Extreme canonical correlations and high-dimensional cointegration analysis

Alexei Onatski[a] and Chen Wang[b]

[a]Faculty of Economics, University of Cambridge, Sidgwick Ave., Cambridge, CB3 9DD, United Kingdom. E-mail: ao319@cam.ac.uk

[b]Department of Statistics and Actuarial Science, Run Run Shaw Building, University of Hong Kong, Pokfulam Road, Hong Kong. E-mail: stacw@hku.hk

January 24, 2019

Abstract

We prove that the extreme squared sample canonical correlations between a random walk and its own innovations almost surely converge to the upper and lower boundaries of the support of the Wachter distribution when the sample size and the dimensionality go to infinity proportionally. This result is used to derive previously unknown analytic expressions for the Bartlett-type correction coefficients for Johansen's trace and maximum eigenvalue tests in a high-dimensional VAR(1). An analysis of cointegration among a large number of log exchange rates illustrates the usefulness of our theoretical results.

KEY WORDS: High-dimensional cointegration, extreme canonical correlations, trace statistic, maximum eigenvalue statistic, Bartlett correction.

GEL CODES: C32, C38.

1 Introduction and the main result

Analysis of cointegration between a large number of time series is a challenging but useful exercise. Its applications include high-dimensional vector error correction modelling for forecasting purposes (Engel et al. (2015)), inference in nonstationary panel data models (Banerjee et al. (2004), Pedroni et al. (2015)), and verification of the assumptions under which composite commodity price indexes satisfy microeconomic laws of demand (Lewbel (1996), Davis (2003)).

A central role in the likelihood-based cointegration analysis is played by the squared sample canonical correlation coefficients between a simple transformation of the levels and the first differences of the data. This paper and its companion Onatski and Wang (2018) (OW18) study such canonical correlations under the *simultaneous asymptotic* regime, where the dimensionality of the data goes to infinity proportionally to the sample size.

OW18 shows that the empirical distribution of the squared sample canonical correlations weakly converges to the so-called *Wachter distribution*. It uses this result to explain the severe over-rejection of the no cointegration hypothesis when the dimensionality of the data is relatively large. In this paper, we show that the extreme squared sample canonical correlations almost surely (a.s.) converge to the upper and lower boundaries of the support of the Wachter distribution.

Our finding yields strong laws of large numbers for the averages of functions of the squared sample canonical correlations that may be discontinuous or unbounded outside an open interval containing the support of the Wachter distribution. In particular, we establish the a.s. limit of the scaled Johansen's (1988, 1991) trace statistic, which has a logarithmic singularity at unity. We use this limit and the limit of the largest squared sample canonical correlation to derive previously unknown explicit expressions for the Bartlett-type correction coefficients for Johansen's trace and maximum eigenvalue tests.

Results of this paper and of OW18 suggest the following quick method for detecting cointegration in high dimensions that may complement more formal tests. Superimpose the graph of the empirical cumulative distribution function (cdf) of the squared sample canonical correlations on the Wachter cdf. If the null hypothesis of no cointegration is correct, a good match is expected in terms of both the supremum distance and the closeness of the extreme values of the two distributions. A poor match signals cointegration. In Section 6, we apply this technique to study cointegration between a large number of exchange rates.

Our setting can be described in the context of the likelihood ratio (LR) test for no cointegration in the model

$$\Delta X_t = \Pi \left(X_{t-1} - t\hat{\rho}_1 \right) + \gamma + \eta_t, \tag{1}$$

where X_t, $t = 1, ..., T+1$, are p-dimensional data, $\Delta X_t = X_t - X_{t-1}$, and $\hat{\rho}_1 = X_{T+1}/\left(T+1\right)$. We will assume that the data are Gaussian and have zero initial value.

Assumption A1 *Random vectors* η_t *with* $t = 1, ..., T+1$ *are i.i.d.* $N\left(0, \Sigma\right)$, *where* Σ *is a* p*-dimensional positive-definite matrix. The initial value* $X_0 = 0$.

Model (1) is similar to Johansen's (1995, eq. 5.14) model H^* :

$$\Delta X_t = \Pi \left(X_{t-1} - t\rho_1 \right) + \gamma + \eta_t, \tag{2}$$

where the deterministic trend is introduced so that there is no quadratic trend in X_t. In (1) ρ_1 is replaced by an estimate $\hat{\rho}_1$. Such a replacement yields the simultaneous diagonalizability of matrices used in the computation of the squared sample canonical correlations, which makes our theoretical analysis possible. We explain this in more detail in Section 4.

As is well known, the LR statistic for testing the null hypothesis that $\Pi = 0$ against $\Pi \neq 0$ equals

$$LR_{\text{trace}} = -\left(T+1\right) \sum_{j=1}^{p} \ln\left(1 - \lambda_{pj}\right), \tag{3}$$

where λ_{pj} is the j-th largest squared sample canonical correlation between demeaned vectors ΔX_t and $X_{t-1} - t\hat{\rho}_1$. When the alternative is restricted so that $\operatorname{rank} \Pi = 1$, the LR test statistic becomes

$$LR_{\max} = -\left(T+1\right) \ln\left(1 - \lambda_{p1}\right). \tag{4}$$

Note that demeaning $X_{t-1} - t\hat{\rho}_1$ and $X_{t-1} - \left(t-1\right)\hat{\rho}_1$ yields the same result. On the other hand, $X_t - t\hat{\rho}_1$ is a p-dimensional random walk detrended so that its last values are tied down to zero. Hence, λ_{pj} can be interpreted as the squared sample canonical correlations between a lagged detrended and demeaned random walk and its demeaned innovations.

In what follows, we will assume that the null hypothesis holds so that the true value of Π is zero. In addition, we will assume that the true value of γ in the data generating process (1) is zero as well.

Assumption A2 *The data generating process (1) has* $\Pi = 0$ *and* $\gamma = 0$.

Consider the simultaneous asymptotic regime where $p, T \to \infty$ so that $p/T \to c_0$. We abbreviate such a regime as $p, T \to_{c_0} \infty$. Without loss of generality, we assume that p is strictly increasing along the sequence, so that T can be viewed as a function of p.

Theorem 1 of OW18 shows that as $p, T \to_{c_0} \infty$ with $c_0 \in (0, 1]$, the empirical distribution of $\lambda_{p1} \geq ... \geq \lambda_{pp}$,

$$F_p(\lambda) \equiv \frac{1}{p} \sum_{i=1}^{p} \mathbf{1}\{\lambda_{pi} \leq \lambda\},$$

a.s. weakly converges[1] to the Wachter distribution W_{c_0} with an atom of size $\max\{0, 2 - 1/c_0\}$ at unity, and density

$$f(\lambda; c_0) = \frac{1 + c_0}{2\pi c_0 \lambda (1 - \lambda)} \sqrt{(b_{0+} - \lambda)(\lambda - b_{0-})} \tag{5}$$

supported on $[b_{0-}, b_{0+}] \subseteq (0, 1]$, where

$$b_{0\pm} = c_0 \left(\sqrt{2} \mp \sqrt{1 - c_0}\right)^{-2}.$$

The main result of this paper strengthens OW18's finding as follows.

Theorem 1 *Suppose that assumptions A1 and A2 hold. Then, for any* $c_0 \in (0, 1/2)$, $\lambda_{p1} \stackrel{\text{a.s.}}{\to} b_{0+}$ *and* $\lambda_{pp} \stackrel{\text{a.s.}}{\to} b_{0-}$ *as* $p, T \to_{c_0} \infty$.

