h(\delta) = \mathrm{E}\,\|\prod_{l=0}^n C_{n-l}\|^\delta$ and since $S = h'(0) < 0$, $h(\delta)$ is decreasing in a neighborhood of zero and $h(0) = 1$, it follows that $h(r) < 1$ for some small $0 < r < s$. Next, consider the drift function $g(\cdot) = 1 + |\cdot|^r$ and by solving (8) recursively we obtain that

$$\mathrm{E}g(\sigma_n^2|\sigma_0^2 = x) \leq 1 + \mathrm{E}\left[\left\|\prod_{l=0}^{n} C_{n-l}\right\|^r\right] |x|^r + \mathrm{E}\left[\sum_{l=0}^{n} \|C_n \cdots C_{n-l+1}\|\right]^r$$

Since the second term on the RHS is bounded by Assumption 3.3, we have that for some compact set $\mathcal{A}$ with $\mu(\mathcal{A}) > 0$, $\mathrm{E}g(\sigma_n^2|\sigma_0^2 = x) \leq \eta g(x)$ for all $x \in \mathcal{A}^C$ and some $\eta \in (0, 1)$. Therefore, $\{\sigma_t^2\}$ is geometrically ergodic and $\mathrm{E}|\sigma_t^2|^r < \infty$ by Theorems 1 and 2 of Feigin and Tweedie (1985). Hence, by Meitz and Saikkonen (2004) we also have that $\{y_t\}$ is geometrically ergodic and $\mathrm{E}|y_t|^r < \infty$.

b) Repeating the arguments of part (a) of the theorem and using Assumption 3.4 with the drift function $g(\cdot) = 1 + |\cdot|^r$ provides the desired result. ■

Proof of Theorem 2.

We first note that $\forall t$ and $\forall \theta \in \Theta$,

$$\det(\tilde{H}_t) = \det(W \tilde{\Sigma}_t W') = \det(W)^2 \prod_{i=1}^{N} \tilde{\Sigma}_{ii} > 0$$

see Lütkepohl (p.48, 1996). Next, leting $\tilde{\lambda}_i(\theta_0)$ be the eigenvalues of $\tilde{H}_t(\theta_0)$, we have for some $r > 0$

$$\mathrm{E}\log\det(\tilde{H}_{t,\theta_0}) \leq \mathrm{E}|\det(\tilde{H}_{t,\theta_0})|^{r/N} \leq \mathrm{E}\prod_{i=1}^{N}\tilde{\lambda}_i^{r/N}(\theta_0) \leq C\mathrm{E}\left\|\tilde{H}_{t,\theta_0}\right\|^r$$

where $C > 0$ and $\mathrm{E}\left\|\tilde{H}_{t,\theta_0}\right\|^r$ is bounded by Assumptions 4.2, 4.3 and Theorem 1. Given these results and Assumptions 4.1-4.4 and by using similar arguments as in Jeantheau (1998), Proposition 2.1, we can show that θ_0 is identifiable. Further, similar to Francq and Zakoian (2004), we can show that for any $\theta \neq \theta_0$ there exists a neighborhood $N(\theta_0)$ such that

$$\liminf_{n\to\infty}\inf_{\bar{\theta}\in N(\theta)\cap\Theta} L_n(\theta) > E\tilde{l}_t(\theta_0).$$

In order to complete the proof it is sufficient to show that $\sup_{\theta\in\Theta}|L_n - \tilde{L}_n| \to 0$ a.s. By Cesaro's mean theorem it suffices to check that $\mathrm{E}\sup_{\theta\in\Theta}|l_t - \tilde{l}_t|^r$ is bounded by a summable sequence in t for some $r > 0$. Then the desired result follows by the Markov inequality and the Borel-Cantelli Lemma. We have

$$H_t - \tilde{H}_t = W(\Sigma_t - \tilde{\Sigma}_t)W' = W(diag(\sigma_t^2 - \tilde{\sigma}_t^2))W'. \tag{11}$$

Since $diag(B^t x) \leq B^t(x, x, \ldots, x)$ for some positive vector x and $\|(x, x, \ldots, x)\| \leq |x|$, we obtain

$$\left\|H_t - \tilde{H}_t\right\| \leq C\|W\|^2\left\|B^t\right\||\tilde{\sigma}_0^2 - \sigma_0^2|.$$

Now $\sup_{\theta\in\Theta}\|B\| < 1$ implies that there exists a $\rho \in (0,1)$ such that $\sup_{\theta\in\Theta}\|B^t\| = O(\rho^t)$. Thus, by the c_r inequality and the compactness assumption

$$\mathrm{E}\sup_{\theta\in\Theta}\left\|H_t - \tilde{H}_t\right\|^r = O(\rho^t) \tag{12}$$

for some $r > 0$. From (6), the c_r inequality and $\sup_{\theta\in\Theta}\|B\| < 1$ it is straightforward that

$$\mathrm{E}\sup_{\theta\in\Theta}\|H_t\|^r < \infty, \quad \mathrm{E}\sup_{\theta\in\Theta}\left\|\tilde{H}_t\right\|^r < \infty \tag{13}$$

From (10) and a first order Taylor expansion we have that

$$\sup_{\theta\in\Theta}|l_t(\theta) - \tilde{l}_t(\theta)|^r \leq \left\{|y_t|^{2r}\left\|\tilde{H}_t^{-1}\otimes\tilde{H}_t^{-1}\right\|^r + \left\|vec(\bar{H}_t^{-1})\right\|^r\right\}\left\|H_t - \tilde{H}_t\right\|^r$$

where $\bar{H}_t$ is evaluated on the chord between H_t and $\tilde{H}_t$. Compactness of Θ, (12) and (13) imply the desired result. ∎

Proof of Theorem 3.

