Alternative asymptotics for cointegration tests in large VARs.

Alexei Onatski and Chen Wang
Faculty of Economics, University of Cambridge

August 23, 2016

Abstract

Johansen's (1988, 1991) likelihood ratio test for cointegration rank of a Gaussian VAR depends only on the squared sample canonical correlations between current changes and past levels of a simple transformation of the data. We study the asymptotic behavior of the empirical distribution of those squared canonical correlations when the number of observations and the dimensionality of the VAR diverge to infinity simultaneously and proportionally. We find that the distribution almost surely weakly converges to the so-called *Wachter distribution*. This finding provides a theoretical explanation for the observed tendency of Johansen's test to find "spurious cointegration". It also sheds light on the workings and limitations of the Bartlett correction approach to the over-rejection problem. We propose a simple graphical device, similar to the scree plot, for a preliminary assessment of cointegration in high-dimensional VARs.

1 Introduction

Johansen's (1988, 1991) likelihood ratio (LR) test for cointegration rank is a very popular econometric technique. However, it is rarely applied to systems of more than three or four variables. On the other hand, there exist many applications involving much larger systems. For example, Davis (2003) discusses a possibility of applying the test to the data on seven aggregated and individual commodity prices to test Lewbel's (1996) generalization of the Hicks-Leontief composite commodity theorem.

In a recent study of exchange rate predictability, Engel, Mark, and West (2015) contemplate a possibility of determining the cointegration rank of a system of seventeen OECD exchange rates. Banerjee, Marcellino, and Osbat (2004) emphasize the importance of testing for no cross-sectional cointegration in panel cointegration analysis (see Breitung and Pesaran (2008) and Choi (2015)), and the cross-sectional dimension of modern macroeconomic panels can easily be as large as forty.

The main reason why the LR test is rarely used in the analysis of relatively large systems is its poor finite sample performance. Even for small systems, the test based on the asymptotic critical values does not perform well (see Johansen (2002)). For large systems, the size distortions become overwhelming, leading to severe over-rejection of the null in favour of too much cointegration as shown in many simulation studies, including Ho and Sorensen (1996) and Gonzalo and Pitarakis (1995, 1999).

In this paper, we study the asymptotic behavior of the sample canonical correlations that the LR statistic is based on, when the number of observations and the system's dimensionality go to infinity simultaneously and proportionally. We show that the empirical distribution of the squared sample canonical correlations almost surely converges to the so-called *Wachter distribution*, originally derived by Wachter (1980) as the limit of the empirical distribution of the eigenvalues of the multivariate beta matrix of growing dimension and degrees of freedom. Our analytical findings explain the observed over-rejection of the null hypothesis by the LR test, shed new light on the workings and limitations of the Bartlett-type correction approach to the problem (see Johansen (2002)), and lead us to propose a very simple graphical device, similar to the scree plot, for a preliminary analysis of the validity of cointegration hypotheses in large vector autoregressions.

The basic framework for our analysis is standard. Consider a p-dimensional VAR in the error correction form

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \tag{1}$$

where D_t and ε_t are vectors of deterministic terms and zero-mean Gaussian errors with unconstrained covariance matrix, respectively. The LR statistic for the test of the null hypothesis of no more than r cointegrating relationships between the p

elements of X_t against the alternative of more than r such relationships is given by

$$LR_{r,p,T} = -T \sum_{i=r+1}^{p} \log\left(1 - \lambda_i\right), \tag{2}$$

where T is the sample size, and $\lambda_1 \geq ... \geq \lambda_p$ are the squared sample canonical correlation coefficients between residuals in the regressions of ΔX_t and X_{t-1} on the lagged differences ΔX_{t-i}, $i = 1, ..., k-1$, and the deterministic terms. In the absence of the lagged differences and deterministic terms, the λ's are the eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$, where S_{00} and S_{11} are the sample covariance matrices of ΔX_t and X_{t-1}, respectively, while S_{01} is the cross sample covariance matrix.

Johansen (1991) shows that the asymptotic distribution of $LR_{r,p,T}$ under the asymptotic regime where $T \to \infty$ while p remains fixed, can be expressed in terms of the eigenvalues of a matrix whose entries are explicit functions of a $p-r$-dimensional Brownian motion. Unfortunately, for relatively large p, this asymptotics does not produce good finite sample approximations, as evidenced by the over-rejection phenomenon mentioned above. Therefore, in this paper, we consider a *simultaneous* asymptotic regime $p, T \to_c \infty$ where both p and T diverge to infinity so that

$$p/T \to c \in (0, 1], \tag{3}$$

while p remains no larger than T. Our Monte Carlo analysis shows that the corresponding asymptotic approximations are relatively accurate even for such small sample sizes as $p = 10$ and $T = 20$.

The basic specification for the data generating process (1) that we consider has $k = 1$. In the next section, we discuss extensions to more general VARs with low-rank Γ_i matrices and additional common factor terms. We also explain there that our main results hold independently from whether a deterministic vector D_t with fixed or slowly-growing dimension is present or absent from the VAR.

Our study focuses on the behavior of the empirical distribution function (d.f.) of the squared sample canonical correlations,

$$F_{p,T}(\lambda) = \frac{1}{p}\sum_{i=1}^{p} \mathbf{1}\left\{\lambda_i \leq \lambda\right\}, \tag{4}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. We find that, under the null of r cointegrating relationships, as $p, T \to_c \infty$ while $r/p \to 0$, almost surely (a.s.),

$$F_{p,T}(\lambda) \Rightarrow W_c(\lambda) \equiv W(\lambda; c/(1+c), 2c/(1+c)), \tag{5}$$

where $\Rightarrow$ denotes the weak convergence of d.f.'s (see Billingsley (1995), p.191), and $W(\lambda; \gamma_1, \gamma_2)$ denotes the *Wachter* d.f. with parameters γ_1 and γ_2, described in detail in the next section.

As explained below, convergence (5) guarantees an a.s. asymptotic lower bound for the scaled LR statistic,

$$\liminf_{p,T\to_c\infty} LR_{r,p,T}/p^2 \geq -\frac{1}{c}\int \log(1-\lambda)\,\mathrm{d}W_c(\lambda). \tag{6}$$

In contrast, we show that under the standard asymptotic regime where $T \to \infty$ while p is held fixed, $LR_{r,p,T}/p^2$ concentrates around 2 for relatively large p. A direct calculation reveals that 2 is smaller than the lower bound (6), for all $c > 0$, with the gap growing as c increases. That is, the standard asymptotic distribution of the LR statistic is centered at a too low level, especially for relatively large p. This explains the tendency of the asymptotic LR test to over-reject the null.

The reason for the poor centering delivered by the standard asymptotic approximation is that it classifies terms $(p/T)^j$ in the asymptotic expansion of the LR statistic as $O(T^{-j})$. When p is relatively large, such terms substantially contribute to the finite sample distribution of the statistic, but are ignored as asymptotically negligible. In contrast, the *simultaneous asymptotics* classifies all terms $(p/T)^j$ as $O(1)$. They are not ignored asymptotically, which improves the centering of the simultaneous asymptotic approximation relative to the standard one.

Using bound (6), we construct a Bartlett-type correction factor for the standard LR test, hence, addressing a long-standing problem (see Johansen (2002)). As we show below, for $p/T < 1/3$, the value of our theoretical correction factor is very close to the simulation-based factor described in Johansen, Hansen and Fachin (2005). However, for larger p/T, the values diverge. Johansen, Hansen and Fachin's (2005) simulations do not consider combinations of p and T with $p/T > 1/3$, and the functional form that they use to fit the simulated correction factors does not work well uniformly in p/T.

The weak convergence result (5) can be put to a more direct use by comparing the quantiles of the empirical distribution of the squared sample canonical correlations with the quantiles of the limiting Wachter distribution. Under the null, the former quantiles plotted against the latter ones should form a 45° line, asymptotically. Deviations of such a *Wachter quantile-quantile (q-q) plot* from the line indicate violations of the null. Creating Wachter plots requires practically no additional computations beyond those needed to compute the LR statistic, and we propose to use this simple graphical device for a preliminary analysis of cointegration in large VARs.

Our study is the first to derive the limit of the empirical d.f. of the squared sample canonical correlations between random walk X_{t-1} and its innovations ΔX_t. Wachter (1980) shows that $W(\lambda; \gamma_1, \gamma_2)$ is the weak limit of the empirical d.f. of the squared sample canonical correlations between q- and m-dimensional independent Gaussian white noises with the size of the sample n, when $q, m, n \to \infty$ so that $q/n \to \gamma_1$ and $m/n \to \gamma_2$. Yang and Pan (2012) show that Wachter's (1980) result holds without the Gaussianity assumption for i.i.d. data with finite second moments. Our proofs do not rely on those previous results. The novelty and difficulty of our setting is that X_t and ΔX_t are not independent processes. This requires original ideas for our proofs.

Our paper opens up a new direction for the asymptotic analysis of panel VAR cointegration tests based on the sample canonical correlations. One such test is developed in Larsson and Lyhagen (2007) (see also Larsson, Lyhagen, and Lothgren (2001) and Groen and Kleibergen (2003)). Larsson and Lyhagen (2007) are reluctant to recommend their test for large VARs. They suggest that for the analysis of relatively large panels it may be better to rely on tighter parameterized models, such as that of Bai and Ng (2004).

We conjecture that the Larsson-Lyhagen test, as well as the LR test, based on the *simultaneous* asymptotics work well in panels with comparable cross-sectional and temporal dimensions. The results of this paper can be used to describe the appropriate centering of the corresponding test statistics. The next step would be to derive the *simultaneous* asymptotic distribution of scaled deviations of such statistics from the centering values. We expect the *simultaneous* asymptotic distribution of $LR_{r,p,T}$ to be Gaussian, as is often the case for averages of regular functions of eigenvalues of large random matrices (see Bai and Silverstein (2010) and Paul and Aue (2014)). We are currently pursuing this line of research.

The rest of this paper is structured as follows. In Section 2, we prove the convergence of $F_{p,T}(\lambda)$ to the *Wachter* d.f. under the simultaneous asymptotics. Section 3 derives the sequential limit of the empirical d.f. of the squared sample canonical correlations as, first $T \to \infty$ and then $p \to \infty$. It then uses differences between the sequential and simultaneous limits to explain the over-rejection phenomenon, and to design a theoretical Bartlett-type correction factor for the LR statistic in high-dimensional VARs. Section 4 contains a Monte Carlo study and illustrates the proposed Wachter q-q plot technique using a macroeconomic panel. Section 5 concludes. All proofs are given in the Appendix and Supplementary Material (SM).

2 Convergence to the Wachter distribution

Consider the following basic version of (1)

$$\Delta X_t = \Pi X_{t-1} + \Phi D_t + \varepsilon_t \tag{7}$$

with d_D-dimensional vector of deterministic regressors D_t. Let R_{0t} and R_{1t} be the vectors of residuals from the OLS regressions of ΔX_t on D_t, and X_{t-1} on D_t, respectively. Define

$$S_{00} = \frac{1}{T}\sum_{t=1}^{T} R_{0t}R_{0t}', \; S_{01} = \frac{1}{T}\sum_{t=1}^{T} R_{0t}R_{1t}', \text{ and } S_{11} = \frac{1}{T}\sum_{t=1}^{T} R_{1t}R_{1t}', \tag{8}$$

and let $\lambda_1 \geq ... \geq \lambda_p$ be the eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$.

The main goal of this section is to establish the a.s. weak convergence of the empirical d.f. of the λ's to the Wachter d.f., under the null of r cointegrating relationships, when $p, T \to_c \infty$. The Wachter distribution with d.f. $W(\lambda; \gamma_1, \gamma_2)$ and parameters $\gamma_1, \gamma_2 \in (0,1)$ has density

$$f_W(\lambda; \gamma_1, \gamma_2) = \frac{1}{2\pi\gamma_1}\frac{\sqrt{(b_+ - \lambda)(\lambda - b_-)}}{\lambda(1-\lambda)} \tag{9}$$

on $[b_-, b_+] \subseteq [0,1]$ with

$$b_\pm = \left(\sqrt{\gamma_1(1-\gamma_2)} \pm \sqrt{\gamma_2(1-\gamma_1)}\right)^2, \tag{10}$$

and atoms of size $\max\{0, 1-\gamma_2/\gamma_1\}$ at zero, and $\max\{0, 1-(1-\gamma_2)/\gamma_1\}$ at unity.

We assume that model (7) may be misspecified in the sense that the true data generating process is described by the following generalization of (1)

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Psi F_t + \varepsilon_t, \tag{11}$$

where ε_t, $t = 1, ..., T$, are still i.i.d. $N(0, \Sigma)$ with arbitrary $\Sigma > 0$, $\operatorname{rank} \Pi = r$, but k is not necessarily unity, and F_t is a d_F-dimensional vector of deterministic or stochastic variables that does not necessarily coincide with D_t. For example, some of the components of F_t may be common factors not observed and not modelled by the econometrician. Further, we do not put any restrictions on the roots of the characteristic polynomial associated with (11). In particular, explosive behavior and seasonal unit roots are allowed. Finally, no constraints on F_t, and the initial values $X_{1-k}, ..., X_0$, apart from the asymptotic requirements on d_F and k as spelled out in the following theorem, are imposed.

Theorem 1 *Suppose that the data are generated by (11), and let* $\Gamma = [\Gamma_1, ..., \Gamma_{k-1}]$. *If*

$$\frac{1}{p}\left(d_D + d_F + r + k + \operatorname{rank} \Gamma\right) \to 0 \tag{12}$$

as $p, T \to_c \infty$ *while* p *remains no larger than* T*, then, almost surely,*

$$F_{p,T}(\lambda) \Rightarrow W_c(\lambda) \equiv W(\lambda; c/(1+c), 2c/(1+c)). \tag{13}$$

Theorem 1 implies that the weak limits of $F_{p,T}(\lambda)$ corresponding to the general model (11) and to the basic model $\Delta X_t = \Pi X_{t-1} + \varepsilon_t$ are the same as long as (12) holds. Condition (12) guarantees that the difference between the general and basic versions of $S_{01} S_{11}^{-1} S_{01}' S_{00}^{-1}$ have rank R that is less than proportional to p (and to T). Then, by the so-called rank inequality (Theorem A43 in Bai and Silverstein (2010)), the Lévy distance between the general and basic versions of $F_{p,T}(\lambda)$ is no larger than R/p, which converges to zero as $p, T \to_c \infty$. Since the Lévy distance metrizes the weak convergence (see Billingsley (1995), problem 14.5), the limiting d.f. is the same for both versions. For further details, see the proof of Theorem 1 in the Appendix.

