ARCH INNOVATIONS AND THEIR IMPACT ON COINTEGRATION RANK TESTING

By Anders Rahbek, Ernst Hansen and Jonathan G. Dennis

Dept. of Statistics and Operations Research, University of Copenhagen
Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

June 11, 2002

Abstract: *In this paper we consider the effect of ARCH innovations on the well-known trace test for the cointegrating rank in VAR models. We show that the cointegration rank test is robust in the sense that the usual assumption of i.i.d. Gaussian errors can be relaxed to a situation where the innovations form an uncorrelated martingale difference sequence with finite second order moments. This includes ARCH and by means of Markov chain theory we can explain findings in simulation studies in the literature on size distortions under ARCH. By the mentioned theory we establish and discuss necessary and sufficient conditions for the multivariate ARCH process to be stationary and geometrically ergodic with finite second order moments. The results are emphasized by simulations of the rank test in the presence of ARCH innovations which illustrates the sufficiency and necessity of the existence of second order moments. Also possible size distortions which can be related to different ways of measuring the impact of ARCH by the size of a certain matrix are discussed. Our focus is on the multivariate BEKK-ARCH process in Engle and Kroner (1995) but we show how the results extend to other types of multivariate ARCH processes appearing in the literature. Finally, rank determination in S&P500 spot and futures data illustrates.*

Keywords: Cointegration, Trace test, Multivariate ARCH, BEKK, Ergodicity.

1. Introduction

The well-known trace test for cointegration rank in the vector autoregressive model is derived in Johansen (1996) under the assumption of independently and identically distributed (i.i.d.) Gaussian errors. Using this test we study the impact of multivariate autoregressive conditional heteroscedastic (ARCH) innovations for cointegration inference in the vector autoregressive (VAR) model. By means of analysis based on Markov chain theory we can explain findings in simulation studies in the literature on size distortions under ARCH. Included simulations illustrate this as well as other points regarding possible size distortions which can be related to measures of size of matrices, see below. Our focus is on the multivariate ARCH process in Engle and Kroner (1995), denoted the BEKK-ARCH process here, but we show how the results extend to the other types of multivariate ARCH processes appearing in the literature.

Denoting the rank test derived as if the errors were i.i.d. Gaussian by the pseudo likelihood ratio or PLR test, we show that indeed – under regularity conditions – the class of ARCH innovations are martingale differences for which the limiting distribution is the same as the one obtained under the i.i.d. Gaussian assumption. The result regarding the limiting distribution when the innovations are martingale differences has been suggested in the literature but to the best of our knowledge nowhere presented with a proof.

What is important is that an essential regularity condition in this respect is the existence of second order moments for the uncorrelated ARCH innovations. We show that the necessary and sufficient condition for the existence of second order moments is that the largest modulus of the eigenvalues or equivalently, that the spectral radius of a certain matrix Φ parametrizing the conditional heteroscedasticity in the BEKK-ARCH process is smaller than one. The proof is partly based on an application of a slightly modified k-step drift criteria in Markov chain theory based on Hansen and Rahbek (1998).

The effect of BEKK-ARCH innovations is illustrated by simulations. In particular, it is illustrated that the PLR statistic indeed diverges as the parameters reach the region where the second order moments cease to exist or, equivalently, the spectral radius of Φ approaches one. We note that an essential step in the proof for geometric ergodicity is based on changing the measure of size of the matrix Φ from norm to spectral radius. The fact that Φ – or any matrix – can have an arbitrarily large norm but small spectral radius is fundamental for the difference between multivariate ARCH processes and, say, univariate ARCH(1) processes. In the latter case the two measures of size are identical. We illustrate in the simulations that the ARCH effect as measured by sprectral radius of Φ is vital, whereas the size of the coefficients as measured by the norm play no role.

This generalizes and at the same time explains findings in the literature where the impact of ARCH has been investigated for the PLR test for the case of diagonal conditional variance with univariate ARCH effects on the diagonal alone as studied in Boswijk, Lucas and Taylor (2002), Lucas (1998) and Ling, Li and Wong (2001). The diagonal cases studied exclude in particular possible ARCH effects involving cross-moments or products, see later.

In addition, our simulations also clearly indicate that the speed of convergence of the PLR test to the limiting distribution as a function of the number of observations is slower than in the case of i.i.d. Gaussian errors. This reflects small-sample problems in addition to the already well-known small sample problems in the i.i.d. case which, under the assumption of Gaussian innovations, recently have lead to Bartlett-type corrections in Johansen (2002).

While our simulations focus on the BEKK-ARCH process, other multivariate ARCH processes in the literature are discussed: The BEKK-GARCH (as opposed to BEKK-ARCH) of Engle and Kroner (1995) for which asymptotic inference in the stationary case is treated in Comte and Lieberman (2001); the constant correlation ARCH process suggested in Bollerslev (1990) and discussed in Jeantheau (1998); and the aforementioned diagonal ARCH process.

Thus the conclusion is that indeed the PLR rank test statistic is robust with respect to ARCH type innovations provided the parameters of the actual ARCH, or martingale difference sequence process in general, are such that second order moments exist. A dominating part of the univariate empirical analyses in the literature find parameter estimates which indeed are in the region where second order moments do not exist (see for instance the massive literature on integrated generalized ARCH, IGARCH). Likewise, the existing and quoted simulations on cointegration rank tests focus on the case of diagonal marginal and conditional variance in the region of IGARCH for each component. On the other hand, other multivariate empirical analyses in the literature as well as our included S&P500 spot and futures illustration, exhibit parameter estimates which appear not to exclude second order moments when applying the results on existence of moments listed here.

This points towards the development of inference for cointegration rank in the case of multivariate ARCH where the ARCH structure is imposed and not ignored when computing rank statistics. In the case of known rank, Li, Ling and Wong (2001) consider estimation of cointegration vectors under the assumption of both diagonal conditional and unconditional variance with ARCH components. Similar to the univariate case of unit-root testing in Seo (1999), their results indicate that rank inference involves limiting distributions with nuisance parameters which are functions of the ARCH parameters. This is also the case for the i.i.d. but non-Gaussian case investigated in Boswijk and Lucas (2002) where more leptokurtosis than Gaussianity is introduced by innovations which are basically mixtures of t type distributions.

With respect to Markov chain theory, note that a k-step type drift criteria similar to Hansen and Rahbek (1998) appears in Tjøstheim (1990) derived under different assumptions. The general idea of using drift criteria to show existence of moments as well as geometric ergodicity for non linear processes is by now well-established in the econometric literature; see for instance Carrasco and Chen (2002) for a recent application to univariate GARCH processes. With respect to the multivariate BEKK-ARCH(1) processes, as opposed to the here considered BEKK-ARCH(p) processes, it first appeared in Hansen and Rahbek (1998).

The point that lack of relevant moments of an underlying ARCH type process leads to distortions of certain test statistics, in particular misspecification tests, is known from the literature on extreme values as in Embrechts, Klüppelberg and Mikosh (1997).

Some notation is used throughout the paper: First regarding norms, with x a vector in $\mathbb{R}^n$, $\|x\|$ denotes the euclidean norm, $\|x\|^2 = \Sigma_{i=1}^n x_i^2$. Denote the space of linear maps from $\mathbb{R}^n \to \mathbb{R}^n$ by $L(\mathbb{R}^n)$. Then with A an $n \times n$ matrix, A represents an element in $L(\mathbb{R}^n)$ and has norm, $\|A\|$ which we choose as $\|A\|^2 = tr(A'A)$. The spectral radius $\rho(A)$ is the maximal modulus of the eigenvalues of A and we note that $\|A\| \geq \rho(A)$. At one or two occasions we shall in addition use the operator norm on matrices as defined by $\|A\|_{\text{op}} := \sup_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|} = \sqrt{\rho(A'A)}$. Also $\|A\|_{\text{op}} \geq \rho(A)$, with equality provided A is symmetric. For linear mappings $\phi : L(\mathbb{R}^n) \to L(\mathbb{R}^m)$ we use the operator norm defined by $\|\phi\| := \sup_{\|X\| \neq 0} \frac{\|\phi(X)\|}{\|X\|}$. As an example, $\phi = (A \otimes A)$ is a linear map from $L(\mathbb{R}^n) \to L(\mathbb{R}^n)$ defined by $(A \otimes A)(X) = AXA'$ for X in $L(\mathbb{R}^n)$. Note that the latter definition closely resembles the well-known practice of using the $vec(\cdot)$ operator.

Second, for any $m \times n$ matrix α of full column rank n, introduce the $m \times (m-n)$ matrix $\alpha_\perp$ of full column rank satisfying $\alpha' \alpha_\perp = 0$.

1.1. Outline of paper

In section 2 we briefly discuss cointegration rank testing in the VAR model and show that the asymptotics hold under general martingale difference assumptions as opposed to the assumption of i.i.d. Gaussian innovations. In section 3 we derive conditions for geometric ergodicity and existence of moments for the BEKK-ARCH process with a thorough discussion of the role of norm and spectral radius. Also simulations are reported in section 3. Section 4 contains a short discussion of other multivariate (G)ARCH processes and mentions empirical (and simulated) results from the literature. An example of rank testing as outlined in the paper is given for S&P500 data. The Appendix contains the majority of proofs.

2. Cointegration test and martingale differences

We consider here the VAR(k) model analysed for cointegration and demonstrate that the analysis in Johansen (1996) inter alia, which is based on i.i.d. Gaussian innovations, is robust to general martingale difference sequences satisfying the regularity condition stated below. In particular, this class of martingale differences will include multivariate ARCH processes considered in the next sections.

2.1. The PLR test and cointegration

The VAR(k) model for the observed p-dimensional process Y_t in error correction form is given by

$$\Delta Y_t = \Pi Y_{t-1} + \textstyle\sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \xi_t \qquad (1)$$

where Π and $(\Gamma_i)_{i=1,\ldots,k-1}$ are $p \times p$ matrices and $t = 1, \ldots, T$. For the derivation of the test statistic and statistical analysis in general the initial values $Y_0^* := (Y_0', Y_{-1}', \ldots, Y_{-k+1}')'$ are considered fixed and the innovations sequence ξ_t is considered to be i.i.d. Gaussian with covariance Ω which is symmetric and positive definite. However, as already emphasized for the asymptotic analysis of test statistics, ξ_t will be a stationary martingale difference sequence satisfying regularity conditions.

For cointegration analysis in the VAR model given by (1) the hypothesis of interest addresses the rank of Π or, equivalently,

$$H(r) : \Pi = \alpha\beta', \qquad (2)$$

where α and β are $p \times r$ matrices not necessarily of full column rank $r \leq p$. Note that here we have not included deterministic terms in the VAR model to keep the presentation simple. This is without loss of generality as all results stated here hold provided the hypothesis of rank reflects the deterministic terms included, see the simulations study later on and Nielsen and Rahbek (2000) for a formulation and discussion of such relevant rank hypotheses.

The PLR test is given by the well-known likelihood ratio test for $H(r)$ against $H(p)$ derived as if ξ_t were i.i.d. Gaussian innovations,

$$PLR := -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i). \qquad (3)$$

Here the eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_r \geq \ldots \geq \hat{\lambda}_p \geq 0$ are the squared empirical canonical correlations between ΔY_t and Y_{t-1} where both have been corrected for lagged differences by ordinary least squares. Equivalently, the $\hat{\lambda}_i$ solve the eigenvalue problem given in equation (22) in the Appendix.

Next, in order to discuss cointegration properties for Y_t under $H(r)$ some assumptions are needed for the autoregressive parameters:

Assumption 1: Assume that $\Pi = \alpha\beta'$ has rank equal to r. Assume further that the characteristic polynomial

$$A(z) = (1-z)I_p - \Pi z - \textstyle\sum_{i=1}^{k-1} \Gamma_i (1-z) z^i, \; z \in \mathbb{C} \tag{4}$$

has exactly $p - r$ roots at $z = 1$ and that the remaining roots are greater than one in absolute value.