By the mean-value theorem we obtain that for the score function around θ_0,

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\theta}\bigg|_{\theta=\hat{\theta}_n} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{\partial l_t}{\partial\theta}-\frac{\partial\tilde{l}_t}{\partial\theta}\right)\bigg|_{\theta=\theta_0} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t}{\partial\theta}\bigg|_{\theta=\theta_0} \\ &+ \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t}{\partial\theta\partial\theta'}\bigg|_{\theta=\theta_0} - J\right)\sqrt{n}(\hat{\theta}_n-\theta_0) \\ &+ \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\theta\partial\theta'}\bigg|_{\theta=\tilde{\theta}_n} - \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t}{\partial\theta\partial\theta'}\bigg\|_{\theta=\theta_0}\right)\sqrt{n}(\hat{\theta}_n-\theta_0) + J\sqrt{n}(\hat{\theta}_n-\theta_0) \end{aligned} \quad (14)$$

for some $\tilde{\theta}_n \in (\hat{\theta}_n, \theta_0)$. We first show that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}$ obeys the CLT. Under Assumption 4.2 the score function is strictly stationary and ergodic. Using (1)-(3), we have that $\mathrm{E}\left[\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}|\mathcal{F}_{t-1}\right] = 0$, where $\mathcal{F}_t$ is the σ-field generated by the past values of y_t's and f_t's, i.e. $\mathcal{F}_t \equiv \sigma(y_t, f_t, y_{t-1}, f_{t-1}, \ldots)$. Thus, the score is also a martingale difference sequence. Therefore, if $\mathrm{E}\left\|\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}'\right\|$ is finite, we can apply the CLT of Scott (1973) and the Cramér-Wold device to establish the asymptotic normality of the score function.

We proceed by calculating the components of the score function. The score with respect to the i-th component of the parameter vector $\theta = (vec(W)', vec(A)', vec(B)')'$ is given by

$$-2\frac{\partial\tilde{l}_t(\theta)}{\partial\theta_i} = Tr[(I_n - y_t y_t'\tilde{H}_t^{-1})\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}],$$

using the notation $\dot{\tilde{H}}_{t,i} = \partial\tilde{H}_t/\partial\theta_i$. We have

$$\frac{\partial\tilde{H}_t}{\partial W_{ij}} = J_{ij}\tilde{\Sigma}_t W' + W\tilde{\Sigma}_t J_{ij}' \quad (15)$$

$$\frac{\partial\tilde{H}_t}{\partial\alpha_{ij}} = W\frac{\partial\tilde{\Sigma}_t}{\partial\alpha_{ij}}W', \quad \frac{\partial\tilde{H}_t}{\partial\beta_{ij}} = W\frac{\partial\tilde{\Sigma}_t}{\partial\beta_{ij}}W' \quad (16)$$

$$\frac{\partial\tilde{\sigma}_t^2}{\partial\alpha_{ij}} = \sum_{k=0}^{\infty}B^k J_{ij}f_{t-k-1}^2 \le C\sum_{k=0}^{\infty}B^k A f_{t-k-1}^2 \quad (17)$$

$$\frac{\partial \tilde{\sigma}_t^2}{\partial \beta_{ij}} = \sum_{k=1}^{\infty} \left(\sum_{l=1}^{k} B^{l-1} J_{ij} B^{k-l} \right) c_{t-k} \leq C \sum_{k=1}^{\infty} k B^k c_{t-k} \tag{18}$$

where J_{ij} is an $N \times N$ matrix containing a one at the ij-th position and zeros elsewhere. Eq. (17)-(18) use the fact that $A_{ij} J_{ij} \leq A$ and $B_{ij} J_{ij} \leq B$. Note that $\sigma_t^2 = \sum_{j=0}^{\infty} B^j c_{t-j} \geq \iota + B^j c_{t-j} \forall j \geq 1$ and for (17) and (18) we show that

$$\begin{aligned} \left\| \dot{\tilde{H}}_{t,i} \tilde{H}_t^{-1} \right\| &\leq \left\| \dot{\tilde{H}}_{t,i} \right\| \left\| \tilde{H}_t^{-1} \right\| \\ &= \left\| W \dot{\tilde{\Sigma}}_{t,i} W' \right\| \left\| W'^{-1} \tilde{\Sigma}_t^{-1} W^{-1} \right\| \\ &\leq C \left\| \dot{\tilde{\Sigma}}_{t,i} \right\| \left\| \tilde{\Sigma}_t^{-1} \right\|, \ i = N^2+1, \ldots, 3N^2, \end{aligned} \tag{19}$$

with $C = \|W\|^2 \|W^{-1}\|^2 < \infty$ and where we use the definition of H_t in (3).

Next we show

A. $\mathrm{E} \sup_{\theta \in N(\theta_0)} \left\| \dot{\tilde{H}}_{t,i} \tilde{H}_t^{-1} \right\|^d < \infty, \ \ d \geq 1$

B. $\mathrm{E} \sup_{\theta \in N(\theta_0)} \left\| y_t y_t' \tilde{H}_t^{-1} \right\|^d < \infty, \ \ d = 2,$

where $N(\theta_0)$ is some arbitrarily small neighborhood around θ_0. This result is stronger than needed, but we will use it later.