The theorem yields the a.s. convergence of the LR statistics (3) and (4), divided by p^2 and p, respectively. Indeed, it guarantees that no squared sample canonical correlation lies outside any open interval covering $[b_{0-}, b_{0+}]$ for sufficiently large p, almost surely. Since F_p a.s. weakly converges to W_{c_0}, any function $f(\cdot)$ that is continuous and bounded on the open interval covering $[b_{0-}, b_{0+}]$, but may have discontinuities or other singularities outside that interval, satisfies the strong law

[1] OW18 establishes the weak convergence $F_p(\lambda) \Rightarrow W_{c_0}(\lambda)$ both for Gaussian and non-Gaussian η. When η is non-Gaussian and has two finite moments, OW18 establishes the weak convergence in probability. When η is Gaussian, the convergence is a.s.

of large numbers

$$\frac{1}{p}\sum_{j=1}^{p} f\left(\lambda_{pj}\right) \overset{\text{a.s.}}{\to} \int_{-\infty}^{\infty} f\left(\lambda\right) \mathrm{d}W_{c_0}(\lambda)$$

as $p, T \to_{c_0} \infty$. In particular, statistic LR_{trace}/p^2, although defined in terms of an unbounded function $\ln\left(1-\lambda\right)$, a.s. converges to a constant because its singularity lies outside $[b_{0-}, b_{0+}]$ for $c_0 \in (0, 1/2)$. The a.s. convergence of $LR_{\max}/p$ follows from Theorem 1 and the continuity of $\ln\left(1-\lambda\right)$ at $\lambda = b_{0+}$.

Corollary 1 *Suppose that assumptions A1 and A2 hold. Then, for any* $c_0 \in (0, 1/2)$, *as* $p, T \to_{c_0} \infty$,

$$LR_{\text{trace}}/p^2 \overset{\text{a.s.}}{\to} LR_{\text{trace,c}_0} \text{ and } LR_{\max}/p \overset{\text{a.s.}}{\to} LR_{\max,\text{c}_0},$$

where

$$\begin{aligned} LR_{\text{trace,c}_0} &= \frac{1+c_0}{c_0^2}\ln\left(1+c_0\right) - \frac{1-c_0}{c_0^2}\ln\left(1-c_0\right) + \frac{1-2c_0}{c_0^2}\ln\left(1-2c_0\right) \text{ and} \\ LR_{\max,\text{c}_0} &= -\frac{1}{c_0}\ln\left(1 - c_0\left(\sqrt{2}-\sqrt{1-c_0}\right)^{-2}\right). \end{aligned}$$

Proof: Corollary 3 of OW18 shows that the expression on the right hand side of the above definition of $LR_{\text{trace,c}_0}$ equals $-\int_{-\infty}^{\infty}\ln\left(1-\lambda\right)\mathrm{d}W_{c_0}\left(\lambda\right)$ (see pages 34–36 of the Supplementary Material to OW18 for a detailed derivation). Since by Theorem 1, λ_{p1} a.s. remains bounded away from unity, the a.s. weak convergence of F_p to W_{c_0} implies that this integral is the a.s. limit of LR_{trace}/p^2. The a.s. convergence of $LR_{\max}/p$ follows from Theorem 1, the continuity of $\ln\left(1-\lambda\right)$ at $\lambda = b_{0+}$, and the definition of b_{0+}. $\square$

Remark 1 *For* $c_0 = 1/2$, $b_{0+} = 1$ *so that the singularity of* $\ln\left(1-\lambda\right)$ *lies at the upper boundary of the support of* W_{c_0}. *For* $c_0 > 1/2$, λ_{p1} *equals 1 with probability 1. Therefore, in high-dimensional environments where* $p > T/2$, *both* LR_{trace} *and* $LR_{\max}$ *statistics are not well defined.*

In the next section we use Corollary 1 to derive previously unknown explicit expressions for the Bartlett-type correction coefficients for Johansen's trace and maximum eigenvalue tests. In Section 3, we describe the setup and outline the proof of Theorem 1. In Section 4 we discuss reasons for working with model (1) rather than (2), and derive some results for (2). Section 5 contains Monte Carlo analysis that confirms the good quality of our asymptotic results in finite samples and under deviations from the Gaussianity. Section 6 illustrates the real world relevance of our theoretical results with an empirical example. Section 7 discusses directions for future work and concludes. All technical proofs are given in the Supplementary Material (SM).

2 Bartlett-type correction

The standard Johansen's LR tests are based on the asymptotic critical values that assume that p is fixed whereas $T \to \infty$. As is well known, the tests perform poorly in finite samples where p is moderately large. Even relatively small p's, such as five or six, lead to substantial over-rejection of the null hypothesis (see Ho and Sorensen (1996) and Gonzalo and Pitarakis (1995, 1999)).

One of the partial solutions to the over-rejection problem is the Bartlett correction of the LR statistics (see Johansen (2002)). The idea is to scale the statistic so that 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

$$E_{p,T} = E_{p,\infty}\left(1 + a(p)/T + o\left(1/T\right)\right), \tag{6}$$

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)).

Theoretical analysis of the adjustment factor $1 + a(p)/T$ is difficult. The exact expression for $a(p)$ is known only for $p = 1$ (see Larsson (1998)). Therefore, Johansen (2002) proposes to approximate the Bartlett correction factor $BC \equiv E_{p,T}/E_{p,\infty}$ numerically. More precisely, Johansen (section 2.5, 2002) describes a correction factor for the trace statistic as a product of two terms, the first of which needs to be evaluated numerically, while the second one estimates an analytic expression derived in that paper's Corollary 1.

Here, we propose an alternative correction factor, equal to the ratio of the limits of LR_{trace}/p^2 under the simultaneous asymptotics $p, T \to_{c_0} \infty$ and under the sequential asymptotics, where first T and then p goes to infinity. We also propose a similar correction factor for the maximum eigenvalue test. It equals the ratio of the limits of $LR_{\max}/p$ under the simultaneous asymptotics and under the sequential asymptotics.

Monte Carlo analysis in OW18 (see their Figure 6) suggests that the simultaneous asymptotic limit LR_{trace,c_0} derived in Corollary 1 provides a very good concentration point for LR_{trace}/p^2, for moderately large p. Similarly, Table 1 below shows the good finite sample quality of the sequential asymptotic approximation for λ_{p1}, which suggests that $LR_{\max,c_0}$ approximates $LR_{\max}/p$ well for moderately large p. From a theoretical perspective, the good finite sample quality of the simultaneous asymptotic approximations can be explained by the fact that, in contrast to the standard asymptotics, the simultaneous one does not neglect terms $(p/T)^j$ of relatively high order.

The following theorem derives the sequential asymptotic limits of LR_{trace}/p^2 and $LR_{\max}/p$ (see SM for a proof).