Under Assumption 1 and provided ξ_t is a stationary and ergodic sequence Y_t has the Granger-Johansen representation below. The representation states that Y_t under $H(r)$ is non-stationary in the sense that it is integrated of order one, I(1), and cointegrated with cointegration vector β.

Theorem 1. *(Granger-Johansen) Assume that the innovations sequence ξ_t is stationary and ergodic. Under H(r) and Assumption 1, then the process $Y_t^* = (Y'_{t-1}\beta, \Delta Y'_{t-1}, ..., \Delta Y_{t-k+1})'$ can be given an initial distribution such that it is stationary and ergodic. Furthermore,*

$$Y_t = C \sum_{i=1}^{t} \xi_i + \nu_t + \nu \tag{5}$$

where $C = \beta_\perp \left(\alpha'_\perp (I - \Sigma_{i=1}^k \Gamma_i) \beta_\perp\right)^{-1} \alpha'_\perp$. The process $\nu_t = \Sigma_{i=0}^\infty C_i \xi_{t-i}$ is stationary and the C_i matrices decrease exponentially. Finally, the vector ν depends on initial values Y_0^* and satisfy $\beta'\nu = 0$.

Proof of Theorem 1: As $(\xi_t)_{t=1,...}$ is a stationary and ergodic sequence it can be extended to the infinite past and the proof follows from Johansen (1996), proof of Theorem 4.2. □

Note that Y_t is non-stationary due to the presence of $\sum_{i=1}^t \xi_i$. In the i.i.d. Gaussian case $\sum_{i=1}^t \xi_i$ is a random walk. In the case of ARCH innovations the process $\sum_{i=1}^t \xi_i$ is a 'weak random walk' as the innovations are identically distributed and uncorrelated but not independent. Note also that the r linear combinations $\beta' Y_t$ are stationary, and therefore cointegrating, as $\beta'\xi_t$ is a linear process in terms of the ξ_t sequence. Hence the interpretation of cointegration closely resembles the well-known theory.

2.2. The martingale differences and Asymptotics

While stationarity and ergodicity for the ξ_t sequence, as well as restrictions on the VAR parameters, are sufficient for the Granger-Johansen representation to hold, other assumptions are needed for the asymptotic analysis. These are given below and imply essentially (i) existence of second order moments of the uncorrelated sequence ξ_t; (ii) that a law of large numbers and a functional central limit theorem applies to ξ_t and finally; (iii) convergence of appropriate sums to stochastic integrals. The conditions we state are quite general and will be sufficient for the treatment of multivariate ARCH processes. While the regularity conditions for e.g. the central limit theory can be relaxed we note that the assumptions on uncorrelatedness and existence of second order moments can not.

Assumption 2.1: Assume that the sequence $(\xi_t)_{t=1,2,\ldots}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_t$, $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$. Assume further that ξ_t has finite second order moment,

$$E\left(\xi_t \xi_t'\right) = \Sigma. \tag{6}$$

Assumption 2.2: Assume that for any mapping $g : \left(\mathbb{R}^p\right)^{\mathbb{N}_0} \to \mathbb{R}$,

$$\tfrac{1}{T}\Sigma_{t=1}^{T} g\left(\xi_t, \xi_{t-1}, \ldots\right) \overset{P}{\to} Eg\left(\xi_t, \xi_{t-1}, \ldots\right)$$

when the expectation is well-defined. Furthermore, assume that a Lindeberg type condition hold for the ξ_t sequence, i.e. for any $\delta > 0$

$$\tfrac{1}{T}\Sigma_{t=1}^{T} E\left(\left\|\xi_t\right\|^2 1\left\{ \left\|\xi_t\right\| > \delta\sqrt{T} \right\} \middle| \mathcal{F}_{t-1} \right) \overset{P}{\to} 0.$$

The use of the individual assumptions is clear from the proof of Theorem 2 below. By Assumption 2.1, ξ_t is an uncorrelated martingale difference sequence, while Assumption 2.2 implies that the law of large numbers holds for functionals involving moments including moments of order two. Furthermore, the Lindeberg condition implies that the functional central limit theorem applies to ξ_t, and finally, by Hansen (1992) convergence to stochastic integrals is ensured by the uncorrelatedness of the martingale difference sequence.

The main result of this section is the following:

Theorem 2. *Under Assumptions 1, 2.1 and 2.2, the PLR statistic has the same asymptotic distribution as in the i.i.d. Gaussian case.*

The proof is in the Appendix.

As will be seen in the next section Theorem 2 states that cointegration analysis as derived in Johansen (1996) is robust to ARCH innovations with second order moments.

As a corollary we note that this result also holds for hypothesis testing on the cointegration vectors β. Denote the likelihood ratio test for a linear hypothesis on β derived with i.i.d. Gaussian innovations by PLR_β (cf. Johansen, 1996).

Corollary 3. *Under Assumptions 1, 2.1 and 2.2 the PLR_β test statistic has the same asymptotic χ^2 distribution as in the i.i.d. Gaussian case.*

3. BEKK-ARCH(m)

In this section we consider the case where the innovations ξ_t follow the BEKK-ARCH(m) process of Engle and Kroner (1995)

$$\xi_t = \Omega_t^{1/2}\varepsilon_t, \tag{7}$$

$$\Omega_t = \Omega + \Sigma_{i=1}^{m}\Sigma_{j=1}^{\tilde{m}} A_{ij}\xi_{t-i}\xi_{t-i}' A_{ij}', \tag{8}$$

where ε_t is assumed to be $p-$dimensional i.i.d. with mean zero, variance I_p and positive continuous[1] density $f(\cdot)$. As a simple and natural example ε_t can be i.i.d. Gaussian, $N(0, I_p)$. The variance parameter Ω is symmetric and positive definite and the A_{ij} matrices are real $p \times p$ matrices. Note that the dependence on $\tilde{m}$ is suppressed in the abbreviation ARCH(m).

We derive necessary and sufficient conditions for existence of second order moments. Furthermore, we establish geometric ergodicity and relate this to Assumptions 2.1 and 2.2 in the previous section. The difference between measuring matrix size by a matrix norm and spectral radius respectively turns out to be of vital importance. This is demonstrated in the proof as well as in the included simulations. Specifically, size distortions and divergence of the PLR test occurs as the spectral radius tends to one. By fixing the spectral radius at a value smaller than one, and letting the norm increase, we see that the latter plays little, if any, role.

3.1. Ergodicity and existence of moments

To establish the conditions under which the ξ_t process satisfies Assumptions 2.1 and 2.2 we proceed by showing geometric ergodicity of the ξ_t sequence using a drift criterion which at the same time ensures existence of second order moments.

[1] or e.g. lower-semicontinuous

Our focus in this section will be on the step in the derivation where the two different measures of matrix size, norm and spectral radius respectively, occur. The additional technical arguments needed are placed in the Appendix. For applications of drift criteria on multivariate chains in general see inter alia Feigin and Tweedie (1985), Pham (1986), Tjøstheim (1990) and with focus on ARCH processes, Hansen and Rahbek (1998).

With ξ_t defined in (7) consider the Markov chain

$$x_t = (\xi_t', \xi_{t-1}', ..., \xi_{t-m+1}')' \tag{9}$$

on $\mathbb{R}^{pm}$ endowed with the Borel σ-algebra. By the factorization of the density of x_{t+m} conditional on x_t,

$$g\left(x_{t+m} \middle| x_t\right) = \Pi_{i=1}^{m} f\left(\xi_{t+i} \middle| \xi_{t-1+i,...,}\xi_{t-m+i}\right), \tag{10}$$

the m-step transition density is continuous and positive as ε_t is assumed to be i.i.d. $(0, I_p)$. By Lemma 11 in the Appendix this implies that x_t is aperiodic, irreducible with respect to the Lebesgue measure and that compact sets in $\mathbb{R}^{pm}$ are small. This again implies that we can use the k-step drift criterion from Hansen and Rahbek (1998) with the drift function

$$v(x) = 1 + \|x\|^2. \tag{11}$$

More precisely, in terms of the present set-up:

Theorem 4. *(Hansen and Rahbek, Tjøstheim) Let $(x_t)_{t=0,1,..}$ be a time homogeneous Markov chain on $\mathbb{R}^d$ endowed with the Borel σ-algebra $\mathbb{B}^d$ for which the m-step transition density is continuous and positive. Let $v : \mathbb{R}^d \mapsto [1, \infty]$ be some drift function. Assume there exists an integer $k \geq 1$, a compact set $K \subset \mathbb{R}^d$ and constants $0 < \gamma < 1$, $g > 0$ such that*

$$E\left(v(x_{t+k}) \middle| x_t = x\right) \leq \gamma v\left(x\right) \tag{12}$$

for x in K^c, while $E\left(v(x_{t+k} \middle| x_t = x\right)$ is bounded by g on K.

Then x_t is geometrically ergodic and x_0 can be given an initial distribution such that x_t is stationary. Furthermore, all moments bounded by $v(\cdot)$ exist.

The proof is given in the Appendix.

So provided we show that the bounds in Theorem 4 hold for the drift function in (11) we immediately have sufficient conditions for Assumptions 2.1 and 2.2 to hold: Specifically, ξ_t is a martingale difference with respect to $\mathcal{F}_t = \sigma(\xi_t, \xi_{t-1}, ..)$ and it has second order moments. Provided $Eg(\xi_t, \xi_{t-1}, ..)$ is finite the law of large numbers holds and, finally, the Lindeberg condition holds by a simple application of the Cauchy-Schwarz inequality.

We first address sufficient conditions in Theorem 5 below, while necessity is discussed seperately in Theorem 6 afterwards.

Theorem 5. *Consider the BEKK-ARCH(m) process ξ_t defined in (7). Then a sufficient condition for geometric ergodicity, stationarity and existence of second order moments of ξ_t is that $\rho(\Phi) < 1$. Here the $mp^2 \times mp^2$ matrix Φ is defined by*

$$\Phi = \begin{pmatrix} \Sigma_{j=1}^{\tilde{m}}(A_{1j} \otimes A_{1j}) \cdots \Sigma_{j=1}^{\tilde{m}}(A_{m-1j} \otimes A_{m-1j}) & \Sigma_{j=1}^{\tilde{m}}(A_{mj} \otimes A_{mj}) \\ (I_{p(m-1)} \otimes I_{p(m-1)}) & (0_{p(m-1)\times p} \otimes 0_{p(m-1)\times p}) \end{pmatrix}. \tag{13}$$

Before turning to the proof note that in the case where $m = \tilde{m} = 1$, then $\Phi = (A_{11} \otimes A_{11})$, while if $m = 2$ and $\tilde{m} = 1$,

$$\Phi = \begin{pmatrix} A_{11} \otimes A_{11} & A_{21} \otimes A_{21} \\ I_p \otimes I_p & 0 \otimes 0 \end{pmatrix}$$

Proof of Theorem 5: As noted above we shall focus on the role of the measure of matrix size by spectral radius and norm respectively.