We see immediately that the derivatives w.r.t. the elements of W multiplied by $\left\| \tilde{\Sigma}_t^{-1} \right\|$, are naturally bounded. Now we will consider the score w.r.t. α_{ij}. Using (17), (19), and the compactness of the parameter space, we obtain that

$$\begin{aligned} \sup_{\theta \in N(\theta_0)} \left\| \dot{\tilde{\Sigma}}_{t,i} \right\| \left\| \tilde{\Sigma}_t^{-1} \right\| &\leq C \sup_{\theta \in N(\theta_0)} \frac{\sum_{k=0}^{\infty} [B^k A f_{t-k-1}^2]_j}{\sum_{k=0}^{\infty} 1 + [B^k c_{t-k}]_{j'}} \\ &\leq C \sup_{\theta \in N(\theta_0)} \frac{\sum_{k=0}^{\infty} [B^k (\iota + A f_{t-k-1}^2)]_j}{\sum_{k=0}^{\infty} [B^k c_{t-k}]_{j'}} \\ &\leq C \sup_{\theta \in N(\theta_0)} \frac{\tilde{\sigma}_{jt}^2}{\tilde{\sigma}_{j't}^2} \\ &\leq C \sup_{\theta \in N(\theta_0)} \frac{\bar{\beta} \sum_{i=1}^{N} \tilde{\sigma}_{i,t-1}^2 + \bar{\alpha} \sum_{i=1}^{N} \tilde{\sigma}_{i,t-1}^2 \xi_{it}^2}{\underline{\beta} \sum_{i=1}^{N} \tilde{\sigma}_{i,t-1}^2 + \underline{\alpha} \sum_{i=1}^{N} \tilde{\sigma}_{i,t-1}^2 \xi_{it}^2} \\ &\leq C \left(\frac{\bar{\beta}}{\underline{\beta}} - \frac{\bar{\alpha}}{\underline{\alpha}} \right) < \infty \end{aligned} \tag{20}$$

where in (20) j and j' are the indices of the maximum and minimum components, respectively, of the vector $B^{k-1}Af^2_{t-k-1}$. The third and fourth inequalities use (6) and (7) respectively, where $\bar{\alpha} = \max(\alpha_{ij})$, $\bar{\beta} = \max(\beta_{ij})$, $\underline{\alpha} = \min(\alpha_{ij})$, $\underline{\beta} = \min(\beta_{ij})$. Due to Assumption 4.5 and Theorem 2, $\underline{\beta} > 0$ and $\underline{\alpha} > 0$.

In the following consider the score w.r.t. β_{ij}. Equations (19) and (20) imply

$$\begin{aligned}
\mathrm{E}\sup_{\theta\in N(\theta_0)}\left|\left\|\dot{\tilde{\Sigma}}_{t,i}\right\|\left\|\tilde{\Sigma}_t^{-1}\right\|\right|^d &\leq C_1\mathrm{E}\sup_{\theta\in N(\theta_0)}\left|\sum_{k=1}^{\infty}\frac{k[B^kc_{t-k}]_j}{1+[B^kc_{t-k}]_{j'}}\right|^d \qquad (21)\\
&\leq C_1\mathrm{E}\sup_{\theta\in N(\theta_0)}\left|\sum_{k=1}^{\infty}\frac{k[B^kc_{t-k}]_{j'}}{1+[B^kc_{t-k}]_{j'}}\frac{[B^kc_{t-k}]_j}{[B^kc_{t-k}]_{j'}}\right|^d\\
&\leq C_1\mathrm{E}\sup_{\theta\in N(\theta_0)}\left|\sum_{k=1}^{\infty}k[B^kc_{t-k}]_{j'}^{s/d}\frac{\max_{uv}[B^k]_{uv}|c_{t-k}|}{\min_{uv}[B^k]_{uv}|c_{t-k}|}\right|^d\\
&\leq C_2\left|\sum_{k=1}^{\infty}k\left(\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|B^kc_{t-k}\right\|^s\right)^{1/d}\right|^d
\end{aligned}$$

$$\begin{aligned}
&\leq C_3\left|\sum_{k=1}^{\bar{m}}k\left\|B^k\right\|\left(\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|c_{t-k}\right\|^s\right)^{1/d}+\sum_{k=\bar{m}+1}^{\infty}k\rho^{ks}\left(\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|c_{t-k}\right\|^s\right)^{1/d}\right|^d\\
&\leq C_4\left|\sum_{k=1}^{\bar{m}}k\left\|B^k\right\|\left(\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|c_{t-k}\right\|^s\right)^{1/d}\right|^d+C_5\left|\sum_{k=\bar{m}+1}^{\infty}k\rho^{ks}\left(\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|c_{t-k}\right\|^s\right)^{1/d}\right|^d<\infty
\end{aligned}$$

where in (21) j and j' are the indices of the maximum and minimum components, respectively, of the vector B^kc_{t-k}. The first and second inequalities are due to direct calculations. The third inequality uses the fact that for $x \geq 0$, $x/(x+1) < x^{s/d}$ for some $s \in (0,1)$ and $d \geq 1$. Note that even though $B^k = O(\rho^k)$, the ratio of $\max_{uv}[B^k]_{uv}$ to $\min_{uv}[B^k]_{uv}$ is $O(1)$. The fourth inequality is implied by the Minkowski inequality. The fact that there exists an $\bar{m} \geq 1$ and $\rho \in (0,1)$ s.t. $\sup_{\theta\in\Theta}\|B^m\| \leq C\rho^m$ for all $m > \bar{m}$, the compactness of parameter set, the c_r inequality and Assumption 4.3 imply the last two inequalities. Hence, A. is established.