Remark 2 *In standard cases where* D_t *is represented by* $(1, t)$, *it is customary to impose restrictions on* Φ *so that there is no quadratic trend in* X_t *(see Johansen (1995), ch. 6.2). Then, the LR test is based on the eigenvalues of* $S_{01}^{*} S_{11}^{*-1} S_{01}^{*\prime} S_{00}^{*-1}$,

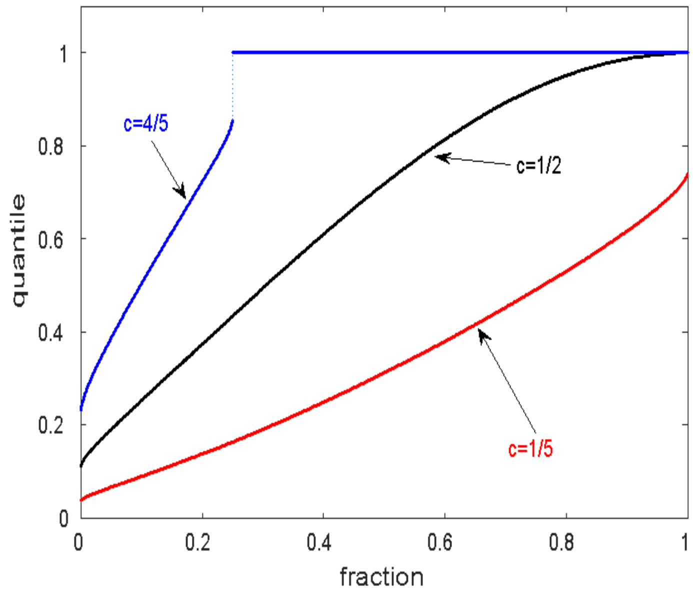

Figure 1: Quantile functions of $W_c(\lambda)$ for $c = 1/5$, $c = 1/2$, and $c = 4/5$.

defined similarly to $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ by replacing X_{t-1} with $\left(X_{t-1}', t\right)'$ and regressing ΔX_t and $\left(X_{t-1}', t\right)'$ on constant only to obtain R_{0t} and R_{1t}. The empirical d.f. of so modified eigenvalues still converges to $W_c(\lambda)$ because the difference between matrices $S_{01}^{*}S_{11}^{*-1}S_{01}^{*\prime}S_{00}^{*-1}$ and $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ has small rank.

Figure 1 shows quantile plots of $W_c(\lambda)$ for different values of c. For $c = 1/5$, the dimensionality of the data constitutes 20% of the sample size. The corresponding Wachter limit of $F_{p,T}(\lambda)$ is supported on $[0.04, 0.74]$. In particular, we expect λ_1 be larger than 0.7 for large p, T even in the absence of any cointegrating relationships. For $c = 1/2$, the upper boundary of support of the Wachter limit is unity. This accords with Gonzalo and Pitarakis' (1995, Lemma 2.3.1) finding that as $T/p \to 2$, $\lambda_1 \to 1$. For $c = 4/5$, the Wachter limit has mass 3/4 at unity.

Wachter (1980) derives $W(\lambda; \gamma_1, \gamma_2)$ as the weak limit of the empirical d.f. of eigenvalues of the p-dimensional beta[1] matrix $B_p(n_1/2, n_2/2)$ with n_1, n_2 degrees of freedom as $p, n_1, n_2 \to \infty$ so that $p/n_1 \to \gamma_1/\gamma_2$ and $p/n_2 \to \gamma_1/(1-\gamma_2)$. The eigenvalues of multivariate beta matrices are related to many important concepts in multivariate statistics, including canonical correlations, multiple discriminant ratios, and MANOVA. In particular, the squared sample canonical correlations between q-

[1] For the definition of the multivariate beta see Muirhead (1982), p.110.

and m-dimensional independent Gaussian samples of size n are jointly distributed as the eigenvalues of $B_q\left(m/2,(n-m)/2\right)$, where $q \leq m$ and $n \geq q+m$. Therefore, their empirical d.f. weakly converges to $W\left(\lambda;\gamma_1,\gamma_2\right)$ with $\gamma_1 = \lim q/n$ and $\gamma_2 = \lim m/n$.

Note that the latter limit coincides with $W_c\left(\lambda\right)$ when $n = T+p$, $q = p$, and $m = 2p$. Hence, Theorem 1 implies that the limiting empirical distribution of the squared sample canonical correlations between T observations of p-dimensional random walk and its own innovations is the same as that between $T+p$ observations of independent p- and $2p$-dimensional white noises. This suggests that there might exist a deep connection between these two settings, which is yet to be discovered.

In the context of multiple discriminant analysis, Wachter (1976) proposes to use a q-q plot, where the multiple discriminant ratios are plotted against quantiles of $W\left(\lambda;\gamma_1,\gamma_2\right)$, as a simple graphical method that helps one "recognize hopeless from promising analyses at an early stage." Nowadays, such q-q plots are called *Wachter plots* (see Johnstone (2001)). Theorem 1 implies that the Wachter plot can be used as a simple preliminary assessment of cointegration hypotheses in large VARs.

As an illustration, Figure 2 shows a Wachter plot of the simulated sample squared canonical correlations corresponding to a 20-dimensional VAR(1) model (7) with $\Pi = \operatorname{diag}\left\{-I_3, 0\times I_{17}\right\}$ so that there are three white noise and seventeen random walk components of X_t. No deterministic terms are included. We set $T = 200$ and $c = 1/10$. The graph clearly shows three canonical correlations that destroy the 45° line fit, so that the null hypothesis of no cointegration is compromised.

Theorem 1 does not provide any explanation to the fact that exactly three canonical correlations deviate from the 45° line in Figure 2. To interpret deviations of the Wachter plots from the 45° line, it is desirable to investigate behavior of $F_{p,T}\left(\lambda\right)$ under various alternatives. So far, we were able to obtain a clear result only for the "extreme" alternative, where X_t is a vector of independent white noises. Under such an alternative,

$$F_{p,T}\left(\lambda\right) \Rightarrow W\left(\lambda; c/(2-c), 1/(2-c)\right). \tag{14}$$

We plan to publish a full proof of this and some related results elsewhere.

The a.s. weak convergence of $F_{p,T}\left(\lambda\right)$ established in Theorem 1 implies the a.s. convergence of bounded continuous functionals of $F_{p,T}\left(\lambda\right)$. An example of such a functional is the scaled Pillai-Bartlett statistic for the null of no more than r cointegrating

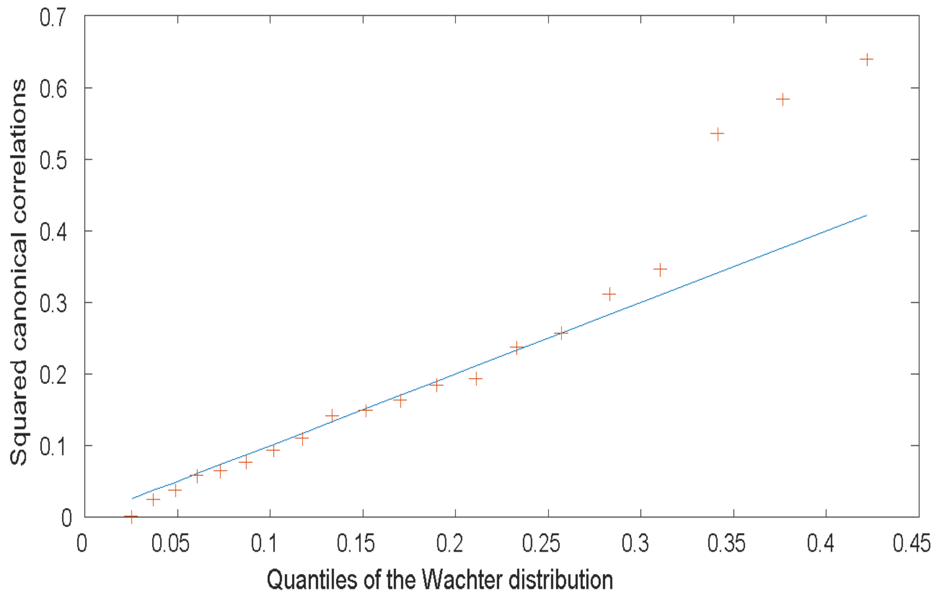

Figure 2: Wachter plot of the squared canonical correlations corresponding to 20-dimensional series with 3 components being white noises and the other components being independent random walks. $p = 20, T = 200$.

relationships (see Gonzalo and Pitarakis (1995))

$$\frac{1}{Tp} PB_{r,p,T} = \frac{1}{p} \sum_{j=r+1}^{p} \lambda_j,$$

which is asymptotically equivalent to the LR statistic under the standard asymptotic regime. Since, by definition, $\lambda_j \in [0, 1]$, we have

$$\frac{1}{Tp} PB_{r,p,T} = \int f(\lambda) \mathrm{d}F_{p,T}(\lambda) - \frac{1}{p} \sum_{j=1}^{r} \lambda_j, \tag{15}$$

where f is the bounded continuous function

$$f(\lambda) = \begin{cases} 0 & \text{for } \lambda < 0 \\ \lambda & \text{for } \lambda \in [0, 1] \\ 1 & \text{for } \lambda > 1. \end{cases} .$$

As long as $r/p \to 0$ as $p, T \to_c \infty$, the second term on the right hand side of (15) converges to zero. Therefore, Theorem 1 implies that $PB/(Tp)$ a.s. converges to $\int f(\lambda)\mathrm{d}W_c(\lambda)$. A direct calculation based on (9), which we report in the SM, yields the following corollary.

Corollary 3 *Under the assumptions of Theorem 1, as $p, T \to_c \infty$, a.s.,*

$$PB_{r,p,T}/(Tp) \to 2c/(1+c) + \max\{0, 2-1/c\}.$$

A similar analysis of the LR statistic (2) is less straightforward because $\log(1-\lambda)$ is unbounded on $\lambda \in [0,1]$. In fact, for $c > 1/2$, $LR_{r,p,T}$ is ill-defined because a non-negligible proportion of the squared sample canonical correlations exactly equal unity. However for $c < 1/2$, we can obtain the a.s. asymptotic lower bound on $LR_{r,p,T}/(Tp)$. Note that for such c, the upper bound of the support of $W_c(\lambda)$ equals $b_+ = c\left(\sqrt{2}-\sqrt{1-c}\right)^{-2} < 1$. Let

$$\overline{\log}(1-\lambda) = \begin{cases} 0 & \text{for } \lambda < 0 \\ \log(1-\lambda) & \text{for } \lambda \in [0, b_+] \\ \log(1-b_+) & \text{for } \lambda > b_+. \end{cases} \tag{16}$$

Clearly, $\overline{\log}(1-\lambda)$ is a bounded continuous function and

$$\frac{1}{Tp}LR_{r,p,T} \geq -\frac{1}{p}\sum_{j=r+1}^{p}\overline{\log}(1-\lambda_j).$$

Hence, we have the following a.s. lower bound on $LR_{r,p,T}/(Tp)$ (the corresponding calculations are reported in the SM).

Corollary 4 *Under the assumptions of Theorem 1, for $c < 1/2$, as $p, T \to_c \infty$, a.s.,*

$$\liminf_{p,T\to_c\infty}\frac{1}{Tp}LR_{r,p,T} \geq \frac{1+c}{c}\ln(1+c) - \frac{1-c}{c}\ln(1-c) + \frac{1-2c}{c}\ln(1-2c).$$

Remark 5 *We conjecture that the above lower bound is, in fact, the a.s. limit of $LR_{r,p,T}/(Tp)$. To prove this conjecture, one needs to show that λ_{r+1} is a.s. bounded away from unity so that the unboundedness of $\log(1-\lambda)$ is not consequential. We leave this for future research.*

Corollary 4 suggests an appropriate "centering point" for the scaled LR statistic when p and T are large. As we show in the next section, the standard asymptotic distribution concentrates around a very different point for large p. To study such a concentration, in the next section, we consider the *sequential* asymptotic regime where first $T \to \infty$, and then $p \to \infty$.

3 A comparison to sequential asymptotics

3.1 Sequential asymptotics

To obtain useful results under the sequential asymptotics, we study eigenvalues of the scaled matrix

$$\frac{T}{p} S_{01} S_{11}^{-1} S_{01}' S_{00}^{-1}. \tag{17}$$

Under the simultaneous asymptotic regime, the behavior of the scaled and unscaled eigenvalues is the same up to the factor $c^{-1} = \lim T/p$. In contrast, as $T \to \infty$ while p remains fixed, the unscaled eigenvalues converge to zero, while scaled ones do not. We shall denote the empirical d.f. of eigenvalues of the scaled matrix as $F_{p,T}^{(s)}(\lambda)$.

Without loss of generality (see Lemmas 10 and 11 in the Appendix), we focus on the case of the simple data generating process

$$\Delta X_t = \varepsilon_t, \; t = 1, ..., T, \; \text{and } X_0 = 0, \tag{18}$$

and on the situation where the econometrician does not include any deterministic regressors in his or her model, that is $d_D = 0$. For simplicity, in the rest of this section, we assume that $r = 0$, and consider statistics $LR_{0,p,T}$ rather than a more general $LR_{r,p,T}$.

Under the above simplifications, Johansen's (1988, 1991) results imply that, as $T \to \infty$ while p is held fixed, the eigenvalues of the scaled matrix (17) jointly converge in distribution to the eigenvalues of

$$\frac{1}{p} \int_0^1 (\mathrm{d}B) B' \left(\int_0^1 BB' \mathrm{d}u \right)^{-1} \int_0^1 B (\mathrm{d}B)', \tag{19}$$

where B is a p-dimensional Brownian motion. We denote the eigenvalues of (19) as $\lambda_j^{(\infty)}$, and their empirical d.f. as $F_{p,\infty}(\lambda)$.

It is reasonable to expect that, as $p \to \infty$, $F_{p,\infty}(\lambda)$ becomes close to the limit of the empirical d.f. of eigenvalues of (17) under a simultaneous, rather than sequential, asymptotic regime $p, T \to_\gamma \infty$, where γ is close to zero. We denote such a limit as $F_\gamma(\lambda)$. This expectation turns out to be correct in the sense that the following theorem holds. Its proof is given in the SM.