For an element X in $L(\mathbb{R}^{pm})$ identify it with the $pm \times pm$ matrix divided into m^2 blocks of size $p \times p$, $X = (X_{ik})_{i,k=1,..,m}$ and define the important mapping $\phi : L(\mathbb{R}^{pm}) \to L(\mathbb{R}^{pm})$ by

$$\phi(X) = \begin{pmatrix} \Sigma_{i=1}^{m}\Sigma_{j=1}^{\tilde{m}} A_{ij} X_{ii} A_{ij}' & 0 \\ 0 & (I_{p(m-1)}, 0) X (I_{p(m-1)}, 0)' \end{pmatrix}. \tag{14}$$

With this definition turn to sufficiency of the condition $\rho(\phi) < 1$: Applying the drift function $v(\cdot)$ in (11) along with the definition of x_t in (9) immediately leads to

$$\begin{aligned} E(v(x_{t+1})|\, x_t) &= 1 + E\left(tr\left\{x_{t+1}x_{t+1}'\right\}\middle|\, x_t\right) = 1 + tr\left\{\phi(x_t x_t')\right\} + tr\{\tilde{\Omega}\} \\ E(v(x_{t+1})|\, x_{t-1}) &= 1 + tr\{\tilde{\Omega}\} + E\left(tr\left\{\phi(x_t x_t')\right\}|\, x_{t-1}\right) \\ &= 1 + tr\{\tilde{\Omega}\} + tr\{\phi(\tilde{\Omega})\} + tr\left\{\phi^2\left(x_{t-1}x_{t-1}'\right)\right\} \end{aligned}$$

where $\tilde{\Omega}$ =blockdiag$\big(\Omega, 0_{p(m-1)\times p(m-1)}\big)$. Clearly this can be used inductively by successive conditioning to obtain

$$\begin{aligned} E(v(x_{t+k})|\, x_t = x) &= 1 + E\left(tr\left\{x_{t+k}x_{t+k}'\right\}\middle|\, x_t = x\right) \\ &= 1 + tr\left\{\phi^k(xx')\right\} + \Sigma_{i=0}^{k-1} tr\{\phi^i(\tilde{\Omega})\}. \end{aligned} \tag{15}$$

The central argument for (12) to hold concerns finding an upper bound for the term $tr\left\{\phi^k(xx')\right\}$. In terms of the operator norm

$$tr\left\{\phi^k(xx')\right\} = \left|tr\left\{\phi^k(xx')\right\}\right| \le \kappa \left\|\phi^k(xx')\right\| \le \kappa \left\|\phi^k\right\| \|xx'\| = \kappa \left\|\phi^k\right\| \|x\|^2$$

for some positive constant κ. We have used that $\phi(xx')$, and hence $\phi^k(xx')$, are symmetric and positive semidefinite.

Obviously, in the first step, where $k = 1$, $tr\{\phi(xx')\}$ is bounded from above by $\gamma \|x\|^2$ by choosing the norm of ϕ small. However, this will lead to a too strong restriction on the parameters. Instead the relevant measure for ϕ is the spectral radius

$$\lim_{k\to\infty} \left\|\phi^k\right\|^{1/k} = \rho(\phi),$$

see Pedersen (1988, Theorem 4.1.13). Hence if $\rho(\phi) < 1$, then for some k large enough the norm of ϕ^k is arbitrarily small irrespectively of the size of $\|\phi\|$.

The proof of sufficiency is now concluded by choosing k such that $\kappa\left\|\phi^k\right\| = \tilde{\gamma} < 1$. Next, define $c(\tilde{\gamma}) = 1 - \tilde{\gamma} + \Sigma_{i=0}^{k-1} tr\{\phi^i(\tilde{\Omega})\}$ and choose $\tilde{\gamma} < \gamma < 1$. Then

$$E\left(v\left(x_{t+k}\right)\mid x_t = x\right) \le \gamma v(x) \quad \text{for } x \in K^c$$

where the compact set K is defined by

$$K = \left\{x \mid v(x) \le c(\tilde{\gamma})/(\gamma - \tilde{\gamma})\right\}.$$

For computational purposes note that $\rho(\phi) = \rho(C_0 \otimes C_0 + \Sigma_{i=1}^{m} \Sigma_{j=1}^{\tilde{m}} (C_{ij} \otimes C_{ij}))$ where the $pm \times pm$ matrices C_{ij} are given by

$$C_{ij} = \begin{pmatrix} 0\cdots 0 & A_{ij} & 0\cdots 0 \\ 0\cdots & & \cdots 0 \end{pmatrix} \quad \text{for } i = 1, 2, .., m \text{ and } j = 1, 2, ..., \tilde{m}$$

such that C_{ij} is the zero matrix except for the entries in the first m rows and columns i to $i + m$ which contain the A_{ij} matrix. C_0 is the zero matrix except for the lower left corner which contains the $p(m-1)$-dimensional identity matrix. Finally note that $\rho(C_0 \otimes C_0 + \Sigma_{i=1}^{m} \Sigma_{j=1}^{\tilde{m}} (C_{ij} \otimes C_{ij})) < 1$ is equivalent to $\rho(\Phi) < 1$ with Φ defined in (13). This ends the proof of Theorem 5. □

Next we turn to necessary conditions for existence of second order moments.

Theorem 6. *Consider the the BEKK-ARCH(m) process ξ_t defined in (7). For ξ_t to have a finite and well-defined finite second order moment, $E(\xi_t \xi_t') = \Sigma$, and thereby be weakly stationary, it is necessary that the spectral radius $\rho(\Phi|_{sym}) < 1$. The spectral radius $\rho(\Phi|_{sym})$ is the maximal absolute value of the eigenvalues of Φ corresponding to (the vec) of symmetric matrices, while Φ is defined in Theorem 5.*

The proof is given in the Appendix.

Recall that in the case of a BEKK-ARCH(1) process with $m = \tilde{m} = 1$ Φ will be identical to $(A \otimes A)$ for some A. For use in the simulations we note that in this case $\rho(\Phi|_{sym}) = \rho(\Phi)$. In other words, the restriction to the space of symmetric matrices does not matter when computing the maximal eigenvalues and hence $\rho(\Phi) < 1$ will be both necessary and sufficient for $\Phi = (A \otimes A)$.

3.2. Simulations

In the simulations we illustrate that as the spectral radius $\rho(\Phi)$ approaches (and exceeds) one, the PLR test diverges and the size is distorted. The simulations also illustrate that the ARCH effect as measured by the matrix norm of Φ is irrelevant for the convergence and only $\rho(\Phi)$ is relevant. Finally, the simulations show a clear shift in small sample properties of the ARCH-model when compared to the i.i.d. case.

3.2.1. Processes and design

For the choice of PLR test and process under the null hypothesis, we want the simplest possible system in which the effect of ARCH, as measured by spectral radius and norm of Φ in Theorem 5 can be studied.

We do not focus on small sample problems created by the VAR parameters as these, under the assumption of i.i.d. Gaussian innovations, are treated in detail by Johansen (2002), who considers Bartlett type corrections of the trace test. Thus to keep size distortions from the VAR parameters minimal we investigate the PLR test of the null of $H(0)$ against the unrestricted alternative $H(p)$ in the VAR model of order one, VAR(1). In the notation of section 2.1 the VAR(1) model is here given by,

$$\Delta Y_t = \Pi(Y_{t-1} - \mu) + \xi_t \tag{16}$$

for $t = 1, ..., T$. Note that a parameter $\mu \in \mathbb{R}^p$ allowing for a non-zero level of the process has been added. The PLR test is still given by (3), but with a small change in the eigenvalue problem in equation (22) in the appendix as a constant is added in the reduced rank regression, see e.g. Johansen (1996). Observe that under $H(0)$, $\Pi = 0$ and $\Delta Y_t = \xi_t$ or $Y_t = \Sigma_{t=1}^{t}\xi_i + Y_0$, while under $H(r)$ and Assumption 1, $\Pi = \alpha\beta'$ and Y_t is non-stationary with cointegration relations $\beta' Y_t$ with mean $E(\beta' Y_t) = \beta'\mu$ which in general is non-zero. As discussed in Nielsen and Rahbek (2000), the trace test is asymptotically similar in this case. Thereby our choice of VAR(1) model, and hence PLR test statistic, in particular reduces the effect or influence of the initial value Y_0 of Y_t under the null hypothesis in the simulations. This will be important for the simulations of the ARCH innovations (as well), see below. We also note that for the i.i.d. Gaussian case the effect of the common trends dimension as given by $p-r$, or p in the case of $r = 0$, is studied as well in Johansen (2002). So we restrict attention the bivariate case where $p = 2$ in order to focus on ARCH effects.

To sum up, in the simulations the trace or PLR test of the null $H(0)$ against the alternative $H(2)$ is computed for the VAR(1) model in (16). The Y_t process is simulated under the null as we are concerned alone with size or rejection probabilities as well as divergence of the test.

Next, turn to the simulations of the BEKK-ARCH innovations ξ_t. We use a bivariate BEKK-ARCH(1) for ξ_t, i.e.

$$\begin{aligned} \xi_t &= \Omega_t^{1/2}\varepsilon_t \\ \Omega_t &= \Omega + A\xi_t\xi_t' A', \end{aligned} \tag{17}$$

where ε_t is an i.i.d. $N(0, I_2)$ sequence. As ARCH loading matrix A we choose an upper–triangular form

$$A = \begin{pmatrix} a_1 & \lambda \\ 0 & a_2 \end{pmatrix}, \tag{18}$$

where a_1, a_2 and λ are real numbers. In terms of the notation of Theorem 5 and Theorem 6

$$\begin{aligned} \Phi &= A \otimes A, \\ \|\Phi\|^2 &= \left(a_1^2 + a_2^2\right)^2 + \lambda^2\left(\lambda^2 + 2\left(a_1^2 + a_2^2\right)\right) \text{ and} \\ \rho(\Phi) &= \rho(\Phi|_{sym}) = \max\left\{a_1^2, a_2^2\right\}. \end{aligned} \tag{19}$$

By the specific choice of A in (18) the spectral radius $\rho(\Phi)$ is independent of λ. At the same time, the ARCH effect as measured by the norm of Φ is increasing in λ, that is, $\|\Phi\| \to \infty$ as $|\lambda| \to \infty$, independently of a_1 and a_2. Together this makes it simple to keep either of the two measures of matrix size fixed while varying the other. Notice that due to norm equivalence we use the norm $\|\Phi\|$ as defined in the introduction and not the operator norm which appears in the proof of Theorem 5.

With respect to Ω in (17) we set

$$\Omega = I_2.$$

This choice is based on two observations: First, clearly Ω, and hence also Ω_t, is positive definite for all t which is vital in the assumptions used for the application of the drift criteria. Second, given that Ω is positive definite (and symmetric) the PLR test does not depend on Ω which can be seen by simply redefining ξ_t (or Y_t) as $\Omega^{-1/2}\xi_t$ (or $\Omega^{-1/2}Y_t$).

When keeping either the spectral radius $\rho(\Phi)$ or the norm $\|\Phi\|$ fixed, the simulated distributions of the PLR test are based on computing the trace statistic for $N = 10000$ realisations of the process $(Y_t)_{t=1,\ldots,T}$ under the null and with a sample length within each realisation of $T = 1000$. The same i.i.d.$N(0, I_2)$ sequence ε_t

has been used with the simulations for fixed spectral radius and norm respectively. For each realisation of ξ_t we simulate a series of length 2000 discarding the initial 1000 observations. This way the ξ_t series can be considered as initiated from its invariant distribution and possible 'burn-in' effects can be ignored. As ξ_0 is a function of ΔY_0 (and of ε_0) this implies especially that Y_t is simulated with fixed initial values Y_0 and ΔY_0, where the latter in general is non-zero by construction of the ξ_t sequence. As mentioned above, the chosen version PLR test is similar with respect to the choice of initial value of Y_t so that we ignore possible effects of initial values. Unreported simulations on the basis of the PLR test in the VAR model with no level parameter were highly sensitive to the choice of initial values of Y_t.

The 90%, 95% and 99% quantiles of the simulated finite sample distribution are compared to asymptotic quantiles listed in MacKinnon et al. (1999) using asymptotic confidence bands from Dudewicz and Mishra (1988). All simulations were performed in RATS 5.0 and a RATS program can be obtained from the authors.

