



SAPIENZA  
UNIVERSITÀ DI ROMA

Dipartimento di Scienze Statistiche  
Sezione di Statistica Economica ed Econometria

**Marco Avarucci   Eric Beutner   Paolo Zaffaroni**

**On moment conditions for Quasi-Maximum  
Likelihood estimation of multivariate ARCH  
models**

*DSS Empirical Economics and Econometrics  
Working Papers Series*

DSS-E3 WP 2012/1

## DSS Empirical Economics and Econometrics Working Papers Series

- 2011/1 Massimo Franchi, Paolo Paruolo "Normal forms of regular matrix polynomials via local rank factorization"
- 2011/2 Francesca Di Iorio, Stefano Fachin "A Sieve Bootstrap range test for poolability in dependent cointegrated panels"
- 2011/3 Maria Grazia Pittau, Shlomo Yitzhaki, Roberto Zelli "The make-up of a regression coefficient: An application to gender"
- 2011/4 Søren Johansen "The analysis of nonstationary time series using regression, correlation and cointegration - with an application to annual mean temperature and sea level"
- 2011/5 Mario Forni, Marc Hallin, Marco Lippi, Paolo Zaffaroni "One-Sided Representations of Generalized Dynamic Factor Models"

Dipartimento di Scienze Statistiche  
Sezione di Statistica Economica ed Econometria  
"Sapienza" Università di Roma  
P.le A. Moro 5 – 00185 Roma - Italia  
<http://www.dss.uniroma1.it>

# On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models

Marco Avarucci<sup>a</sup>, Eric Beutner<sup>a</sup>, and Paolo Zaffaroni<sup>b</sup>

<sup>a</sup>Maastricht University

<sup>b</sup>Imperial College London and University of Rome “La Sapienza”

October 26, 2011

## Abstract

This paper questions whether it is possible to derive consistency and asymptotic normality of the Gaussian quasi-maximum likelihood estimator (QMLE) for possibly the simplest VEC-GARCH model, namely the multivariate ARCH(1) model of the BEKK form, under weak moment conditions similar to the univariate case. In contrast to the univariate specification, we show that the expectation of the loglikelihood function is unbounded, away from the true parameter value, if (and only if) the observable has unbounded second moment. Despite this non-standard feature, consistency of the Gaussian QMLE is still warranted. The same moment condition proves to be necessary and sufficient for the stationarity of the score, when evaluated at the true parameter value. This explains why high moment conditions, typically bounded sixth moment and above, have been used hitherto in the literature to establish the asymptotic normality of the QMLE in the multivariate framework.

## 1 Introduction

Whereas the probabilistic and statistical theory of univariate GARCH models can be considered firmly established, there are still many open problems for multivariate GARCH (MGARCH) models. In the univariate framework, necessary and sufficient conditions for strict stationarity have been established by Nelson (1990) and Bougerol and Picard (1992a,b). On the other hand, in a multivariate framework only sufficient conditions are known; see, for example,

---

Address correspondence to Paolo Zaffaroni, Imperial College Business School, Imperial College London, South Kensington campus, London SW7 2AZ, UK. Tel.: +44 (0) 207 594 9186; E-mail: p.zaffaroni@imperial.ac.uk. Research supported by the ESRC Grant RES-000-22-3219.

Boussama et al. (2011). Berkes et al. (2003) derived consistency and asymptotic normality of the Gaussian QMLE for univariate GARCH( $p, q$ ) processes under mild conditions, extending the GARCH(1, 1) results of Lumsdaine (1996) and Lee and Hansen (1994). In particular, their results imply that the existence of the second moment of the observed variable, typically a financial return, is not required. Francq and Zakoïan (2004) and Robinson and Zaffaroni (2006) obtained the same results, the former assuming slightly weaker conditions, the latter considering a more general model. Jensen and Rahbek (2004a,b) showed that the parameters of the ARCH(1) and GARCH(1,1) models, with a known intercept, can be consistently estimated and exhibits a conventional Gaussian asymptotic distribution with a standard rate of convergence, even if the parameters are outside the strict stationarity region. From a theoretical point of view, these results show that the presence of a unit, and even explosive, root in the autoregressive equation does not imply a knife-edge result on the asymptotic distribution of the QMLE. From an empirical point of view, conditions requiring only the existence of low order (fractional) moments of the observed variable are more likely met by financial time series, characterized by a fat-tail distribution.

Let us now turn back to the multivariate framework. When several times series displaying temporal or contemporaneous dependencies are under investigation, it is important to analyze them in a multivariate framework. In particular, understanding the co-movements of financial returns is of great importance, for example, for asset pricing, risk management, and asset allocation. However, unlike ARMA models, the GARCH model specification does not suggest a natural extension to the multivariate framework (Francq and Zakoïan, 2010a, Chapter 11). Two recent surveys by Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2008) offer a comprehensive review of the most common MGARCH models. Since these models may differ substantially from each other, there is no general result covering all MGARCH models. Most of the results in the literature are valid only for particular models; see Francq and Zakoïan (2010a), Section 11.5. for a complete and updated bibliography. For instance, Hafner and Preminger (2009a) and Francq and Zakoïan (2010b) derived asymptotic normality of the QMLE estimator for a full factor GARCH model and the CCC (constant conditional correlation) GARCH model of Bollerslev (1990) respectively, under conditions that do not involve the existence of the second-order moment of the observable process. However, for the general MGARCH model known as the VEC-GARCH model, which was introduced by Bollerslev et al. (1988), the more general results are by Comte and Lieberman (2003) and Hafner and Preminger (2009b) who showed asymptotic normality of the QMLE assuming the eighth and sixth moment, respectively, of the observable process. Bardet and Wintenberger (2009) provide asymptotic results for the QMLE of a class of multidimensional causal processes that include multivariate ARCH(1) processes as a special case. Asymptotic normality is obtained assuming that the fourth moment of the observable variables is bounded, among other less primitive conditions (Bardet and Wintenberger, 2009, Sec. 4.3).

It is worth recalling that the VEC-GARCH model can be viewed as the most direct generalization of the univariate GARCH. Moreover, this model is very general, and it contains as important special case the BEKK model by Engle and Kroner (1995). The main advantage of the BEKK specification is that it automatically ensures that the conditional covariance matrices are almost surely positive definite. The relation between the VEC and the BEKK parametrization is analyzed by Stelzer (2008). Given the discrepancy between the strength of the regularity conditions employed to derive the asymptotics of the Gaussian QMLE in the univariate and multivariate case, it is natural to ask whether this discrepancy is due to some fundamental difference or just because of technical difficulties that have not been solved so far. To analyze this question we shall consider one of the simplest models (see Equation (1)) contained in the VEC-GARCH model.

The core results of the paper are the following. First, for the MGARCH model under consideration, we show that there are parameters for which the Gaussian loglikelihood function has unbounded mean *if and only if* the observable has an unbounded second moment. This holds under a rather weak set of conditions. Roughly speaking it excludes that the observable vector has a singular distribution and that the parameter space is too small. Instead, the mean is finite at the true parameter value under an extremely mild moment condition. This non standard behavior does not preclude and in fact, if any, facilitates the proof for consistency of the Gaussian QMLE. Second, we establish that boundedness of the second moment of the observable is also necessary and sufficient for boundedness of the variance matrix of the score function. Thus, when the second moment is unbounded, one of the major steps in proving asymptotic normality of the QMLE is ruled out. This shows that, although the moment conditions imposed in the literature might be stronger than necessary, certainly asymptotic normality of the Gaussian QMLE cannot be established simply assuming a mild fractional moment condition of the observable, as occurring in the univariate case.

The rest of the article is organized as follows. In Section 2 we briefly recall the model we shall study, introduce some notation and assumptions. Section 3 investigates the behavior of the expectation of the likelihood function paying particular attention to the behavior of the expectation of the likelihood function if the observable variables do not have a finite second moment. The ideas we use to show this behavior will also be the basis for proving that the second moments of the observable variables are necessary to ensure that the score function has a finite variance, as illustrated in Section 4. In the Appendix we give some auxiliary results that are used in the proofs.

## 2 Model, notation and assumptions

We denote the set of real  $p \times p$  real matrices by  $\mathcal{M}_p$ , the linear subspace of symmetric matrices by  $\mathcal{S}_p$ , the cone of positive semidefinite matrices by  $\mathcal{S}_p^{\dagger}$  and its interior (positive definite matrices) by  $\mathcal{S}_p^{\ddagger}$ . For  $B \in \mathcal{M}_p$  we denote by  $\text{tr}(B)$

and  $B'$  the trace and the transpose of  $B$ , respectively;  $\det(B)$  is the determinant and  $B^{-1}$  the inverse of  $B$ . The scalar  $\lambda_i(B)$  is the  $i$ th eigenvalue of the matrix  $B$ ,  $\lambda_{\min}$  and  $\lambda_{\max}$  denote the minimum and the maximum eigenvalue, respectively. The selection vector  $e_i^{(p)}$  is defined as the  $i$ th column of the identity matrix  $I_p$ . The  $\text{vec}(\cdot)$  operator stacks columns of a matrix one under another in a single column; the  $\text{vech}(\cdot)$  of  $B \in \mathcal{S}_p$  does the same thing but starting each column at its diagonal elements. We use the notation  $\mathbb{E}[\cdot]$  and  $\|\cdot\|$  for the expectation and the Euclidean norm, respectively. We denote by  $\xrightarrow{a.s.}$  almost sure convergence. For a random variable  $X$  we denote by  $X^+$  its positive part, and by  $X^-$  its negative part. As usual, we say that  $\mathbb{E}[X]$  is defined if either  $\mathbb{E}[X^+]$  or  $\mathbb{E}[X^-]$  is finite. If both are finite, we say that  $\mathbb{E}[X]$  exists.

Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a sequence of random vectors with values in  $\mathbb{R}^m$  and let  $\mathcal{F}_{t-1}$  be the  $\sigma$ -field generated by past  $X_t$ 's, i.e.  $\mathcal{F}_{t-1} := \{X_{t-1}, X_{t-2}, \dots\}$ . We assume that

$$X_t = H_t^{1/2}(\theta_0)\varepsilon_t, \quad t \in \mathbb{Z},$$

where  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ ,  $\varepsilon_t \in \mathbb{R}^m$ , is a sequence of unobservable and real-valued random vectors and the conditional variance-covariance matrix  $H_t(\theta_0)$  evolves according to the equation

$$H_t(\theta_0) = C_0 + A_0 X_{t-1} X_{t-1}' A_0', \quad (1)$$

with  $C_0 \in \mathcal{S}_m^\dagger$ ,  $A_0 \in \mathcal{M}_m$ . The true, but unknown  $v$ -dimensional vector  $\theta_0$  on the left-hand side of model (1), is given by

$$\theta_0 = (\text{vech}(C_0)', \text{vec}(A_0)')',$$

where  $v = m(3m + 1)/2$ . Without loss of generality,  $H_t^{1/2}(\theta_0)$  is chosen to be symmetric and positive definite; the choice is then unique by Theorem 7.2.6. in Horn and Johnson (1985). Under the standard assumptions (cf. Assumption 2.1 below) it follows that  $H_t(\theta_0)$  is the conditional covariance matrix of  $X_t$  given  $\mathcal{F}_{t-1}$ , that is

$$\mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[X_t X_t' | \mathcal{F}_{t-1}] = H_t(\theta_0).$$

The normal density based quasi-maximum likelihood estimator  $\hat{\theta}_n$  is defined as minimizing

$$\mathcal{L}_n(\theta) = \frac{1}{2n} \sum_{t=2}^n \ell_t(\theta), \quad (2)$$

with

$$\ell_t(\theta) = X_t' H_t^{-1}(\theta) X_t + \log(\det(H_t(\theta))). \quad (3)$$

Minimization is with respect to the set of matrices with elements

$$H_t(\theta) = C(\theta) + A(\theta) X_{t-1} X_{t-1}' A(\theta),$$

where  $C(\theta) \in \mathcal{S}_m^\dagger$  and  $A(\theta) \in \mathcal{M}_m$ . The parameter  $\theta$  and the matrices  $C(\theta)$  and  $A(\theta)$  are related by  $\theta := (\text{vech}(C)', \text{vec}(A)')$ . The main advantage of model

(1) is that the inverse of  $H_t(\theta)$  is easy to compute using the Sherman-Morrison-Woodbury formula (Harville, 1997, Corollary 18.2.10.):

$$H_t^{-1}(\theta) = C(\theta)^{-1} - \frac{C(\theta)^{-1}A(\theta)X_{t-1}X'_{t-1}A(\theta)'C(\theta)^{-1}}{1 + X'_{t-1}A'(\theta)C(\theta)^{-1}A(\theta)X_{t-1}}.$$

In the original paper of Engle and Kroner (1995) the constant matrix we denoted  $C_0$  is the product of a lower triangular matrix and its transpose. This, of course, ensures positive definiteness. Moreover, this parametrization has advantages for estimation but it is not necessary from a theoretical point of view and it is less convenient in terms of notation.

At several places later on we need one or more of the following assumptions. Because we are only interested in the expectation of the likelihood function and the variance of the score, we do not state our assumptions for the whole sequence  $\{X_t\}$  and do not incorporate conditions like stationarity, necessary to prove consistency and asymptotic normality.

**Assumption 2.1** (a)  $\varepsilon_t$  has zero mean and identity variance-covariance matrix.

(b)  $\mathbb{E}[\|X_{t-1}\|^s] < \infty$  for some  $s \in (0, 1)$ .

(c)  $\Theta \subset \mathbb{R}^v$  is a compact set, the parametrization is canonical so that  $C(\theta)$  and  $A(\theta)$  are continuous functions of  $\theta$ , and for all  $\theta \in \Theta$ :  $C(\theta) \in \mathcal{S}_m^\ddagger$ .

(d) For  $\theta \neq \theta_0$  if  $\mathbb{E}[\ell_t(\theta)] < \infty$ , we have

$$H_t(\theta)H_t^{-1}(\theta_0) \neq I \text{ a.s.}$$

(e) The components of  $\varepsilon_t$  are independent and  $\mathbb{E}\varepsilon_{it}^4 = \mu_4 < \infty$  for all  $i = 1, \dots, m$ .

(f) The vector  $\varepsilon_t$  is independent of  $\mathcal{F}_{t-1}$ .

Assumption 2.1 (e) is used in the proof of Theorem 4.1. However it could be weakened by assuming that the fourth moment of  $\varepsilon_{it}$  equals  $\mu_{4,i}$ ,  $i = 1, \dots, m$ .

**Remark 2.2** Identification of model (1) has been assumed in Assumption 2.1. (d) for sake of simplicity. We will show that the aforementioned condition is fulfilled by the model imposing further assumptions. Assume that (i) a given  $(i, j)$  entry of the matrix  $A(\theta)$  is restricted to be positive for every  $\theta \in \Theta$  and (ii)  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  is a non-degenerate and i.i.d. (independent and identically distributed) sequence. We need to prove that for a fixed  $t$ ,  $H_t(\theta) = H_t(\theta_0) P_{\theta_0}$  - a.s. implies that  $\theta = \theta_0$ . Suppose that for some  $\theta \neq \theta_0$ ,  $H_t(\theta) = H_t(\theta_0) P_{\theta_0}$  - a.s. It implies that

$$AX_{t-1}X'_{t-1}A' - A_0X_{t-1}X'_{t-1}A'_0 = CC' - C_0C'_0$$

or equivalently, taking the vec transformation,

$$B_t \text{vec}(\varepsilon_{t-1}\varepsilon'_{t-1}) = d, \tag{4}$$

with

$$\begin{aligned} B_t &= \left[ \left( AH_{t-1}^{1/2}(\theta_0) \otimes AH_{t-1}^{1/2}(\theta_0) \right) - \left( A_0 H_{t-1}^{1/2}(\theta_0) \otimes A_0 H_{t-1}^{1/2}(\theta_0) \right) \right] \\ d &= \text{vec}(C) - \text{vec}(C_0). \end{aligned}$$

Let  $B_t^*$  be the Moore-Penrose inverse of  $B_t$ . Equation (4) implies either

$$B_t \text{vec}(\varepsilon_{t-1} \varepsilon'_{t-1}) = B_t B_t^* d \quad (5)$$

or

$$B_t = d = 0 \quad (6)$$

Equality (5) implies that  $\text{vec}(\varepsilon_{t-1} \varepsilon'_{t-1})$  is a function  $\varepsilon_{t-2}, \varepsilon_{t-3}, \dots, \varepsilon_1$ . This is impossible by (ii). Assumption (i) rules out the case  $A = -A_0$ , so that (6) is satisfied if and only if  $\theta = \theta_0$ , which proves the statement. Restrictions similar to (i) are used by Engle and Kroner (1995) (see their Proposition 2.1) to exclude equivalent representations. Assumption (ii) is standard in the literature and clearly implies (f).

### 3 The expectation of the likelihood function

We start by analyzing the expectation of the likelihood function. The first three parts of the next theorem show that the expectation of the likelihood function has all the properties necessary to show consistency along the lines of Francq and Zakoïan (2004) (see our Remark 3.4). Anyhow, the fourth and the fifth part of it demonstrate already that the second moments of the observable variables play an important role in a multivariate framework. Moreover, as mentioned in the introduction, part four and five of the following theorem indicate that there is a fundamental difference between the univariate and the multivariate framework. To see this, recall that for a univariate GARCH process the expectation of the likelihood function is finite at every  $\theta$  in a compact parameter space  $\Theta$  that rules out zero coefficients; see, for instance the proof of Lemma 5.4 in Berkes et al. (2003).

**Theorem 3.1** (i) Under Assumption 2.1 part (c) we have

$$\mathbb{E} \left[ \inf_{\theta \in \Theta} \ell_t(\theta) \right] > -\infty,$$

implying that  $\mathbb{E}[\ell_t(\theta)]$  is defined for all  $\theta \in \Theta$ .

(ii) If part (a) and (b) of Assumption 2.1 hold true, we have

$$\mathbb{E}[|\ell_t(\theta_0)|] < \infty.$$

(iii) If part (a), (b), (d) and (f) of Assumption 2.1 hold true, then whenever  $\theta \neq \theta_0$

$$\mathbb{E}[\ell_t(\theta)] > \mathbb{E}[\ell_t(\theta_0)].$$



(iv) If part (c) of Assumption 2.1 holds true and if we assume that  $\mathbb{E} [\|X_t\|^2] < \infty$ , then we indeed have that  $\sup_{\theta \in \Theta} \mathbb{E} [\ell_t(\theta)] < \infty$ .

(v) Assume part (a) and (f) of Assumption 2.1 and that  $\mathbb{E} [\|X_t\|^2] = \infty$ , and additionally that for some  $\bar{\theta} \in \Theta$  we have that  $A(\bar{\theta})x$  and  $A(\theta_0)x$  are linearly independent  $\forall x \neq 0$ . Then  $\mathbb{E} [\ell_t(\bar{\theta})] = \infty$ .

**Proof** (i) From the definition of  $\ell_t$  (see Equation (3)) we have for every  $\omega \in \Omega$

$$\inf_{\theta \in \Theta} \ell_t(\theta) \geq \inf_{\theta \in \Theta} \log [\det (H_t(\theta))].$$

where  $\Omega$  denotes the sample space and where we suppressed the dependence on  $\omega$ . Moreover,

$$\det (C(\theta) + A(\theta)X_{t-1}(\omega)X'_{t-1}(\omega)A'(\theta)) \geq \det (C(\theta)), \quad \forall \omega \in \Omega.$$

The inequality follows from the fact that

$$\lambda_k(A + B) \geq \lambda_k(B), \quad k = 1, 2, \dots, p, \quad A \in \mathcal{S}_p, \quad B \in \mathcal{S}_p^\dagger \quad (7)$$

(Lütkepohl, 1996, Sec. 5.3.2, Result (4)). Hence,

$$\mathbb{E} \left[ \inf_{\theta \in \Theta} \ell_t(\theta) \right] \geq \inf_{\theta \in \Theta} \log (\det(C(\theta))) > K,$$

where  $K > -\infty$  by Assumption 2.1, part (c).