Theorem 6 *Let $F_0(\lambda)$ be the weak limit of $F_\gamma(\lambda)$ as $\gamma \to 0$. Then, as $p \to \infty$, $F_{p,\infty}(\lambda) \Rightarrow F_0(\lambda)$, in probability. The d.f. $F_0(\lambda)$ corresponds to a distribution supported on $[a_-, a_+]$ with*

$$a_\pm = \left(1 \pm \sqrt{2}\right)^2, \tag{20}$$

and having density

$$f(\lambda) = \frac{1}{2\pi} \frac{\sqrt{(a_+ - \lambda)(\lambda - a_-)}}{\lambda}. \tag{21}$$

A reader familiar with Large Random Matrix Theory (see Bai and Silverstein (2010)) might recognize that $F_0(\lambda)$ is the d.f. of the continuous part of a special case of the Marchenko-Pastur distribution (Marchenko and Pastur (1967)). The general Marchenko-Pastur distribution has density

$$f_{MP}\left(\lambda; \kappa, \sigma^2\right) = \frac{1}{2\pi\sigma^2\kappa} \frac{\sqrt{(a_+ - \lambda)(\lambda - a_-)}}{\lambda}$$

over $[a_-, a_+]$ with $a_\pm = \sigma^2 (1 \pm \sqrt{\kappa})^2$ and a point mass $\max\{0, 1 - 1/\kappa\}$ at zero. Density (21) is two times $f_{MP}(\lambda; \kappa, \sigma^2)$ with $\kappa = 2$ and $\sigma^2 = 1$. The multiplication by two is needed because the mass $1/2$ at zero is not a part of the distribution F_0.

According to Theorem 6, for any $\delta_1, \delta_2 > 0$ and all sufficiently large p,

$$\Pr\left(\frac{1}{p}\sum_{j=1}^{p} \lambda_j^{(\infty)} \geq \int \lambda \mathrm{d}F_0(\lambda) - \delta_1\right) \geq 1 - \delta_2. \tag{22}$$

A direct calculation, which we report in the SM, shows that $\int \lambda \mathrm{d}F_0(\lambda) = 2$. On the other hand, as $T \to \infty$ while p remains fixed,

$$LR_{0,p,T} \xrightarrow{d} p \sum_{j=1}^{p} \lambda_j^{(\infty)} \text{ as } T \to \infty. \tag{23}$$

Hence, we have the following corollary.

Corollary 7 *As first $T \to \infty$, and then $p \to \infty$, the lower probability bound on $LR_{0,p,T}/\left(2p^2\right)$ is unity in the following sense. As $T \to \infty$ while p is held fixed, $LR_{0,p,T}/\left(2p^2\right)$ converges in distribution to $\sum_{j=1}^{p} \lambda_j^{(\infty)}/\left(2p\right)$. Further, for any $\delta_1, \delta_2 >$ 0 and all sufficiently large p, the probability $\Pr\left\{\sum_{j=1}^{p} \lambda_j^{(\infty)}/\left(2p\right) \geq 1 - \delta_1\right\}$ is no smaller than $1 - \delta_2$.*

The reason why we claim only the lower bound on $LR_{0,p,T}/\left(2p^2\right)$ is that Theorem 6 is silent about the behavior of the individual eigenvalues $\lambda_j^{(\infty)}$, the largest of which may, in principle, quickly diverge to infinity. We suspect that 2 is not just the lower bound, but also the probability limit of $\sum_{j=1}^{p} \lambda_j^{(\infty)}/p$, so that the sequential probability limit of $LR_{0,p,T}/\left(2p^2\right)$ is unity. Verification of this conjecture requires more work, similar to that discussed in Remark 5.

Corollary 7 is consistent with the numerical finding of Johansen, Hansen and Fachin (2005, Table 2) that, as T becomes large while p is being fixed, the sample mean of the LR statistic is well approximated by a polynomial $2p^2 + \alpha p$ (see also Johansen (1988) and Gonzalo and Pitarakis (1995)). The value of α depends on how many deterministic regressors are included in the VAR. Our theoretical result captures the 'highest order' sequential asymptotic behavior of the LR statistic, which remains (bounded below by) $2p^2$ independent on the number of the deterministic regressors.

The concentration of the LR statistic around $2p^2$ explains why the critical values of the LR test are so large for large values of p. For example, MacKinnon, Haug and Michelis (1999) report the 5% critical value 311.09 for $p = 12$. The transformation

$$LR_{0,p,T} \mapsto LR_{0,p,T}/p - 2p$$

makes the LR statistic 'well-behaved' under the sequential asymptotics and leads to more conventional critical values. The division by p reduces the 'second order behavior' to $O_{\mathrm{P}}(1)$, while subtracting $2p$ eliminates the remaining explosive 'highest order term'. We report the corresponding transformed 95% critical values alongside the original ones in Table 1.

The transformed critical values resemble 97-99 percentiles of $N(0,1)$. Since the LR test is one-sided, the resemblance is coincidental. However, we do expect the sequential asymptotic distribution of the transformed LR statistic (as well as its simultaneous asymptotic distribution) to be normal (possibly with non-zero mean

p	Unadjusted CV	CV/$p-2p$	p	Unadjusted CV	CV/$p-2p$
1	4.13	2.13	7	111.79	1.97
2	12.32	2.16	8	143.64	1.96
3	24.28	2.09	9	179.48	1.94
4	40.17	2.04	10	219.38	1.94
5	60.06	2.01	11	263.25	1.93
6	83.94	1.99	12	311.09	1.92

Table 1: The 95% asymptotic critical values (CV) for Johansen's LR test. The unadjsuted values are taken from the first column of Table II in MacKinnon, Haug and Michelis (1999).

and non-unit variance). A formal analysis of this conjecture is left for future research.

3.2 The over-rejection, and the Bartlett correction

In this subsection, let us assume that the following conjecture holds.

Conjecture 8 *The simultaneous and sequential asymptotic lower bounds for the scaled LR statistics derived in Corollaries 4 and 7 represent the corresponding simultaneous and sequential asymptotic limits. Specifically, for* $c < 1/2$*,*

$$\lim_{p,T\to_c\infty} \frac{1}{2p^2} LR_{0,p,T} = \frac{1+c}{2c^2}\ln(1+c) - \frac{1-c}{2c^2}\ln(1-c) + \frac{1-2c}{2c^2}\ln(1-2c), \quad (24)$$

$$\operatorname*{plim}_{p\to\infty} \lim_{T\to\infty} \frac{1}{2p^2} LR_{0,p,T} = 1. \quad (25)$$

Figure 3 plots the right hand side of (24). It is larger than unity (dashed line) for all $c \in (0, 1/2)$ with the gap increasing in c. The Monte Carlo analysis in the next section shows that 'typical' values of the LR statistic in finite samples with comparable p and T are concentrated around the solid line. In contrast, the 'standard' asymptotic critical values are concentrated around the dashed line. Hence, the standard asymptotic distribution of the LR statistic is centered at a too low level. This explains the over-rejection of the null of no cointegration by the standard asymptotic LR test.

A popular approach to addressing the over-rejection problem is based on the Bartlett-type correction of the LR statistic. It was explored in much detail in various studies, including Johansen (2002). The idea is to scale the LR statistic so that

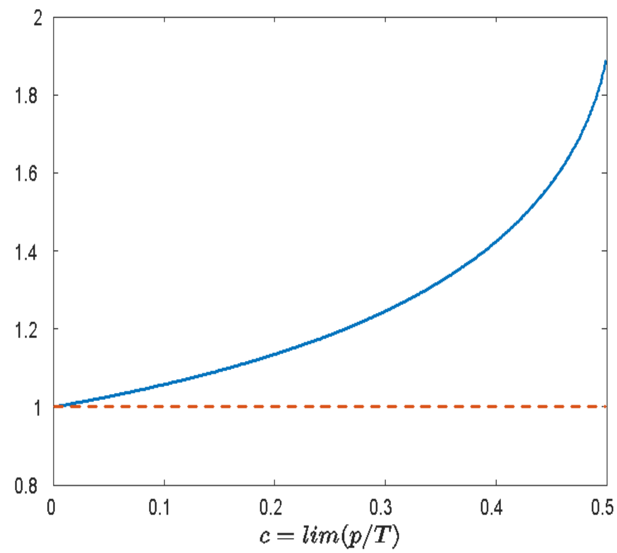

Figure 3: The asymptotic limits (under Conjecture 8) of the scaled LR statistic $L_{0,p,T}/\left(2p^2\right)$. Dashed line: sequential asymptotic limit. Solid line: simultaneous asymptotic limit.

its finite sample distribution better fits the asymptotic distribution of the unscaled statistic. Specifically, let $E_{p,\infty}$ be the mean of the asymptotic distribution under the fixed-p, large-T asymptotic regime. Then, if the finite sample mean, $E_{p,T}$, satisfies

$$E_{p,T} = E_{p,\infty}\left(1 + a(p)/T + o\left(1/T\right)\right), \tag{26}$$

the scaled statistic is defined as $LR/\left(1 + a(p)/T\right)$. By construction, the fit between the scaled mean and the original asymptotic mean is improved by an order of magnitude. Although, as shown by Jensen and Wood (1997) in the context of unit root testing, the fit between higher moments does not improve by an order of magnitude, it may become substantially better (see Nielsen (1997)).

A theoretical analysis of the adjustment factor $1 + a(p)/T$ can be rather involved. In general, $a(p)$ will depend not only on p, but also on all the parameters of the VAR. However, for Gaussian VAR(1) without deterministic terms, under the null of no cointegration, $a(p)$ depends only on p.

For $p = 1$, the exact expression for $a(p)$ was derived in Larsson (1998). Given the difficulty of the theoretical analysis of $a(p)$, Johansen (2002) proposes to numerically evaluate the Bartlett correction factor $BC_{p,T} \equiv E_{p,T}/E_{p,\infty}$ by simulation. Johansen,

Hansen and Fachin (2005) simulate $BC_{p,T}$ for various values of $p \leq 10$ and $T \leq 3000$ and fit a function of the form

$$BC^*_{p,T} = \exp\left\{a_1 p/T + a_2 (p/T)^2 + \left[a_3 (p/T)^2 + b\right]/T\right\}$$

to the obtained results. For relatively large values of T, the term $\left[a_3 (p/T)^2 + b\right]/T$ in the above expression is small. When it is ignored, the fitted function becomes particularly simple:

$$\widetilde{BC}_{p,T} = \exp\left\{0.549 p/T + 0.552 (p/T)^2\right\}.$$

Our simultaneous and sequential asymptotic results shed light on the workings of $\widetilde{BC}_{p,T}$. Given that Conjecture 8 holds,

$$\frac{\lim_{p,T\to_c\infty} LR_{0,p,T}}{p \lim_{T\to\infty,p\to\infty} LR_{0,p,T}} = \frac{1+c}{2c^2}\ln(1+c) - \frac{1-c}{2c^2}\ln(1-c) + \frac{1-2c}{2c^2}\ln(1-2c).$$

Therefore, for non-negligible p/T, we expect $BC_{p,T}$ to be well approximated by

$$\widehat{BC}_{p,T} = \frac{1+\hat{c}}{2\hat{c}^2}\ln(1+\hat{c}) - \frac{1-\hat{c}}{2\hat{c}^2}\ln(1-\hat{c}) + \frac{1-2\hat{c}}{2\hat{c}^2}\ln(1-2\hat{c}),$$

where $\hat{c} = p/T$ is the finite sample analog of c.

Figure 4 superimposes the graphs of $\widehat{BC}_{p,T}$ and $\widetilde{BC}_{p,T}$ as functions of $\hat{c}$. For $p/T \leq 0.3$, there is a strikingly good fit between the two curves, with the maximum distance between them 0.0067. For $p/T > 0.3$ the quality of the fit quickly deteriorates. This can be explained by the fact that all p, T-pairs used in Johansen, Hansen and Fachin's (2005) simulations are such that $p/T < 0.3$.

4 Monte Carlo and some examples

4.1 Monte Carlo experiments

Throughout this section, the analysis is based on 1000 Monte Carlo replications. First, we generate pure random walk data with zero starting values for $p = 10, T = 100$ and $p = 10, T = 20$. Figure 5 shows the Tukey boxplots summarizing the MC distribution of each of the λ_i, $i = 1, ..., 10$ (sorted in the ascending order throughout this section).

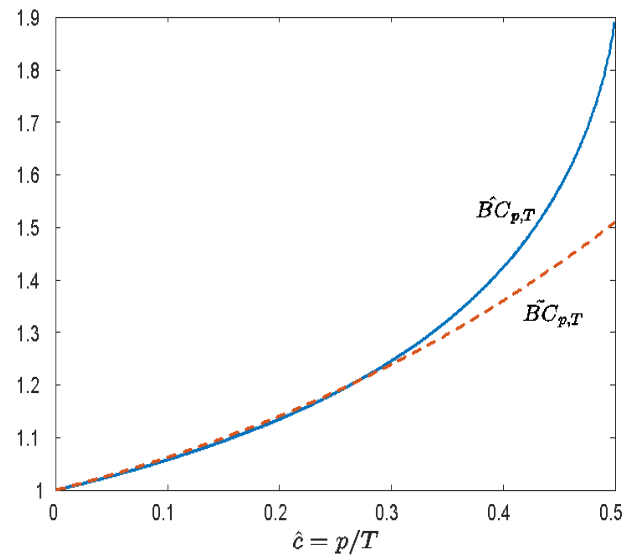

Figure 4: Bartlett correction factors as functions of p/T. Solid line: the factor based on simultaneous asymptotics. Dashed line: numerical approximation from Johansen, Hansen and Fachin (2005).

The boxplots are superimposed with the quantile function of the Wachter limit with $c = 1/10$ for the left panel and $c = 1/2$ for the right panel. Precisely, for $x = i$, we show the value the $100\,(i - 1/2)\,/p$ quantile of the Wachter limit. For $i = 1, 2, ..., 10$, these are the 5-th,15-th,...,95-th quantiles of $W\left(\lambda; c/\left(1+c\right), 2c/\left(1+c\right)\right).$ Even for such small values of p and T, the theoretical quantiles track the location of the MC distribution of the empirical quantiles very well.

The dispersion of the MC distributions around the theoretical quantile is quite large for the chosen small values of p and T. To see how such a dispersion changes when p and T increase while p/T remains fixed, we generated pure random walk data with $p = 20, T = 200$ and $p = 100, T = 1000$ for $p/T = 1/10$, and with $p = 20, T = 40$ and $p = 100, T = 200$ for $p/T = 1/2$. Instead of reporting the Tukey boxplots, we plot only the 5-th and 95-th percentiles of the MC distributions of the λ_i, $i = 1, ..., p$ against $100\,(i - 1/2)\,/p$ quantiles of the corresponding Wachter limit. The plots are shown on Figure 6. We see that the [5%,95%] ranges of the MC distributions of λ_i are still considerably large for $p = 20$. These ranges become much smaller for $p = 100$.