Next we show B. We can write (1) as

$$y_t = \tilde{H}_t^{1/2}\xi_t = W\tilde{\Sigma}_{0t}^{1/2}\xi_t, \tag{22}$$

and therefore

$$y_t y_t' \tilde{H}_t^{-1} = (W\tilde{\Sigma}_{0t}^{1/2}\xi_t\xi_t'\Sigma_{0t}^{1/2}W')\tilde{H}_t^{-1}. \tag{23}$$

Since by Theorem 2, $\hat{\theta}_n \rightarrow_{a.s.} \theta_0$, we can define an arbitrary small neighborhood $N(\theta_0)$ around θ_0 such that $\beta_{ij}/\beta_{0,ij} \leq (1+\delta)$, $\delta > 0$ and $(1+\delta)\rho < 1$, where $\rho = \sup_{\theta\in\Theta}\|B\|$ and $\beta_{0,ij}$ is the true parameter value of β_{ij}. Using that $\left\|\tilde{H}_t^{-1}\right\| \leq C\left\|\tilde{\Sigma}_t^{-1}\right\|$,

$$\sup_{\theta\in N(\theta_0)}\left\|W\Sigma_{0t}^{1/2}\xi_t\xi_t'\tilde{\Sigma}_{0t}^{1/2}W'\right\|\left\|H_t^{-1}\right\| \leq \sup_{\theta\in N(\theta_0)} C\left\|\tilde{\Sigma}_{0t}\right\||\xi_t|^2\left\|\tilde{\Sigma}_t^{-1}\right\|$$

and for some $s \in (0,1)$, we get

$$\begin{aligned}
\sup_{\theta\in N(\theta_0)}\left\|\tilde{\Sigma}_{0t}\right\|\left\|\tilde{\Sigma}_t^{-1}\right\| &\leq \sup_{\theta\in N(\theta_0)}\sum_{k=0}^{\infty}\frac{[B_0^k c_{t-k}]_j}{1+[B^k c_{t-k}]_{j'}} \\
&= \sup_{\theta\in N(\theta_0)}\sum_{k=0}^{\infty}\frac{[B^k c_{t-k}]_j}{1+[B^k c_{t-k}]_{j'}}\frac{[B_0^k c_{t-k}]_j}{[B^k c_{t-k}]_j} \\
&\leq \sup_{\theta\in N(\theta_0)}\sum_{k=0}^{\infty}\frac{[B^k c_{t-k}]_{j'}}{\iota+[B^k c_{t-k}]_{j'}}\frac{\max_{uv}[B^k]_{uv}|c_{t-k}|}{\min_{uv}[B^k]_{uv}|c_{t-k}|}(1+\delta)^k \\
&\leq \sup_{\theta\in N(\theta_0)} C\sum_{k=0}^{\infty}[B^k c_{t-k}]_{j'}^s(1+\delta)^k \\
&\leq C\sum_{k=0}^{\infty}[\rho(1+\delta)]^k\sup_{\theta\in N(\theta_0)}|c_{t-k}|^s.
\end{aligned} \tag{24}$$

By choosing $s > 0$ such that $\mathrm{E}|c_t|^{2s} < \infty$, we have that $\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\tilde{\Sigma}_{0t}\right\|^2\left\|\tilde{\Sigma}_t^{-1}\right\|^2 < \infty$ and the desired results is implied by Assumptions 4.3 and 4.6.

We have that

$$\mathrm{E}\left|\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta_i}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta_j}\right| \leq C\mathrm{E}\left(\left\|y_t y_t'\tilde{H}_{t,\theta_0}^{-1}\right\|^2\left\|\tilde{H}_{t,i,\theta_0}\tilde{H}_{0t}^{-1}\right\|\left\|\tilde{H}_{t,j,\theta_0}\tilde{H}_{t,\theta_0}^{-1}\right\|\right) \tag{25}$$

$$\leq C\mathrm{E}\left(\left\|\tilde{H}_{t,\theta_0}^{1/2}\xi_t\xi_t'\tilde{H}_{t,\theta_0}^{-1/2}\right\|^2\left\|\tilde{H}_{t,i,\theta_0}\tilde{H}_{t,\theta_0}^{-1}\right\|\left\|\tilde{H}_{t,j,\theta_0}\tilde{H}_{t,\theta_0}^{-1}\right\|\right)$$

$$\leq C\mathrm{E}(|\xi_t|^4)\left(\mathrm{E}\left(\left\|\tilde{H}_{t,i,\theta_0}\tilde{H}_{t,\theta_0}^{-1}\right\|\right)^2\right)^{1/2}\left(\mathrm{E}\left(\left\|\tilde{H}_{t,j,\theta_0}\tilde{H}_{t,\theta_0}^{-1}\right\|\right)^2\right)^{1/2} < \infty$$

The first inequality uses the fact that if A and B are $d \times d$ matrices, then $|tr(AB)| \leq d||A|| \cdot ||B||$ see Lütkepohl (1996, p.111). The second inequality results from (22) and we note that $\left\|H_t^{1/2}\xi_t\xi_t'H_t^{-1/2}\right\|^2 \leq C_1\left\|\xi_t\xi_t'\right\|^2 \leq C_2\left|\xi_t\right|^4$ a.s., see Lütkepohl (1996, pp. 42-44). This result and the Cauchy-Schwarz inequality imply the third inequality. Assumption 4.6, B and that ξ_t is independent of $\tilde{H}_t$ and its derivatives imply the last inequality. These results also imply that V exists. Hence, the second term in (14) converges in distribution to the normal distribution.