(ii) Similar as in part (i) one sees that there is a uniform lower bound on  $\ell_t^-(\theta_0)$ . It only remains to prove that  $\mathbb{E}[\ell_t^+(\theta_0)] < \infty$ . Since the first part of the right-hand side in Equation (3) equals  $\varepsilon_t' \varepsilon_t$  for  $\theta_0$  we need only to show that  $\mathbb{E} [\log^+ (\det (H_t(\theta_0)))] < \infty$ . Using  $\det (H_t(\theta_0)) \leq \prod_{i=1}^m h_{ii,t}(\theta_0)$  (cf. Lütkepohl (1996), Sec. 5.3.1, Result (7)), it is thus sufficient to prove that  $\mathbb{E} [\log^+ (h_{ii,t}(\theta_0))] < \infty$ . Here  $h_{ii,t}(\theta_0)$  denote the diagonal elements of  $H_t(\theta_0)$ . Taking into account that  $h_{ii,t}(\theta_0) = e_i^{(m)'} C_0 e_i^{(m)} + (e_i^{(m)'} A_0 X_{t-1})^2$ , we obtain

$$\begin{aligned} \mathbb{E} [\log^+ (h_{ii,t}(\theta_0))] &= \mathbb{E} [\log (h_{ii,t}(\theta_0)) \mathbb{1}_{\{h_{ii,t}(\theta_0) \geq 1\}}] \\ &= \mathbb{E} \left[ \frac{2}{s} \log \left( \left( e_i^{(m)'} C_0 e_i^{(m)} + (e_i^{(m)'} A_0 X_{t-1})^2 \right)^{\frac{s}{2}} \right) \mathbb{1}_{\{h_{ii,t}(\theta_0) \geq 1\}} \right] \\ &\leq \mathbb{E} \left[ \frac{2}{s} \left( e_i^{(m)'} C_0 e_i^{(m)} + \text{tr} \left( e_i^{(m)'} A_0 X_{t-1} X'_{t-1} A_0 e_i^{(m)} \right) \right)^{\frac{s}{2}} \mathbb{1}_{\{h_{ii,t}(\theta_0) \geq 1\}} \right] \\ &\leq \mathbb{E} \left[ \frac{2}{s} \left( e_i^{(m)'} C_0 e_i^{(m)} \right)^{\frac{s}{2}} \mathbb{1}_{\{h_{ii,t}(\theta_0) \geq 1\}} \right] \\ &\quad + \mathbb{E} \left[ \frac{2}{s} \text{tr} \left( e_i^{(m)'} A_0 X_{t-1} X'_{t-1} A_0 e_i^{(m)} \right)^{\frac{s}{2}} \mathbb{1}_{\{h_{ii,t}(\theta_0) \geq 1\}} \right] \\ &\leq K_1 + \mathbb{E} \left[ \frac{2}{s} \text{tr} \left( A_0 e_i^{(m)} e_i^{(m)'} A_0 \right)^{\frac{s}{2}} \text{tr} \left( X_{t-1} X'_{t-1} \right)^{\frac{s}{2}} \right] \\ &\leq K_1 + K_2 \mathbb{E} [\|X_{t-1}\|^s]. \end{aligned}$$

The first inequality uses  $\log(x) \leq x$  for any  $x > 0$ , the second make use of Loeve's  $c_r$  inequality, and the third uses that

$$\text{tr}(AB) \leq \text{tr}(A) \text{tr}(B), \quad A, B \in \mathcal{S}_p^{\dagger} \quad (8)$$

(Abadir and Magnus, 2005, Exercise 12.14 (a)) and  $\text{tr}(AB) = \text{tr}(BA)$  (Lütkepohl, 1996, Sec. 4.1.1, Result (8)).

(iii) If  $\mathbb{E}[\ell_t(\theta)] = \infty$  the result is obvious from the finiteness of  $\mathbb{E}[\ell_t(\theta_0)]$ .

Otherwise

$$\begin{aligned} & \mathbb{E}[\ell_t(\theta)] - \mathbb{E}[\ell_t(\theta_0)] \\ &= \mathbb{E}[\varepsilon_t'(H_t^{\frac{1}{2}}(\theta_0)H_t^{-1}(\theta)H_t^{\frac{1}{2}}(\theta_0) - I)\varepsilon_t] + \mathbb{E}[\log(\det(H_t(\theta)H_t^{-1}(\theta_0)))] \\ &= \mathbb{E}[\text{tr}((H_t^{\frac{1}{2}}(\theta_0)H_t^{-1}(\theta)H_t^{\frac{1}{2}}(\theta_0) - I)\varepsilon_t\varepsilon_t')] - \mathbb{E}[\log(\det(H_t^{\frac{1}{2}}(\theta_0)H_t^{-1}(\theta)H_t^{\frac{1}{2}}(\theta_0)))] \\ &= \mathbb{E} \left[ \text{tr} \left( H_t^{\frac{1}{2}}(\theta_0)H_t^{-1}(\theta)H_t^{\frac{1}{2}}(\theta_0) \right) - m - \log \left( \det(H_t^{\frac{1}{2}}(\theta_0)H_t^{-1}(\theta)H_t^{\frac{1}{2}}(\theta_0)) \right) \right] > 0, \end{aligned}$$

because  $\log \det(A) \leq \text{tr}(A) - n$  for any matrix  $A \in \mathcal{S}_p^{\dagger}$ , with equality if and only if  $A = I_p$  (Lütkepohl, 1996, Sec. 4.1.2, Result (10)).

(iv) We have

$$\sup_{\theta \in \Theta} \mathbb{E} [|\ell_t(\theta)|] \leq \sup_{\theta \in \Theta} \mathbb{E} \left[ |X_t' H_t^{-1}(\theta) X_t| \right] + \sup_{\theta \in \Theta} \mathbb{E} \left[ |\log[\det(H_t(\theta))]| \right]. \quad (9)$$

We consider the two terms on the right-hand side separately. Using (8) and (7), the first term can be bounded as follows

$$\mathbb{E} [\text{tr}(X_t' H_t^{-1}(\theta) X_t)] \leq \mathbb{E} [\text{tr}(H_t^{-1}(\theta)) \text{tr}(X_t X_t')] \leq \text{tr}(C(\theta))^{-1} \mathbb{E} [\|X_t\|^2]. \quad (10)$$

For the second term similar as in the proof of part (ii) we find, using condition (c), for some continuous and positive functions  $K_1 : \Theta \rightarrow \mathbb{R}$ , and  $K_2 : \Theta \rightarrow \mathbb{R}$ , and some  $s' < 2$

$$\mathbb{E} [\log(h_{ii,t}(\theta))] \leq K_1(\theta) + K_2(\theta) \mathbb{E} \left[ \|X_{t-1}\|^{s'} \right]. \quad (11)$$

By our assumption  $\mathbb{E} \|X_{t-1}\|^2 < \infty$  and the compactness of  $\Theta$  the upper bounds in (10) and (11) are finite, implying that the right-hand side of (9) is finite as well.

(v) To abbreviate the notation we let  $\bar{A} = A(\bar{\theta})$ ,  $\bar{C} = C(\bar{\theta})$ . We now show that  $\mathbb{E}[\ell_t^+(\bar{\theta})] = \infty$ . By the arguments in part (i) we have

$$\ell_t^+(\bar{\theta}) \geq X_t' H_t^{-1}(\bar{\theta}) X_t - K, \quad 0 \leq K < \infty. \quad (12)$$

Thus it is sufficient to show that  $\mathbb{E}[X_t' H_t^{-1}(\bar{\theta}) X_t] = \infty$ . First we note that by part (f) of Assumption 2.1

$$\begin{aligned} \mathbb{E} \left[ X_t' H_t^{-1}(\bar{\theta}) X_t | \mathcal{F}_{t-1} \right] &= \mathbb{E} \left[ \text{tr}(H_t^{\frac{1}{2}}(\theta_0) H_t^{-1}(\bar{\theta}) H_t^{\frac{1}{2}}(\theta_0) \varepsilon_t \varepsilon_t') | \mathcal{F}_{t-1} \right] \\ &= \text{tr} (H_t(\theta_0) H_t^{-1}(\bar{\theta})). \end{aligned}$$

Next  $H_t(\theta_0)H_t^{-1}(\bar{\theta})$  is computed explicitly for a fixed but arbitrary realization of  $X_{t-1}$ . To simplify the notation we suppress the subscript  $t-1$  and denote this realization just by  $x$ . By the Sherman-Morrison-Woodbury formula applied to  $H_t^{-1}(\bar{\theta})$

$$\begin{aligned} & \text{tr}(H_t(\theta_0)H_t^{-1}(\bar{\theta})) \\ &= \text{tr}(C_0\bar{C}^{-1}) + \frac{x'A_0'\bar{C}^{-1}A_0x - x'\bar{A}'\bar{C}^{-1}C_0\bar{C}^{-1}\bar{A}x}{1 + x'\bar{A}'\bar{C}^{-1}\bar{A}x} \end{aligned} \quad (13)$$

$$+ \frac{(x'A_0'\bar{C}^{-1}A_0x)(x'\bar{A}'\bar{C}^{-1}\bar{A}x) - (x'A_0'\bar{C}^{-1}\bar{A}x)^2}{1 + x'\bar{A}'\bar{C}^{-1}\bar{A}x}. \quad (14)$$