The behavior of the smallest squared canonical correlation λ_1 in Figures 5 and 6 is special in that its MC distribution lies below the corresponding Wachter quantile. This does not contradict our theoretical results because a weak limit of the empirical

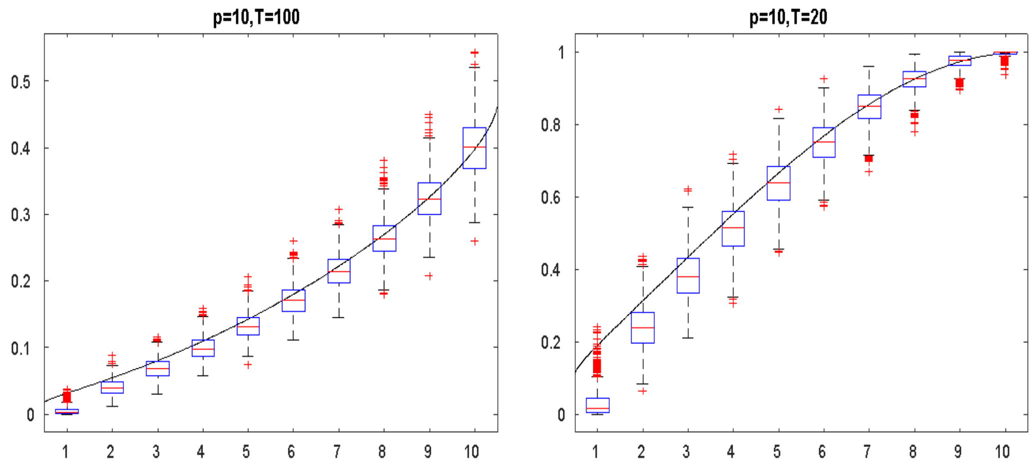

Figure 5: The Tukey boxplots for 1000 MC simulations of ten sample squared canonical correlations correponding to pure random walk data. The boxplots are superimposed with the quantile function of the Wachter limit.

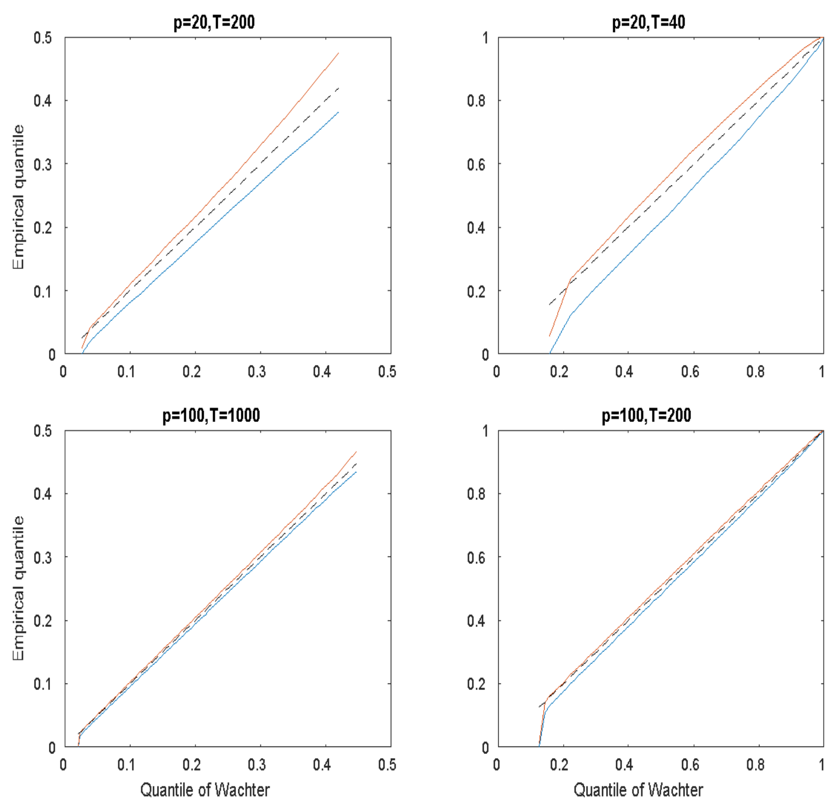

Figure 6: The q-q Wachter plots for pure random walk data. The dashed line is the 45° line. The solid lines are the 5-th and the 95-th percentiles of the MC distributions of λ_i, which are plotted against $100(i-1/2)/p$ quantiles of the Wachter limit.

distribution of λ's is not affected by an arbitrary change in a finite (or slowly growing) number of them. In fact, we find it somewhat surprising that only the distribution of λ_1 is not well-alligned with the derived theoretical limit. Our proofs are based on several low rank alterations of the matrix $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$, and there is nothing in them that guarantee that only one eigenvalue of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ behaves in a "special" way. In future work, it would be interesting to investigate behavior of the extreme eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$ theoretically.

Our next Monte Carlo experiment simulates data that are not random walk. Instead, the data are stationary VAR(1) with zero mean, zero initial value, and $\Pi = \rho I_p$. We consider three cases of ρ : 0, 0.5, and 0.95. Figure 7 shows the Wachter plots with solid lines representing 5th and 95-th percentiles of the MC distributions of λ_i plotted against the $100(i - 1/2)/p$ quantiles of the corresponding Wachter limit. The dashed line correspond to the null case where the data are pure random walk (shown for comparison).

The lower panel of the figure corresponds to the most persistent alternative with $\rho = 0.95$. Samples with $p = 20$ seem to be too small to generate substantial differences in the behavior of Wachter plots under the null and under such persistent alternatives. The less persistent alternative with $\rho = 0.5$ is easily discriminated against by the Wachter plot for $p/T = 1/10$ (left panel). The discrimination power of the plot for $p/T = 1/5$ (central panel) is weaker. For $p/T = 1/2$ there is still some discrimination power left, but the location of the Wachter plot under alternative "switches" the side relative to the 45° line.

The plots easily discriminate against white noise ($\rho = 0$) alternative for $c = 1/10$ and $c = 1/5$, but not for $c = 1/2$. This accords with (14), which implies that the empirical d.f.'s of the squared sample canonical correlations based on random walk and on white noise data converge to the same limit when $c = 1/2$.

Results reported in Figure 7 indicate that for relatively small p and p/T, Wachter plots can be effective in discriminating against alternatives to the null of no cointegration, where the cointegrating linear combinations of the data are not very persistent. Further, tests of no cointegration hypothesis that may be developed using simultaneous asymptotics would probably need to be two-sided. It is because the location of the Wachter plot under the alternative may "switch sides" relative to the 45° depending on the persistence of the data under the alternative. Finally, cases with c close to 1/2 must be analyzed with much care. For such cases, the behavior of the

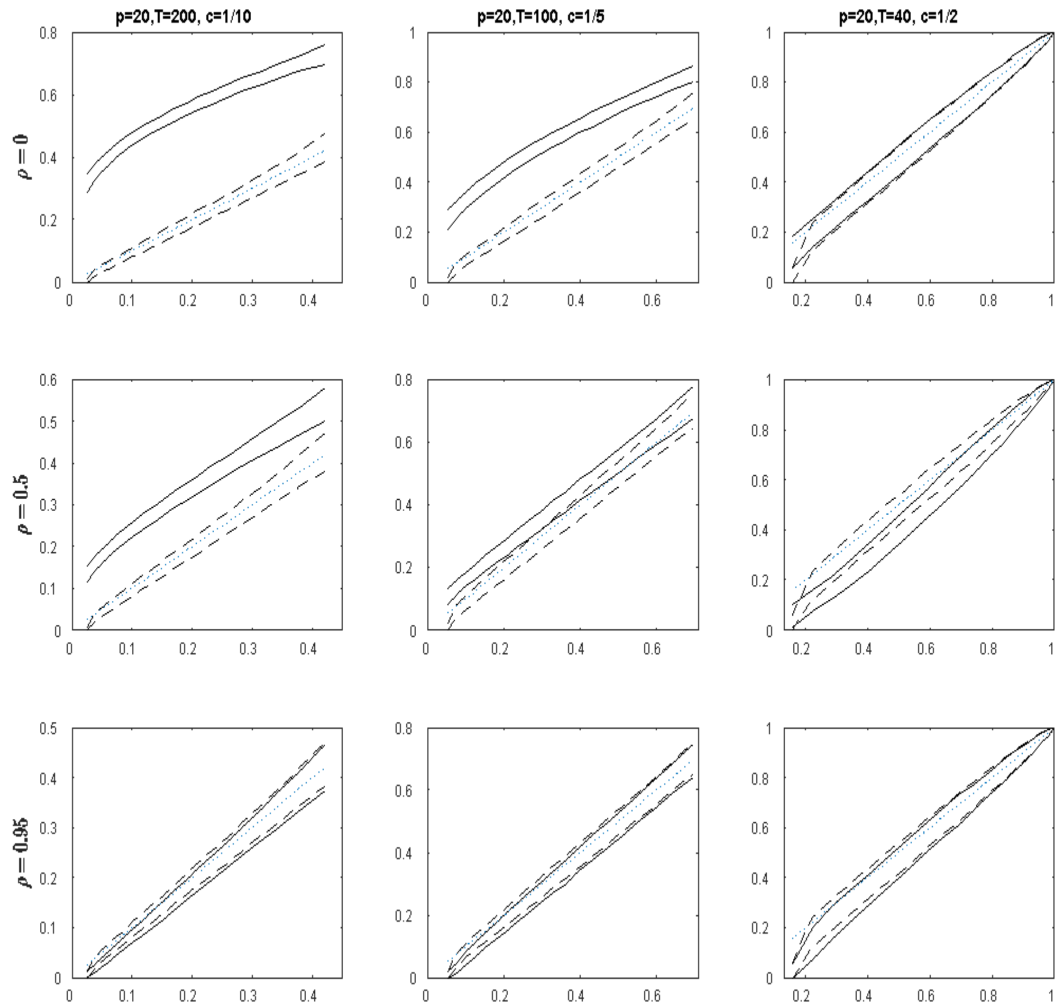

Figure 7: The q-q Wachter plots for stationary data $X_t = \rho X_{t-1} + \varepsilon_t$. Solid lines: 5 and 95 percentiles of the MC distribution of λ_i plotted against $100(i - 1/2)/p$ quantile of the Wachter limit. Dashed lines correspond to 5 and 95 percentiles of the MC distribution of λ_i for pure random walk data (the null).

sample canonical correlations become similar under extremely different random walk and white noise data generating processes.

Our final MC experiment studies the finite sample behavior of $LR_{0,p,T}/\left(2p^2\right)$. We simulate pure random walk data with $p = 10$ and $p = 100$ and T varying so that p/T equals 1/10,2/10,...,5/10. Corollary 4 shows that the simultaneous asymptotic lower bound on $LR_{0,p,T}/\left(2p^2\right)$ has form

$$\frac{1+c}{2c^2}\ln\left(1+c\right)-\frac{1-c}{2c^2}\ln\left(1-c\right)+\frac{1-2c}{2c^2}\ln\left(1-2c\right). \tag{27}$$

Figure 8 shows the Tukey boxplots of the MC distributions of $LR_{0,p,T}/\left(2p^2\right)$ corresponding to $p/T = 1/10, ..., 5/10$ with $p = 10$ (left panel), and $p = 100$ (right panel). The boxplots are superimposed with the plot of (27) with c replaced by p/T. For $p = 10$, we also show (horizontal dashed line) the standard 95% asymptotic critical value (scaled by $1/(2p^2)$) taken from MacKinnon, Haug and Michelis (1999, Table II). For $p = 100$, the standard critical values are not available, and we show the dashed horizontal line at unit height instead. This is the sequential asymptotic lower bound on $LR_{0,p,T}/\left(2p^2\right)$ as established in Corollary 7. The reported results support our conjecture that the simultaneous asymptotic lower bound (27) is, in fact, the simultaneous asymptotic limit of $LR_{0,p,T}/\left(2p^2\right)$ for $c < 1/2$.

The left panel of Figure 8 illustrates the "over-rejection phenomenon". The horizontal dashed line that corresponds to the standard 95% critical value is just above the interquartile range of the MC distribution of $LR_{0,p,T}/\left(2p^2\right)$ for $c = 1/10$, is below this range for $c \geq 3/10$, and is below all 1000 MC replications of the scaled LR statistic for $c = 5/10$.

Although the lower bound (27) seems to provide a very good centering point for the scaled LR statistic, the MC distribution of this statistic is quite dispersed around such a center for $p = 10$. As discussed above, we suspect that the scaled statistic centered by (27) and appropriately rescaled has Gaussian simultaneous asymptotic distribution. Supporting this conjecture, the Tukey plots on Figure 8, that correspond to $c < 1/2$, look reasonably symmetric although some skewness is present for the left panel where $p = 10$.

4.2 Examples

Our first example uses $T = 103$ quarterly observations (1973q2-1998q4, with the initial observation 1973q1) on bilateral US dollar log nominal exchange rates for $p = 17$ OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Japan, Italy, Korea, Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom. The data are as in Engel, Mark, and West (2015), and were downloaded from Charles Engel's website at http: // www.ssc.wisc.edu / ~cengel /. That data are available for a longer time period up to 2008q1, but we have chosen to use only the "early sample" that does not include the Euro period.

Engel, Mark, and West (2015) point out that log nominal exchange rates are well modelled by random walk, but may be cointegrated, which can be utilized to improve individual exchange rate forecasts relative to the random walk forecast benchmark. They propose to estimate the common stochastic trends in the exchange rates by extracting a few factors from the panel. In principle, the number of factors to extract can be determined using Johansen's test for cointegrating rank, but Engel, Mark, and West (2015) do not exploit this possibility, referring to Ho and Sorensen (1996) that reports poor performance of the test for large p.

Figure 9 shows the Wachter plot for the log nominal exchange rate data. The squared sample canonical correlations are computed as the eigenvalues of $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$, where S_{ij} are defined as in (8) with R_{0t} and R_{1t} being the demeaned changes and the lagged levels of the log exchange rates, respectively. The dashed lines correspond to the 5-th and 95-th percentiles of the MC distribution of the squared canonical correlation coefficients under the null of no cointegration. To obtain these percentiles, we generated data from model (7) with $p = 17$, $T = 103$, $\Pi = 0$, $D_t = 1$, and Φ being i.i.d. $N(0, I_p)$ vectors across the MC repetitions. Log exchange rates for 1973q1 was used as the initial value of the generated series.