Now, we show that the third term in (14) converges a.s. to zero. The second derivative is given by

$$-2\frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j} = Tr[(I_N - y_ty_t'\tilde{H}_t^{-1})(\ddot{\tilde{H}}_{t,ij}\tilde{H}_t^{-1} - \dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}) + y_ty_t'\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}] \tag{26}$$

where we use the notation $\ddot{\tilde{H}}_{t,ij} = \frac{\partial^2\tilde{H}_t}{\partial\theta_i\partial\theta_j}$.

The components of $\ddot{\tilde{H}}_{t,ij}$ are given in the following.

$$\frac{\partial^2\tilde{H}_t}{\partial W_{ij}\partial W_{rs}} = J_{ij}\tilde{\Sigma}_tJ_{rs}' + J_{rs}\tilde{\Sigma}_tJ_{ij}' \tag{27}$$

$$\frac{\partial^2\tilde{H}_t}{\partial X_{ij}\partial W_{rs}} = J_{rs}\frac{\partial\tilde{\Sigma}_t}{\partial X_{ij}}W' + W\frac{\partial\tilde{\Sigma}_t}{\partial X_{ij}}J_{ij}' \tag{28}$$

$$\frac{\partial^2\tilde{H}_t}{\partial X_{ij}\partial Y_{rs}} = W\frac{\partial^2\tilde{\Sigma}_t}{\partial X_{ij}\partial Y_{rs}}W' \tag{29}$$

$$\frac{\partial^2\tilde{\sigma}_t^2}{\partial\alpha_{ij}\partial\alpha_{rs}} = 0 \tag{30}$$

$$\frac{\partial^2\tilde{\sigma}_t^2}{\partial\alpha_{ij}\partial\beta_{rs}} = \sum_{k=2}^{\infty}\left(\sum_{l=1}^{k-1}B^{l-1}J_{rs}B^{k-l-1}\right)J_{ij}f_{t-k}^2 \leq C\sum_{k=2}^{\infty}(k-1)B^{k-1}Af_{t-k}^2 \tag{31}$$

$$\frac{\partial^2 \tilde{\sigma}_t^2}{\partial \beta_{ij} \partial \beta_{rs}} = \sum_{k=2}^{\infty} \left[\sum_{l=2}^{k} \left(\sum_{t=1}^{l-1} B^{t-1} J_{rs} B^{l-1-t} \right) J_{ij} B^{k-l} + \sum_{l=1}^{k-1} B^{l-1} J_{ij} \left(\sum_{t=1}^{k-l} B^{t-1} J_{rs} B^{k-l-t} \right) \right] c_{t-k} \tag{32}$$

$$\leq C \sum_{k=2}^{\infty} k(k-1) B^k c_{t-k}$$

where $X, Y = \alpha, \beta$.

It is straightforward to see that (27), when multiplied by $\left\| \tilde{\Sigma}_t^{-1} \right\|$ is naturally bounded. We will show that

C. $\mathrm{E}\sup_{\theta \in N(\theta_0)} \left\| \ddot{\tilde{H}}_{t,ij} \tilde{H}_t^{-1} \right\|^d < \infty, \;\; d \geq 1$

The proof is similar to the one used to establish A., and hence omitted. Note that this result is stronger than needed, but will be used later. We have by (26), A-C, the Cauchy-Schwarz inequality and the Minkowski inequality that

$$\mathrm{E}\left| \frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta_i \partial \theta_j} \right| \leq C\mathrm{E}\left[\left\| y_t y_t' \tilde{H}_{t,\theta_0}^{-1} \right\| \left(\left\| \ddot{\tilde{H}}_{t,ij,\theta_0} \tilde{H}_{t,\theta_0}^{-1} \right\| + 2 \left\| \dot{\tilde{H}}_{t,i,\theta_0} \tilde{H}_{t,\theta_0}^{-1} \right\| \cdot \left\| \dot{\tilde{H}}_{t,j,\theta_0} \tilde{H}_{t,\theta_0}^{-1} \right\| \right) \right] \tag{33}$$

$$\leq C \left\{ \left(\mathrm{E}\left(\left\| y_t y_t' \tilde{H}_{t,\theta_0}^{-1} \right\|^2 \right) \right)^{1/2} \left[\left(\mathrm{E}\left(\left\| \ddot{\tilde{H}}_{t,ij,\theta_0} \tilde{H}_{t,\theta_0}^{-1} \right\|^2 \right) \right)^{1/2} + \right. \right.$$

$$\left. 2 \left(\mathrm{E}\left(\left\| \dot{\tilde{H}}_{t,i,\theta_0} \tilde{H}_{t,\theta_0}^{-1} \right\|^4 \right) \right)^{1/4} \cdot \left(\mathrm{E}\left(\left\| \dot{\tilde{H}}_{t,j,\theta_0} \tilde{H}_{t,\theta_0}^{-1} \right\|^4 \right) \right)^{1/4} \right] < \infty$$