Theorem 2 *Suppose that assumptions A1 and A2 hold. Then as first T and then p go to infinity, $LR_{\text{trace}}/p^2 \to 2$ and $LR_{\max}/p \to 3+2\sqrt{2}$, where both convergences are in probability.*

Theorem 2 and Corollary 1 yield the following analytic expressions for the proposed Bartlett-type correction factors

$$\widehat{BC}_{\text{trace}} = \frac{1+c}{2c^2}\ln(1+c) - \frac{1-c}{2c^2}\ln(1-c) + \frac{1-2c}{2c^2}\ln(1-2c)\,, \text{ and} \quad (7)$$

$$\widehat{BC}_{\max} = -\frac{1}{\left(3+2\sqrt{2}\right)c}\ln\left(1-c\left(\sqrt{2}-\sqrt{1-c}\right)^{-2}\right), \quad (8)$$

where $c \equiv p/T$. Expressions (7) and (8) are elementary, and easy to compute and analyze. They do not depend on details of any numerical experiments, and the range of their applicability covers all $c < 1/2$.

We are unaware of the previously obtained BC factors for the maximal eigenvalue test. For the trace test, a numerical approximation to BC is obtained in Johansen (2002). That paper simulates BC for various values of $p \leq 10$ and $T \leq 3000$ and fits a function of the form $\widetilde{BC} = 1 + a_1c + a_2c^2 + a_3c^3 + b/T$ to the obtained results. Its Table I (row $n_d = 1$) reports the following parameter values

coefficient:	a_1	a_2	a_3	b
value:	0.541	0.625	1.077	-1.518

.

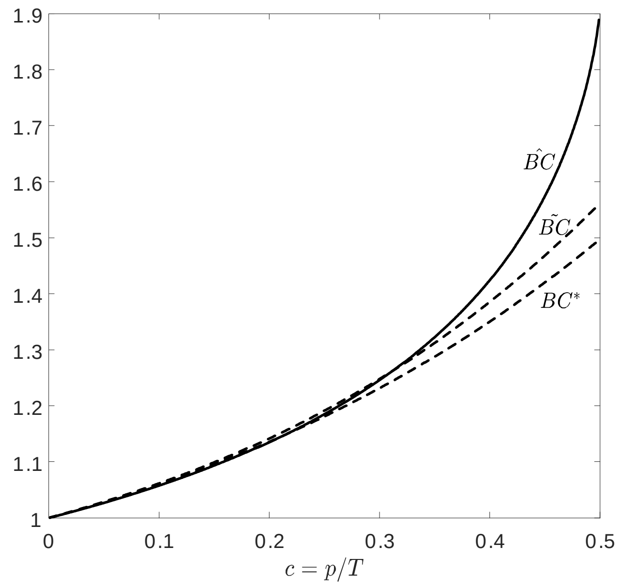

Figure 1: Bartlett correction factors as functions of p/T. Solid line: the factor based on the ratio of the simultaneous and sequential limits of LR_{trace}/p^2. Upper dashed line: Johansen's (2002) approximation. Lower dashed line: Johansen et al.'s (2005) approximation.

Johansen et al. (2005) reports a "programming error in the original simulations" and fits function $BC^* = \exp\left(a_1 c + a_2 c^2 + (a_3 c^2 + b)/T\right)$ to the new simulations. Its Table 3 (row $n_d = 1$) reports the following parameter values

coefficient:	a_1	a_2	a_3	b
value:	0.523	0.569	15.712	-0.238

.

For relatively large values of T, the terms b/T and $\left(a_3 c^2 + b\right)/T$ in the above expressions for $\widetilde{BC}$ and BC^* are small. When they are ignored, the fitted functions become particularly simple:

$$\begin{aligned}\widetilde{BC} &= 1 + 0.541c + 0.625c^2 + 1.077c^3, \\ BC^* &= \exp\left(0.523c + 0.569c^2\right).\end{aligned}$$

Figure 1 shows the graphs of $\widehat{BC}_{\text{trace}}$, $\widetilde{BC}$, and BC^* as functions of c. For $c \leq 0.3$, there is a good fit between the three curves. For $c > 0.3$, the quality of the fit quickly deteriorates. This can be partially explained by the fact that all (p, T)-pairs used in Johansen's (2002) and Johansen et al.'s (2005) simulations are such that $c < 0.3$, so the corresponding numerical approximations do not cover cases with $c > 0.3$.

The terms b/T and $(a_3 c^2 + b)/T$, ignored above, may substantially influence the values of the correction factors $\widetilde{BC}$ and BC^* for relatively small T. We compare the performance of $\widehat{BC}_{\text{trace}}$, $\widetilde{BC}$, and BC^* when all terms in the latter two correction factors are taken into account using Monte Carlo simulations in Section 5.

3 Setup and proof of Theorem 1

In this section, we introduce the setup and give an outline of the proof of Theorem 1. Details of the proof can be found in SM. Let ΔX, X_{-1} and η be $p \times (T+1)$ matrices with columns ΔX_t, $X_{t-1} - t\hat{\rho}_1$, and η_t, respectively. Further, let l be a $(T+1)$-vector of ones, $M_l = I_{T+1} - ll'/(T+1)$ be the projection on the space orthogonal to l, and let U be the $(T+1) \times (T+1)$ upper triangular matrix with ones above the main diagonal and zeros on the diagonal. Then under the null hypothesis

$$\Delta X M_l = \eta M_l \text{ and } X_{-1} M_l = \eta M_l U M_l, \tag{9}$$

where the second equality is derived as follows. Let $\tau = (1, 2, ..., T+1)'$. Note that $\tau' = l'U + l'$ and $\hat{\rho}_1 = \gamma + \eta l/(T+1)$. Therefore,

$$X_{-1} M_l = (\eta U - \hat{\rho}_1 \tau') M_l = \eta U M_l - \frac{1}{T+1} \eta l l' U M_l = \eta M_l U M_l.$$

Equations (9) imply that the squared sample canonical correlations λ_{pj}, $j = 1, ..., p$, between demeaned ΔX_t and demeaned $X_{t-1} - t\hat{\rho}_1$ can be interpreted as the eigenvalues of the product $P_1 P_2$, where P_1 and P_2 are projections on the column spaces of $M_l U' M_l \eta'$ and $M_l \eta'$, respectively. Clearly, λ_{pj}'s are invariant with respect to right-multiplication of η' by any invertible matrix. Hence, without loss of generality, we will assume that η_t are i.i.d. $N(0, I_p)$ vectors.

An equivalent interpretation of λ_{pj}, $j = 1, ..., p$, views them as the eigenvalues of matrix $S_{01} S_{11}^{-1} S_{10} S_{00}^{-1}$, where $S_{10} = S_{01}'$ and

$$S_{01} = \eta M_l U' M_l \eta', \ S_{11} = \eta M_l U M_l U' M_l \eta', \ S_{00} = \eta M_l \eta'. \tag{10}$$

As shown in Lemma 3 of the Supplementary Material to OW18 (OWSM18), $M_l U' M_l$, $M_l U M_l U' M_l$ and M_l are circulant matrices, that is, their (i_1, j_1)-th and (i_2, j_2)-th elements are equal as long as $i_1 - j_1$ equals $i_2 - j_2$ modulo $T+1$.

As is well known (e.g. Golub and Van Loan (1996), ch. 4.7.7), circulant matrices are simultaneously diagonalizable. Precisely, if V is a $(T+1)\times(T+1)$ circulant matrix with the first column v, then $V=\mathcal{F}^{*}\operatorname{diag}(\mathcal{F}v)\,\mathcal{F}/(T+1)$, where $\mathcal{F}$ is the Discrete Fourier Transform matrix with elements

$$\mathcal{F}_{st}=\exp\left\{-\mathrm{i}2\pi\left(s-1\right)\left(t-1\right)/\left(T+1\right)\right\},$$

and the superscript '$*$' denotes transposition and complex conjugation. Here and in what follows, 'i' denotes the imaginary unit $\sqrt{-1}$. This yields the following lemma.