3.2.2. The role of spectral radius

Turning first to the role of the spectral radius of Φ we consider simulations of ξ_t in (17) for fixed norm $\|\Phi\| = 1$ and varying spectral radius $\rho(\Phi) \leq 1$. For $\rho(\Phi)$ exceeding 1 we set $\|\Phi\| = \rho(\Phi)$ i.e. the smallest compatible norm.

	$\rho(\Phi)$									
Level	i.i.d.	0.1	0.2	0.4	0.6	0.8	1.0	1.1	1.2	1.3
1%	0.98	1.21	1.01	1.26	1.34	1.92	3.30	4.39	5.83	7.04
5%	5.14	5.63	5.47	5.73	6.04	6.69	8.90	10.35	11.99	13.81
10%	10.56	10.80	10.11	10.98	11.29	12.07	14.63	16.17	17.96	20.49

Table 1: *Simulated rejection probabilities based on* 10000 *simulations of the PLR-test for $H(0)$ against $H(2)$. Sample size is $T = 1000$. More simulations are located in Table 6.*

The simulation results are reported in Table 1 and clearly show how size distortion, as measured by rejection probabilities, grows when increasing the ARCH-effect. In addition the QQ-plot in Figure 1 illustrates convergence of the PLR test for $\rho(\Phi) < 1$ and divergence otherwise. This demonstrates the necessity of the condition in Theorem 5.

Unreported simulations confirm these findings in case of unit root testing in one dimension where $\rho(\Phi) = ||\Phi|| = |a_1|$. This corresponds to the simulations of near-integrated models in e.g. Kim and Schmidt (1993) and Lucas (1998).

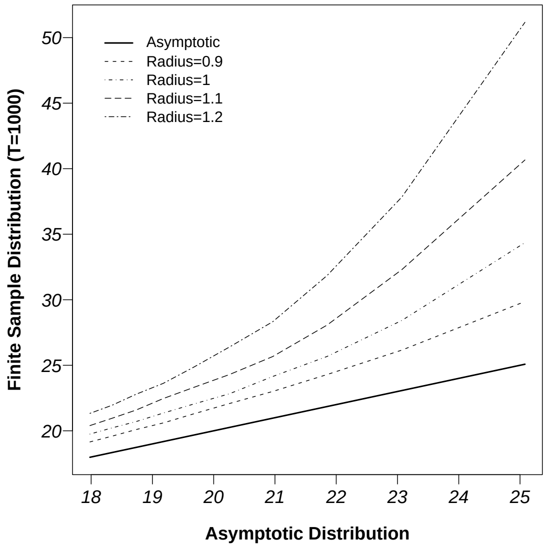

Figure 1: *Finite sample distribution of the simulated PLR test statistic for* $H(0)$ *against* $H(2)$ *plotted for increasing values of* $\rho(A \otimes A)$ *(denoted 'Radius' in graph header) and* $T = 1000$. *The graphed interval covers the 90% to 99% quantiles. Note that the difference between the asymptotic and the finite sample distribution is increasing as a function of the spectral radius. For radii larger than one there is no sign of convergence.*

3.2.3. The role of the norm

In order to investigate the effect of the norm $||\Phi||$ we consider simulations of ξ_t in (17) where $a_1 = a_2 = 1/2$, such that the spectral radius is fixed at $\rho(\Phi) = 1/4$ but with λ, and hence the norm, varying.

In Table 2 the size of the PLR test are reported for increasing values of the norm and a sample size of $T = 1000$ and we note in particular that for any value of the norm, the simulations indicate no distortion of the test size as measured by simulated rejection probabilities. For smaller samples rejection probabilities are reported in Table 7, showing that the distortion is negligible and that for $T = 100$ or more, the test has the correct rejection probability.

This is also illustrated in Figure 2 which shows QQ-plots of the simulated distribution for different values of the norm plotted against the asymptotic distribution where we in particular note the lack of effect of the norm of Φ.

Hence the simulations clearly suggest that the norm of Φ plays little, if any, role with regards to the size and convergence of the trace test.

	$\|\Phi\|$									
Level	i.i.d.	1	2	4	6	8	10	20	30	40
1%	0.98	1.09	1.12	1.29	1.08	1.09	1.10	1.09	1.17	1.29
5%	5.14	5.24	5.49	5.74	5.35	5.16	5.32	5.16	5.27	5.49
10%	10.56	10.08	10.51	10.78	10.42	10.37	10.52	10.37	10.64	10.87

Table 2: *Empirical rejection probabilities based on* 10000 *simulations of the PLR-test for* $H(0)$ *against* $H(2)$. *Sample size is* $T = 1000$. *More simulations are located in Table 7.*

3.2.4. Small sample properties

Figure 3 shows how simulated 95% quantiles converge to the asymptotic 95% quantile as a function of the log of the sample size T. Table 3 reports the number of observations needed for the asymptotic quantile to lie in the confidence band of the simulated quantile. For comparsion with the i.i.d. Gaussian case the case of $\rho(\Phi) = 0$ or equivalently $A = 0$ in (17) is included. Both Figure 3 and Table 3 are based on the simulations in Table 5.

	$\rho(\Phi)$			
Quantile	0.0	0.1	0.5	1.0
90%	60	225	+1 000	+1 000
95%	60	250	650	+1 000
99%	70	275	+1 000	+1 000

Table 3: *Smallest sample for which the asymptotic quantiles are within the* 95% *confidence interval of the simulated quantiles for increasing values of* $\rho(\Phi)$ *and fixed* $\|\Phi\| = 1$.

As illustrated in Figure 3 as well as in Table 3 comparing $\rho(\Phi) = 0$ with the cases of $\rho(\Phi) > 0$ clearly indicate small sample problems which can be attributed to ARCH effects alone. Measured in terms of the aforementioned asymptotic confidence bands the simulations suggest that for a sample size of 800 or more, the empirical and asymptotic 95% quantiles of the PLR test distribution are indistiguishable, see Table 5.

4. Comments on other ARCH processes and empirical illustrations

Various multivariate ARCH type models have been suggested in the literature of which the predominant ones are the BEKK-GARCH of Engle and Kroner (1995),

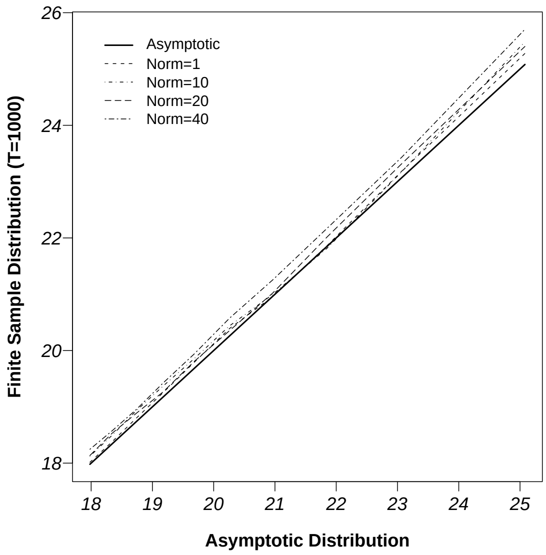

Figure 2: *Simulated distribution of the PLR test statistic for* $H(0)$ *against* $H(2)$ *plotted for increasing values of* $\|A \otimes A\|$ *(denoted 'Norm' in graph header) and sample length* $T = 1000$. *The graphed interval covers the 90% to 99% quantiles. Note that lack of difference between the asymptotic and the simulated finite sample distribution regardless of the size of the norm.*

Constant Correlations of Bollerslev (1990) and Diagonal ARCH models studied in e.g. Li, Ling and Wong (2001). We shall briefly discuss these; in each case with focus on which matrix plays the role of Φ, the matrix parametrizing the ARCH effects discussed above for the BEKK-ARCH process. With relation to existence of second order moments and geometric ergodicity, we briefly comment on illustrations in the literature as well as the included S&P500 illustration. In the S&P500 illustration we also illustrate cointegration rank testing.

4.1. Moments and ergodicity

Asymptotic theory for the BEKK-GARCH process suggested in Engle and Kroner (1995) has recently been studied in Comte and Lieberman (2001). The BEKK-GARCH is given by a natural extension of the conditional variance in (8) as

$$\begin{aligned}\xi_t &= \Omega_t^{1/2}\varepsilon_t,\\ \Omega_t &= \Omega + \Sigma_{i=1}^{m}\Sigma_{j=1}^{\tilde{m}}(A_{ij}\xi_{t-i}\xi_{t-i}'A_{ij}' + B_{ij}\Omega_{t-i}B_{ij}')\end{aligned} \tag{20}$$

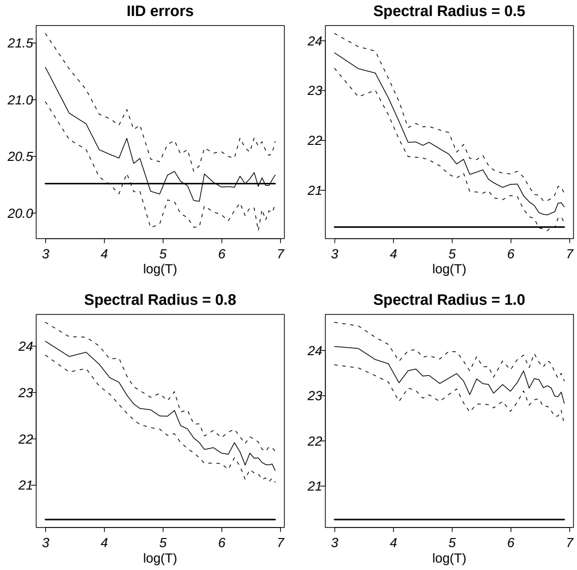

Figure 3: *The 95% quantile of the simulated distribution of the PLR test of* $H(0)$ *against* $H(2)$ *plotted as a function of* $\log T$, $T = 10, \ldots, 1\,000$ *for different values of the spectral radius* $\rho(\Phi)$*. The dotted line is the asymptotic 95% quantile.*

where as before ε_t is i.i.d. $(0, I_p)$, the variance parameter Ω is symmetric and positive definite and A_{ij} and B_{ij} are real $p \times p$ matrices. Consider Φ as given by (13) but with each entry $(A_{ij} \otimes A_{ij})$ replaced by

$$(A_{ij} \otimes A_{ij}) + (B_{ij} \otimes B_{ij}).$$

One finds, provided the drift criterion can be applied, that $\rho(\Phi)$ should be smaller than one for existence of second order moments. In this case ξ_t will also be geometrically ergodic. Comte and Lieberman (2001) quote results in Boussama (1998) where in fact a similar condition appears for geometric ergodicity.

As for the diagonal ARCH process in Li, Ling and Wong (2001) and diagonal GARCH process used for simulations in Boswijk, Lucas and Taylor (2000) these are straightforward examples of the BEKK-ARCH and BEKK-GARCH processes. Specifically, Ω_t in (8) and (20) is assumed to be diagonal with each element on the diagonal, σ_t^{ii} say, a univariate ARCH or GARCH process respectively.