The figure shows a mild evidence for cointegration in the data with the largest five λ's being close to the corresponding 95-th percentiles of the MC distributions. Recall, however, that the ability of the Wachter plot to differentiate against highly persistent cointegration alternatives with $p/T \approx 1/5$ is very low, so there well may be more than five cointegrating relationships in the data. Whatever such relationships are, the deviations from the corresponding long-run equilibrium are probably highly persistent as no dramatic deviations from the $45°$ line are present in the Wachter plot.

Very different Wachter plots (shown in Figure 10) correspond to the log industrial

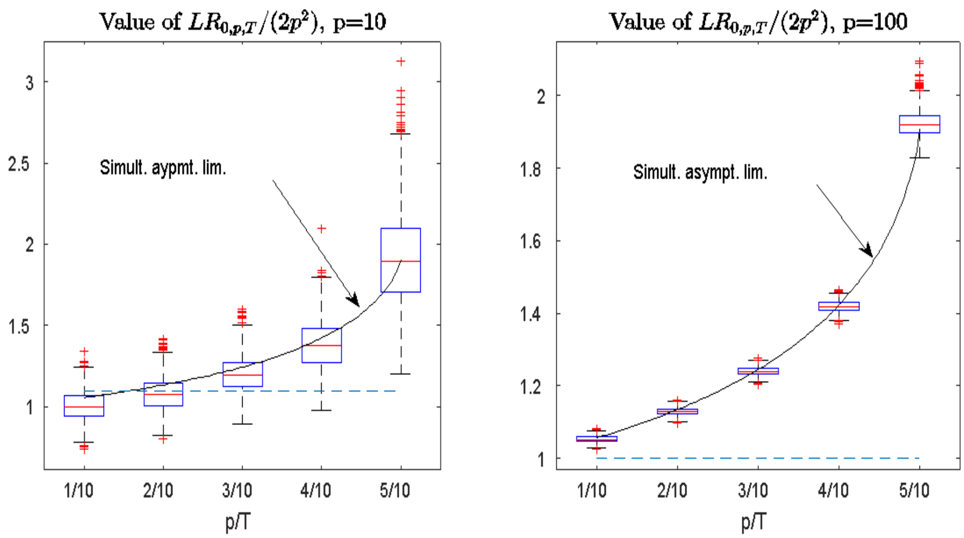

Figure 8: The Tukey boxplots for the MC distributions of $LR_{0,p,T}/\left(2p^2\right)$ for various p/T ratios. The boxplots are superimposed with the simultaneous asymptotic lower bound on $LR_{0,p,T}/\left(2p^2\right)$. Dashed line in the left panel correspond to 95% critical value for the satandard asymptotic LR test (taken from MacKinnon, Haug and Michelis (1999, Table II)). Dashed line in the right panel has ordinate equal one.

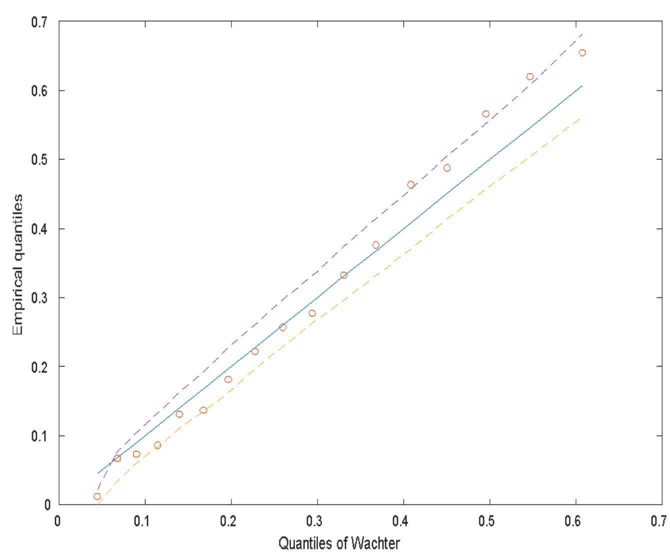

Figure 9: The Wachter plot for the bilateral US log nominal excahnge rates of 17 OECD countries. Dashed lines: 5% and 95% quantiles of the MC distribution of the squared sample canonical correlations under the null of no cointegration.

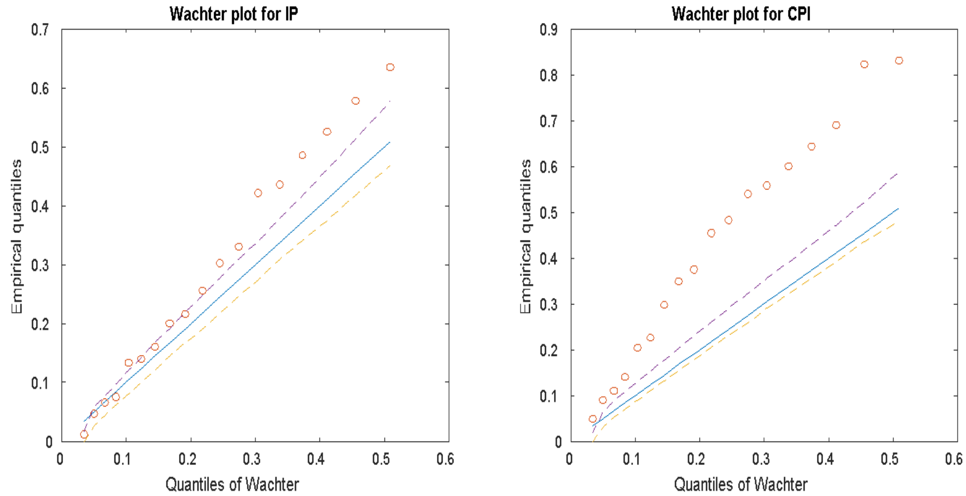

Figure 10: The Wachter plots for the industrial production indices and consumer price indices of 18 OECD countries. Dashed lines: 5% and 95% quantiles of the MC distribution of the squared sample canonical correlations under the null of no cointegration.

production (IP) index data and the log consumer price index (CPI) data for the same countries plus the US. These data are still the same as in Engel, Mark, and West (2015). We used the long sample 1973q2:2008q1 ($T = 140$) because the IP and CPI data are not affected by the introduction of the Euro to the same degree as the exchange rate data. For the CPI data, we included both intercept and trend in model (7) for the first differences because the level data seem to be quadratically trending. The plots clearly indicate that the IP and CPI data are either stationary or cointegrated with potentially many cointegrating relationships, short run deviations from which are not very persistent.

5 Conclusion

In this paper, we consider the simultaneous, large-p, large-T, asymptotic behavior of the squared sample canonical correlations between p-dimensional random walk and its innovations. We find that the empirical distribution of these squared sample canonical correlations almost surely weakly converges to the so-called *Wachter distribution* with parameters that depend only on the limit of p/T as $p, T \to_c \infty$. In contrast, under the sequential asymptotics, when first $T \to \infty$ and then $p \to \infty$, we establish the convergence in probability to the so-called Marchenko-Pastur distribution. The differences between the limiting distributions allow us to explain from a theoretical point of view the tendency of the LR test for cointegration to severely over-reject the null when the dimensionality of the data is relatively large. Furthermore, we derive a simple analytic formula for the Bartlett-type correction factor in systems with relatively large p/T ratio.

We propose a quick graphical method, the Wachter plot, for a preliminary analysis of cointegration in large-dimensional systems. The Monte Carlo analysis shows that the quantiles of the Wachter distribution constitute very good centering points for the finite sample distributions of the corresponding squared sample canonical correlations. The quality of the centering is excellent even for such small p and T as $p = 10$ and $T = 20$. However, for such small values of p and T, the empirical distribution of the squared sample canonical correlation can considerably fluctuate around the Wachter limit. As p increases to 100, the fluctuations become numerically very small.

This paper opens up many directions for future research. First, it is important to study the fluctuations of the empirical distribution around the Wachter limit. We conjecture that linear combinations of reasonably smooth functions of the squared sample canonical correlations, including the $\log(1 - \lambda)$ used by the LR statistic, are asymptotically Gaussian after appropriate centering and scaling. A proof would require different methods from those used here. We are currently investigating this research direction.

Further, it would be desirable to remove the Gaussianity assumption on the data. We believe that the existence of the finite fourth moments is sufficient for the validity of the Wachter limit. Next, it is interesting to study the simultaneous asymptotic behavior of a few of the largest sample canonical correlations. This may lead to a modification of Johansen's maximum eigenvalue test. Last, but not least, a study of the quality of bootstrap when p is large is needed. Our own very preliminary analysis indicates that the currently available non-parametric bootstrap procedures (see, for example, Cavaliere, Rahbek, and Taylor (2012)) might not work well for p/T as large as, say, 1/3. We hope that our paper would stimulate further research along these and other directions.

6 Appendix. Proof of Theorem 1.

6.1 Reduction to pure random walk data.

Let $G(\lambda)$ and $\tilde{G}(\lambda)$ be d.f.'s that may depend on p and T and are possibly random. We shall call them asymptotically equivalent if the a.s. weak convergence $G(\lambda) \Rightarrow F(\lambda)$ to some non-random d.f. $F(\lambda)$ implies similar a.s. weak convergence for $\tilde{G}(\lambda)$, and vice versa. Let S_i and $\tilde{S}_i$ with $i = 0, 1, 2$ be, possibly random, matrices that may depend on p and T such that S_i and $\tilde{S}_i$ are a.s. positive definite for $i = 0, 1$. Below, we shall often refer to the following auxiliary lemma.

Lemma 9 *If $\frac{1}{p}\operatorname{rank}\left(S_i - \tilde{S}_i\right) \to 0$, a.s., as $p, T \to_c \infty$ for $i = 0, 1, 2$, then $G(\lambda)$ and $\tilde{G}(\lambda)$ are asymptotically equivalent, where $G(\lambda)$ and $\tilde{G}(\lambda)$ are the empirical d.f.'s of eigenvalues of $S_2 S_1^{-1} S_2' S_0^{-1}$ and $\tilde{S}_2 \tilde{S}_1^{-1} \tilde{S}_2' \tilde{S}_0^{-1}$, respectively.*

Proof of Lemma 9. Let $R = \operatorname{rank}\left(S_2 S_1^{-1} S_2' S_0^{-1} - \tilde{S}_2 \tilde{S}_1^{-1} \tilde{S}_2' \tilde{S}_0^{-1}\right)$. The a.s. convergence $\frac{1}{p}\operatorname{rank}\left(S_i - \tilde{S}_i\right) \to 0$ implies the a.s. convergence $R/p \to 0$. On the other hand, by the rank inequality (Theorem A43 in Bai and Silverstein (2010)), $\mathcal{L}\left(G, \tilde{G}\right) \leq R/p$, where $\mathcal{L}\left(G, \tilde{G}\right)$ is the Lévy distance between $G(\lambda)$ and $\tilde{G}(\lambda)$. Since the Lévy distance metrizes the weak convergence, the a.s. convergence $\mathcal{L}\left(G, \tilde{G}\right) \to 0$ yields the asymptotic equivalence of $G(\lambda)$ and $\tilde{G}(\lambda)$.□

Now, let $S_0 = S_{00}, S_1 = S_{11}$, and $S_2 = S_{01}$, and let

$$\tilde{S}_0 = \frac{1}{T}\sum_{t=1}^{T} \Delta X_t \Delta X_t', \quad \tilde{S}_1 = \frac{1}{T}\sum_{t=1}^{T} X_{t-1} X_{t-1}', \text{ and } \tilde{S}_2 = \frac{1}{T}\sum_{t=1}^{T} \Delta X_t X_{t-1}'.$$

Since R_{0t} and R_{1t}, which enter the definition (8) of S_{ij}, are the residuals in the regressions of ΔX_t on D_t and X_{t-1} on D_t, respectively, we have $\max_{i=0,1,2} \operatorname{rank}\left(S_i - \tilde{S}_i\right) \leq d_D$. By assumption, $d_D/p \to 0$ as $p, T \to_c \infty$, so that by Lemma 9, $F_{p,T}(\lambda)$ is asymptotically equivalent to the empirical d.f. of eigenvalues of $\tilde{S}_2 \tilde{S}_1^{-1} \tilde{S}_2' \tilde{S}_0^{-1}$. Therefore, we may and will replace R_{0t} and R_{1t} in the definitions (8) of S_{ij} by ΔX_t and X_{t-1}, respectively, without loss of generality. Furthermore, scaling S_{ij} by T does not change the product $S_{01}S_{11}^{-1}S_{01}'S_{00}^{-1}$, and thus, in the rest of the proof, we work with

$$S_{00} = \sum_{t=1}^{T} \Delta X_t \Delta X_t', \;\; S_{01} = \sum_{t=1}^{T} \Delta X_t X_{t-1}', \;\text{ and } S_{11} = \sum_{t=1}^{T} X_{t-1} X_{t-1}'. \tag{28}$$

Next, we show that, still without loss of generality, we may replace the data generated process (11) by a pure random walk with zero initial value. Indeed, let $X = [X_{-k+1}, ..., X_T]$, where $X_{-k+1}, ..., X_0$ are arbitrary and X_t with $t \geq 1$ are generated by (11). Further, let $\tilde{X}_{-k+1}, ..., \tilde{X}_0$ be zero vectors, $\tilde{X}_t = \sum_{s=1}^{t} \varepsilon_t$ for $t \geq 1$, and $\tilde{X} = [\tilde{X}_{-k+1}, ..., \tilde{X}_T]$.

Lemma 10 rank $\left(X - \tilde{X}\right) \leq 2\left(r + \operatorname{rank}\Gamma + k + d_F\right)$.

A proof of this lemma is given in the SM. It is based on the representation of X_t as a function of the initial values, ε and F (see Theorem 2.1 in Johansen (1995)), and requires only elementary algebraic manipulations. Lemmas 10 and 9 together with assumption (12) imply that replacing ΔX_t and X_{t-1} in (28) by $\Delta\tilde{X}_t$ and $\tilde{X}_{t-1}$, respectively, does not change the weak limit of $F_{p,T}(\lambda)$. Hence, in the rest of the proof of Theorem 1, without loss of generality, we assume that the data are generated by

$$\Delta X_t = \varepsilon_t, \;\; t = 1, ..., T, \;\text{ with } X_0 = 0. \tag{29}$$

Since the sample canonical correlations are invariant with respect to the multiplication of the data by any invertible matrix, we assume without loss of generality that the variance of ε_t equals $\Sigma = I_p/T$. Further, we assume that T is even. The case of odd T can be analyzed similarly, and we omit it to save space.