Lemma 1 *Let $\omega_s=2\pi s/(T+1)$ and*

$$\hat{\nabla}=\operatorname{diag}\left\{\left(e^{\mathrm{i}\omega_1}-1\right)^{-1},...,\left(e^{\mathrm{i}\omega_T}-1\right)^{-1}\right\}. \tag{11}$$

Further, let $\hat{\eta}=\eta\mathcal{F}^{*}$ be a $p\times(T+1)$ matrix whose rows are the discrete Fourier transforms at frequencies $0,\omega_1,...,\omega_T$ of the rows of η, and let $\hat{\eta}_{-0}$ be the $p\times T$ matrix obtained from $\hat{\eta}$ by removing its first column, corresponding to zero frequency. Then

$$\begin{aligned}S_{01} &= \hat{\eta}_{-0}\hat{\nabla}\hat{\eta}_{-0}^{*}/\left(T+1\right), S_{11}=\hat{\eta}_{-0}\hat{\nabla}^{*}\hat{\nabla}\hat{\eta}_{-0}^{*}/\left(T+1\right),\\ S_{10} &= \hat{\eta}_{-0}\hat{\nabla}^{*}\hat{\eta}_{-0}^{*}/\left(T+1\right),\ \textit{and } S_{00}=\hat{\eta}_{-0}\hat{\eta}_{-0}^{*}/\left(T+1\right).\end{aligned}$$

The diagonal of $\hat{\nabla}$ consists of the reciprocals of the values of the transfer function (see e.g. Brillinger (1981) eq. 2.7.9) of the "leaded" first-difference filter

$$X_t\mapsto\Delta X_{t+1}\equiv X_{t+1}-X_t \tag{12}$$

at frequencies ω_s, $s\neq 0$. Hence λ_{pj} can also be viewed as the sample squared canonical correlations between discrete Fourier transforms of $\hat{\eta}_t$ and their products with the inverse of the transfer function of filter (12).

Below we work with real-valued sin and cos Fourier transforms of η. In addition, we interchange the order of the frequencies so that ω_{s_1} and ω_{s_2} with $s_1+s_2=T+1$ become adjacent pairs. Specifically, let T be even (the case of odd T can be analyzed similarly) and let $P=\{p_{st}\}$ be a $T\times T$ permutation matrix with elements

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

Define a unitary matrix

$$W = I_{T/2} \otimes \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ \mathrm{i}/\sqrt{2} & -\mathrm{i}/\sqrt{2} \end{pmatrix},$$

where $\otimes$ denotes the Kronecker product.

Note that $WP'\hat{\nabla}^* PW^* = \nabla \equiv \operatorname{diag}\{\nabla_1, ..., \nabla_{T/2}\}$ with

$$\nabla_j = -\frac{1}{2}\begin{pmatrix} 1 & -\cot(\omega_j/2) \\ \cot(\omega_j/2) & 1 \end{pmatrix}.$$

A direct calculation shows that $\nabla\nabla' = \nabla'\nabla$ is a diagonal matrix

$$\nabla\nabla' = \nabla'\nabla = \operatorname{diag}\left\{r_1^{-1} I_2, ..., r_{T/2}^{-1} I_2\right\} \text{ with } r_j = 4\sin^2(\omega_j/2).$$

Defining $\varepsilon = \hat{\eta}_{-0} PW^*/\sqrt{T(T+1)}$ and using Lemma 1, we obtain the following lemma.

Lemma 2 *The columns of* ε *are i.i.d.* $N(0, I_p/T)$ *vectors. Matrix* $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ *equals* $CD^{-1}C'A^{-1}$ *where*

$$C = \varepsilon\nabla'\varepsilon', D = \varepsilon\nabla\nabla'\varepsilon', \text{ and } A = \varepsilon\varepsilon'.$$

This lemma yields yet another interpretation of λ_{pj}, $j = 1, ..., p$. They can be thought of as the eigenvalues of matrix

$$CD^{-1}C'A^{-1} \equiv (\varepsilon\nabla'\varepsilon')(\varepsilon\nabla\nabla'\varepsilon')^{-1}(\varepsilon\nabla\varepsilon')(\varepsilon\varepsilon')^{-1}.$$

The convenience of this interpretation stems from the block-diagonality of ∇ and the diagonality of $\nabla\nabla'$.

Let $\varepsilon_{(j)}$ be a $p \times 2$ matrix that consists of the $(2j-1)$-th and the $2j$-th columns of ε. In particular, $\varepsilon = \left[\varepsilon_{(1)}, ..., \varepsilon_{(T/2)}\right]$. The key advantage of studying C, D, A as opposed to S_{01}, S_{11}, and S_{00} is that C, D, A can be represented as sums of *independent* components of rank two. Specifically,

$$C = \sum_{j=1}^{T/2} \varepsilon_{(j)}\nabla_j'\varepsilon_{(j)}', D = \sum_{j=1}^{T/2} r_j^{-1}\varepsilon_{(j)}\varepsilon_{(j)}', \text{ and } A = \sum_{j=1}^{T/2} \varepsilon_{(j)}\varepsilon_{(j)}'.$$

OW18 exploits these representations to derive the limit of the empirical distribution F_p of the eigenvalues of $CD^{-1}C'A^{-1}$ (the detailed derivation starts on page 16 of OWSM18). That paper proves the convergence of F_p to W_{c_0} by establishing the convergence of the Stieltjes transform of F_p, defined as

$$m_p(z) \equiv \int_{-\infty}^{\infty} (\lambda - z)^{-1} \mathrm{d}F_p(\lambda) = \mathrm{tr}\left(CD^{-1}C'A^{-1} - zI_p\right)^{-1}/p$$

for any z from the upper half of the complex plane $\mathbb{C}^+$.

Our proof of Theorem 1 relies on some of the results of OW18. Therefore, to complete the setup of the analysis below, we now briefly outline the relevant findings of that paper.

First, OW18 derive the following identities (see eq. (38-41) in OWSM18)

$$m_p(z) = \frac{T}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{T/2}\frac{1}{(1-z)^2}\mathrm{tr}\left(\left[I_2, r_j\nabla_j'\right]\Omega_j^{(q)}\left[I_2, r_j\nabla_j'\right]'\right), \quad (13)$$

$$\frac{T}{p} + zm_p(z) = \frac{T}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{T/2}\frac{1}{(1-z)^2}\mathrm{tr}\left(\left[I_2, r_jz\nabla_j'\right]\Omega_j^{(q)}\left[I_2, zr_j\nabla_j'\right]'\right) \quad (14)$$

$$1 + zm_p(z) = \frac{T}{p}\frac{1}{1-z} - \frac{1}{p}\sum_{j=1}^{T/2}\frac{1}{(1-z)^2}\mathrm{tr}\left(\left[I_2, r_jz\nabla_j'\right]\Omega_j^{(q)}\left[I_2, r_j\nabla_j'\right]'\right), \quad (15)$$

$$0 = \frac{1}{p}\sum_{j=1}^{T/2}\frac{1}{1-z}\mathrm{tr}\left(\left[0, I_2\right]\Omega_j^{(q)}\left[I_2, r_j\nabla_j'\right]'\right), \quad (16)$$

where

$$\Omega_j^{(q)} \equiv \Omega_{pj}^{(q)}(z) = \begin{pmatrix} \frac{1}{1-z}I_2 + v_j^{(q)}(z) & \frac{r_j}{1-z}\nabla_j' + u_j^{(q)\prime}(z) \\ \frac{r_j}{1-z}\nabla_j + u_j^{(q)}(z) & \frac{r_jz}{1-z}I_2 + z\tilde{v}_j^{(q)}(z) \end{pmatrix}^{-1}. \quad (17)$$