Finally, Jeantheau (1995) and Ling and McAleer (2001) study the Constant Correlation GARCH process suggested in Bollerslev (1990). In this case the p-dimensional conditional variance in (20) is given by the entries

$$(\Omega_t)_{ij} = \begin{cases} \sigma_{it}\sigma_{jt}\sigma_{ij}^2 & \text{for } i \neq j \\ \sigma_{it}^2 & \text{for } i = j \end{cases}$$

where σ_{ij}^2 is strictly positive and σ_{it}^2 is a univariate GARCH process

$$\sigma_{it}^2 = \sigma_i^2 + \Sigma_{j=1}^m (a_{ij}\xi_{t-j}^2 + b_{ij}\sigma_{it-j}^2),$$

where a_{ij} and b_{ij} are greater than or equal to zero. Note that by definition the conditional correlation between ξ_{it} and ξ_{jt} is given by σ_{ij}^2. Define the $p \times p$ diagonal matrices A_j =diag$(a_{ij})_{i=1,\dots,p}$ and B_j =diag$(b_{ij})_{i=1,\dots,p}$ for $j = 1, .., m$. Then the Φ for which $\rho(\Phi) < 1$ is the relevant condition can be written in the form

$$\Phi = \begin{pmatrix} (A_1 + B_1) \cdots (A_{m-1} + B_{m-1}) & (A_m + B_m) \\ I_{p(m-1)\times p(m-1)} & 0_{p(m-1)\times p} \end{pmatrix}.$$

For the case of no lagged variances, i.e. all $B_j = 0$, the Constant Correlations ARCH process, or rather the implied Markov chain, satisfies the regularity conditions for applying our drift criterion in Theorem 4. Applying simple algebra will lead to the condition that $\rho(\Phi) < 1$ for the just given Φ matrix. Due to the nature of the state space inherited from the univariate specification of the $\sigma_{it}'s$ the drift criterion can also be applied for the case of Constant Correlation GARCH, see also Carrasco and Chen (2002) and Ling and MacAleer (1999).

4.2. Brief comments on empirical illustrations

Using the spectral radius of the individual $\Phi's$ reported above as an indicator for the existence of second order moments, we find mixed results in the empirical applications of multivariate (G)ARCH processes quoted in the literature. In the univariate cases, and therefore not surprisingly for the case of Diagonal as well as Constant Correlations GARCH models, we observe that $\rho(\Phi)$ with Φ replaced by the estimated values are usually close to one corresponding to integrated GARCH, see for instance results in Tse (2000) and Bollerslev (1990). Correspondingly most simulations with regards to cointegration, including the ones in Boswijk, Lucas and Taylor (2000), use Diagonal (parameter region close to) integrated GARCH. On the other hand, e.g. Li, Ling and Wong (2001) use Diagonal ARCH and there $\rho(\Phi)$ seems much smaller than one.

Below we consider estimation and rank testing in a VAR(1) model with BEKK-ARCH(1) innovations for S&P500 data and we find that $\rho(\hat{\Phi}) < 1$.

Specifically, we consider daily observations of S&P500 index and the S&P futures contract traded at Chicago Mercantile Exchange. Data covers the period 1.1.98 to 31.1.02 totalling 1003 observations. On the basis of the daily prices a futures series F_t for the entire period is obtained by rolling over the price series of the individual contracts using open interest. As in Kessler and Rahbek (2001) the well-known no-arbitrage, cost-of-carry argument suggests that the log of the discounted futures series, denoted f_t, should cointegrate with the logarithmic spot

$H(r)$	PLR	95%
0	675.4	20.0
1	5.2	9.2

Table 4: *PLR test for the cointegrating rank.*

price s_t. For discounting we use interpolated[2] USD LIBOR rates for approximating the continuously compounded rate.

With $Y_t = (s_t, f_t)'$ we use the bivariate VAR(1) model in (16) as given by,

$$\Delta Y_t = \alpha(\beta' Y_{t-1} - \mu) + \xi_t$$

under $H(r)$. The extension with the level parameter μ does not change anything with respect to the conclusions but allows for (the observed) non-zero initial level of Y_t. Using the PLR test and Table 15.2 in Johansen (1996) we find that the hypothesis of rank cointegration $r = 1$ is accepted; see Table 4. Also the hypothesis of cointegration vector $\beta = (1, -1)'$ based on the PLR_β test is accepted.

As mentioned above, the VAR(1) model is estimated with a BEKK-ARCH(1) specification applied to ξ_t:

$$\xi_t = \Omega + A\xi_{t-1}\xi_{t-1}'A'.$$

The maximum likelihood estimators and their corresponding empirical standard errors of the VAR parameters under the hypothesis of rank $r = 1$ are in this case given by:

$$\hat{\alpha}(\hat{\beta}' Y_t - \hat{\mu}) = \begin{pmatrix} \underset{[.0039]}{0.0229} \\ \underset{[.0338]}{0.962} \end{pmatrix} (s_t - \underset{[.0006]}{0.9968} f_t + \underset{[.0043]}{0.0226})$$

while the significant maximum likelihood estimators of the ARCH parameters are as follows

$$\hat{\Omega} = \hat{C}\hat{C}', \ \hat{C} = \begin{bmatrix} \underset{[.0005]}{0.0035} & \underset{[.0005]}{0.0055} \\ \underset{[.0003]}{-0.0123} & \underset{[.0004]}{-0.0119} \end{bmatrix} \text{ where } \hat{A} = \begin{bmatrix} \underset{[.0975]}{0.1642} & \underset{[.1048]}{-0.4347} \\ \underset{[.0417]}{-0.0700} & \underset{[.0592]}{0.5249} \end{bmatrix}.$$

The spectral radius of $\hat{\Phi}$, $\rho(\hat{\Phi})$, equals $0.3546 < 1$ while the norm $||\hat{\Phi}||$ equals 0.4963. Thus both values seem to indicate that the rank determination can be based on the PLR test.

[2]Data and further details can be obtained from the authors.

5. Conclusion

The outcome of our analysis is that indeed cointegration rank testing in the VAR model based on well-known asymptotic inference for the PLR test is valid for ARCH processes and in general for martingale differences. We demonstrate through analysis and simulations that apart from uncorrelatedness of the innovations in the VAR model, the cruical assumptions are that the innovations pocess second order moments and that the spectral radius of the Φ matrix parametrizing the ARCH effect should be smaller than one. Based on the comments on empirical work it seems that estimation of ARCH models such as the Diagonal and Constant Correlations types, which as argued are essentially univariate, lead to parameter estimation in the region where second order moments do not exist hence invalidating the asymptotic inference. On the other hand, estimation of multivariate ARCH models such as the BEKK model indicate parameter estimates that do imply second order moments for the underlying ARCH process. As already mentioned, this points toward rank inference based on full maximum likelihood analysis.

As a curiosity, with respect to the BEKK-ARCH process the regularity condition addressing existence of second order moments is slightly different from the one stated in Engle and Kroner (1995) for general lags. More precisely we can state the following corollary of Theorem 5:

Corollary 7. *The condition that* $\rho(\Phi) < 1$ *in Theorem 5 is equivalent to*

$$\left| I_{p^2} - \Sigma_{i=1}^{m} \Sigma_{j=1}^{\tilde{m}} \left(A_{ij} \otimes A_{ij} \right) z^i \right| = 0 \Rightarrow |z| > 1$$

The condition stated in Engle and Kroner (1995), based on an argument for the $m = 1$ case, is that

$$\left| I_{p^2} - \Sigma_{i=1}^{m} \Sigma_{j=1}^{\tilde{m}} \left(A_{ij} \otimes A_{ij} \right) z \right| = 0 \Rightarrow |z| > 1.$$

We note that the two conditions are equivalent for the univariate ARCH(2) case and indeed for the case of $m = 1$ in the BEKK-ARCH process specification, but have failed to show the equivalence in the general case.

Appendix

A. Proof of Theorem 2:

The proof of Theorem 2 follows closely Johansen (1988, 1996). The main difference is that the innovations ξ_t considered here are not i.i.d. Gaussian, but instead uncorrelated and identically distributed with mean zero and variance Σ. We state in Lemmas 8, 9 and 10 results which are sufficient for mimicking the proofs in Johansen (1996). We emphasize the importance of the uncorrelatedness of the ξ_t's together with the existence of second order moments, Σ.

A.1. Computation of the PLR test

Consider briefly the definitions needed for an explicit form of the PLR test for the hypothesis $H(r)$. With indices $0, 1$ and z referring to the first differences ΔY_t, lagged levels Y_{t-1} and lagged differences $Z_t := (\Delta Y_{t-1}', \dots, \Delta Y_{t-k+1}')'$ respectively, define the product moment matrices

$$S_{ij} = M_{ij} - M_{iz}M_{zz}^{-1}M_{zj} \quad \text{for } i, j = 0, 1 \tag{21}$$

and where, say,

$$M_{oz} = \frac{1}{T}\sum_{t=1}^{T} \Delta Y_t Z_t' \text{ and } M_{1z} = \frac{1}{T}\sum_{t=1}^{T} Y_{t-1} Z_t'$$

Then the eigenvalue problem solved to compute the PLR test in (3) is given by

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0. \tag{22}$$

A.2. Asymptotics

First we state a central result involving functional central limit theory:

Lemma 8. *Under Assumptions 2.1 and 2.2*

$$\left(\frac{1}{\sqrt{T}}\Sigma_{t=1}^{[T\cdot]}\xi_t, \frac{1}{T}\Sigma_{t=1}^{T}\Sigma_{i=1}^{t-1}\xi_i\xi_t'\right) \xrightarrow{D} \left(B(\cdot), \int_0^1 BdB'\right)$$

where $B = \Sigma^{1/2}W$ is a p-dimensional Brownian motion with variance Σ and W a $p-$dimensional standard Brownian motion.

Proof of Lemma 8: As ξ_t is a martingale difference we apply the central limit theorem for martingale differences in Brown (1971, Theorem 3) to see that $\frac{1}{\sqrt{T}}\Sigma_{t=1}^{[T\cdot]}\xi_t \xrightarrow{D} B(\cdot)$ under the Lindeberg condition in Assumption 2.2. The joint convergence to the stochastic integral follows by Theorem 2.1 in Hansen (1992) since, trivially, $\sup_T \frac{1}{T}\Sigma_{t=1}^{T}E(\xi_t\xi_t') < \infty$. □

Next, recall that by Theorem 1, Y_t has the representation in (5) and that ΔY_t is stationary. Let the index β correspond to $\beta' Y_{t-1}$ and define the unconditional moments Ω_{ij} for $i,j = 0, \beta$ and z where, say,

$$\Omega_{oz} = E\left(\Delta Y_t Z_t'\right) \text{ and } \Omega_{\beta 0} = E\left(\beta' Y_{t-1}\Delta Y_t'\right) \tag{23}$$

which are well-defined by Assumption 2.1. Define in terms of these

$$\Sigma_{ij} := \Omega_{ij} - \Omega_{iz}\Omega_{zz}^{-1}\Omega_{zj}. \tag{24}$$

Finally, define the $(p-r)$-dimensional Brownian motion on $[0,1]$,

$$F(u) := \beta_\perp' C B(u), \tag{25}$$

where $B(\cdot)$ is defined in Lemma 8.

We now establish convergence of relevant empirical moments to the moments defined in (24) and demonstrate some vital identities which hold in terms of these:

Lemma 9. *Under the assumptions of Theorem 2 and Assumptions 2.1 and 2.1 then as $T \to \infty$,*

$$S_{00} \xrightarrow{p} \Sigma_{00},\ \beta' S_{10} \xrightarrow{p} \Sigma_{\beta 0} \text{ and } \beta' S_{11}\beta \xrightarrow{p} \Sigma_{\beta\beta}. \tag{26}$$

In terms of these the following identities hold,

$$\Sigma_{00} = \alpha\Sigma_{\beta 0} + \Sigma,\ \Sigma_{0\beta} = \alpha\Sigma_{\beta\beta} \tag{27}$$

and finally,

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\alpha(\alpha'\Sigma_{00}^{-1}\alpha)^{-1}\alpha'\Sigma_{00}^{-1} = \alpha_\perp(\alpha_\perp'\Sigma\alpha_\perp)^{-1}\alpha_\perp'. \tag{28}$$

Proof of Lemma 9: Consider $\beta' S_{10} = \beta' M_{10} - \beta' M_{1z}M_{zz}^{-1}M_{z0}$, where

$$\beta' M_{10} = \tfrac{1}{T}\sum_{t=1}^{T}\beta' Y_{t-1}\Delta Y_t = \tfrac{1}{T}\sum_{t=1}^{T}\beta'\nu_{t-1}\Delta Y_t.$$

As $\beta'\nu_t$ and ΔY_t are stationary and ergodic processes with finite second order moments the result that $\beta' S_{10} \xrightarrow{p} \Sigma_{\beta 0}$ follows by the law of large numbers. Likewise the other results in (26) hold.