6.2 Block-diagonalization

Let $\varepsilon = [\varepsilon_1, ..., \varepsilon_T]$ and let U be the upper-triangular matrix with ones above the main diagonal and zeros on the diagonal. Then $\varepsilon U = [X_0, ..., X_{T-1}]$ so that

$$S_{00} = \varepsilon\varepsilon', \;\; S_{01} = \varepsilon U'\varepsilon', \;\text{ and } S_{11} = \varepsilon U U'\varepsilon'. \tag{30}$$

We shall show that the empirical d.f. of the λ's, $F_{p,T}(\lambda)$, is asymptotically equivalent to the empirical d.f. $\hat{F}_{p,T}(\lambda)$ of eigenvalues of $CD^{-1}C'A^{-1}$, where

$$C = \varepsilon\Delta_2'\varepsilon',\ D = \varepsilon\Delta_1\varepsilon',\ \text{and } A = \varepsilon\varepsilon',$$

Δ_1 is a diagonal matrix,

$$\Delta_1 = \text{diag}\left\{r_1^{-1}I_2, ..., r_{T/2}^{-1}I_2\right\}, \tag{31}$$

and Δ_2 is a block-diagonal matrix,

$$\Delta_2 = \text{diag}\left\{r_1^{-1}(R_1 - I_2), ..., r_{T/2}^{-1}\left(R_{T/2} - I_2\right)\right\}. \tag{32}$$

Here I_2 is the 2-dimensional identity matrix, and r_j, R_j are defined as follows. Let $\theta = -2\pi/T$. Then for $j = 1, 2, ..., T/2 - 1$,

$$r_{j+1} = 2 - 2\cos j\theta,\ R_{j+1} = \begin{pmatrix} \cos j\theta & -\sin j\theta \\ \sin j\theta & \cos j\theta \end{pmatrix},$$

whereas $r_1 = 4$, $R_1 = -I_2$.

Lemma 11 *The d.f.'s $F_{p,T}(\lambda)$ and $\hat{F}_{p,T}(\lambda)$ are asymptotically equivalent.*

Proof of Lemma 11. Let V be the circulant matrix (see Golub and Van Loan (1996, p.201)) with the first column $v = (-1, 1, 0, ..., 0)'$. Direct calculations show that $UV = I_T - le_T'$ and $VU = I_T - e_1l'$, where e_j is the j-th column of I_T, and l is the vector of ones. Using these identities, it is straightforward to verify that

$$U = (V + e_1e_1')^{-1} - le_1',\ \text{and} \tag{33}$$

$$UU' = \left(V'V - (e_1 - e_T)(e_1 - e_T)' + e_Te_T'\right)^{-1} - ll'. \tag{34}$$

Now, let us define

$$C_1 = \varepsilon(U + le_1')'\varepsilon' \text{ and } D_1 = \varepsilon(UU' + ll')\varepsilon'.$$

Using identities (30) for S_{ij} and Lemma 9, we conclude that $F_{p,T}(\lambda)$ is asymptotically equivalent to $F_{p,T}^{(1)}(\lambda)$, where $F_{p,T}^{(1)}(\lambda)$ is the empirical d.f. of the eigenvalues of

$C_1 D_1^{-1} C_1' A^{-1}$.

Further, (33) and (34) yield

$$\begin{aligned} C_1 &= \varepsilon \left(V + e_1 e_1'\right)^{-1} \varepsilon' \text{ and} \\ D_1 &= \varepsilon \left(V'V - \left(e_1 - e_T\right)\left(e_1 - e_T\right)' + e_T e_T'\right)^{-1} \varepsilon'. \end{aligned}$$

Applying Lemma 9 one more time, we obtain the asymptotic equivalence of $F_{p,T}^{(1)}(\lambda)$ and $F_{p,T}^{(2)}(\lambda)$, where $F_{p,T}^{(2)}(\lambda)$ is the empirical d.f. of the eigenvalues of $C_2 D_2^{-1} C_2' A^{-1}$ with

$$C_2 = \varepsilon V^{-1} \varepsilon' \text{ and } D_2 = \varepsilon \left(V'V\right)^{-1} \varepsilon'. \tag{35}$$

As is well known (see, for example, Golub and Van Loan (1996), chapter 4.7.7), $T \times T$ circulant matrices can be expressed in terms of the discrete Fourier transform matrices $\mathcal{F} = \{\exp\left(\mathrm{i}\theta\left(s-1\right)\left(t-1\right)\right)\}_{s,t=1}^{T}$ with $\theta = -2\pi/T$. Precisely,

$$V = \frac{1}{T}\mathcal{F}^* \operatorname{diag}\left(\mathcal{F}v\right)\mathcal{F}, \text{ and } V'V = \frac{1}{T}\mathcal{F}^* \operatorname{diag}\left(\mathcal{F}w\right)\mathcal{F},$$

where $w = (2, -1, 0, ..., 0, -1)'$ and the star superscript denotes transposition and complex conjugation. For the s-th diagonal elements of $\operatorname{diag}(\mathcal{F}v)$ and $\operatorname{diag}(\mathcal{F}w)$, we have

$$\operatorname{diag}\left(\mathcal{F}v\right)_s = -1 + \exp\left\{\mathrm{i}\theta\left(s-1\right)\right\}, \text{ and } \operatorname{diag}\left(\mathcal{F}w\right)_s = 2 - 2\cos\left(s-1\right)\theta.$$

Note that $\operatorname{diag}(\mathcal{F}w)_s = \operatorname{diag}(\mathcal{F}w)_{T+2-s}$ for $s = 2, 3, ...$ Define a permutation matrix P so that the equal diagonal elements of $P' \operatorname{diag}(\mathcal{F}w) P$ are grouped in adjacent pairs. Precisely, let $P = \{p_{st}\}$, where

$$p_{st} = \begin{cases} 1 \text{ if } t = 2s - 1 \text{ for } s = 1, ..., T/2 \\ 1 \text{ if } t = 2\left(T - s + 2\right) \operatorname{mod} T \text{ for } s = T/2 + 1, ..., T \\ 0 \text{ otherwise} \end{cases}$$

and let W be the unitary matrix

$$W = \begin{pmatrix} I_2 & 0 \\ 0 & I_{T/2} \otimes Z \end{pmatrix} \text{ with } Z = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{pmatrix},$$

where $\otimes$ denotes the Kronecker product. Further, let $Q = \frac{1}{\sqrt{T}} W P' \mathcal{F}$. As is easy to check, Q is an orthogonal matrix. Furthermore,

$$V = Q' \left(\Delta_2^{-1} + 2 e_1 e_1' \right) Q, \text{ and } VV' = Q' \left(\Delta_1^{-1} - 4 e_1 e_1' \right) Q,$$

where Δ_1 and Δ_2 are as defined in (31) and (32). Combining this with (35) and using Lemma 9 once again, we obtain the asymptotic equivalence of $F_{p,T}^{(2)}(\lambda)$ and $F_{p,T}^{(3)}(\lambda)$, where $F_{p,T}^{(3)}(\lambda)$ is the empirical d.f. of the eigenvalues of $C_3 D_3^{-1} C_3' A^{-1}$ with

$$C_3 = \varepsilon Q' \Delta_2 Q \varepsilon' \text{ and } D_3 = \varepsilon Q' \Delta_1 Q \varepsilon'.$$

Because of the rotational invariance of the Gaussian distribution, the distributions of $\varepsilon Q'$ and ε are the same. Hence, $F_{p,T}^{(3)}(\lambda)$ is asymptotically equivalent to $\hat{F}_{p,T}(\lambda)$, and thus, $\hat{F}_{p,T}(\lambda)$ is asymptotically equivalent to $F_{p,T}(\lambda)$.□

6.3 A system of equations for the Stieltjes transform

Our proof of the a.s. weak convergence of $\hat{F}_{p,T}(\lambda)$ to the Wachter distribution consists of showing that the Stieltjes transform of $\hat{F}_{p,T}(\lambda)$,

$$\hat{m}_{p,T}(z) = \int \frac{1}{\lambda - z} \hat{F}_{p,T}(\mathrm{d}\lambda), \tag{36}$$

a.s. converges pointwise in $z \in \mathbb{C}^+ = \{\zeta : \Im\zeta > 0\}$, where $\Im\zeta$ denotes the imaginary part of a complex number ζ, to the Stieltjes transform $m(z)$ of the Wachter distribution. To establish such a convergence, we show that, if m is a limit of $\hat{m}_{p,T}(z)$ along any *subsequence* of $p, T \to_c \infty$, then it must satisfy a system of equations with unique solution given by $m(z)$. The a.s. convergence of $\hat{F}_{p,T}(\lambda)$ (and thus, also of $F_{p,T}(\lambda)$) to the Wachter distribution follows then from the Continuity Theorem for the Stieltjes transforms (see, for example, Corollary 1 in Geronimo and Hill (2003)).

We shall write $\hat{m}$ for the Stieltjes transform $\hat{m}_{p,T}(z)$ to simplify notation. Let

$$M = CD^{-1}C' - zA \text{ and } \tilde{M} = C'A^{-1}C - zD. \tag{37}$$

Then by definition (36), $\hat{m}$ must satisfy the following equations

$$\hat{m} = \frac{1}{p}\operatorname{tr}\left[AM^{-1}\right], \tag{38}$$

$$\hat{m} = \frac{1}{p}\operatorname{tr}\left[D\tilde{M}^{-1}\right]. \tag{39}$$

Let us study the above traces in detail. Define

$$\varepsilon_{(j)} = \left[\varepsilon_{2j-1}, \varepsilon_{2j}\right],\ j = 1, ..., T/2.$$

We now show that the traces in (38) and (39) can be expressed as functions of the terms having form $\varepsilon'_{(j)}\Omega_j\varepsilon_{(j)}$, where Ω_j is independent from $\varepsilon_{(j)}$. Then, we argue that

$$\varepsilon'_{(j)}\Omega_j\varepsilon_{(j)} - \frac{1}{T}\operatorname{tr}\left[\Omega_j\right] I_2$$

a.s. converge to zero, and use this fact to derive equations that the limit of $\hat{m}$, if it exists, must satisfy.

First, consider (38). Note that

$$\frac{1}{p}\operatorname{tr}\left[AM^{-1}\right] = \frac{1}{p}\sum_{j=1}^{T/2}\operatorname{tr}\left[\varepsilon'_{(j)}M^{-1}\varepsilon_{(j)}\right]. \tag{40}$$

Let us introduce new notation:

$$\Delta_{1j} = r_j^{-1}I_2,\ \Delta_{2j} = r_j^{-1}\left(R_j - I_2\right),$$

$$C_j = C - \varepsilon_{(j)}\Delta'_{2j}\varepsilon'_{(j)},\ D_j = D - \varepsilon_{(j)}\Delta_{1j}\varepsilon'_{(j)},$$

$$A_j = A - \varepsilon_{(j)}\varepsilon'_{(j)},\ \text{and}\ M_j = C_jD_j^{-1}C'_j - zA_j.$$

In addition, let

$$\begin{aligned} s_j &= \varepsilon'_{(j)}D_j^{-1}\varepsilon_{(j)}, u_j = \varepsilon'_{(j)}D_j^{-1}C'_jM_j^{-1}\varepsilon_{(j)}, \\ v_j &= \varepsilon'_{(j)}M_j^{-1}\varepsilon_{(j)},\ \text{and} \\ w_j &= \varepsilon'_{(j)}D_j^{-1}C'_jM_j^{-1}C_jD_j^{-1}\varepsilon_{(j)}. \end{aligned}$$

A straightforward algebra that involves multiple use of the Sherman-Morrison-

Woodbury formula (see Golub and Van Loan (1996), p.50)

$$(V + XWY)^{-1} = V^{-1} - V^{-1}X\left(W^{-1} + YV^{-1}X\right)^{-1}YV^{-1}, \tag{41}$$

and the identity

$$\Delta_{2j}\Delta'_{2j} = \Delta'_{2j}\Delta_{2j} = \Delta_{1j}, \tag{42}$$

establishes the following equality

$$\varepsilon'_{(j)}M^{-1}\varepsilon_{(j)} = v_j - [v_j, u'_j]\Omega_j[v_j, u'_j]', \tag{43}$$

where

$$\Omega_j = \begin{pmatrix} \frac{1}{1-z}I_2 + v_j & \frac{1}{1-z}r_j\Delta'_{2j} + u'_j \\ \frac{1}{1-z}r_j\Delta_{2j} + u_j & \frac{z}{1-z}r_jI_2 - s_j + w_j \end{pmatrix}^{-1}.$$

A derivation of (43) can be found in the SM.

Let us define

$$\begin{aligned} \hat{s} &= \frac{1}{T}\operatorname{tr}\left[D^{-1}\right],\ \hat{u} = \frac{1}{T}\operatorname{tr}\left[D^{-1}C'M^{-1}\right], \\ \hat{v} &= \frac{1}{T}\operatorname{tr}\left[M^{-1}\right],\ \text{and} \\ \hat{w} &= \frac{1}{T}\operatorname{tr}\left[D^{-1}C'M^{-1}CD^{-1}\right]. \end{aligned}$$

We have the following lemma, where $\|\cdot\|$ denotes the spectral norm. Its proof is given in the SM.

Lemma 12 *For all $z \in \mathbb{C}^+$, as $p, T \to_c \infty$, we have*

$$\max_{j=1,\ldots,T/2}\|s_j - \hat{s}I_2\| \stackrel{a.s.}{\to} 0,\quad \max_{j=1,\ldots,T/2}\|u_j - \hat{u}I_2\| \stackrel{a.s.}{\to} 0$$
$$\max_{j=1,\ldots,T/2}\|v_j - \hat{v}I_2\| \stackrel{a.s.}{\to} 0,\quad \max_{j=1,\ldots,T/2}\|w_j - \hat{w}I_2\| \stackrel{a.s.}{\to} 0.$$

The lemma yields an approximation to the right hand side of (43), which we use in (40) and (38) to obtain the following result.

Proposition 13 *There exists $\zeta > 0$ such that, for any z with zero real part, $\Re z = 0$,* *and the imaginary part satisfying $\Im z > \zeta$, we have*

$$\hat{m} = \frac{1}{2\pi c}\int_0^{2\pi} \frac{f_1(\varphi)}{(1-z)\, f_1(\varphi) + f_2(\varphi)} \mathrm{d}\varphi + o(1), \ \textit{where} \tag{44}$$

$$\begin{aligned} f_1(\varphi) &= \left(\hat{w} - \hat{s} - 4\sin^2\varphi\right)\hat{v} - \hat{u}^2, \\ f_2(\varphi) &= \hat{w} - \hat{s} - 4\sin^2\varphi\,(1 - \hat{u} - \hat{v}), \end{aligned}$$

and $o(1) \overset{a.s}{\to} 0$, as $p, T \to_c \infty$.