and the desired result follows from the ergodic theorem. Now we show that $\mathrm{E}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta \partial \theta'}$ is positive definite (p.d.). Note that $\mathrm{E}(\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial \theta \partial \theta'} | \mathrm{F}_{t-1}) = Tr(\dot{\tilde{H}}_{t,i} \tilde{H}_t^{-1} \dot{\tilde{H}}_{t,j} \tilde{H}_t^{-1})$. Comte and Lieberman (2003) show that this matrix is positive definite almost surely, as otherwise the model is not identifiable. We proceed by establishing that

$$\left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \tilde{l}_t}{\partial \theta \partial \theta'} \bigg|_{\theta = \tilde{\theta}_n} - \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \tilde{l}_t}{\partial \theta \partial \theta'} \bigg|_{\theta = \theta_0} \right\| \to_{a.s.} 0 \tag{34}$$

By using Theorem 2 and similar arguments as in Francq and Zakoian (2004), a sufficient condition is given by

$$\mathrm{E} \sup_{\theta \in N(\theta_0)} \left| \frac{\partial^3 \tilde{l}_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| < \infty \;\; \forall\, i, j, k. \tag{35}$$

Next, we take the third derivatives.

$$-2\frac{\partial^3 \tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} = Tr\left[(I_N - y_t y_t'\ \tilde{H}_t^{-1})\left\{\dddot{\tilde{H}}_{t,ijk}\ \tilde{H}_t^{-1} - \dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\ddot{\tilde{H}}_{t,jk}\tilde{H}_t^{-1}\right.\right.$$
$$-\ddot{\tilde{H}}_{t,ik}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1} - \ddot{\tilde{H}}_{t,ij}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,k}\tilde{H}_t^{-1}$$
$$\left.+\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,k}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1} + \dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,k}\tilde{H}_t^{-1}\right\}$$
$$+y_t y_t'\tilde{H}_t^{-1}\left\{\ddot{\tilde{H}}_{t,jk}\ \tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1} + \dot{\tilde{H}}_{t,k}\tilde{H}_t^{-1}\ddot{\tilde{H}}_{t,ij}\tilde{H}_t^{-1}\right.$$
$$-\dot{\tilde{H}}_{t,k}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1} - \dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,k}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}$$
$$\left.\left.-\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,k}\tilde{H}_t^{-1} - \dot{\tilde{H}}_{t,k}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,i}\tilde{H}_t^{-1}\dot{\tilde{H}}_{t,j}\tilde{H}_t^{-1}\right\}\right] \tag{36}$$

where we use the notation $\dddot{\tilde{H}}_{t,ijk} = \frac{\partial^3 \tilde{H}_t}{\partial\theta_i\partial\theta_j\partial\theta_k}$ with components of $\dddot{\tilde{H}}_{t,ijk}$ given by

$$\frac{\partial^3 \tilde{H}_t}{\partial W_{ij}\partial W_{rs}\partial W_{uv}} = 0 \tag{37}$$

$$\frac{\partial^3 \tilde{H}_t}{\partial X_{ij}\partial W_{rs}\partial W_{uv}} = J_{rs}\frac{\partial\tilde{\Sigma}_t}{\partial X_{ij}}J_{uv}' + J_{uv}\frac{\partial\tilde{\Sigma}_t}{\partial X_{ij}}J_{rs}' \tag{38}$$

$$\frac{\partial^3 \tilde{H}_t}{\partial X_{ij}\partial Y_{rs}\partial W_{uv}} = J_{uv}\frac{\partial^2\tilde{\Sigma}_t}{\partial X_{ij}\partial Y_{rs}}W' + W\frac{\partial^2\tilde{\Sigma}_t}{\partial X_{ij}\partial Y_{rs}}J_{uv}' \tag{39}$$

$$\frac{\partial^3 \tilde{H}_t}{\partial X_{ij}\partial Y_{rs}\partial Z_{uv}} = W\frac{\partial^3\tilde{\Sigma}_t}{\partial X_{ij}\partial Y_{rs}\partial Z_{uv}}W'\frac{\partial^3\tilde{\sigma}_t^2}{\partial\alpha_{ij}\partial\beta_{rs}\partial\beta_{uv}} \tag{40}$$

$$= \sum_{k=2}^{\infty}\left[\sum_{l=2}^{k}\left\{\sum_{t=1}^{l-1}B^{t-1}J_{uv}B^{l-1-t}\right\}J_{rs}B^{k-l}\right.$$
$$\left.+\sum_{l=1}^{k-1}B^{l-1}J_{rs}\left\{\sum_{t=1}^{k-l}B^{t-1}J_{uv}B^{k-l-t}\right\}\right]J_{ij}f_{t-k}^2$$
$$\leq C\sum_{k=1}^{\infty}(k-1)(k-2)B^{k-1}Af_{t-k}^2 \tag{41}$$

$$\frac{\partial^3\tilde{\sigma}_t^2}{\partial\beta_{ij}\partial\beta_{rs}\partial\beta_{uv}} = \sum_{k=3}^{\infty}\left[\sum_{l=3}^{k}\left(\sum_{t=2}^{l-1}\left\{\sum_{n=1}^{t-1}B^{n-1}J_{uv}B^{t-1-n}\right\}J_{rs}B^{l-1-t}\right.\right.$$