The identities in (27) follow by postmultiplying in equation (1) by $\xi_t', \Delta Y_t'$ and $Z_t' = (\Delta Y_{t-1}', \ldots, \Delta Y_{t-k+1}')'$ respectively and taking averages. Specifically, rewrite (1) as

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \Psi Z_t + \xi_t.$$

Then by the law of large numbers

$$\Omega_{00} = \alpha\Omega_{\beta 0} + \Psi\Omega_{z0} + \Sigma, \;\; \Omega_{0z} = \alpha\Omega_{\beta z} + \Phi\Omega_{zz} \;\; \text{and} \;\; \Omega_{0\beta} = \alpha\Omega_{\beta\beta} + \Psi\Omega_{z\beta}$$

since ξ_t and Z_t as well as ξ_t and $\beta' Y_{t-1}$ are uncorrelated. Substituting for Ψ and using the definitions in (24) the results in (27) hold.

To prove the identity in (28) use the projection identity

$$I_p = \Sigma_{00}^{-1}\alpha(\alpha'\Sigma_{00}^{-1}\alpha)^{-1}\alpha' + \alpha_\perp(\alpha_\perp'\Sigma_{00}\alpha_\perp)^{-1}\alpha_\perp'\Sigma_{00}$$

and $\alpha_\perp'\Sigma_{00} = \alpha_\perp'\Sigma$, see (27). □

Note that if the sequence ξ_t was not uncorrelated then these identities would not hold.

Lemma 10. *Under the assumptions of Theorem 2 then as $T \to \infty$,*

$$\tfrac{1}{\sqrt{T}}\beta_\perp' Y_{[Tu]} \overset{w}{\to} F(u) \;, \tag{29}$$

$$\beta_\perp' S_{10}\alpha_\perp = \beta_\perp{}' S_{1x}\alpha_\perp \overset{w}{\to} \int_0^1 F dB'\alpha_\perp \;, \tag{30}$$

$$\tfrac{1}{T}\beta_\perp' S_{11}\beta_\perp \overset{w}{\to} \int_0^1 F(u)F(u)'du \;, \tag{31}$$

and furthermore,

$$\sqrt{T}\beta' S_{10}\alpha_\perp = \sqrt{T}\beta' S_{1\xi}\alpha_\perp \overset{w}{\to} N_{r\times p-r}(0, \Sigma_{\beta\beta}\otimes\alpha_\perp'\Sigma\alpha_\perp), \tag{32}$$

$$\beta' S_{11}\beta_\perp \in O_p(1). \tag{33}$$

Proof of Lemma 10: The invariance principle for Y_t in (29) follows by the invariance principle for ξ_t as Y_t has the representation in (5). Result (31) then holds by the continuous mapping theorem.
To prove (30) note that

$$\beta_\perp' S_{1\xi} = \beta_\perp' M_{1\xi} - \beta_\perp' M_{1z} M_{zz}^{-1} M_{z\xi}.$$

Turn first to $\beta'_\perp M_{1\xi}$ and use the representation of Y_t to see that

$$\begin{aligned}\beta'_\perp M_{1\xi} &= \tfrac{1}{T}\textstyle\sum_{t=1}^T \beta'_\perp Y_{t-1}\xi'_t \\ &= \tfrac{1}{T}\left(\beta'_\perp C \textstyle\sum_{t=1}^T (\sum_{t=1}^{t-1}\xi_i)\xi'_t + \beta'_\perp \sum_{t=1}^T \nu_{t-1}\xi'_t + \beta'_\perp \nu \sum_{t=1}^T \xi'_t\right),\end{aligned}$$

which by Lemma 8, the law of large numbers and the fact that ξ_t and ξ_{t-1} are uncorrelated converges weakly to $\int_0^1 FdB'$. Next, $M_{\xi z} = \frac{1}{T}\sum_{t=1}^T \xi_t Z'_t$ tends to zero in probability by the law of large numbers as ξ_t and Z_t are uncorrelated. Since $\beta'_\perp M_{1z} \in O_p(1)$ and M_{zz} converges in probability by the law of large numbers,

$$\beta'_\perp S_{1\xi} \xrightarrow{w} \int_0^1 FdB'$$

which shows (30).

That $\beta'_\perp M_{1z}$ and similarly $\beta'_\perp S_{11}\beta$ are $O_p(1)$ follows as they converge weakly to stochastic integrals by the results in Phillips and Solo (1992) applied to linear processes in terms of the martingale difference sequence ξ_t. Finally (32) holds by applying the central limit theorem to the martingale difference sequence $\nu_{t-1}\xi'_t$ rewriting $S_{1\xi}$ as above. □

By mimicking the proof of Theorem 11.1 in Johansen (1996), the results in Lemma 10 imply immediately that the PLR statistic has the desired limit distribution in terms of the standard Brownian motion,

$$(\beta'_\perp C\Sigma C'\beta_\perp)^{-1/2}F(\cdot)$$

where F is defined in (25). □

Proof of Corollary 3: With $\hat{\beta} = (v_1, ..., v_r)$, where v_i is the eigenvector corresponding to the ith eigenvalue λ_i in (22), consider the normalized version defined by

$$\tilde{\beta} = \hat{\beta}(\bar{\beta}'\hat{\beta})^{-1}$$

where $\bar{\beta} = \beta(\beta'\beta)^{-1}$. Then by mimicking the proof of Lemma 13.2 in Johansen (1996), it follows that

$$\begin{aligned} T\beta'_\perp(\tilde{\beta}-\beta) &\xrightarrow{w} (\int_0^1 F(u)F(u)'du)^{-1}\int_0^1 FdV' \\ V(u) &= (\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}B(u).\end{aligned}$$

As V and F are uncorrelated, $\tilde{\beta}$ is asymptotically mixed Gaussian. Hence mimicking the proof of Theorem 4 in Johansen (1988) shows that the PLR test for a linear hypothesis on β is asymptotically χ^2 distributed. □

B. Markov Chain results

Lemma 11. *The Markov chain given by (9) is aperiodic, irreducible with respect to the Lebesgue measure and compact sets in $\mathbb{R}^{p\times m}$ are 'small'*

Proof of Lemma 11: Denote for any t the conditional probabilities of $x_{t+n} \in A$ given $x_t = x$ by $(P_x^n(A))_{x\in\mathbb{R}^{pm}}$.

(i): Irreducibility with respect to λ follows by Proposition 4.2.1 (ii) in Meyn and Tweedie (1993) if $\sum_{n=1}^{\infty} P_x^n(A) > 0$ for all $x \in \mathbb{R}^{pm}$ and sets A in the Borel $\sigma-$algebra, $\mathbb{B}^{pm}$, with $\lambda(A) > 0$. Note that

$$\sum_{n=1}^{\infty} P_x^n(A) \geq P^m(A|x) = \int_A f(y|x)\, dy > 0$$

by continuity and positivity of f and the result follows.

(ii): An irreducible chain is periodic if it has period $d > 1$ and aperiodic if it has period $d = 1$. If it has period $d > 1$ then by Theorem 5.4.4 in Meyn and Tweedie (1993) there exists disjoint sets $D_0, D_1, ..., D_{d-1}$ in $\mathbb{B}^k$ such that

$$P_x^1(D_{i+1}) = 1 \text{ for } x \in D_i \text{ and } i = 0, 1, .., d-1 \pmod{d} \tag{34}$$

and furthermore $\psi\left(\bigcup_{i=1}^{d} D_{i-1}\right)^c = 0$, where ψ is a maximal irreducibility measure. Now, by Proposition 4.2.2 (ii) in Meyn and Tweedie (1993) the Lebegue measure λ is absolutely continuous with respect to ψ and therefore also $\lambda\left(\bigcup_{i=1}^{d} D_{i-1}\right)^c = 0$. For this to hold at least one of the sets, D_1 say, must have $\lambda(D_1) > 0$. Hence as above, $P^m(D_1|x) > 0$ for all $x \in \mathbb{R}^{pm}$. But iterating (34) m times one gets for some j the contradiction

$$P_x^m(D_1) = 0 \text{ with } x \in \bigcup_{i\neq j} D_i.$$

Hence the chain has period $d = 1$ and is therefore said to be aperiodic.

(iii): If K is a compact set $f(\cdot|\cdot)$ attains its minimum on $K \times K$ which is strictly positive since $f > 0$. In other words

$$f(y|x) \geq \delta$$

for some $\delta > 0$ and $(x, y) \in K \times K$. For any $x \in K$ and any $A \in \mathbb{B}^{pm}$

$$P_x^m(A) \geq P_x^m(A \cap K) = \int_{A\cap K} f(y|x)dy \geq \delta\lambda(A \cap K).$$

Hence for all $x \in K$, $P_x^m(\cdot)$ is minorized by $\lambda(\cdot \cap K)$ and therefore K is by definition small, cf. p.106 in Meyn and Tweedie (1993). This ends the proof of Lemma 11. □

B.1. Proof of Theorem 4:

Before stating the result we need to introduce the v-norm associated with the drift function v. For any signed measure μ on $\mathbb{B}^{pm}$

$$||\mu||_v := \sup\{\left|\int g(x)d\lambda(x)\right| \;\mid\; |g(x)| \leq v(x),\; x \in \mathbb{R}^{pm}\}$$

defines a norm on $\mathbb{R}^{pm}$; if $v(x) = 1$ then this is the total variation norm. As drift functions are larger than 1, note that convergence in v-norm implies convergence in total variation norm and hence in particular convergence in distribution.

Lemma 12. *(Meyn and Tweedie) Let $(x_t)_{t=0,1,..}$ be a time homogenous Markov chain on $(\mathbb{R}^q, \mathbb{B}^q)$ which is aperiodic and irreducible with respect to the Lebesgue measure λ. If there is a drift function v_1 satisfying a 1-step drift criterion in Theorem 4 with a 'small' set K not necessarily compact, then there is an invariant probability measure π. The function v_1 is π-integrable and there is a constant $\delta > 0$ such that x_t is geometric ergodic,*

$$||P_x^n(\cdot) - \pi||_{v_1} \leq \delta v_1(x)\gamma^n \text{ for all } x \in \mathbb{R}^q,\, n \geq 1. \tag{35}$$

Furthermore, if $x_0(P) = \pi$, then the process is stationary.

Proof of Lemma 12: The existence of the invariant distribution and the result on convergence speed is contained in the geometric ergodic theorem of Meyn and Tweedie (1993, Theorem 15.0.1(iii)). □

Applying Lemma 12, Theorem 4 holds by the following lemma:

Lemma 13. *Under the assumptions of Theorem 4 with k-step drift function v, there is a drift function $v_1 \geq v$ satisfying a one-step drift criterion with a 'small' set K.*

Proof: The k-step drift function v is a 1-step drift function for the k-skeleton chain,$(x_{kt})_{t=0,1,..}$ which by Lemma 11 is aperiodic and irreducible.

Hence by Lemma 12 there is an invariant distribution π for the k-skeleton chain and

$$\left\|P_x^{nk}(\cdot) - \pi\right\|_v \leq \delta\gamma^n v(x) \text{ for all } n \geq 1 \text{ and } x \in \mathbb{R}^q.$$

Note also that v is π integrable. As $\gamma < 1$ choose d so that for some γ

$$\delta\gamma^d < \gamma < 1.$$

Next define the drift function

$$v_1(x) := \sum_{i=0}^{dk-1} \gamma^{-i/dk} E\left(v(x_t)|\, x_{t-i} = x\right)$$

such that in particular $v_1(x) \geq v(x)$.