The 2×2 matrices $v_j^{(q)} \equiv v_j^{(q)}(z)$, $u_j^{(q)} \equiv u_j^{(q)}(z)$, and $\tilde{v}_j^{(q)} \equiv \tilde{v}_j^{(q)}(z)$ are defined as follows. Let

$$\begin{aligned} A_j &= A - \varepsilon_{(j)}\varepsilon_{(j)}', C_j = C - \varepsilon_{(j)}\nabla_j'\varepsilon_{(j)}',\ D_j = D - r_j^{-1}\varepsilon_{(j)}\varepsilon_{(j)}', \\ M_j &= C_jD_j^{-1}C_j' - zA_j,\ \text{and}\ \tilde{M}_j = C_j'A_j^{-1}C_j - zD_j. \end{aligned}$$

Then,

$$v_j^{(q)} = \varepsilon_{(j)}'M_j^{-1}\varepsilon_{(j)},\ u_j^{(q)} = \varepsilon_{(j)}'D_j^{-1}C_j'M_j^{-1}\varepsilon_{(j)},\ \text{and}\ \tilde{v}_j^{(q)} = \varepsilon_{(j)}'\tilde{M}_j^{-1}\varepsilon_{(j)}.$$

The entries of these matrices are quadratic forms in the columns of $\varepsilon_{(j)}$. In what follows, we use superscript '(q)' to denote matrices that involve quadratic forms in the columns of $\varepsilon_{(j)}$ to distinguish them from similarly defined matrices that do not involve such quadratic forms.

The next step in OW18 (see Section 2.1.5 of OWSM18) is to approximate $\Omega_j^{(q)}$ by matrix Ω_j, which is obtained from $\Omega_j^{(q)}$ by replacing $v_j^{(q)}(z)$, $u_j^{(q)}(z)$, and $\tilde{v}_j^{(q)}(z)$ in (17) with $v_p(z)I_2$, $u_p(z)I_2$, and $\tilde{v}_p(z)I_2$, respectively, where

$$v_p(z) = \operatorname{tr}\left(M^{-1}\right)/T,\ u_p(z) = \operatorname{tr}\left(D^{-1}C'M^{-1}\right)/T,\ \text{and } \tilde{v}_p(z) = \operatorname{tr}\left(\tilde{M}^{-1}\right)/T.$$

Here $M = CD^{-1}C' - zA$ and $\tilde{M} = C'A^{-1}C - zD$. To simplify notation, we will suppress the dependence of $v_p(z), u_p(z)$, and $\tilde{v}_p(z)$ on p and z. It is straightforward to verify that matrix Ω_j has the following explicit form

$$\Omega_j = \frac{1-z}{\delta_j}\begin{pmatrix} \frac{z}{1-z}r_jI_2 + z\tilde{v}I_2 & -\frac{1}{1-z}r_j\nabla_j' - uI_2 \\ -\frac{1}{1-z}r_j\nabla_j - uI_2 & \frac{1}{1-z}I_2 + vI_2 \end{pmatrix}, \tag{18}$$

where

$$\delta_j = z\tilde{v}\left(1 + v - zv\right) + r_j\left(u + zv - 1\right) - (1-z)\,u^2. \tag{19}$$

Taking traces in equations (13-16), after approximating $\Omega_j^{(q)}$ by Ω_j, yields equations

$$m_p(z) = \frac{1}{c\,(1-z)} - \frac{2}{cT}\sum_{j=1}^{T/2}\frac{z\tilde{v} + r_j\left(u + v - 1\right)}{(1-z)\,\delta_j} + e_1(z), \tag{20}$$

$$\frac{1}{c} + zm_p(z) = \frac{1}{c\,(1-z)} - \frac{2}{cT}\sum_{j=1}^{T/2}\frac{z\tilde{v} + r_jz\left(u + zv - 1\right)}{(1-z)\,\delta_j} + e_2(z), \tag{21}$$

$$1 + zm_p(z) = \frac{1}{c\,(1-z)} - \frac{2}{cT}\sum_{j=1}^{T/2}\frac{z\tilde{v} + r_j\left(u\left(1+z\right)/2 + zv - 1\right)}{(1-z)\,\delta_j} + e_3(z), \tag{22}$$

$$0 = \frac{2}{cT}\sum_{j=1}^{T/2}\frac{-u - r_jv/2}{\delta_j} + e_4(z), \tag{23}$$

where $e_k(z)$, $k = 1, ..., 4$, are the approximation errors

$$e_1(z) = \frac{1}{p}\sum_{j=1}^{T/2} \frac{1}{(1-z)^2} \operatorname{tr}\left(\left[I_2, r_j \nabla_j'\right] \left(\Omega_j - \Omega_j^{(q)}\right) \left[I_2, r_j \nabla_j'\right]' \right), \tag{24}$$

$$e_2(z) = \frac{1}{p}\sum_{j=1}^{T/2} \frac{1}{(1-z)^2} \operatorname{tr}\left(\left[I_2, r_j z \nabla_j'\right] \left(\Omega_j - \Omega_j^{(q)}\right) \left[I_2, z r_j \nabla_j'\right]' \right), \tag{25}$$

$$e_3(z) = \frac{1}{p}\sum_{j=1}^{T/2} \frac{1}{(1-z)^2} \operatorname{tr}\left(\left[I_2, r_j z \nabla_j'\right] \left(\Omega_j - \Omega_j^{(q)}\right) \left[I_2, r_j \nabla_j'\right]' \right), \tag{26}$$

$$e_4(z) = -\frac{1}{p}\sum_{j=1}^{T/2} \frac{1}{1-z} \operatorname{tr}\left([0, I_2] \left(\Omega_j - \Omega_j^{(q)}\right) \left[I_2, r_j \nabla_j'\right]' \right). \tag{27}$$

Finally, Lemma 10 of OWSM18 shows that the errors $e_k(z)$, $k = 1, ..., 4$, converge to zero pointwise over z from a compact subset of $\mathbb{C}^+$. This allows OW18 to argue that $m_p(z)$ converges to $\bar{m}_0(z)$ uniformly over this compact subset, where $\bar{m}_0(z)$ satisfies the "limiting version" of system (20-23) that sets $e_k(z)$ to zeros. Solving the limiting system, OW18 shows (see Section 2.1.6 of OWSM18) that $\bar{m}_0(z)$ is the Stieltjes transform of W_{c_0}, which yields the convergence of F_p to W_{c_0}.

Our proof of Theorem 1 starts from the system (20-23). It amounts to establishing fast convergence of the errors $e_k(z)$, $k = 1, ..., 4$, to zero as z runs over a sequence z_p with $\operatorname{Im} z_p \to 0$ and $\operatorname{Re} z_p$ bounded away from the support of the Wachter distribution W_{c_0}.