$$+\sum_{t=1}^{l-2} B^{t-1} J_{rs} \left\{ \sum_{n=1}^{l-1-t} B^{n-1} J_{uv} B^{l-1-t-n} \right\} \Bigg) J_{ij} B^{k-1}$$

$$+\sum_{l=3}^{k-1} \left\{ \sum_{t=1}^{l-1} B^{t-1} J_{rs} B^{l-1-t} \right\} J_{ij} \left\{ \sum_{n=1}^{k-l} B^{n-1} J_{uv} B^{k-l-n} \right\}$$

$$+\sum_{l=2}^{k-1} \left\{ \sum_{n=1}^{l-1} B^{n-1} J_{uv} B^{l-1-n} \right\} J_{ij} \left\{ \sum_{t=1}^{k-l} B^{t-1} J_{rs} B^{k-l-t} \right\}$$

$$+\sum_{l=1}^{k-1} B^{l-1} J_{ij} \Bigg(\sum_{t=1}^{k-l} \left\{ \sum_{n=1}^{t-1} B^{n-1} J_{uv} B^{t-1-n} \right\} J_{rs} B^{k-l-t}$$

$$+\sum_{t=1}^{k-l-1} B^{t-1} J_{rs} \left\{ \sum_{n=1}^{k-l-t} B^{n-1} J_{uv} B^{k-l-t-n} \right\} \Bigg) \Bigg] c_{t-k} \leq C \sum_{k=3}^{\infty} k(k-1)(k-2) B^k c_{t-k} \quad (42)$$

where $X, Y, Z = \alpha, \beta$.

Using similar arguments which were already used to prove A. and C., we can show that

D. $\mathrm{E} \sup_{\theta \in N(\theta_0)} \left\| \tilde{H}_{t,ijk} \tilde{H}_t^{-1} \right\|^d < \infty, \ \ d \geq 1$

where A.-D., (36), repeated applications of the Cauchy-Schwarz inequality and the Minkowski inequality and some tedious calculations allow us to establish (35).

Now we show that for some $r > 0$, there exists a $\rho \in (0, 1)$ such that

$$E \left| \frac{\partial l_t(\theta_0)}{\partial \theta_i} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta_i} \right|^r = O(\rho^t). \quad (43)$$

Given the score function, we have that (the argument θ_0 is omitted for simplicity).

$$\begin{aligned}
& \left\| \dot{H}_{t,i} H_t^{-1}(I_N - y_t y_t' H_t^{-1}) - \dot{\tilde{H}}_{t,i} H_t^{-1}(I_N - y_t y_t' H_t^{-1}) \right\| \\
= & \left\| (\dot{H}_{t,i} H_t^{-1} - \dot{\tilde{H}}_{t,i} \tilde{H}_t^{-1})(I_N - y_t y_t' H_t^{-1}) + \dot{\tilde{H}}_{t,i} \tilde{H}_t^{-1} y_t y_t' (\tilde{H}_t^{-1} - H_t^{-1}) \right\| \\
= & \left\| \left[(\dot{H}_{t,i} - \dot{\tilde{H}}_{t,i}) H_t^{-1} + \dot{\tilde{H}}_{t,i} (H_t^{-1} - \tilde{H}_t^{-1}) \right] (I_N - y_t y_t' H_t^{-1}) + \dot{\tilde{H}}_{t,i} \tilde{H}_t^{-1} y_t y_t' (\tilde{H}_t^{-1} - H_t^{-1}) \right\| \\
= & \left\| \left[(\dot{H}_{t,i} - \dot{\tilde{H}}_{t,i}) H_t^{-1} + \dot{\tilde{H}}_{t,i} \tilde{H}_t^{-1} (\tilde{H}_t - H_t) H_t^{-1} \right] (I_N - y_t y_t' H_t^{-1}) \right. \\
+ & \left. \dot{\tilde{H}}_{t,i} \tilde{H}_t^{-1} y_t y_t' H_t^{-1} (H_t - \tilde{H}_t) \tilde{H}_t^{-1} \right\| \qquad (44)
\end{aligned}$$

Now from (11) and (17), we get

$$\left\|\dot{H}_{t,i} - \dot{\tilde{H}}_{t,i}\right\| = \|W\|^2 \left\|\frac{\partial B^t}{\partial \beta_{ij}}\right\| \left\|\sigma_0^2 - \tilde{\sigma}_0^2\right\| \leq C\rho^t \quad a.s. \tag{45}$$

In addition, note that if λ is an eigenvalue of a nonsingular matrix A, then λ^{-1} is an eigenvalue of A^{-1} see Lütkepohl (1996, p. 64). Let $\lambda_1(A), \ldots, \lambda_m(A)$ denote the eigenvalues of an $m \times m$ matrix A in ascending order. Then,

$$\left\|H_t^{-1}\right\| = \lambda_1^{-1}(W\Sigma_t W') < \infty \tag{46}$$

since $W\Sigma_t W'$ is positive definite and, hence, $\lambda_1(W\Sigma_t W') > 0$. Similarly, we can show that $\left\|\tilde{H}_t^{-1}\right\|$ is bounded and under Assumptions 4.1-4.3 for some $r > 0$, $\mathrm{E}|y_t|^{2r}$ is finite.