By iterated expectations

$$E(\, v_1(x_t)|\, x_{t-1} = x) = \sum_{i=0}^{dk-1} \gamma^{-i/dk} E(\, v(x_t)|\, x_{t-i-1} = 0).$$

Use that

$$\begin{aligned} E(\, v(x_{t+dk})|\, x_t = x) &= \int v(y) dP_x^{dk}(y) \\ &\leq |\int v(y) dP_x^{dk}(y) - \int v(y) d\pi(y)| + \int v(y) d\pi(y) \\ &\leq \left\| P^{dk}(\cdot) - \pi \right\|_v + \int v(y) d\pi(y) \leq \gamma v(x) + \kappa \end{aligned} \tag{36}$$

with κ finite as v is π integrable. Applying (36) together with the definition of v_1 it follows that

$$\begin{aligned} E(\, v_1(x_{t+1})|\, x_t = x) &= \gamma^{1/dk}(v_1(x) - v(x)) + \gamma^{1/dk-1} E(\, v(x_{t+qk})|\, x_t = x) \\ &\leq \gamma^{1/dk}(v_1(x) - v(x)) + \gamma^{1/dk} v(x) + \gamma^{1/dk-1}\kappa \\ &= \gamma^{1/dk} v_1(x) + \tilde{\kappa}. \end{aligned}$$

Note that $\gamma^{1/dk} < 1$ and $\tilde{\kappa} < \infty$.

Hence v_1 will satisfy a one-step drift criterion with a set K_1 of the form $K_1 := \{x \in \mathbb{R}^q | v_1(x) \leq \alpha\}$ for a suitable $\alpha > 0$. To see that K_1 is 'small' with respect to the full Markov chain x_{kt} we consider the set $K_v := \{x \in \mathbb{R}^q | v(x) \leq \alpha\}$. By Lemma 15.2.2(ii) in Meyn and Tweedie (1993) K_v is 'small' relative to the k-skeleton chain x_{kt}, and therefore also 'small' relative to the full Markov chain. Finally, note that by $v_1 \geq v$, $K_1 \subseteq K_v$ and therefore K_1 is also 'small' relative to the full Markov chain. □

C. Proof of Theorem 6

Proof: Consider first the case of $m = \tilde{m} = 1$. From (7) it follows that $\Sigma = V(\xi_t) = E(\xi_t \xi_t')$ satisfies

$$\Sigma = \phi(\Sigma) + \Omega \tag{37}$$

where the mapping ϕ in this case is defined by $\phi(\Sigma) = A\Sigma A'$, see equation (14) and Ω is positive definite, $\Omega > 0$.

With B some $p \times p$ positive semi-definite matrix, $B \geq 0$, then by Lemma 14 below there exists a small positive real number γ for which

$$\Omega \geq \gamma B.$$

This gives the inequalities

$$\Omega \geq \Omega - \gamma B \geq 0$$

which as ϕ is linear and maps positive semi-definite matrices into positive semi-definite matrices gives

$$\phi^n(\Omega) \geq \phi^n(\Omega) - \gamma\phi^n(B) \geq 0. \tag{38}$$

By iterating in (37) it follows that for all m,

$$\Sigma_{n=0}^m \phi^n(\Omega) \leq \Sigma.$$

This implies that

$$x'\phi^n(\Omega)x \to 0 \text{ as } n \to \infty \text{ and for all } x \in \mathbb{R}^p$$

and hence, using (38),

$$x'\phi^n(B)x \to 0 \text{ as } n \to \infty \text{ and for all } x \in \mathbb{R}^p$$

which means that $\phi^n(B)$ converges to zero in the weak operator topology, see Pedersen (1988, section 4.6.1). By Rudin, (1991, Theorem 1.21) there is only one vector space topology on any finite dimensional vector space and therefore in particular for the operator norm we have,

$$\|\phi^n(B)\| \to 0 \text{ as } n \to \infty. \tag{39}$$

Next, we want to show that the same holds for any symmetric matrix C. To do so, decompose C as the sum of two positive semi-definite matrices

$$C = B^+ - B^-, \quad B^+ \geq 0, \; B^- \geq 0.$$

Immediately we get that for all C symmetric,

$$\|\phi^n(C)\| = \left\|\phi^n(B^+) - \phi^n(B^-)\right\| \to 0 \text{ as } n \to \infty \tag{40}$$

The convergence in (40) for any symmetric C means that $\phi^n(\cdot)$ when restricted to the space of symmetric matrices converges in the strong operator topology, see Pedersen (1988, Section 4.6.1). Similar to before introducing the operator norm,

$$\left\| \phi|_{sym}^n \right\| := \sup_{C symmetric} \frac{\|\phi^n(C)\|}{\|C\|} \tag{41}$$

it holds by Rudin (1991, Theorem 1.21) that,

$$\left\| \phi|_{sym}^{n} \right\| < 1 \text{ for some } n \text{ large enough.}$$

Next, using submultiplicativity of the operator norm in (41) we get by Pedersen (1988, 4.1.13),

$$\begin{aligned} \rho\left(\phi|_{sym}\right) &= \lim_{m\to\infty} \left\| \phi|_{sym}^{m} \right\|^{1/m} = \lim\sup_{m\to\infty} \left\| \phi|_{sym}^{mn} \right\|^{1/mn} \\ &\leq \lim\sup_{m\to\infty} \left\| \phi|_{sym}^{n} \right\|^{1/n} < 1 \end{aligned}$$

Identifying $\rho(\Phi)$ with $\rho(\phi)$ as in the proof of Theorem 5 finishes the proof for the case of $m = \tilde{m} = 1$.

For general $m, \tilde{m}$ use that $\tilde{\Sigma} = E(x_t x_t')$ with x_t defined in (9) satisfies

$$\tilde{\Sigma} = \phi(\tilde{\Sigma}) + \tilde{\Omega} = \phi^m(\tilde{\Sigma}) + \Sigma_{i=0}^{m-1}\phi^i(\tilde{\Omega}) := \phi^*(\tilde{\Sigma}) + \Omega^*$$

where ϕ is defined in (14) and Ω^* is positive definite, see (10). Using the arguments from before we see that

$$\|\phi^{mn}\| < 1 \text{ for some } n \text{ large enough.}$$

and therefore that $\rho(\phi|_{sym}) < 1$. □

Lemma 14. *Assume that Ω is a symmetric and positive definite $n \times n$ matrix. Then for any $n \times n$ positive semi-definite matrix B there exists a constant $\gamma > 0$ such that*

$$\Omega - \gamma B > 0$$

Proof: For any $x \neq 0$ in $\mathbb{R}^n$,

$$\begin{aligned} x'(\Omega - \gamma B)x &= \left(\tfrac{x'\Omega x}{x'x} - \gamma\tfrac{x'Bx}{x'x}\right)x'x \\ &\geq [\lambda_{\min}(\Omega) - \gamma\lambda_{\max}(B)]\,\|x\|^2 > 0 \end{aligned}$$

for γ small enough. Here $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimal, respectively maximal, modulus of the eigenvalues of the matrix involved. □

D. Tables

Table 5: *Smallest sample for which the asymptotic quantiles are within the 95% confidence interval of the simulated quantiles for varying $\rho(\Phi)$ and fixed $\|\Phi\|$ (left columns) and varying $\|\Phi\|$ for fixed $\rho(\Phi)$ (right columns). In brackets are reported the asymptotic confidence band for the simulated quantile. The asymptotic quantiles from McKinnon et al. (1999) are quoted in the 'a.q.' row.*

	Quantile				*Quantile*		
$\rho(\Phi)$	90%	95%	99%	$\|\Phi\|$	90%	95%	99%
a.q.	17.98	20.26	25.08		17.98	20.26	25.08
i.i.d.	60 [17.94, 18.38]	60 [20.25, 20.83]	70 [24.76, 26.00]	0.5	225 [17.93, 18.39]	250 [20.20, 20.70]	160 [24.94, 26.20]
0.1	225 [17.85, 18.23]	250 [20.10, 20.66]	275 [24.88, 26.31]	1	350 [17.91, 18.37]	300 [20.25, 20.89]	200 [24.96, 26.64]
0.2	250 [17.90, 18.36]	250 [20.15, 20.77]	250 [24.94, 26.62]	2	800 [17.95, 18.41]	850 [20.22, 20.79]	700 [25.06, 26.45]
0.3	650 [17.91, 18.38]	650 [20.21, 20.73]	750 [24.86, 26.08]	4	550 [17.94, 18.39]	700 [20.24, 20.66]	550 [24.89, 26.28]
0.4	+1 000	900 [20.23, 20.78]	800 [24.79, 26.04]	6	600 [17.84, 18.26]	750 [20.17, 20.78]	900 [24.97, 26.09]
0.5	+1 000	650 [20.24, 20.90]	+1 000	8	700 [17.97, 18.42]	750 [20.19, 20.79]	800 [24.88, 26.25]
0.6	+1 000	+1 000	+1 000	10	1 000 [17.96, 18.32]	800 [20.24, 20.80]	750 [24.93, 26.05]
0.7	+1 000	+1 000	+1 000	20	700 [17.90, 18.40]	850 [20.24, 20.91]	800 [25.07, 26.41]
0.8	+1 000	+1 000	+1 000	30	750 [17.95, 18.39]	1 000 [20.14, 20.80]	750 [25.08, 26.16]
0.9	+1 000	+1 000	+1 000	40	+1 000	+1 000	850 [25.02, 26.55]
1.0	+1 000	+1 000	+1 000				

Table 6: *Empirical rejection probability of the PLR test for varying spectral radius, $\rho(\Phi)$. For each value of $\rho(\Phi)$ the first row is the 1% nominal level, the second the 5% nominal level and the third is the 10% nominal level.*