The general strategy of our proof is similar to that used in Bai and Silverstein's (1998) (BS98) study of the asymptotic behavior of the extreme eigenvalues of sample covariance matrices. The main ideas are as follows. Consider a sequence $\{z_p\}$ such that (s.t.)

$$x_p \equiv \operatorname{Re} z_p \in [0,1] \qquad \text{and} \qquad y_p \equiv \operatorname{Im} z_p = y_0 p^{-\alpha} \tag{28}$$

with $\alpha \geq 0$ and $y_0 \in (0, 1]$ that are independent from p. We study the behavior of $m_p(z_p)$ as $p, T \to_{c_0} \infty$.

Let $m_0(z)$ be the Stieltjes transform of W_c, where W_c is obtained from the limiting distribution W_{c_0} by replacing c_0 with $c \equiv p/T$. Consider an interval $[a, b]$ outside the supports of W_c and W_{c_0} for all large p. Since F_p consists of masses $1/p$

at λ_{pj}, and since $W_c([a,b]) = 0$, we have the following decomposition

$$\operatorname{Im}\left(m_p(z_p) - m_0(z_p)\right) = \sum_{\lambda_{pj} \in [a,b]} \frac{1}{p} \frac{y_p}{\left(\lambda_{pj} - x_p\right)^2 + y_p^2} + \int_{[a,b]^c} \frac{y_p \mathrm{d}(F_p(\lambda) - W_c(\lambda))}{\left(\lambda - x_p\right)^2 + y_p^2}. \tag{29}$$

The existence of $\lambda_{pj} \in [a,b]$ puts an upper bound on the speed of the a.s. convergence

$$\sup_{x_p \in [a,b]} |m_p(z_p) - m_0(z_p)| \to 0 \tag{30}$$

as $p, T \to_{c_0} \infty$. This bound is linked to the speed of convergence $y_p \to 0_+$ via the first term on the right hand side of (29). Proving that convergence (30) is faster than that bound shows that there are no λ_{pj} in $[a,b]$ for all sufficiently large p.

The analysis of the speed of convergence of (30) is done in several steps.

1. We show that the expected number of eigenvalues in $[a,b]$ cannot grow faster than p^β with $\beta < 1$ as $p \to \infty$.

2. We use 1. to derive an upper bound on the speed of convergence $m_p(z_p) - \mathbb{E}m_p(z_p) \to 0$ of the "stochastic part" of $m_p(z_p) - m_0(z_p)$.

3. We derive an upper bound on the speed of convergence $\mathbb{E}m_p(z_p) - m_0(z_p) \to 0$ of the "deterministic part" of $m_p(z_p) - m_0(z_p)$, and combine the results.

An implementation of these three steps requires a non-trivial extension of BS98. The fact that we have to deal with the product of four dependent stochastic matrices, $CD^{-1}C'A^{-1}$, presents substantial challenges, relative to the case of a sample covariance matrix, that we overcome. The key is to establish fast convergence of the errors $e_k(z_p)$ defined in (24-27) to zero, which requires detailed analysis of matrices Ω_j, $\Omega_j^{(q)}$, and their difference $\Omega_j - \Omega_j^{(q)}$. A detailed proof of Theorem 1 can be found in SM.

4 Johansen's H^* model

If the data generating process is described by Johansen's H^* model (2) rather than (1), the LR statistics for testing the null hypothesis that $\Pi = 0$ still have forms (3) and (4). However now, λ_{pj}'s equal the eigenvalues of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$, where $\tilde{S}_{ij}$ are defined differently from S_{ij} given in (10). Specifically, they correspond to sample covariance and cross-covariance matrices of the demeaned processes ΔX_t and $\left(X'_{t-1}, t\right)'$ (see Johansen (1995, ch. 6.2)). That is, in contrast to (10),

$$\tilde{S}_{01} = \begin{pmatrix} \eta M_l U'\eta', & \eta M_l \tau \end{pmatrix} \text{ and } \tilde{S}_{11} = \begin{pmatrix} \eta U M_l U'\eta' & \eta U M_l \tau \\ \tau' M_l U'\eta' & \tau' M_l \tau \end{pmatrix},$$

while similarly to above, $\tilde{S}_{10} = \tilde{S}'_{01}$ and $\tilde{S}_{00} = \eta M_l \eta'$. Here τ denotes the time trend, $\tau = (1, 2, ..., T+1)'$.

In contrast to matrices S_{01}, S_{11}, and S_{00} given in (10), matrices $\tilde{S}_{01}, \tilde{S}_{11}$, and $\tilde{S}_{00}$ cannot be simultaneously rotated to the form $\varepsilon' W \varepsilon$, where W is a block-diagonal matrix. Therefore, in the case of H^* model, there is no convenient frequency domain reformulation of Johansen's test, and the above analysis will not go through. It is however possible to show that at most one eigenvalue of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$ remains above and separated from b_{0+} and at most one eigenvalue remains below and separated from b_{0-}, asymptotically. Hence, the second largest and smallest eigenvalues of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$ a.s. converge to b_{0+} and b_{0-}.

Recall that the eigenvalues of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ equal those of P_1P_2, where P_1 and P_2 are projections on the column spaces of $Y \equiv M_l U' M_l \eta'$ and $Z \equiv M_l \eta'$, respectively. Similarly, the eigenvalues of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$ equal those of $\tilde{P}_1P_2$, where $\tilde{P}_1$ is the projection on the column space of $\tilde{Y} \equiv \begin{pmatrix} M_l U'\eta', & M_l \tau \end{pmatrix}$.

Note that $\tilde{Y}$ has $p+1$ columns whereas Y has p columns. Let us augment Y by a zero column to obtain $\bar{Y} \equiv \begin{pmatrix} M_l U' M_l \eta', & 0 \end{pmatrix}$. Obviously, projections on the columns of Y and $\bar{Y}$ coincide and equal P_1. Further,

$$\tilde{Y} - \bar{Y} = \begin{pmatrix} M_l U' l l' \eta' / (T+1), & M_l \tau \end{pmatrix} = \begin{pmatrix} M_l \tau l' \eta' / (T+1), & M_l \tau \end{pmatrix}, \quad (31)$$

and matrix $\begin{pmatrix} M_l \tau l' \eta' / (T+1), & M_l \tau \end{pmatrix}$ has rank one.

Lemma 3 *Let Y_1 and Y_2 be $n \times m$ matrices and let P_{Y_1} and P_{Y_2} be projections on the spaces spanned by the columns of Y_1 and Y_2, respectively. If* $\operatorname{rank}(Y_1 - Y_2) = r$, *then there exist $n \times r$ matrices y_1 and y_2 such that $P_{Y_1} - P_{Y_2} = P_{y_1} - P_{y_2}$, where P_{y_1} and P_{y_2} are projections on the spaces spanned by the columns of y_1 and y_2, respectively. In particular,* $\operatorname{rank}(P_{Y_1} - P_{Y_2}) \leq 2r$.