These results, the c_r inequality, A.-B., (12), (44) and (45) imply (43). The Markov inequality and (43) imply that

$$\left| n^{-1/2} \sum_{t=1}^{n} \frac{\partial l_t(\theta_0)}{\partial \theta_i} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial \theta_i} \right| = o_{\mathrm{P}}(1) \tag{47}$$

The same method can be applied to show

$$\mathrm{E} \sup_{\theta \in N(\theta_0)} \left| \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial \theta \partial \theta'} \right| = O(\rho^t).$$

Using similar arguments as in Theorem 2, we can establish that

$$\mathrm{E} \sup_{\theta \in N(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial \theta \partial \theta'} \right| = o_{a.s.}(1) \tag{48}$$

Theorem 2, (33), (34), (47) and (48) imply that the first term in (14) is $o_{\mathrm{P}}(1)$ and the third and fourth terms are $o_{a.s.}(1)$. Since we have shown that the second term in (14) obeys the CLT, the desired result follows from the Slutsky Theorem. ∎

References

Alexander C.O. (2001). *Orthogonal GARCH*, in *Mastering Risk*, Vol. II, edited by C.O. Alexander, pp. 21-38, Prentice Hall.

Andrews D. W.K. (1999). Estimation when a parameter is on a boundary. *Econometrica*, 67, 1341-1383.

Baillie R.T. and Chung H. (2001). Estimation of GARCH models from the autocorrelations of the squares of a process. *Journal of Time Series Analysis*, 22, 631-650.

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

Berkes I., Horvath L. and Kokoszka P.S. (2003). GARCH processes: structure and estimation. *Bernoulli*, 9, 201-227.

Bollerslev T. (1990). Modelling the coherence on short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and statistics*, 72, 498-505.

Bollerslev T. and Wooldridge J. (1992). Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*, 11, 143-72.

Bollerslev T., Engle R.F. and Wooldridge J.M. (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96, 116-131.

Bougerol P. and Picard N. (1992a). Strict stationarity of generalized autoregressive processes, *The Annals of Probability*, 20, 1714-1730.

Bougerol P. and Picard N. (1992b). Stationarity of GARCH processes and of some nonnegative time series, *Journal of Econometrics*, 52, 115-127.

Boussama F. (1998). Ergodicity, Mixing and Estimation in GARCH models, Ph.D. dissertation, University Paris 7.

Boussama F. (2000). Asymptotic normality for the quasi-maximum likelihood estimator of a GARCH model, C.R. Acad. Sci. Paris Series I 331, 81-84.

Carrasco M. and Chen X. (2002). Mixing and moments properties of various GARCH and stochastic models. *Econometric Theory*, 18, 17-39.

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

Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1997). Adaptive estimation in time series models. *Annals of Statistics* **25**, 786-817.

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

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

Feigin P.D. and Tweedie R.L. (1985). Random coefficient autoregressive processes a Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis*, 6, 1-14.

Francq C. and Zakoian J-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH Processes. *Bernoulli,* 10, 605–637.

Francq C. and Zakoian J-M. (2005). Mixing properties of a general class of GARCH(1,1) models without moment assumptions. Manuscript, GREMARS.

Francq C. and Zakoian J-M. (2006). Quasi-likelihood inference in GARCH processes when some coefficients are equal to zero. Manuscript, GREMARS.

Goldsheid I. Ya. (1991). Lyapounov exponents and asymptotic behavior of the product of random matrices. In: *Lecture Notes in Mathematics* 1486, 23–37. Springer, Berlin.

Gourieroux C. (1997). *ARCH Models and Financial Applications.* Springer Verlag.

Hafner C.M. (2003). Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics*, 1, 26-54.

Hafner C.M. and Rombouts J. (2006). Semiparametric multivariate volatility models. *Econometric Theory*, forthcoming.

Jeantheau T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory*, 14, 70-86.

Jones G.L. (2004). On the Markov chain central limit theorem. *Probabilities Surveys*, 1, 299-320.

Kingman J.F.C. (1973). Subadditive ergodic theory. *Annals of Probability*, 1, 883-909.

Kristensen D. (2005). Geometric ergodicity of a class of Markov chains with applications to time series models. Working paper, University of Wisconsin.

Kristensen D. and Linton O. (2006). A closed form estimator for the GARCH (1,1) model, *Econometric Theory*, 22, 323-337.

Lanne, M. and Saikkonen, P. (2005). A multivariate generalized orthogonal factor GARCH model. Working paper, Helsinki Center of Economic Research.

Lee S-W. and Hansen B.E. (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10, 29-52.

Ling S. and McAleer M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, 19, 280-310.

Lumsdaine R.L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH (1,1) models. *Econometrica*, 64, 575-596.

Lütkepohl H. (1996). *Handbook of Matrices.* New York: Wiley.

Meyn S.P. and R.L. Tweedie (1993). *Markov Chains and Stochastic Stability*. New York: Springer-Verlag.

Meitz M. and Saikkonen P. (2004). Ergodicity, mixing and existence of moments of a class of Markov models with applications to GARCH and ACD models. Working Paper Series in Economics and Finance, Stockholm School of Economics.

Nelson D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6,318-334.

Preminger A. and Storti G. (2006). A GARCH(1,1) estimate with (almost) no moment conditions on the error term, CORE Discussion paper 68, Université catholique de Louvain.

Scott D.J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability*, 5, 119-137.

Storti G. (2006). Minimum distance estimation of GARCH(1,1) models. C*omputational Statistics and Data Analysis,* forthcoming.

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

Vrontos I.D., Dellaportas P. and Politis D.N. (2003). A full-factor multivariate GARCH model. *Econometrics Journal*, 6, 311-333.