	T										
$\rho(\Phi)$	20	40	60	80	100	200	300	400	600	800	1 000
	1.93	1.22	1.23	1.03	1.14	0.87	1.02	0.95	1.01	1.15	0.98
i.i.d.	6.94	5.95	5.41	5.59	5.36	5.05	5.16	4.99	5.00	5.06	5.14
	13.39	11.45	10.49	10.43	10.41	10.02	10.01	10.00	9.89	10.03	10.56
	2.95	1.97	1.71	1.76	1.57	1.32	1.15	1.16	1.17	1.22	1.21
0.1	10.21	7.15	6.68	6.41	6.57	5.85	5.08	5.31	5.33	5.38	5.63
	16.85	13.16	12.36	11.49	11.64	10.98	10.50	10.10	10.23	10.72	10.80
	3.20	2.13	1.82	1.59	1.67	1.36	1.05	1.13	1.00	0.90	1.01
0.2	10.10	7.78	7.16	7.13	6.25	5.85	5.70	5.29	4.93	5.17	5.47
	17.17	14.13	12.87	12.47	11.31	10.75	10.61	10.31	10.15	10.26	10.11
	3.92	2.75	2.28	2.06	1.89	1.45	1.48	1.48	1.22	1.05	1.04
0.3	11.62	8.92	8.02	7.36	6.65	6.21	5.92	5.83	5.55	5.28	5.35
	18.58	14.93	13.81	13.13	12.45	11.80	11.73	11.31	10.68	10.23	10.27
	4.49	3.49	2.94	2.95	2.39	1.84	1.62	1.60	1.45	1.07	1.26
0.4	12.67	9.71	9.18	8.41	7.94	6.66	6.56	6.34	5.81	5.56	5.73
	19.59	16.19	15.42	14.92	13.91	12.43	12.19	11.86	10.98	11.03	10.98
	3.72	3.27	2.65	2.31	2.15	1.79	1.46	1.46	1.35	1.31	1.27
0.5	10.83	9.63	8.23	7.61	7.61	6.82	6.56	6.55	5.75	5.43	5.62
	18.17	15.93	13.86	13.42	13.52	12.53	12.63	12.17	10.95	10.78	10.99
	4.04	3.27	2.66	2.61	2.36	2.06	2.00	1.76	1.41	1.38	1.34
0.6	11.70	9.46	8.53	7.81	7.98	7.30	6.88	6.59	5.96	6.04	6.04
	19.11	15.97	14.41	13.87	13.56	13.40	12.25	12.10	11.23	11.11	11.29
	4.23	3.24	3.02	2.60	2.62	2.44	2.22	1.71	1.68	1.70	1.60
0.7	12.30	9.70	8.52	8.06	7.91	7.44	7.29	6.31	6.69	6.59	5.94
	18.88	15.91	14.36	13.87	13.94	12.84	12.39	12.09	11.92	11.90	11.43
	4.00	3.74	3.58	3.29	2.98	2.53	2.28	2.10	2.32	2.13	1.92
0.8	11.93	10.28	9.37	8.86	8.90	7.81	7.52	7.27	6.98	6.78	6.69
	18.71	16.29	15.41	14.81	15.10	13.80	13.11	12.51	12.20	11.98	12.07
	4.11	3.59	3.43	3.38	3.43	2.95	2.87	2.92	2.80	2.38	2.45
0.9	11.50	9.97	9.40	9.34	8.88	8.52	8.14	7.97	7.95	7.56	7.68
	18.66	16.64	15.84	15.30	14.88	14.45	13.55	13.33	13.04	13.16	13.21
	4.13	3.98	3.52	3.83	3.66	3.51	3.48	3.51	3.71	3.48	3.30
1.0	11.39	10.47	9.51	9.52	9.51	9.50	9.08	9.22	9.36	8.65	8.90
	18.11	16.32	15.27	15.80	15.63	14.92	15.17	15.01	14.54	14.38	14.63
	4.53	4.47	4.09	4.35	4.40	4.33	4.40	4.38	4.84	4.59	4.39
1.1	11.95	11.00	10.14	10.56	10.38	10.34	10.41	10.81	10.60	10.34	10.35
	18.65	16.92	16.02	16.54	16.62	16.29	16.52	16.84	16.17	15.90	16.17
	4.80	4.95	4.54	4.89	5.34	5.28	5.48	5.52	6.12	5.97	5.83
1.2	12.48	11.66	10.89	11.52	11.55	11.37	11.93	12.26	12.34	12.14	11.99
	19.28	17.36	16.96	17.61	17.49	17.55	17.90	18.49	18.06	17.68	17.96
	5.21	5.45	5.04	5.42	6.05	6.24	6.74	6.87	7.37	7.28	7.04
1.3	13.17	12.35	11.56	12.36	12.56	12.78	13.54	13.65	13.89	13.86	13.81
	19.76	18.31	17.80	18.58	18.84	18.90	19.35	20.17	20.21	19.660	20.49

Table 7: *Empirical rejection probability of the PLR test for varying norm,* $\|\Phi\|$*. For each value of* $\|\Phi\|$ *the first row is the 1% nominal level; the second the 5% nominal level; and the third is the 10% nominal level. The spectral radius is fixed at* $\rho(\Phi) = 0.25$*, the number of replications is* $N = 10\,000$*.*

	T										
$\|\Phi\|$	20	40	60	80	100	200	300	400	600	800	1 000
	1.93	1.22	1.23	1.03	1.14	0.87	1.02	0.95	1.01	1.15	0.98
i.i.d.	6.94	5.95	5.41	5.59	5.36	5.05	5.16	4.99	5.00	5.06	5.14
	13.39	11.45	10.49	10.43	10.41	10.02	10.01	10.00	9.89	10.03	10.56
	2.57	1.78	1.76	1.77	1.71	1.15	1.14	1.16	1.13	1.16	1.28
0.25	9.58	7.48	6.93	6.73	6.38	5.52	5.22	5.79	5.32	5.43	5.29
	15.95	13.34	12.84	12.51	11.75	10.87	10.66	10.66	10.59	10.68	10.34
	3.57	2.71	2.14	1.82	1.82	1.17	1.23	1.28	1.26	1.27	1.09
1	10.73	8.879	7.52	7.01	6.81	5.82	5.42	5.49	5.30	5.18	5.24
	17.93	15.41	13.24	12.89	12.18	10.97	10.96	10.84	10.66	10.29	10.08
	3.96	2.94	2.47	2.24	1.91	1.46	1.26	1.31	1.32	1.11	1.12
2	11.28	9.07	8.02	7.86	7.04	5.74	5.95	5.55	5.63	5.54	5.49
	18.27	15.18	14.44	13.71	13.17	11.81	11.46	10.77	10.95	10.50	10.51
	4.25	3.08	2.53	2.21	2.05	1.64	1.29	1.36	1.20	1.34	1.29
4	12.59	9.09	7.78	7.41	7.10	6.05	5.94	5.86	5.60	5.63	5.74
	19.30	15.13	13.71	13.32	13.05	11.40	11.02	11.06	10.50	10.78	10.78
	4.38	3.07	2.51	2.08	1.99	1.50	1.36	1.35	1.32	1.36	1.08
6	12.19	9.25	8.07	7.48	6.93	5.98	5.80	5.81	5.55	5.68	5.35
	19.78	15.56	13.76	13.04	12.47	11.38	11.07	10.98	10.02	10.25	10.42
	4.36	3.07	2.45	2.07	1.77	1.21	1.29	1.37	1.19	1.14	1.09
8	12.13	8.80	8.04	7.32	7.08	5.49	5.77	5.89	5.23	5.17	5.26
	19.54	15.20	13.71	13.22	12.58	11.49	11.24	10.88	10.69	10.57	10.45
	4.53	3.04	2.55	2.18	1.80	1.39	1.31	1.31	1.27	1.91	1.10
10	12.37	9.24	8.29	7.60	6.81	5.92	5.83	6.11	5.60	5.39	5.32
	19.57	15.37	13.95	13.01	12.59	10.95	11.47	11.17	10.83	10.77	10.52
	4.39	3.07	2.40	2.07	1.69	1.36	1.39	1.33	1.17	1.20	1.09
20	12.40	8.95	7.75	7.68	6.90	5.87	5.74	5.84	5.45	5.46	5.16
	16.55	15.04	13.92	13.54	12.52	11.28	10.84	10.82	10.89	10.31	10.37
	4.38	3.09	2.40	1.90	1.64	1.32	1.30	1.26	1.07	1.17	1.17
30	12.20	9.00	7.93	7.28	6.47	5.83	5.81	5.68	5.78	5.71	5.27
	19.52	15.13	13.98	13.10	12.52	11.30	10.80	11.80	10.80	10.53	10.64
	4.39	3.09	2.43	2.00	1.93	1.33	1.35	1.35	1.20	1.32	1.29
40	12.32	9.00	8.04	7.28	6.68	5.79	5.67	5.75	5.65	5.95	5.49
	19.67	15.13	14.05	13.07	12.60	11.42	10.88	11.20	11.20	10.86	10.87

References

Bollerslev, T. (1990): Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model, *Review of Economics and Statistics,* 72 : 498-505.

Brown, B.M. (1971): Martingale Central Limit Theorems, *The Annals of Mathematical Statistics,* 42 : 59-66.

Boswijk, H.P. and A. Lucas (2002): Semi-Nonparametric Cointegration Testing, *Forthcoming in Journal of Econometrics.*

Boswijk, H.P., A. Lucas and N. Taylor (2002): A Comparison of parametric, Semi-Nonparametric, Adaptive and Nonparametric Cointegration Tests, *in: Advances in Econometrics: Applying Kernel and Nonparametric Estimation to Economic Topics. T.B. Fomby and R.C. Hill (eds.), Stamford, JAI Press,* 14 : 25-47.

Carrasco, M. and X. Chen (2002): Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, *Econometric Theory,* 18 : 17-39.

Comte, F. and O. Lieberman (2001): Asymptotic Theory for Multivariate GARCH Processes, *working paper, CREST-ENSAE.*

Dudewicz, E.J. and S.N. Mishra (1988): *Modern Mathematical Statistics, John Wiley & Sons, New York.*

Embrechts, P., C. Kluppelberg and T. Mikosch 1997): *Modelling Extremal Events for Insurance and Finance, Applications of Mathematics 33, Springer-Verlag.*

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

Feigin, P.D. and R.L. Tweedie (1985): Random Coefficient Autoregressive Processes: A Markov Chain Analysis of Stationarity and Finiteness of Moments, *Journal of Time Series Analysis,* 6 : 1-14.

Hansen, B.E. (1992): Convergence to Stochastic Integrals for Dependent Heterogenous Processes, *Econometric Theory,* 8 : 489-500.

Hansen, E. and A. Rahbek (1998): Stationarity and Asymptotics of Multivariate ARCH Time Series with an Application to Robustness of Cointegration Analysis, *preprint, Department of Theoretical Statistics, University of Copenhagen. CAF working paper series, 22.*

Jeantheau, T. (1998): Strong Consistency of Estimators for Multivariate GARCH Models, *Econometric Theory,* 14 : 70-86.

Johansen, S. (1988): Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control,* 12 : 231-254.

——— (1996): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press, Oxford, UK, 2nd edn.*

——— (2002): A Small Sample Correction of the Test for Cointegrating Rank in the Vector Autoregressive Model, *EUI Working Paper ECO No. 2000/15. Forthcoming in Econometrica.*

Kessler, M. and A. Rahbek (2001): Identification and Inference for Cointegrated Gaussian Diffusions, *Preprint no. 3, Department of Statistics and Operations Research, University of Copenhagen.*

Kim, K. and P. Schmidt (1993): Unit Roots Tests with Conditional Heteroscedasticity, *Journal of Econometrics,* 59 : 287-300.

Li, W.K., S. Ling and H. Wong (2001): Estimation for Partially Nonstationary Multivariate Autoregressive Models with Conditional Heteroscedasticity, *Biometrika,* 88 : 1135-1152.

Ling, S. and M. McAleer (2001): Asymptotic Theory for a Vector ARMA-GARCH Model, *Discussion paper no. 549, The Institute of Social and Economic Research, Osaka University. Forthcoming in Econometric Theory.*

Lucas, A. (1998): Inference on Cointegration Ranks Using LR And LM Tests Based on Pseudo-Likelihoods, *Econometric Reviews,* 17 : 185-214.

MacKinnon, J.G., A.A. Haug and L. Michelis (1999): Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration, *Journal of Applied Econometrics,* 14 : 563-577.

Meyn, S.P. and R.L. Tweedie (1993): *Markov Chains and Stochastic Stability, Communications and Control Engineering Series, Springer-Verlag.*

Nielsen, B. and A. Rahbek (2000): Similarity Issues in Cointegration Analysis, *Oxford Bulletin of Economics and Statistics,* 62 : 5-22.

Pedersen, G.K. (1988): *Analysis Now, Springer-Verlag.*

PHAM, D.T. (1986): The Mixing Property of Bilinear and Generalised Random Coefficient Autoregressive Models, *Stochastic Processes and their Applications,* 23 : 291-300.

PHILIPS, P.C.B. AND V. SOLO (1992): Asymptotics For Linear Processes, *The Annals of Statistics,* 20 : 970-1001.

RUDIN, W. (1991): *Functional Analysis, McGraw-Hill Inc., New York*

SEO, B. (1999): Distribution Theory for Unit Root Tests with Conditional Heteroscedasticity, *Journal of Econometrics,* 91 : 113-144.

TJØSTHEIM, D. (1990): Non-linear Time Series and Markov Chains, *Advances in Applied Probability,* 22 : 587-611.

TSE, Y.K. (2000): A test for constant correlations in a multivariate GARCH model, *Journal of Econometrics,* 98 : 107-127.