It turns out that (13) and (14) are both non-negative. In fact, regarding (13),

$$\begin{aligned} & \text{tr}(C_0\bar{C}^{-1}) + \frac{x'A_0'\bar{C}^{-1}A_0x - x'\bar{A}'\bar{C}^{-1}C_0\bar{C}^{-1}\bar{A}x}{1 + x'\bar{A}'\bar{C}^{-1}\bar{A}x} \\ &= \frac{\text{tr}(C_0\bar{C}^{-1}) + \text{tr}(C_0\bar{C}^{-1})x'\bar{A}'\bar{C}^{-1}\bar{A}x + x'A_0'\bar{C}^{-1}A_0x - x'\bar{A}'\bar{C}^{-1}C_0\bar{C}^{-1}\bar{A}x}{1 + x'\bar{A}'\bar{C}^{-1}\bar{A}x} \\ &= \frac{\text{tr}(C_0\bar{C}^{-1})}{1 + x'\bar{A}'\bar{C}^{-1}\bar{A}x} + \frac{x'A_0'\bar{C}^{-1}A_0x}{1 + x'\bar{A}'\bar{C}^{-1}\bar{A}x} \end{aligned} \quad (15)$$

$$+ \frac{\text{tr}(C_0\bar{C}^{-1})\text{tr}(\bar{C}^{-1}\bar{A}xx'\bar{A}') - \text{tr}(\bar{C}^{-1}C_0\bar{C}^{-1}\bar{A}xx'\bar{A}')}{1 + x'\bar{A}'\bar{C}^{-1}\bar{A}x}. \quad (16)$$

Looking at the three terms after the last equality, the first is larger than 0 and less than or equal to  $\text{tr}(C_0\bar{C}^{-1})$ , the second term is non-negative since  $\bar{C}$  is positive definite, and the third term is non-negative, because (8) implies

$$\begin{aligned} & \text{tr}(C_0\bar{C}^{-1})\text{tr}(\bar{C}^{-1}\bar{A}xx'\bar{A}') - \text{tr}(\bar{C}^{-1}C_0\bar{C}^{-1}\bar{A}xx'\bar{A}') \\ &= \text{tr}(C_0\bar{C}^{-1})\text{tr}(\bar{C}^{-1}\bar{A}xx'\bar{A}') - \text{tr}\left[\left(\bar{C}^{-\frac{1}{2}}C_0\bar{C}^{-\frac{1}{2}}\right)\left(\bar{C}^{-\frac{1}{2}}\bar{A}xx'\bar{A}'\bar{C}^{-\frac{1}{2}}\right)\right] \\ &\geq \text{tr}(C_0\bar{C}^{-1})\text{tr}(\bar{C}^{-1}\bar{A}xx'\bar{A}') - \text{tr}\left(\bar{C}^{-\frac{1}{2}}C_0\bar{C}^{-\frac{1}{2}}\right)\text{tr}\left(\bar{C}^{-\frac{1}{2}}\bar{A}xx'\bar{A}'\bar{C}^{-\frac{1}{2}}\right) \\ &= \text{tr}(C_0\bar{C}^{-1})\text{tr}(\bar{C}^{-1}\bar{A}xx'\bar{A}') - \text{tr}(C_0\bar{C}^{-1})\text{tr}(\bar{C}^{-1}\bar{A}xx'\bar{A}') = 0. \end{aligned}$$

Thus, the random variable in (13) is non-negative. Moreover, noting that the assumption of linear independence between  $A(\theta_0)x$  and  $A(\bar{\theta})x$  implies that  $A(\bar{\theta})$  has full rank, Remark 3.2, part (ii) shows that (13) is even bounded.

We now turn to (14). Notice first that (14) is non-negative by the Cauchy-Schwarz inequality. We next show that there is a constant  $K > 0$  such that

$$\begin{aligned} & (x'A_0'\bar{C}^{-1}A_0x)(x'\bar{A}'\bar{C}^{-1}\bar{A}x) - (x'A_0'\bar{C}^{-1}\bar{A}x)^2 \\ & \geq K(x'A_0'\bar{C}^{-1}A_0x)(x'\bar{A}'\bar{C}^{-1}\bar{A}x), \quad \forall x, \|x\| \geq 1. \end{aligned} \quad (17)$$

Consider an arbitrary vector  $x$  with  $\|x\| = 1$ . Then

$$(x'A_0'\bar{C}^{-1}A_0x)(x'\bar{A}'\bar{C}^{-1}\bar{A}x) - (x'A_0'\bar{C}^{-1}\bar{A}x)^2 > 0,$$

because by our assumption the vectors  $A_0x$  and  $\bar{A}x$  are linearly independent  $\forall x \neq 0$ , implying that  $x'A_0'\bar{C}^{-\frac{1}{2}}$  and  $x'\bar{A}'\bar{C}^{-\frac{1}{2}}$  are independent  $\forall x \neq 0$ . Moreover, by the compactness of the unit sphere

$$\inf_{x, \|x\|=1} [(x'A_0'\bar{C}^{-1}A_0x)(x'\bar{A}'\bar{C}^{-1}\bar{A}x) - (x'A_0'\bar{C}^{-1}\bar{A}x)^2] > \tilde{K}, \quad (18)$$

where  $\tilde{K} > 0$ . To see that this implies (17) for  $x$  with  $\|x\| = 1$ , notice first that (17) holds with  $K = 1$  if  $x' A_0' \bar{C}^{-1} \bar{A} x = 0$ . Moreover, for all  $x$  with  $\|x\| = 1$  and  $x' A_0' \bar{C}^{-1} \bar{A} x \neq 0$  we obtain from the uniform lower bound in (18)

$$(x' A_0' \bar{C}^{-1} A_0 x)(x' \bar{A}' \bar{C}^{-1} \bar{A} x) > \tilde{K} + (x' A_0' \bar{C}^{-1} \bar{A} x)^2$$

which in turn implies

$$\frac{(x' A_0' \bar{C}^{-1} A_0 x)(x' \bar{A}' \bar{C}^{-1} \bar{A} x)}{(x' A_0' \bar{C}^{-1} \bar{A} x)^2} > \frac{\tilde{K}}{\sup_{x, \|x\|=1} (x' A_0' \bar{C}^{-1} \bar{A} x)^2} + 1 \geq \tilde{K}_1$$

with  $\tilde{K}_1 > 1$ . Hence

$$\tilde{K}_1^{-1} (x' A_0' \bar{C}^{-1} A_0 x)(x' \bar{A}' \bar{C}^{-1} \bar{A} x) > (x' A_0' \bar{C}^{-1} \bar{A} x)^2.$$

It follows that

$$\begin{aligned} & (x' A_0' \bar{C}^{-1} A_0 x)(x' \bar{A}' \bar{C}^{-1} \bar{A} x) - (x' A_0' \bar{C}^{-1} \bar{A} x)^2 \\ & > (1 - \tilde{K}_1^{-1})(x' A_0' \bar{C}^{-1} A_0 x)(x' \bar{A}' \bar{C}^{-1} \bar{A} x). \end{aligned}$$

Setting  $K := 1 - \tilde{K}_1^{-1} > 0$ , the latter inequality proves (17) for  $x$  with  $\|x\| = 1$ . For  $x$  with  $\|x\| > 1$  inequality (17) follows now from the fact that  $x = \|x\| \frac{x}{\|x\|}$ . Hence, we obtain

$$\begin{aligned} \mathbb{E} [X_t' H_t^{-1}(\bar{\theta}) X_t] & \geq \mathbb{E} \left[ \frac{K (X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1})(X_{t-1}' \bar{A}' \bar{C}^{-1} \bar{A} X_{t-1})}{1 + X_{t-1}' \bar{A}' \bar{C}^{-1} \bar{A} X_{t-1}} \mathbb{1}_{\{\|X_{t-1}\| \geq 1\}} \right] \\ & \geq K \bar{K} \mathbb{E} [X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1} \mathbb{1}_{\{\|X_{t-1}\| \geq 1\}}]. \end{aligned}$$

Here  $\bar{K} = \inf_{x, \|x\| \geq 1} \frac{x' \bar{A}' \bar{C}^{-1} \bar{A} x}{1 + x' \bar{A}' \bar{C}^{-1} \bar{A} x} > 0$ . We have  $\bar{K} > 0$  since  $\bar{A}$  has full rank by assumption and  $\bar{C}^{-1}$  is positive definite. Now the claim follows from the fact that  $\mathbb{E}[\|X_t\|^2] = \infty$  implies that

$$\mathbb{E} [X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1}] = \infty \quad (19)$$

which clearly implies that  $\mathbb{E} [X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1} \mathbb{1}_{\{\|X_{t-1}\| \geq 1\}}] = \infty$ . To see that (19) holds true, the following result (Magnus and Neudecker (1999), p.204) is useful:

$$\lambda_{\min}(A) \operatorname{tr}(B) \leq \operatorname{tr}(AB) \leq \lambda_{\max}(A) \operatorname{tr}(B), \quad \forall A \in \mathcal{S}_p, B \in \mathcal{S}_p^\dagger. \quad (20)$$

Then we have

$$\begin{aligned} \mathbb{E} [\operatorname{tr}(X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1})] & = \mathbb{E} [\operatorname{tr}(\bar{C}^{-1} A_0 X_{t-1} X_{t-1}' A_0')] \\ & \geq \lambda_{\min}(\bar{C}^{-1}) \mathbb{E} [\operatorname{tr}(A_0 X_{t-1} X_{t-1}' A_0')]. \end{aligned}$$

Because  $\bar{C}$  has full rank this implies (19) if our assumption implies  $\mathbb{E} [\operatorname{tr}(A_0 X_{t-1} X_{t-1}' A_0')] = \infty$ . The latter follows from

$$\mathbb{E} [\|X_t\|^2] = \mathbb{E} [\operatorname{tr}(H_t)] = \operatorname{tr}(C_0) + \mathbb{E} [\operatorname{tr}(A_0 X_{t-1} X_{t-1}' A_0')],$$

which finishes the proof.  $\square$

Before commenting on the consistency, there are four questions arising immediately from part (v) of the above theorem. Firstly, why can we bound the expectation of the likelihood at  $\theta_0$  imposing mild moment conditions on  $X_t$ ? Secondly, why is the expectation of the likelihood function in the univariate case finite for all  $\theta \in \Theta$  if we exclude zero coefficients in  $\theta$  but the second moment is allowed to be infinite? Thirdly, which terms in  $\ell_t(\theta)$  cause the undesirable behavior? Fourthly, can the phenomenon of part (v) of Theorem 3.1 be excluded if, for instance,  $A(\theta_0)$  has reduced rank in which case linear independence cannot hold for all  $x \neq 0$ ? The first question is easy to answer, because in part (v) we showed that  $\mathbb{E}[X_t H_t^{-1}(\theta) X_t]$  does not exist for some  $\theta \neq \theta_0$ . Clearly, for  $\theta = \theta_0$  this expectation exists. The second and third questions are addressed in the next Remark 3.2 part (i) and (ii), respectively, and the fourth question in Remark 3.3.