Proof: Assume that $Y_1 - Y_2 = ab$, where a is $n \times r$ and $b = \begin{pmatrix} 0, & I_r \end{pmatrix}$. This assumption does not lead to a loss of generality because P_{Y_1} and P_{Y_2} are invariant with respect to multiplication of Y_1 and Y_2 from the right by arbitrary invertible

$m \times m$ matrices. The above form of b can be achieved by such a multiplication. Let us partition Y_1 and Y_2 as $[Y_{11}, Y_{12}]$ and $[Y_{21}, Y_{22}]$, where Y_{12} and Y_{22} are the last r columns of Y_1 and Y_2, respectively. We have

$$Y_{21} = Y_{11} \text{ and } Y_{22} + a = Y_{12}.$$

Denote $I_m - P_{Y_{21}}$ as M_1, where $P_{Y_{21}}$ is the projection on the space spanned by the columns of Y_{21}, and let $y_2 = M_1 Y_{22}$. Note that

$$P_{Y_2} = P_{[Y_{21}, y_2]} = P_{Y_{21}} + P_{y_2},$$

where the second equality holds because Y_{21} is orthogonal to y_2. Similarly, we have

$$P_{Y_1} = P_{Y_{11}} + P_{y_1} = P_{Y_{21}} + P_{y_1},$$

where $y_1 = M_1 Y_{12}$. Therefore, $P_{Y_1} - P_{Y_2} = P_{y_1} - P_{y_2}$. $\square$

Lemma 3 and equality (31) imply that there exist no more than one eigenvalue of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$ that is larger than the largest eigenvalue of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ and no more than one eigenvalue of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$ that is smaller than the smallest eigenvalue of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$. Indeed, note that the eigenvalues of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$, which equal those of $\tilde{P}_1 P_2$, coincide with the eigenvalues of a symmetric matrix $P_2\tilde{P}_1P_2$. Similarly, the eigenvalues of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ coincide with the eigenvalues of a symmetric matrix $P_2P_1P_2$. By Lemma 3,

$$P_2\tilde{P}_1P_2 - P_2P_1P_2 = P_2P_{y_1}P_2 - P_2P_{y_2}P_2,$$

where P_{y_1} and P_{y_2} are projections on one-dimensional spaces. Hence, our statement concerning the eigenvalues of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$ and $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ follows from Weyl's inequalities for eigenvalues of a sum of symmetric matrices (see e.g. Horn and Johnson (1985, Theorem 4.3.1)).

We have conducted a small-scale Monte Carlo study which suggests that the largest and the smallest eigenvalues of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$ do converge to b_{0+} and b_{0-}, respectively, similarly to the largest and the smallest eigenvalues of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$. A formal analysis of the extreme eigenvalues of $\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10}\tilde{S}_{00}^{-1}$ would amount to studying low-rank perturbations of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$. There exists a large literature on the low rank perturbations of classical random matrix ensembles (see e.g. Capitaine and Donati-Martin (2016) and references therein). However, this literature is not directly applicable to $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$. We leave analysis of small rank perturbations of such a matrix for future research.

5 Monte Carlo

To assess the quality of our asymptotic results in finite samples, we simulate p-dimensional random walks with $p = 2$, 5, 10, 20, 50, and 100. The values of the sample size T are chosen from a wide range, starting from 20 and going to $2,000$. Overall, we consider 20 different (p, T)-pairs. The number of the Monte Carlo (MC) replications is set to $10,000$. The data generating process is (1) with $\Pi = 0$. As explained in Section 3, the canonical correlations are invariant with respect to the choice of the constant γ and the error variance Σ in (1). Therefore, we set $\gamma = 0$ and $\Sigma = I_p$ without loss of generality.

According to our theory, $\lambda_{p1} \overset{a.s.}{\to} b_{0+}$ and $\lambda_{pp} \overset{a.s.}{\to} b_{0-}$, where b_{0+} and b_{0-} are the upper and lower boundaries of the support of the Wachter distribution W_{c_0} with $c_0 = \lim p/T$. Table 1 reports the MC mean, 2.5% quantile, $q_{2.5}$, and 97.5% quantile, $q_{97.5}$, for λ_{p1} and λ_{pp}. They should be compared to the sample counterparts of $b_{0\pm}$, denoted in the table as $b_{\pm}$ and defined as the upper and lower boundaries of the support of the Wachter distribution W_c with $c = p/T$.

The MC mean of the largest squared sample canonical correlation is very close to the theoretical prediction b_+ for most of the (p, T)-pairs. Surprisingly, b_+ is quite close to the MC mean of λ_{p1} even for $p = 2$. However, for such an extremely small value of p, the spread of the MC distribution of λ_{p1} is very large. The spread remains relatively large for $p = 5$. For the larger values of p, the MC distribution of λ_{p1} is reasonably tightly concentrated around b_+. Even so, for all the (p, T)-pairs, b_+ is somewhat above the MC mean and is closer to $q_{97.5}$ than to $q_{2.5}$.

The theoretical prediction b_- for the smallest squared sample canonical correlation λ_{pp} as well as the MC mean of λ_{pp} are generally close to zero. The quality of the match between them is on par with that between the MC mean of λ_{p1} and b_+. However, in contrast to λ_{p1}, the MC mean of λ_{pp} is always somewhat larger than its theoretical prediction. Hence, overall, there is a tendency for the empirical distribution of λ_{pj}, $j = 1, ..., p$, to have all its mass inside the interval $[b_-, b_+]$.

In SM, we verify that for most of the (p, T)-pairs, the supremum distance between the entire empirical cumulative distribution function (cdf) of λ_{pj}, $j = 1, ..., p$ and the cdf of W_c is small for relatively large p. This suggests that, in

		Maximum eigenvalue λ_{p1}				Minimum eigenvalue λ_{pp}			
p	T	b_+	mean	$q_{2.5}$	$q_{97.5}$	b_-	mean	$q_{2.5}$	$q_{97.5}$
2	20	0.461	0.358	0.184	0.593	0.018	0.142	0.057	0.285
2	100	0.111	0.086	0.039	0.163	0.003	0.030	0.011	0.065
5	20	0.832	0.749	0.593	0.889	0.048	0.148	0.067	0.258
5	100	0.259	0.223	0.150	0.318	0.009	0.029	0.013	0.053
10	25	0.978	0.954	0.899	0.989	0.083	0.162	0.082	0.254
10	50	0.740	0.691	0.596	0.791	0.038	0.077	0.038	0.124
10	100	0.461	0.422	0.344	0.514	0.018	0.037	0.018	0.061
20	45	0.993	0.986	0.970	0.997	0.095	0.142	0.087	0.199
20	60	0.933	0.912	0.871	0.950	0.067	0.101	0.062	0.144
20	150	0.571	0.542	0.485	0.604	0.024	0.037	0.023	0.053
20	500	0.212	0.199	0.172	0.232	0.007	0.011	0.006	0.016
50	105	0.999	0.998	0.995	0.999	0.104	0.127	0.094	0.157
50	150	0.933	0.923	0.902	0.943	0.067	0.082	0.061	0.103
50	200	0.832	0.817	0.788	0.848	0.048	0.059	0.044	0.074
50	500	0.461	0.447	0.419	0.480	0.018	0.022	0.016	0.028
50	1000	0.259	0.250	0.232	0.272	0.009	0.011	0.008	0.014
100	300	0.933	0.927	0.914	0.940	0.067	0.076	0.063	0.088
100	400	0.832	0.823	0.805	0.842	0.048	0.054	0.045	0.063
100	1000	0.461	0.453	0.435	0.473	0.018	0.020	0.017	0.024
100	2000	0.259	0.253	0.242	0.266	0.009	0.010	0.008	0.012

Table 1: The MC means, 2.5% quantiles, and 97.5% quantiles of λ_{p1} and λ_{pp}. The upper and lower boundaries of the support of the Wachter distribution $W_{p/T}$ are reported as b_+ and b_-, respectively. The number of the MC replications is 10,000.

practice, the hypothesis of no cointegration in high-dimensional VAR(1) may be quickly assessed by visually comparing the empirical cdf with that of W_c. If the hypothesis is correct, a good match is expected in terms of both the supremum distance and the closeness of the extreme values of the two distributions.