Proof of Proposition 13. Consider a 2×2 matrix $\hat{S}_j$ that is obtained from $\varepsilon'_{(j)} M^{-1} \varepsilon_{(j)}$ by replacing s_j, v_j, u_j and w_j in (43) with $\hat{s}I_2, \hat{v}I_2, \hat{u}I_2$, and $\hat{w}I_2$, respectively. We have

$$\hat{S}_j = \hat{v}I_2 - [\hat{v}I_2, \hat{u}I_2]\hat{\Omega}_j[\hat{v}I_2, \hat{u}I_2]',$$

where

$$\hat{\Omega}_j = \begin{pmatrix} \frac{1}{1-z}I_2 + \hat{v}I_2 & \frac{1}{1-z}r_j\Delta'_{2j} + \hat{u}I_2 \\ \frac{1}{1-z}r_j\Delta_{2j} + \hat{u}I_2 & \frac{z}{1-z}r_jI_2 + (\hat{w} - \hat{s})I_2 \end{pmatrix}^{-1}.$$

A simple algebra and the identity $\Delta_{2j} + \Delta'_{2j} = -I_2$ yield

$$\hat{\Omega}_j = \frac{1-z}{\delta_j}\tilde{\Omega}_j, \ \text{where} \tag{45}$$

$$\tilde{\Omega}_j = \begin{pmatrix} \frac{z}{1-z}r_jI_2 + (\hat{w} - \hat{s})I_2 & -\frac{1}{1-z}r_j\Delta'_{2j} - \hat{u}I_2 \\ -\frac{1}{1-z}r_j\Delta_{2j} - \hat{u}I_2 & \frac{1}{1-z}I_2 + \hat{v}I_2 \end{pmatrix}, \tag{46}$$

and

$$\delta_j = (\hat{w} - \hat{s})(1 + \hat{v} - z\hat{v}) + r_j(\hat{u} + z\hat{v} - 1) - (1 - z)\hat{u}^2.$$

By definition,

$$\begin{aligned} |\hat{s}| &\le \frac{p}{T}\left\|D^{-1}\right\|, |\hat{u}| \le \frac{p}{T}\operatorname{tr}\left\|D^{-1}C'M^{-1}\right\|, \\ |\hat{v}| &\le \frac{p}{T}\left\|M^{-1}\right\|, \ \text{and} \ |\hat{w}| \le \frac{p}{T}\operatorname{tr}\left\|D^{-1}C'M^{-1}CD^{-1}\right\|. \end{aligned}$$

In the proof of Lemma 12, we show that the norms $\|D^{-1}\|$, $\|D^{-1}C'\|$, and $\|M^{-1}\|$ a.s. remain bounded as $p, T \to_c \infty$. Hence, $\hat{s}$, $\hat{u}$, $\hat{v}$, and $\hat{w}$ are also a.s. bounded.

Further, by definition,

$$r_j\Delta_{2j} = R_j - I_2 \text{ and } r_j\Delta'_{2j} = R'_j - I_2,$$

where R_j is an orthogonal matrix, so that $\|r_j\Delta_{2j}\|$ and $\left\|r_j\Delta'_{2j}\right\|$ are clearly bounded uniformly in j. Therefore, the norm of matrix $\tilde{\Omega}_j$ a.s. remains bounded as $p,T \to_c \infty$, uniformly in j. Regarding δ_j, which appear in the denominator on the right hand side of (45), the SM establishes the following result.

Lemma 14 *There exists $\zeta > 0$ such that, for any z with $\Re z = 0$ and $\Im z > \zeta$, a.s.,*

$$\liminf_{p,T\to_c\infty} \max_{j=1,\dots,T/2} |\delta_j| > c^2/\left(1-c^2\right).$$

The above results imply that, for z with $\Re z = 0$ and $\Im z > \zeta$, $\left\|\hat{\Omega}_j\right\|$ a.s. remains bounded as $p,T \to_c \infty$, uniformly in j. Therefore, by Lemma 12,

$$\varepsilon'_{(j)}M^{-1}\varepsilon_{(j)} = \hat{S}_j + o(1), \tag{47}$$

where $o(1) \overset{a.s.}{\to} 0$ as $p,T \to_c \infty$, uniformly in j.

A straightforward algebra reveals that

$$\hat{S}_j = \frac{(\hat{w} - \hat{s} - r_j)\,\hat{v} - \hat{u}^2}{\delta_j}.$$

Using this in equations (47) and (40), we obtain

$$\begin{aligned}
\hat{m} &= \frac{2}{p}\sum_{j=0}^{T/2-1} \frac{(\hat{w} - \hat{s} - r_{j+1})\,\hat{v} - \hat{u}^2}{\delta_{j+1}} + o(1) \\
&= \frac{2}{p}\sum_{j=1}^{T/2-1} \frac{f_1(j\pi/T)}{(1-z)\,f_1(j\pi/T) + f_2(j\pi/T)} + o(1),
\end{aligned}$$

where, in the latter expression, the term corresponding to $j = 0$ is included in the $o(1)$ term to take into account the special definition of r_1.

As follows from Lemma 14 and the boundedness of $\hat{s}, \hat{u}, \hat{v}$, and $\hat{w}$, the derivative

$$\frac{\mathrm{d}}{\mathrm{d}\varphi}\frac{f_1(\varphi)}{(1-z)\,f_1(\varphi) + f_2(\varphi)}$$

a.s. remains bounded by absolute value as $p, T \to_c \infty$, uniformly in $\varphi \in [0, 2\pi]$. Therefore

$$\frac{2}{p}\sum_{j=1}^{T/2-1}\frac{f_1(j\pi/T)}{(1-z)f_1(j\pi/T)+f_2(j\pi/T)} = \frac{2}{\pi c}\int_0^{\pi/2}\frac{f_1(\varphi)\,\mathrm{d}\varphi}{(1-z)f_1(\varphi)+f_2(\varphi)} + o(1).$$

The statement of Proposition 13 now follows by noting that the latter integral is one quarter of the integral over $[0, 2\pi]$.□

A similar analysis of equation (39) gives us another proposition, describing $\hat{m}$ as function of $\tilde{s}, \tilde{u}, \tilde{v}$, and $\tilde{w}$, where

$$\begin{aligned}
\tilde{s} &= \frac{1}{T}\operatorname{tr}\left[A^{-1}\right],\ \tilde{u} = \frac{1}{T}\operatorname{tr}\left[A^{-1}C\tilde{M}^{-1}\right], \\
\tilde{v} &= \frac{1}{T}\operatorname{tr}\left[\tilde{M}^{-1}\right], \text{ and} \\
\tilde{w} &= \frac{1}{T}\operatorname{tr}\left[A^{-1}C\tilde{M}^{-1}C'A^{-1}\right].
\end{aligned}$$

We omit the proof because it is very similar to that of Proposition 13.

Proposition 15 *There exists $\zeta > 0$ such that, for any z with $\Re z = 0$ and $\Im z > \zeta$, we have*

$$\hat{m} = \frac{1}{2\pi c}\int_0^{2\pi}\frac{g_1}{(1-z)g_1 + g_2(\varphi)}\mathrm{d}\varphi + o(1), \text{ where} \tag{48}$$

$$\begin{aligned}
g_1 &= (\tilde{w} - \tilde{s} - 1)\tilde{v} - \tilde{u}^2, \\
g_2(\varphi) &= \tilde{v} - 4\sin^2\varphi\,(\tilde{s} + 1 - \tilde{u} - \tilde{w}),
\end{aligned}$$

and $o(1) \overset{a.s.}{\to} 0$, as $p, T \to_c \infty$.

Although we now have two asymptotic equations for $\hat{m}$, (44) and (48), they contain eight unknowns: $\hat{s}, \hat{u}, \hat{v}, \hat{w}$, and the corresponding variables with tildes. Using a simple algebra, we establish the following relationships between the unknowns with hats and tildes. A proof can be found in the SM.

Lemma 16 *We have the following three identities*

$$\hat{u} = \tilde{u},\ z\tilde{v} + \hat{s} = \hat{w}, \text{ and } z\hat{v} + \tilde{s} = \tilde{w}. \tag{49}$$

The identities (49) imply the following equality

$$(1-z)\,f_1(\varphi)+f_2(\varphi)=(1-z)\,g_1+g_2(\varphi)\,.$$

We denote the reciprocal of the common value of the right and left hand sides of this equality as $\hat{h}(z,\varphi)$. A direct calculation shows that

$$\hat{h}(z,\varphi)=\left((1-z)\left(z\tilde{v}\hat{v}-\hat{u}^2\right)+z\tilde{v}+4\sin^2\varphi\,(z\hat{v}+\hat{u}-1)\right)^{-1}, \tag{50}$$

and the asymptotic relationships (44) and (48) can be written in the following form

$$\begin{cases} \hat{m}=\frac{1}{2\pi c}\int_0^{2\pi}\hat{h}(z,\varphi)\left(\left(z\tilde{v}-4\sin^2\varphi\right)\hat{v}-\hat{u}^2\right)\mathrm{d}\varphi+o(1) \\ \hat{m}=\frac{1}{2\pi c}\int_0^{2\pi}\hat{h}(z,\varphi)\left((z\hat{v}-1)\,\tilde{v}-\hat{u}^2\right)\mathrm{d}\varphi+o(1) \end{cases}. \tag{51}$$

This can be viewed as an asymptotic system of two equations with four unknowns: $\hat{m},\tilde{v},\hat{v}$, and $\hat{u}$. We shall now complete the system by establishing the other two asymptotic relationships connecting these unknowns.

Multiplying both sides of the identity

$$MA^{-1}=CD^{-1}C'A^{-1}-zI_p \tag{52}$$

by AM^{-1}, taking trace, dividing by p, and rearranging terms, we obtain

$$1+z\hat{m}=\frac{1}{p}\operatorname{tr}\left[CD^{-1}C'M^{-1}\right]. \tag{53}$$

Next, we analyze (53) similarly to the above analysis of (38). That is, first, we note that

$$\frac{1}{p}\operatorname{tr}\left[CD^{-1}C'M^{-1}\right]=\frac{1}{p}\sum_{j=1}^{T/2}\operatorname{tr}\left[\Delta_{2j}'\varepsilon_{(j)}'D^{-1}C'M^{-1}\varepsilon_{(j)}\right]. \tag{54}$$

Then elementary algebra, based on the Sherman-Morrison-Woodbury formula (41), yields

$$\begin{aligned} \varepsilon_{(j)}'D^{-1}C'M^{-1}\varepsilon_{(j)} &= r_j\left(r_jI_2+s_j\right)^{-1}s_j\Delta_{2j}\left(v_j-\left[v_j,u_j'\right]\Omega_j\left[v_j,u_j'\right]'\right) \\ &\quad +r_j\left(r_jI_2+s_j\right)^{-1}\left(u_j-\left[u_j,w_j\right]\Omega_j\left[v_j,u_j'\right]'\right). \end{aligned} \tag{55}$$

Multiplying both sides of (55) by Δ'_{2j} and replacing s_j, u_j, v_j, and w_j by $\hat{s}I_2, \hat{u}I_2, \hat{v}I_2$, and $\hat{w}I_2$, respectively, yields an asymptotic approximation to $\Delta'_{2j}\varepsilon'_{(j)}D^{-1}C'M^{-1}\varepsilon_{(j)}$, which can be used in (54) and (53) to produce the following result. Its proof, as well as the proof of (55), are given in the SM.

Proposition 17 *There exists $\zeta > 0$ such that, for any z with $\Re z = 0$ and $\Im z > \zeta$, we have*

$$1 + z\hat{m} = \frac{1}{2\pi c}\int_0^{2\pi} \hat{h}(z,\varphi)\left(2\hat{u}\sin^2\varphi + z\tilde{v}\hat{v} - \hat{u}^2\right)\mathrm{d}\varphi + o(1), \text{ where} \tag{56}$$

$o(1) \overset{a.s.}{\to} 0$, as $p, T \to_c \infty$.

One might think that the remaining asymptotic relationship can be obtained by using the identity

$$\tilde{M}D^{-1} = C'A^{-1}CD^{-1} - zI_p, \tag{57}$$

which parallels (52). Unfortunately, following this idea delivers a relationship equivalent to (56). Therefore, instead of using (57), we consider the identity

$$\frac{1}{p}\operatorname{tr}\left[C'M^{-1}\right] = \frac{1}{p}\operatorname{tr}\left[DD^{-1}C'M^{-1}\right], \tag{58}$$

which yields

$$\frac{1}{p}\sum_{j=1}^{T/2}\operatorname{tr}\left[\Delta_{2j}\varepsilon'_{(j)}M^{-1}\varepsilon_{(j)}\right] = \frac{1}{p}\sum_{j=1}^{T/2}\operatorname{tr}\left[\Delta_{1j}\varepsilon'_{(j)}D^{-1}C'M^{-1}\varepsilon_{(j)}\right]. \tag{59}$$

Then, we proceed as in the above analysis of (54) and (40) to obtain the remaining asymptotic relationship. The proof of the following proposition is given in the SM.

Proposition 18 *There exists $\zeta > 0$ such that, for any z with $\Re z = 0$ and $\Im z > \zeta$, we have*

$$0 = \frac{1}{2\pi c}\int_0^{2\pi} \hat{h}(z,\varphi)\left(4\hat{v}\sin^2\varphi + 2\hat{u}\right)\mathrm{d}\varphi + o(1), \text{ where} \tag{60}$$

$o(1) \overset{a.s.}{\to} 0$, as $p, T \to_c \infty$.