**Remark 3.2** (i) *We showed that the expectation of  $\ell_t(\theta)$  does not exist by proving that the expectation of the conditional expectation of  $X_t' H_t^{-1}(\theta) X_t$  does not exist which in turn followed from the fact that the expectation of (14) does not exist. In the univariate case for every  $\theta \in \Theta$  the term (14) does not appear in the likelihood function. Moreover, in the univariate case the second term in the conditional expectation of  $X_t' H_t^{-1}(\theta) X_t$ , i.e. (13), can be bounded under mild conditions; see Berkes et al. (2003), Lemma 5.1.*

(ii) *We have seen in the proof of part (v) of Theorem 3.1 that it is at least the term (14) that has an undesirable behavior. If  $\bar{A}$  has full rank then this is the only term with a non-existing expected value. Indeed, the term (13) can be rewritten in terms of (15) and (16), and both are bounded. Because for the second term in (15) we find (using the convention  $0/0 = 0$ )*

$$\frac{X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1}}{1 + X_{t-1}' \bar{A}' \bar{C}^{-1} \bar{A} X_{t-1}} \leq \frac{Y_t' D Y_t}{Y_t' Y_t} \leq \lambda_{\max}(D)$$

with  $D = \bar{C}^{1/2} \bar{A}^{-1} A_0' \bar{C}^{-1} A_0 \bar{A}^{-1} \bar{C}^{1/2}$ ,  $Y_t = \bar{C}^{-1/2} \bar{A} X_{t-1}$  (see Lütkepohl (1996), Sec. 5.2.2, Result 2 for the second inequality), and (16) is smaller than  $\text{tr}(C_0 \bar{C}^{-1})$ . If we allow  $\bar{A}$  to have reduced rank, the expectation of (15) can still be shown to be finite imposing further restrictions on the density of  $X_{t-1}$ . Because

$$\frac{X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1}}{1 + X_{t-1}' \bar{A}' \bar{C}^{-1} \bar{A} X_{t-1}} \leq \frac{(X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1})^u}{(1 + X_{t-1}' \bar{A}' \bar{C}^{-1} \bar{A} X_{t-1})^v} \quad (21)$$

$$< \frac{(X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1})^u}{(X_{t-1}' \bar{A}' \bar{C}^{-1} \bar{A} X_{t-1})^v} \quad (22)$$

with  $u = 1$ ,  $v \leq u$ , we can apply the results in Roberts (1995) to bound the expectation of the ratio of the quadratic forms. For instance, assuming that  $X_{t-1}$  has an elliptically symmetric distribution, with the  $2(u - v)$ th moments existing, the expectation of (21) exists whenever  $\text{rank}(\bar{A}' \bar{C}^{-1} \bar{A}) > 2v$ . Choosing  $v = 1/2 - \delta$ ,  $\delta \in (0, 1/2)$  only the existence of  $\mathbb{E} \|X_{t-1}\|^{1+2\delta}$  is required when

$\text{rank}(\bar{A})=1$ . Further sets of conditions can be given to ensure that the expectation of (21) exists; see the theorem in Section 7.2.2. of Roberts (1995).

The first part of the next remark shows that the proof of part (v) of Theorem 3.1 only needs a slight modification and a very weak assumption on the distribution of  $X_{t-1}$  to carry over to the case where  $A_0$  has reduced rank. The second and third part of it deal with the modification of the proof of part (v) of Theorem 3.1 for the case where all the  $\bar{A}$ 's have reduced rank. This shows that the problem cannot be overcome by simply restricting the parameter space to matrices  $A(\theta)$  that have reduced rank.

**Remark 3.3** (i) If  $A_0$  has reduced rank, (17) is not satisfied by the  $x$ 's lying in the null space of  $A_0$ . In the following, it will be shown that if the distribution of  $X_{t-1}$  is not concentrated in a neighborhood of the null space of  $A_0$ , (19) remains true. To make the above arguments more precise, let  $V = \{x \in \mathbb{R}^m | A_0 x = 0\} \subset \mathbb{R}^m$  and denote by  $q$  the dimension of  $V$ . To simplify the exposition we focus on the case  $q = 1$ . Take  $v_1 \in V$  with  $\|v_1\| = 1$  and denote its polar coordinates by  $\phi_1^{v_1}, \dots, \phi_{m-1}^{v_1}$ , where  $\phi_i^{v_1} \in (0, \pi), i = 1, \dots, m-2$ , and  $\phi_{m-1}^{v_1} \in (0, 2\pi)$ . If we assume that  $X_t$  has a density with respect to Lebesgue measure denoted by  $f^X$ , then

$$\begin{aligned} & \int_{\mathbb{R}^m} g(x_1, \dots, x_m) d\lambda^m(x_1, \dots, x_m) \\ &= \int_S \rho^{m-1} \sin^{m-2} \phi_1 \cdots \sin \phi_{m-2} g(t(\rho, \phi_1, \dots, \phi_{m-1})) d\lambda^m(\rho, \phi_1, \dots, \phi_{m-1}), \end{aligned} \quad (23)$$

where  $g(x_1, \dots, x_m) = (x_1^2 + \dots + x_m^2) f^X(x_1, \dots, x_m)$ ,  $\lambda^m$  denotes Lebesgue measure on  $\mathbb{R}^m$ ,  $S = (0, \infty) \times (0, \pi)^{m-2} \times (0, 2\pi)$ , and  $t$  denotes the canonical diffeomorphism mapping  $S$  onto  $\mathbb{R}^m \setminus H_m$ , where  $H_m$  is the hyperplane defined by  $H_m = \{x \in \mathbb{R}^m | x_{m-1} \geq 0, x_m = 0\}$ . If  $X_t$  does not have a finite second moment, then both sides in (23) are infinite. Moreover, on the right-hand side in (23) we can replace integration over  $S$  by integration over  $(0, \infty) \times [0, \pi]^{m-2} \times [0, 2\pi]$ . Thus, the right-hand side in (23) is, for  $\epsilon > 0$  sufficiently small, equal to

$$\begin{aligned} & \int_{S_\epsilon} g(t(\rho, \phi_1, \dots, \phi_{m-1})) \rho^{m-1} \sin^{m-2} \phi_1 \cdots \sin \phi_{m-2} d\lambda^m(\rho, \phi_1, \dots, \phi_{m-1}) \\ &+ \int_{S_\epsilon^C} g(t(\rho, \phi_1, \dots, \phi_{m-1})) \rho^{m-1} \sin^{m-2} \phi_1 \cdots \sin \phi_{m-2} d\lambda^m(\rho, \phi_1, \dots, \phi_{m-1}), \end{aligned} \quad (24)$$

where  $S_\epsilon = (0, \infty) \times (\phi_1^{v_1} - \epsilon, \phi_1^{v_1} + \epsilon) \times \cdots \times (\phi_{m-1}^{v_1} - \epsilon, \phi_{m-1}^{v_1} + \epsilon)$  and  $S_\epsilon^C$  denotes the complement of  $S_\epsilon$  in  $(0, \infty) \times [0, \pi]^{m-2} \times [0, 2\pi]$ . To proceed we make the following assumption: There is an  $\epsilon > 0$  such that the first integral in (24) is finite.

Then, if  $X_t$  does not have a finite second moment the second integral in (24) must be infinite. Now fix  $\rho = 1$  and notice that the complement of  $(\phi_1^{v_1} - \epsilon, \phi_1^{v_1} + \epsilon) \times \cdots \times (\phi_{m-1}^{v_1} - \epsilon, \phi_{m-1}^{v_1} + \epsilon)$  in  $[0, \pi]^{m-2} \times [0, 2\pi]$  is compact. Assuming that  $A_0 x$

and  $\bar{A}x$  (for some  $\bar{A}$ ) are linearly independent for all  $x$  of radius  $\rho = 1$  that are in the complement of  $S_{\epsilon,1} = (\phi_1^{v_1} - \epsilon, \phi_1^{v_1} + \epsilon) \times \cdots \times (\phi_{m-1}^{v_1} - \epsilon, \phi_{m-1}^{v_1} + \epsilon)$ , we obtain on  $S_{\epsilon,1}^C$  the uniform lower bound (18) by Cauchy-Schwarz inequality. As in Theorem 3.1 part (v) we now obtain inequality (17) which now holds for all  $x \in S_{\epsilon,\rho \geq 1}^C$ , where  $S_{\epsilon,\rho \geq 1} = [1, \infty) \times ((\phi_1^{v_1} - \epsilon, \phi_1^{v_1} + \epsilon) \times \cdots \times (\phi_{m-1}^{v_1} - \epsilon, \phi_{m-1}^{v_1} + \epsilon))$ . Here and in the following  $x \in S_{\epsilon,\rho \geq 1}^C$  (or  $X_{t-1} \in S_{\epsilon,\rho \geq 1}^C$ ) means that the polar coordinates of  $x$  are in  $S_{\epsilon,\rho \geq 1}^C$  (or that the polar coordinates of  $X_{t-1}$  are in  $S_{\epsilon,\rho \geq 1}^C$ ). Thus, we obtain (recall that  $\bar{A}$  was assumed to have full rank)

$$\begin{aligned} \mathbb{E} [X_t' H_t^{-1}(\bar{\theta}) X_t] &\geq \mathbb{E} \left[ \frac{K(X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1})(X_{t-1}' \bar{A}' \bar{C}^{-1} \bar{A} X_{t-1})}{1 + X_{t-1}' \bar{A}' \bar{C}^{-1} \bar{A} X_{t-1}} \mathbb{1}_{\{X_{t-1} \in S_{\epsilon,\rho \geq 1}^C\}} \right] \\ &\geq K \bar{K} \mathbb{E} [X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1} \mathbb{1}_{\{X_{t-1} \in S_{\epsilon,\rho \geq 1}^C\}}]. \end{aligned}$$

The claim follows now from the facts that we obtained (19) under the assumptions that  $X_t$  does not have a finite second moment and that  $\bar{C}$  has full rank, and that (19) implies  $\mathbb{E} [X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1} \mathbb{1}_{\{X_{t-1} \in S_{\epsilon,\rho \geq 1}^C\}}] = \infty$  under the assumption that there is an  $\epsilon > 0$  such that the first integral in (24) is finite. Indeed, we have

$$\begin{aligned} &\mathbb{E} [X_{t-1}' A_0' \bar{C}^{-1} A_0 X_{t-1} \mathbb{1}_{\{X_{t-1} \in S_\epsilon\}}] \\ &= \mathbb{E} [\text{tr}(A_0' \bar{C}^{-1} A_0 X_{t-1} X_{t-1}' \mathbb{1}_{\{X_{t-1} \in S_\epsilon\}})] \\ &\leq \lambda_{\max}(A_0' \bar{C}^{-1} A_0) \text{tr}(\mathbb{E} [X_{t-1} X_{t-1}' \mathbb{1}_{\{X_{t-1} \in S_\epsilon\}}]) < \infty. \end{aligned}$$

The first inequality follows from (20). This shows our claim.

(ii) In the proof of part (v) of Theorem 3.1 we assumed that  $\bar{A}$  has full rank to conclude that  $\bar{K} = \inf_{x, \|x\| \geq 1} \frac{x' \bar{A}' \bar{C}^{-1} \bar{A} x}{1 + x' \bar{A}' \bar{C}^{-1} \bar{A} x} > 0$ . Of course, if the rank of  $\bar{A}$  was smaller than  $m$ , the conclusion would be wrong. However, introducing similar to part (i) the set  $\bar{V} = \{x \in \mathbb{R}^m \mid \bar{A}x = 0\} \subset \mathbb{R}^m$ , assuming for simplicity that its dimension is equal to 1 and denoting the polar coordinates of  $v_1 \in \bar{V}$  with  $\|v_1\| = 1$  by  $\gamma_1^{v_1}, \dots, \gamma_{m-1}^{v_1}$ , then

$$\bar{K} = \inf_{x \in \bar{S}_{\epsilon,\rho \geq 1}^C} \frac{x' \bar{A}' \bar{C}^{-1} \bar{A} x}{1 + x' \bar{A}' \bar{C}^{-1} \bar{A} x} > 0,$$

where  $\bar{S}_{\epsilon,\rho \geq 1}^C$  is the complement of  $\bar{S}_{\epsilon,\rho \geq 1} = [1, \infty) \times (\gamma_1^{v_1} - \epsilon, \gamma_1^{v_1} + \epsilon) \times \cdots \times (\gamma_{m-1}^{v_1} - \epsilon, \gamma_{m-1}^{v_1} + \epsilon)$ . Now (if we assume that  $A_0$  has full rank), the restriction on the distribution becomes

$$\text{tr} \left( \mathbb{E} [X_{t-1} X_{t-1}' \mathbb{1}_{\{X_{t-1} \in \bar{S}_{\epsilon,\rho \geq 1}^C\}}] \right) < \infty$$

for some  $\epsilon > 0$ .

(iii) By combining (i) and (ii) one can obviously carry over part (v) of Theorem 3.1 to the case where  $A_0$  and all the  $\bar{A}$ 's have reduced rank.

As mentioned above before turning to the variance of the score in the next section we now briefly discuss consistency of the QMLE.

**Remark 3.4** *It is worth pointing out that under Assumptions 2.1 (a)–(d) and (f), if  $\{X_t\}$  is a stationary ergodic process we obtain*

$$\hat{\theta}_n \xrightarrow{a.s.} \theta_0 \text{ as } n \rightarrow \infty, \quad (25)$$

where  $\theta_n$  is a measurable minimizer of  $\mathcal{L}(\theta)$ . The proof can be carried out along the lines of the proof of Theorem 5.14 in van der Vaart (1998) by using the ergodic theorem (Billingsley, 1995, p. 284 and 495) or along the lines of the proof of Theorem 2.1 in Francq and Zakoian (2004).

## 4 Variance of the score

The next theorem shows that the variance of the score is infinite if  $A_0$  has full rank and the second moment of  $X_{t-1}$  does not exist. The former assumption can easily be relaxed; see Remark 4.2. Although not essential for the next theorem, it is worth recalling that the expectation of  $\partial \ell_t(\theta_0)/\partial \theta$ , denoting the score evaluated at  $\theta = \theta_0$ , is equal to zero; see, for example, (Comte and Lieberman, 2003, Section 4).

**Theorem 4.1** *Under Assumptions 2.1 (a), (e), and (f) if  $A_0$  has full rank*

$$\mathbb{E} \left[ \frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)}{\partial \theta'} \right] \quad (26)$$

*is finite if and only if  $\mathbb{E} [\|X_{t-1}\|^2] < \infty$ .*

**Proof** Before turning to the actual proof we introduce some notation and give some useful results that we shall use in the actual proof for sufficiency and necessity. Notice first that the vector  $\partial \ell_t(\theta)/\partial \text{vech}(C(\theta))$  is constant; therefore we focus on the score with respect to the sub-vector  $\text{vec}(A(\theta))$  of  $\theta$ . Let  $[B]_{ij}$  denote the  $(i, j)$ th entry of a matrix  $B$  and set to abbreviate the notation

$$\dot{H}_{0t,ij}^{(a)} = \frac{\partial H_t(\theta_0)}{\partial [A(\theta)]_{ij}}, \text{ and } H_{0t} = H_t(\theta_0).$$

Using Lemma A.1. in Comte and Lieberman (2003), which is possible due to Assumptions 2.1 (a) and (e), we have

$$\begin{aligned} & \mathbb{E} \left[ \left( \frac{\partial \ell_t(\theta_0)}{\partial [A(\theta)]_{ij}} \right)^2 \middle| \mathcal{F}_{t-1} \right] \\ &= (\mu_4 - 3) \sum_{k=1}^m \left( H_{0t}^{-\frac{1}{2}} \dot{H}_{0t,ij}^{(a)} H_{0t}^{-\frac{1}{2}} \right)_{(k,k)}^2 + 2 \text{tr} \left( \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \right) \end{aligned} \quad (27)$$

From Comte and Lieberman (2003) (Proof of B2, p.78) the expectation of the first term on the right-hand side is bounded by the expectation of the second



summand. Concerning the latter term, let  $\dot{A}_{ij} = \partial A(\theta)/\partial [A]_{ij}$ , and notice that  $\dot{A}_{ij} = e_i^{(m)} e_j^{(m) \prime}$  independently of  $\theta$ . Moreover, we note that

$$\dot{H}_{0t,ij}^{(a)} = \dot{A}_{ij} X_{t-1} X_{t-1}' A_0' + A_0 X_{t-1} X_{t-1}' \dot{A}_{ji}.$$

From Appendix A we have:

$$\begin{aligned} & \text{tr} \left( \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \right) \\ &= 2 \frac{\left( X_{t-1}' A_0' C_0^{-1} A_0 X_{t-1} \right) \left( X_{t-1}' \dot{A}_{ij}' C_0^{-1} \dot{A}_{ij} X_{t-1} \right) - \left( X_{t-1}' \dot{A}_{ij}' C_0^{-1} A_0 X_{t-1} \right)^2}{1 + X_{t-1}' A_0' C_0^{-1} A_0 X_{t-1}} \\ & \quad + 2 \frac{\left( X_{t-1}' \dot{A}_{ij}' C_0^{-1} A_0 X_{t-1} \right)^2}{\left( 1 + X_{t-1}' A_0' C_0^{-1} A_0 X_{t-1} \right)^2}. \end{aligned} \quad (28)$$

*Sufficiency:* We now first show that the existence of the second moment implies existence of the information matrix. The expectation of the information matrix exists if the expectations of the diagonal elements exist. Using that the first summand is non-negative by Cauchy-Schwarz we obtain

$$\begin{aligned} & \text{tr} \left( H_{0t}^{-1} \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \dot{H}_{0t,ij}^{(a)} \right) \\ & \leq 2 \frac{\left( X_{t-1}' A_0' C_0^{-1} A_0 X_{t-1} \right) \left( X_{t-1}' \dot{A}_{ij}' C_0^{-1} \dot{A}_{ij} X_{t-1} \right)}{1 + X_{t-1}' A_0' C_0^{-1} A_0 X_{t-1}} \\ & \quad + 2 \frac{\left( X_{t-1}' A_0' C_0^{-1} A_0 X_{t-1} \right) \left( X_{t-1}' \dot{A}_{ij}' C_0^{-1} \dot{A}_{ij} X_{t-1} \right)}{\left( 1 + X_{t-1}' A_0' C_0^{-1} A_0 X_{t-1} \right)} \\ & \leq 4 \left( X_{t-1}' \dot{A}_{ij}' C_0^{-1} \dot{A}_{ij} X_{t-1} \right), \end{aligned}$$

where we applied Cauchy-Schwarz also to the second summand on the right-hand side in (28). Moreover, for any  $i, j$ , if  $\mathbb{E} [\|X_{t-1}\|^2] < \infty$ ,

$$\mathbb{E} \left[ X_{t-1}' \dot{A}_{0,ij}' C_0^{-1} \dot{A}_{0,ij} X_{t-1} \right] \leq \lambda_{\max}(C_0^{-1}) \max_{s \in \{1, \dots, m\}} \mathbb{E} (x_{t-1,s})^2 < \infty,$$

where  $x_{t-1,s}$  denotes the  $s$ th element of the random vector  $X_{t-1}$ . The off-diagonal elements of the information matrix are bounded by the Cauchy-Schwarz inequality.