Summing up the results in Propositions 13, 15, 17, and 18, the unknowns $\hat{m}, \hat{v}, \tilde{v}$, and $\hat{u}$ must satisfy the following system of asymptotic equations

$$\begin{cases} \hat{m} = \frac{1}{2\pi c} \int_0^{2\pi} \hat{h}(z,\varphi) \left(\left(z\tilde{v} - 4\sin^2\varphi \right) \hat{v} - \hat{u}^2 \right) \mathrm{d}\varphi + o(1) \\ \hat{m} = \frac{1}{2\pi c} \int_0^{2\pi} \hat{h}(z,\varphi) \left((z\hat{v} - 1)\tilde{v} - \hat{u}^2 \right) \mathrm{d}\varphi + o(1) \\ 1 + z\hat{m} = \frac{1}{2\pi c} \int_0^{2\pi} \hat{h}(z,\varphi) \left(2\hat{u}\sin^2\varphi + z\tilde{v}\hat{v} - \hat{u}^2 \right) \mathrm{d}\varphi + o(1) \\ 0 = \frac{1}{2\pi c} \int_0^{2\pi} \hat{h}(z,\varphi) \left(4\hat{v}\sin^2\varphi + 2\hat{u} \right) \mathrm{d}\varphi + o(1) \end{cases}. \tag{61}$$

6.4 Solving the system

The definition (36) of $\hat{m}$ implies that $|\hat{m}|$ is bounded by $(\Im z)^{-1}$. Further, as shown in the proof of Proposition 13, $\hat{u}$ and $\hat{v}$ are a.s. bounded by absolute value, and it can be similarly shown that $\tilde{v}$ is a.s. bounded by absolute value. Therefore, there exists a subsequence of p, T along which $\hat{m}, \hat{v}, \tilde{v}$, and $\hat{u}$ a.s. converge to some limits m, v, y, and u.

These limits must satisfy a non-asymptotic system of equations

$$\begin{cases} m = \frac{1}{2\pi c} \int_0^{2\pi} h(z,\varphi) \left(\left(zy - 4\sin^2\varphi \right) v - u^2 \right) \mathrm{d}\varphi \\ m = \frac{1}{2\pi c} \int_0^{2\pi} h(z,\varphi) \left((zv - 1) y - u^2 \right) \mathrm{d}\varphi \\ 1 + zm = \frac{1}{2\pi c} \int_0^{2\pi} h(z,\varphi) \left(2u\sin^2\varphi + zvy - u^2 \right) \mathrm{d}\varphi \\ 0 = \frac{1}{2\pi c} \int_0^{2\pi} h(z,\varphi) \left(2v\sin^2\varphi + u \right) \mathrm{d}\varphi \end{cases}, \tag{62}$$

where

$$h(z,\varphi) = \left[(1-z) \left(zvy - u^2 \right) + zy + 4\sin^2\varphi \, (zv + u - 1) \right]^{-1}.$$

Let us consider, until further notice, only such z that $\Re z = 0$ and $\Im z > \zeta$, for some $\zeta > 0$. Let us solve system (62) for m. Adding two times the last equation to the first one, and subtracting the second equation we obtain

$$0 = \frac{1}{2\pi c} \int_0^{2\pi} h(z,\varphi) (y + 2u) \, \mathrm{d}\varphi. \tag{63}$$

Note that $\int_0^{2\pi} h(z,\varphi) \, \mathrm{d}\varphi \neq 0$. Otherwise, from the second equation of (62), we have $m = 0$, which cannot be true. Indeed, for any $0 \leq \lambda \leq 1$ and z with $\Re z = 0$,

$$\Im \left(\frac{1}{\lambda - z} \right) = \frac{\Im z}{\lambda^2 + (\Im z)^2} \geq \frac{\Im z}{1 + (\Im z)^2}.$$

Therefore, $\Im \hat{m} \geq \Im z / \left(1 + (\Im z)^2 \right)$, and $\hat{m}$ cannot converge to $m = 0$.

Since $\int_0^{2\pi} h(z,\varphi)\,\mathrm{d}\varphi \neq 0$, (63) yields

$$y + 2u = 0 \tag{64}$$

with $y \neq 0$ and $u \neq 0$ (if one of them equals zero, the other equals zero too, and $m = 0$ by the second equation of (62), which is impossible). Since $u \neq 0$, the last equation implies that $v \neq 0$ as well.

Further, subtracting from the third equation the sum of z times the second and u/v times the last equation, and using (64), we obtain

$$1 = \frac{1}{2\pi c}\int_0^{2\pi} h(z,\varphi)\,\frac{u}{v}\,(2zv + u)(zv - v - 1)\,\mathrm{d}\varphi. \tag{65}$$

This equation, together with the second equation of (62) yield

$$m = \frac{v(2zv + u - 2)}{(1 + v - zv)(2zv + u)}. \tag{66}$$

Next, for the integrand in the last equation of (62), we have

$$\begin{aligned} h(z,\varphi)\left(2v\sin^2\varphi + u\right) &= \frac{1}{2}\frac{v}{zv + u - 1} \\ +h(z,\varphi)\,\frac{u}{2}&\left(\frac{(1-z)v(2zv+u) + 2(2zv+u-1)}{zv+u-1}\right). \end{aligned} \tag{67}$$

This assumes that

$$zv + u - 1 \neq 0. \tag{68}$$

If not, then

$$h(z,\varphi) = \left[(1-z)\left(zvy - u^2\right) + zy\right]^{-1}$$

would not depend on φ and the last equation of (62) would imply that $u + v = 0$. The latter equation and the equality $zv + u - 1 = 0$ would yield $v = -(1-z)^{-1}$, which when combined with the second equation of (62) would give us $m = -c^{-1}(1-z)^{-1}$. This cannot be true because m, being a limit of $\hat{m}$, must satisfy $\Im m \geq 0$ for $\Im z > 0$.

Equations (65), (67), and the last equation of (62) imply that

$$u = \frac{2c}{2c - 1 - (1-z)v(1-c)} - 2zv. \tag{69}$$

Combining this with (66) yields

$$m = v\frac{1-c}{c}. \tag{70}$$

Finally, elementary calculations given in the SM show that

$$\left(\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{x+2\sin^2\varphi}\mathrm{d}\varphi\right)^2 = \frac{1}{x\left(x+2\right)}, \tag{71}$$

where $x \in \mathbb{C}\backslash\left[-2,0\right]$. Using (71), (65), and the definition of $h\left(z,\varphi\right)$, we obtain the following relationship

$$\begin{aligned} &\left(\frac{2cv\left(zv+u-1\right)}{u\left(2zv+u\right)\left(zv-v-1\right)}\right)^2 \\ = &\frac{4\left(zv+u-1\right)^2}{u\left(\left(1-z\right)\left(-2zv-u\right)-2z\right)\left(-u+uz+2\right)\left(u+2vz-2\right)}, \end{aligned} \tag{72}$$

that holds as long as

$$\frac{u\left(\left(1-z\right)\left(-2zv-u\right)-2z\right)}{2\left(zv+u-1\right)} \in \mathbb{C}\backslash\left[-2,0\right].$$

The latter inclusion holds because otherwise $h\left(z,\varphi\right)$ is not a bounded function of φ, which would contradict Lemma 14.

Using (69) in (72), and simplifying, we find that there exist only three possibilities. Either

$$v = -\frac{1}{1-z}, \tag{73}$$

or

$$1-\left(c+cz-1\right)v+z\left(1-z\right)\left(1-c\right)v^2 = 0, \tag{74}$$

or

$$\frac{c}{1-c}-\left(c+cz-z\right)v+z\left(1-z\right)\left(1-c\right)v^2 = 0. \tag{75}$$

Equation (73) cannot hold because otherwise, (70) would imply that $\Im m < 0$, which is impossible as argued above. Equation (74) taken together with (69) implies that

$$u+zv-1 = 0,$$

which was ruled out above. This leaves us with (75), so that, using (70), we get

$$m = \frac{-\left(z - c - cz\right) \pm \sqrt{\left(z - c - cz\right)^2 - 4c\left(1 - z\right)z}}{2z\left(1 - z\right)c}. \tag{76}$$

For $z \in \mathbb{C}^+$ with $\Re z = 0$, the imaginary part of the right hand side of (76) is negative when '$-$' is used in front of the square root. Here we choose the branch of the square root, with the cut along the positive real semi-axis, which has positive imaginary part. Since $\Im m$ cannot be negative, we conclude that

$$m = \frac{-\left(z - c - cz\right) + \sqrt{\left(z - c - cz\right)^2 - 4c\left(1 - z\right)z}}{2z\left(1 - z\right)c}. \tag{77}$$

But the right hand side of the above equality is the value of the limit of the Stieltjes transforms of the eigenvalues of the multivariate beta matrix $B_p\left(p, \left(T - p\right)/2\right)$ as $p, T \to_c \infty$. This can be verified directly by using the formula for such a limit, given for example in Theorem 1.6 of Bai, Hu, Pan and Zhou (2015). As follows from Wachter (1980), the weak limit of the empirical distribution of the eigenvalues of the multivariate beta matrix $B_p\left(p, \left(T - p\right)/2\right)$ as $p, T \to_c \infty$ equals $W\left(\lambda; c/\left(1 + c\right), 2c/\left(1 + c\right)\right)$.

Equation (77) shows that, for z with $\Re z = 0$ and $\Im z > \zeta$, any converging subsequence of $\hat{m}$ converges to the same limit. Hence, $\hat{m}$ a.s. converges for all z with $\Re z = 0$ and $\Im z > \zeta$. Note that $\hat{m}$ is a sequence of bounded analytic functions in the domain $\{z : \Im z > \delta\}$, where δ is an arbitrary positive number. Therefore, by Vitaly's convergence theorem (see Titchmarsh (1939), p.168) $\hat{m}$ a.s. converges to m, described by (77), for any $z \in \mathbb{C}^+$. The a.s. convergence of $\hat{F}_{p,T}\left(\lambda\right)$ (and thus, also of $F_{p,T}\left(\lambda\right)$) to the Wachter distribution follows from the Continuity Theorem for the Stieltjes transforms (see, for example, Corollary 1 in Geronimo and Hill (2003)).

References

[1] Bai, Z.D., J. Hu, G. Pan, and W. Zhou (2015) "Convergence of the Empirical Spectral Distribution Function of Beta Matrices," *Bernoulli* 21, 1538-1574.

[2] Bai, J. and S. Ng (2004) "A PANIC Attack on Unit Roots and Cointegration," *Econometrica* 72, 1127-1177.

[3] Bai, Z. D. and J. W. Silverstein, (2010) *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer Verlag, New York.

[4] Banerjee, A., Marcellino, M., and Osbat, C. (2004) "Some cautions on the use of panel methods for integrated series of macroeconomic data." *The Econometrics Journal* 7 (2), 322–340.

[5] Billingsley, P. (1995) *Probability and Measure,* 3rd Ed., John Wiley & Sons.

[6] Breitung, J., and M.H. Pesaran (2008) "Unit Roots and Cointegration in Panels," in the Econometrics of Panel Data, edited by L. Matyas and P. Sevestre, 279-322. Berlin: Springer.

[7] Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2012) "Bootstrap Determination of the Co-integration Rank in Vector Autoregressive Models," *Econometrica* 80, 1721-1740.

[8] Choi, I. (2015) Panel Cointegration. In Baltagi B. H. edt. The Oxford Handbook of Panel Data. Oxford University Press.

[9] Davis, G.C. (2003) "The generalized composite commodity theorem: Stronger support in the presence of data limitations," *Review of Economics and Statistics* 85, 476-480.

[10] Engel, C., N.C. Mark, and K.D. West (2015) "Factor Model Forecasts of Exchange Rates," *Econometric Reviews* 34, 32-55.

[11] Geronimo, J.S. and T.P. Hill (2003) "Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform," *Journal of Approximation Theory* 121, 54-60.

[12] Golub, G.H. and C. F. Van Loan (1996) *Matrix Computations*, The John Hopkins University Press.

[13] Gonzalo, J. and J-Y Pitarakis (1995) "Comovements in Large Systems," Working Paper 95-38, Statistics and Econometrics Series 10, Universidad Carlos III de Madrid.

[14] Gonzalo, J. and J-Y Pitarakis (1999) "Dimensionality Effect in Cointegration Analysis," Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive WJ Granger, Oxford University Press, Oxford, 212-229

[15] Groen, J.J. and F.R. Kleibergen (2003) "Likelihood-based cointegration analysis in panels of vector error correction models," *Journal of Business and Economic Statistics* 13, 27-36.

[16] Ho, M.S., and B.E. Sorensen (1996) "Finding Cointegration Rank in High Dimensional Systems Using the Johansen Test: An Illustration Using Data Based Mote Carlo Simulations," *Review of Economics and Statistics* 78, 726-732.

[17] Jensen, J.L. and A.T.A. Wood (1997) "On the non-existence of a Bartlett correction for unit root tests," *Statistics and Probability Letters* 35, 181-187.

[18] Johansen, S. (1988) "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control* 12, 231-254

[19] Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica* 59, 1551-1580.

[20] Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

[21] Johansen, S. (2002) "A small sample correction for the test of cointegrating rank in the vector autoregressive model," *Econometrica* 70, 1929-1961.

[22] Johansen, S., H. Hansen, and S. Fachin (2005) "A simulation study of some functionals of random walk," manuscript available at http://www.math.ku.dk/~sjo/.

[23] Johnstone, I.M. (2001) "On the distribution of the largest eigenvalue in principal components analysis," *Annals of Statistics* 29, 295–327.

[24] Larsson, R. (1998) "Bartlett Corrections for Unit Root Test Statistics," *Journal of Time Series Analysis* 19, 426-238.

[25] Larsson, R. and J. Lyhagen (2007). "Inference in Panel Cointegration Models with Long Panels," *Journal of Business & Economic Statistics* 25, 473-483.

[26] Larsson, R., J. Lyhagen, and M. Lothgren (2001) "Likelihood-based Cointegration Tests in Heterogeneous Panels," *Econometric Journal* 4, 109-142.

[27] Lewbel, A. (1996) "Aggregation without separability: a generalized composite commodity theorem," *American Economic Review* 86, 524-543.

[28] 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.

[29] Marchenko, V. A. and Pastur, L. A. (1967) "Distribution for some sets of random matrices," *Math. USSR-Sbornik* 1, 457–483.

[30] Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*, John Wiley & Sons.

[31] Nielsen, B. (1997). "Bartlett correction of the unit root test in autoregressive models," *Biometrika* 84 , 500–504.

[32] Paul, D. and A. Aue (2014) "Random matrix theory in statistics: A review," *Journal of Statistical Planning and Inference* 150, 1-29.

[33] Titchmarsh, E. C. (1939). *The Theory of Functions*, second edition. Oxford University Press, London.

[34] Wachter, K. W. (1976). "Probability Plotting of Multiple Discriminant Ratios," *Proceedings of the Social Statistics Section of the American Statistical Association, Part II* 830-833. Prindle, Weber and Schmidt, Boston.

[35] Wachter, K. W. (1980). "The limiting empirical measure of multiple discriminant ratios," *Annals of Statistics* 8, 937–957.

[36] Yang, Y. and G. Pan (2012) "The convergence of the empirical distribution of canonical correlation coefficients," *Electronic Journal of Probability* 64, 1-13.