*Necessity:* We next show the necessity. If

$$\mathbb{E} \left[ \text{tr} \left( \frac{\partial \ell_t(\theta_0)}{\partial \text{vec}(A)} \frac{\partial \ell_t(\theta_0)}{\partial \text{vec}'(A)} \right) \right] = \infty, \quad (29)$$

the information matrix does not exist. Since an arbitrary element of the  $m^2$  elements of the diagonal of the matrix  $\frac{\partial \ell_t(\theta_0)}{\partial \text{vec}(A)} \frac{\partial \ell_t(\theta_0)}{\partial \text{vec}'(A)}$  is given by (28), we have

that the left-hand side in (29) is infinite, if the expectation of

$$\begin{aligned}
& 2 \sum_{i=1}^m \sum_{j=1}^m \frac{(X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1}) \left( X'_{t-1} \dot{A}'_{ij} C_0^{-1} \dot{A}_{ij} X_{t-1} \right) - \left( X'_{t-1} \dot{A}'_{ij} C_0^{-1} A_0 X_{t-1} \right)^2}{1 + X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1}} \\
& + 2 \sum_{i=1}^m \sum_{j=1}^m \frac{\left( X'_{t-1} \dot{A}'_{ij} C_0^{-1} A_0 X_{t-1} \right)^2}{\left( 1 + X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1} \right)^2} \tag{30}
\end{aligned}$$

is infinite. We focus on the first term in (30) since the second is non-negative. Again we consider a fixed but arbitrary realization of  $X_{t-1}$ . To do so, notice that  $h_{ij} : \mathbb{R}^m \rightarrow \mathbb{R}$  defined by

$$h_{ij}(x) = (x' A'_0 C_0^{-1} A_0 x) \left( x' \dot{A}'_{ij} C_0^{-1} \dot{A}_{ij} x \right) - \left( x' \dot{A}'_{ij} C_0^{-1} A_0 x \right)^2$$

is continuous, and hence so is  $\sum_{i=1}^m \sum_{j=1}^m h_{ij}$ . It follows from Appendix B and the Cauchy-Schwarz inequality that  $\sum_{i=1}^m \sum_{j=1}^m h_{ij}(x) > 0, \forall x$  with  $\|x\| = 1$ . Hence,

$$\inf_{x, \|x\|=1} \sum_{i=1}^m \sum_{j=1}^m \left[ (x' A'_0 C_0^{-1} A_0 x) \left( x' \dot{A}'_{ij} C_0^{-1} \dot{A}_{ij} x \right) - \left( x' \dot{A}'_{ij} C_0^{-1} A_0 x \right)^2 \right] > \tilde{K}.$$

Similar as in the proof of part (v) of Theorem 3.1 we obtain for some  $K > 0$ :

$$\begin{aligned}
& \sum_{i=1}^m \sum_{j=1}^m (x' A'_0 C_0^{-1} A_0 x) \left( x' \dot{A}'_{ij} C_0^{-1} \dot{A}_{ij} x \right) - \left( x' \dot{A}'_{ij} C_0^{-1} A_0 x \right)^2 \\
& > K \sum_{i=1}^m \sum_{j=1}^m (x' A'_0 C_0^{-1} A_0 x) \left( x' \dot{A}'_{ij} C_0^{-1} \dot{A}_{ij} x \right).
\end{aligned}$$

By assumption  $A_0$  has full rank and therefore

$$\bar{K} = \inf_{x, \|x\|=1} \left[ (x' A'_0 C_0^{-1} A_0 x) / (1 + x' A'_0 C_0^{-1} A_0 x) \right] > 0.$$

Thus,

$$\begin{aligned}
& \sum_{i=1}^m \sum_{j=1}^m K \frac{(x' A'_0 C_0^{-1} A_0 x) \left( x' \dot{A}'_{ij} C_0^{-1} \dot{A}_{ij} x \right)}{1 + x' A'_0 C_0^{-1} A_0 x} \\
& > \bar{K} K \sum_{i=1}^m \sum_{j=1}^m \left( x' \dot{A}'_{ij} C_0^{-1} \dot{A}_{ij} x \right). \tag{31}
\end{aligned}$$

Finally, notice that by using the left-hand side inequality in (20) we obtain

$$\begin{aligned}
\mathbb{E} \left[ \sum_{i=1}^m \sum_{j=1}^m \left( X'_{t-1} \dot{A}'_{ij} C_0^{-1} \dot{A}_{ij} X_{t-1} \right) \right] & \geq \left[ \sum_{i=1}^m \sum_{j=1}^m \lambda_{\min}(C_0^{-1}) \mathbb{E} \left( X'_{t-1} \dot{A}'_{ij} \dot{A}_{ij} X_{t-1} \right) \right] \\
& \geq \lambda_{\min}(C_0^{-1}) \sum_{i=1}^m \mathbb{E}[x_{t-1,i}^2],
\end{aligned}$$

where  $x_{t-1,i}$  is the  $i$ th component of the random vector  $X_{t-1}$ . Thus, the result follows by using (30), (31) and that  $C_0$  is positive definite.  $\square$

**Remark 4.2** Notice that we did not use that  $A_0$  has full rank when we showed that the existence of the second moment of  $X_{t-1}$  is sufficient for the information matrix to exist. In the necessity part we did use this. However, using the idea of Remark 3.3 the above proof can also be used if  $A_0$  has reduced rank.

**Remark 4.3** Similar to Remark 3.2, we note that the first summand in (28) vanishes when  $X_t$  is a scalar. Concerning the second summand, define

$$\tilde{A} = \frac{1}{2} \dot{A}'_{ij} C_0^{-1} A_0 + \frac{1}{2} A'_0 C_0^{-1} \dot{A}_{ij}.$$

Clearly,  $x' \dot{A}_{ij} C_0^{-1} A_0 x = x' \tilde{A} x$ ,  $\forall x \in \mathbb{R}^m$ . If  $A_0$  has full rank, then the upper bound is given  $\lambda_{\max}(\tilde{D})$ ,  $\tilde{D} = C_0^{1/2} A_0^{-1} \tilde{A} A_0^{-1} C_0^{1/2}$ . Next we consider the case  $\text{rank}(A_0) = r < m$ . Denote by  $USV'$  the singular value decomposition of  $C_0^{-1/2} A_0$ . Let  $\Lambda = S^2$  be the diagonal matrix containing the eigenvalues of the matrix  $A'_0 C_0^{-1} A_0$  in a decreasing order;  $V$  is the matrix of the corresponding eigenvectors. Define the  $m \times r$  matrix  $V_1$ , the columns of which are comprised of the first  $r$  eigenvectors associated to the nonzero eigenvalues, so that  $V$  is partitioned as  $(V_1 : V_2)$ . The inequality

$$\frac{(X'_{t-1} \tilde{A} X_{t-1})^2}{(1 + X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1})^2} \leq \frac{(X'_{t-1} \tilde{A} X_{t-1})^u}{(X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1})^v} \quad (32)$$

is clearly satisfied by  $u = 2, v \leq 2$ , the expectation of the right-hand side of inequality (32) can be bounded using again the theorem in Section 7.2.2. in Roberts (1995) (see also Remark 3.2). Because  $V'_1 \tilde{A} V_2 \neq 0$  but  $V'_2 \tilde{A} V_2 = 0$ , choosing  $v = 3/2 - \delta$ ,  $\delta \in (0, 1/2)$  the inequality  $r > 2v - u$  is satisfied for any  $r \geq 1$ . Under mild conditions on the density of  $X_{t-1}$ , implying the existence of the moment of order  $1 + 2\delta$  (see also Remark 3.2), the theorem in Section 7.2.2. in Roberts (1995) implies that the expectation of the right-hand side of (32) exists.

**Remark 4.4** Establishing asymptotic normality of an extremum estimator requires two steps, asymptotic normality of the score (suitably normalized) at the true parameter value and convergence of the Hessian to a positive definite matrix uniformly in the parameter space. Comte and Lieberman (2003) and Hafner and Preminger (2009b) provide regularity conditions for the QMLE estimator of the VEC-GARCH and BEKK model, which include boundedness of the eighth and sixth moment, respectively, of the observable. These conditions are sufficient to establish uniform convergence of the Hessian, although they could be stronger than effectively needed. On the other hand Theorem 4.1 implies that to establish asymptotic normality of the score one would require at least covariance stationarity of the observables. Further calculations show that boundedness of the Hessian, for all but the true parameter value, require a bounded fourth moment of the observable, casting serious doubts on the possibility of establishing

asymptotic normality with a weaker moment condition. This contrasts with the univariate case when a low order fractional moment condition for the observable suffices for establishing asymptotic normality of the QMLE.

## 5 Conclusion

We have seen that for the simplest BEKK model the finiteness of the variance of the score requires stronger moment assumptions on the observable than we could have expected from the univariate case. However, the normal density based QMLE is not the only possibility. In the univariate case Berkes and Horváth (2004) consider estimators based on other densities than the normal density and show that the moment condition on the unobservable can be weakened compared to the normal density based QMLE. The results in this paper clearly show the necessity to investigate other than the normal density based QMLE in the multivariate case. That will be the subject of a future paper by the authors.

## A Computation of $\text{tr} \left( \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \right)$

Put  $a_0 = A_0 X_{t-1}$  and  $a_{ij} = \dot{A}_{ij} X_{t-1}$ . By the Sherman-Morrison-Woodbury formula

$$\begin{aligned}
& (1 + a_0' C_0^{-1} a_0)^2 \left( \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \right) \\
&= \left[ (\dot{a}_{ij} a_0' C_0^{-1} + a_0 \dot{a}_{ij}' C_0^{-1} + a_0 \dot{a}_{ij}' C_0^{-1} (a_0' C_0^{-1} a_0) - a_0 (\dot{a}_{ij}' C_0^{-1} a_0) a_0' C_0^{-1}) \right]^2 \\
&= \dot{a}_{ij} a_0' C_0^{-1} \dot{a}_{ij} a_0' C_0^{-1} + \dot{a}_{ij} a_0' C_0^{-1} a_0 \dot{a}_{ij}' C_0^{-1} (1 + a_0' C_0^{-1} a_0) - \dot{a}_{ij} a_0' C_0^{-1} a_0 a_0' C_0^{-1} (\dot{a}_{ij}' C_0^{-1} a_0) \\
&+ (1 + a_0' C_0^{-1} a_0) [a_0 \dot{a}_{ij}' C_0^{-1} \dot{a}_{ij} a_0' C_0^{-1} + a_0 \dot{a}_{ij}' C_0^{-1} a_0 \dot{a}_{ij}' C_0^{-1} (1 + a_0' C_0^{-1} a_0) \\
&- a_0 \dot{a}_{ij}' C_0^{-1} a_0 a_0' C_0^{-1} (\dot{a}_{ij}' C_0^{-1} a_0)] \\
&- (\dot{a}_{ij}' C_0^{-1} a_0) [a_0 a_0' C_0^{-1} \dot{a}_{ij} a_0' C_0^{-1} + a_0 a_0' C_0^{-1} a_0 \dot{a}_{ij}' C_0^{-1} (1 + a_0' C_0^{-1} a_0) \\
&- a_0 a_0' C_0^{-1} a_0 a_0' C_0^{-1} (\dot{a}_{ij}' C_0^{-1} a_0)].
\end{aligned}$$

Taking the trace and setting

$$b_0 = a_0' C_0^{-1} a_0, \quad b_{0,ij} = \dot{a}_{ij}' C_0^{-1} a_0, \quad d_{0,ij} = \dot{a}_{ij}' C_0^{-1} a_{ij},$$

yields

$$\begin{aligned}
& (1 + b_0)^2 \text{tr} \left( \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \right) \\
&= b_{0,ij}^2 + (1 + b_0) b_0 d_{0,ij} - b_0 b_{0,ij}^2 + (1 + b_0) b_0 d_{0,ij} + (1 + b_0)^2 b_{0,ij}^2 \\
&- (1 + b_0) b_0 b_{0i}^2 - b_0 b_{0,ij}^2 - (1 + b_0) b_0 b_{0,ij}^2 + b_0^2 b_{0,ij}^2
\end{aligned}$$

which simplifies to

$$\begin{aligned}
& \text{tr} \left( \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \dot{H}_{0t,ij}^{(a)} H_{0t}^{-1} \right) = 2 \frac{b_0}{(1+b_0)} d_{0,ij} + 2 \frac{(1-b_0)}{(1+b_0)^2} b_{0,ij}^2 \\
& = 2 \frac{X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1}}{(1 + X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1})} \left( X'_{t-1} \dot{A}'_{ij} C_0^{-1} \dot{A}_{0,ij} X_{t-1} \right) \\
& + 2 \frac{(1 - X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1})}{(1 + X'_{t-1} A'_0 C_0^{-1} A_0 X_{t-1})^2} \left( X'_{t-1} \dot{A}'_{ij} C_0^{-1} A_0 X_{t-1} \right)^2.
\end{aligned}$$

## B Some linear algebra

**Proposition B.1** *Assume that  $A_0$  has full rank and let  $\dot{A}_{ij} = e_i^{(m)} e_j^{(m)'}.$  Then for every  $x \neq 0$  there is at least one pair  $(i, j), i, j \in \{1, \dots, m\}$  such that the vectors  $A_0 x$  and  $\dot{A}_{ij} x$  are linearly independent.*

**Proof** It is sufficient to consider an arbitrary  $x$  with unit norm. Define

$$I := \{i \in \{1, \dots, m\} | (A_0 x)_i \neq 0\} \quad \text{and} \quad J := \{j \in \{1, \dots, m\} | x_j \neq 0\},$$

where  $z_i$  denotes the  $i$ th element of the vector  $z$ . Notice that both sets are nonempty due to the facts that  $\|x\| = 1$  and that  $A_0$  has full rank. Then for  $I \cap J = \emptyset$  we have that  $\dot{A}_{0,\underline{j}\underline{j}} x$ , where  $\underline{j} \in J$ , and  $A_0 x$  are linearly independent. For  $|I \cap J| = 1$  distinguish the cases that  $|I| = 1$  and that  $|I| > 1$ , where  $|\cdot|$  denotes the cardinality of a set. In the former case take  $\dot{A}_{0,i^c \underline{j}}$ , where  $i^c \notin I$  and  $\underline{j} \in J$ . Then  $\dot{A}_{0,i^c \underline{j}} x$  and  $A_0 x$  are linearly independent. In the latter case take  $\dot{A}_{0,j' \underline{j}}$ , where  $j' \in \{1, \dots, m\}$  is arbitrary and  $\underline{j} \in J$ . With this choice  $\dot{A}_{0,j' \underline{j}} x$  and  $A_0 x$  are linearly independent. Finally, if  $|I \cap J| \geq 2$ , then this implies in particular that  $|I| \geq 2$ . Thus, as before we can take  $\dot{A}_{0,j' \underline{j}}$ , where  $j' \in \{1, \dots, m\}$  is arbitrary and  $\underline{j} \in J$ , and obtain again that  $\dot{A}_{0,j' \underline{j}} x$  and  $A_0 x$  are linearly independent. □

## References

- ABADIR, K. M. AND J. R. MAGNUS (2005): *Matrix Algebra*, Econometric Exercises 1., Cambridge University Press.
- BARDET, J. M. AND O. WINTENBERGER (2009): “Asymptotic normality of quasi-maximum likelihood estimator for multidimensional causal processes,” *Annals of Statistics*, 37, 2730–2759.
- BAUWENS, L., S. LAURENT, AND J. ROMBOUST (2006): “Multivariate GARCH models: a survey,” *Journal of Applied Econometrics*, 21, 79–109.

- BERKES, I. AND L. HORVÁTH (2004): “The efficiency of the estimators of the parameters in GARCH processes,” *Annals of Statistics*, 32, 633–655.
- BERKES, I., L. HORVÁTH, AND P. KOKOSZKA (2003): “Garch processes: structure and estimation,” *Bernoulli*, 9, 201–227.
- BILLINGSLEY, P. (1995): *Probability and Measure*, New York: Wiley.
- BOLLERSLEV, T. (1990): “Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH,” *Review of Economics and Statistics*, 498–505.
- BOLLERSLEV, T., R. ENGLE, AND J. WOOLRIDGE (1988): “A capital asset pricing model with time varying covariances,” *Journal of Political Economy*, 96, 116–131.
- BOUGEROL, P. AND N. PICARD (1992a): “Stationarity of GARCH processes and of some nonnegative time series,” *Journal of Econometrics*, 52, 115–127.
- (1992b): “Strict stationarity of generalized autoregressive processes,” *Annals of Probability*, 20, 1714–1729.
- BOUSSAMA, F., F. FUCHS, AND R. STELZER (2011): “Stationarity and geometric ergodicity of BEKK multivariate GARCH models,” *Stochastic Processes and their Applications*, doi:10.1016/j.spa.2011.06.001, to appear.
- COMTE, F. AND O. LIEBERMAN (2003): “Asymptotic theory for multivariate GARCH processes,” *Journal of Multivariate Analysis*, 84, 61–84.
- ENGLE, R. F. AND K. F. KRONER (1995): “Multivariate Simultaneous Generalized ARCH,” *Econometric Theory*, 11, 122–150.
- FRANCQ, C. AND J.-M. ZAKOÏAN (2004): “Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes,” *Bernoulli*, 10, 605–637.
- (2010a): *GARCH models. Structure, statistical inference and financial applications*, Chichester: John Wiley & Sons.
- (2010b): “QMLE estimation of a class of multivariate GARCH models without moment conditions on the observed process,” MPRA paper 20779, University Library of Munich, Germany.
- HAFNER, C. M. AND A. PREMINGER (2009a): “Asymptotic theory for a factor GARCH model,” *Econometric Theory*, 25, 336–363.
- (2009b): “On asymptotic theory for multivariate GARCH models,” *Journal of Multivariate Analysis*, 100, 2044–2054.
- HARVILLE, D. (1997): *Matrix algebra from a statistician’s perspective*, New York: Springer-Verlag.

- HORN, R. AND C. JOHNSON (1985): *Topics in matrix analysis*, Cambridge: Cambridge University Press.
- JENSEN, S. AND A. RAHBEK (2004a): “Asymptotic inference for nonstationary GARCH,” *Econometric Theory*, 20, 1203–1226.
- (2004b): “Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case,” *Econometrica*, 72, 641–646.
- LEE, S. AND B. HANSEN (1994): “Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator,” *Econometric Theory*, 10, 29–52.
- LUMSDAINE, R. (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*, Chichester: John Wiley & Sons Ltd.
- MAGNUS, J. AND H. NEUDECKER (1999): *Matrix differential calculus with application in Statistics and Econometrics*, New York: John Wiley & Sons.
- NELSON, D. (1990): “Stationarity and persistence in the GARCH(1,1) model,” *Econometric Theory*, 6, 318–334.
- ROBERTS, L. A. (1995): “On the existence of moments of ratios of quadratic forms,” *Econometric Theory*, 11, 750–774.
- ROBINSON, P. M. AND P. ZAFFARONI (2006): “Pseudo-Maximum likelihood estimation for ARCH( $\infty$ ) models,” *Annals of Statistics*, 34, 1049–1074.
- SILVENNOINEN, A. AND T. TERÄSVIRTA (2008): “Multivariate GARCH models,” in *Handbook of Financial Time Series*, ed. by T. Andersen, R. Davis, and J.-P. Stărică, New York: Springer.
- STELZER, R. (2008): “On the relation between the VEC and the BEKK multivariate GARCH models,” *Econometric Theory*, 24, 1131–1136.
- VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge: Cambridge University Press.