For such a quick assessment to be effective, the two distributions must diverge under alternatives where cointegration does take place. However, if the cointegration rank is small, the divergence can manifest itself only in the differences between the upper extreme values of the distributions. Therefore, our next MC experiment checks the sensitivity of λ_{p1} to the existence of a single cointegrating relationship.

Specifically, we replace one component of the simulated p-dimensional random walk data by an AR(1) series with the autoregressive parameter ρ, for $\rho = 0$, 0.3, 0.5, 0.6, 0.7, and 0.9. Figure 2 reports the MC means of λ_{p1} as functions of ρ for the different (p, T)-pairs. The pairs are as in Table 1. They are not explicitly shown, but can be identified by looking up the value of $c = p/T$ at the right edge of the

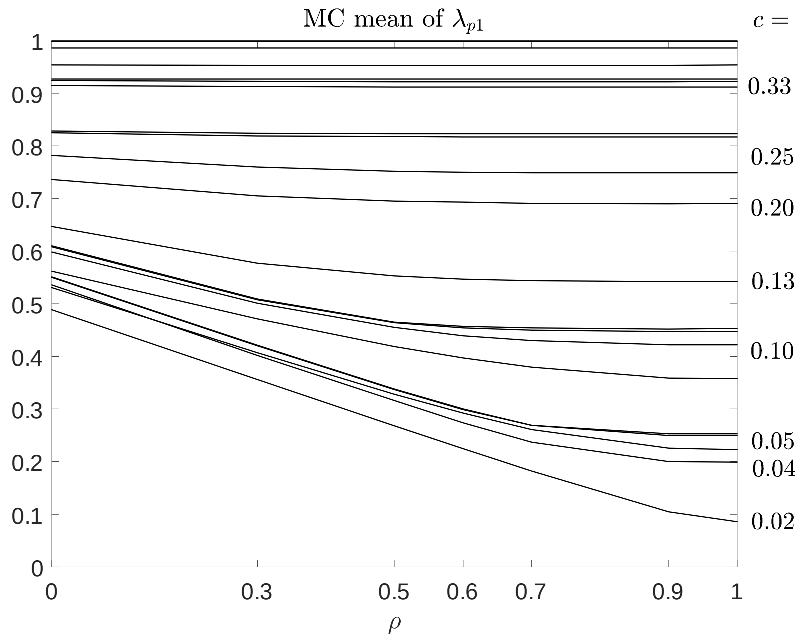

Figure 2: MC mean of λ_{p1} as functions of ρ. The number of MC replications is 10,000. The different graphs correspond to 20 different (p, T)-pairs. The pairs can be partially identified using the value of $c = p/T$ shown at the right edge of the figure.

Figure. For pairs that correspond to the same c, the values of the MC mean for λ_{p1} at $\rho = 1$ (no cointegration) turn out to be smaller for smaller p. For example, there are four (p, T)-pairs and four lines corresponding to $c = 0.1$. The lowest of these lines corresponds to $(p, T) = (2, 20)$, and the highest one corresponds to $(p, T) = (100, 1000)$. We do not label lines corresponding to $c > 0.33$.

In general, the MC mean of λ_{p1} is decreasing as a function of ρ. This accords with intuition that the more persistent the process, the harder to distinguish it from a random walk. However, the MC mean of λ_{p1} is not sensitive to ρ for relatively large values of c. Therefore, for such c, the quick visual inspection of the match between the empirical distribution of λ_{pj} and W_c cannot replace more formal test procedures.

We now verify the accuracy of the proposed Bartlett-type corrections for Johansen's maximum eigenvalue and trace tests. Table 2 reports the actual size of the uncorrected (raw) and Bartlett-corrected tests of the asymptotic size 5%. For the trace test, we report the sizes of the Bartlett-corrected tests that use our correction formula, the formula reported in Table I (row $n_d = 1$) of Johansen (2002) (column J02 of Table 2), and the formula reported in Table 3 (row $n_d = 1$) of

			Max-eval test		Trace test			
p	T	p/T	raw	corrected	raw	corrected	J02	JHF05
2	20	0.10	7.36	5.04	7.31	4.97	8.02	5.18
2	100	0.02	4.73	4.36	4.72	4.43	4.87	4.49
5	20	0.25	27.67	6.51	36.59	7.09	15.40	5.49
5	100	0.05	6.93	5.31	7.42	5.24	6.24	5.32
10	25	0.40	95.78	12.65	99.80	17.71	50.24	14.56
10	50	0.20	29.37	6.26	51.81	6.07	10.48	5.49
10	100	0.10	11.54	5.11	17.84	4.93	6.77	4.89
20	45	0.44	100.0	15.69	100.0	26.67	89.05	59.56
20	60	0.33	98.16	7.89	100.0	9.84	26.25	10.64
20	150	0.13	24.42	4.60	59.42	5.49	6.91	5.01
20	500	0.04	8.18	5.20	12.42	4.83	5.27	4.75
50	105	0.48	100.0	19.90	100.0	41.80	100.0	100.0
50	150	0.33	100.0	7.44	100.0	9.52	28.77	26.95
50	200	0.25	99.74	5.90	100.0	6.37	7.94	6.81
50	500	0.10	32.88	5.03	94.62	4.92	4.54	4.01
50	1000	0.05	12.95	5.18	45.65	4.85	4.63	4.23
100	300	0.33	100.0	7.46	100.0	10.51	38.09	71.30
100	400	0.25	100.0	6.48	100.0	6.96	4.85	10.78
100	1000	0.10	63.10	5.31	100.0	5.62	3.08	3.30
100	2000	0.05	22.51	5.35	93.00	5.23	3.98	4.07

Table 2: The size of the un-corrected (raw) and Bartlett-corrected maximum-eigenvalue (Max-eval) and trace tests. Columns 'J02' and 'JHF05' correspond to the test corrected using the correction factors reported in Table I (case $n_d = 1$) of Johansen (2002) and Table 3 (case $n_d = 1$) of Johansen et al. (2005), respectively. The number of the MC replications is 10,000.

Johansen et al. (2005) (column JHF05 of Table 2).

The data are generated in the same way as for Table 1. The trace and the maximum eigenvalue statistics are computed on the basis of Johansen's H^* model, which is discussed in Section 4. The asymptotic 5% critical values for $p \leq 12$ are available from Table IV (column $k = 0$) of MacKinnon et al. (1999). For $p = 20, 50$, and 100, we compute the asymptotic critical values using the response surface approach of the latter paper.

As is well known, the uncorrected tests severely over-reject in the high-dimensional environment. This is confirmed by the columns of Table 2 which are labeled 'raw'. The Bartlett-corrected tests are much better sized. Their actual size is reasonably close to 5% for all (p, T)-pairs with $p/T < 0.3$. For $p/T \geq 0.4$, even the Bartlett-corrected tests substantially over-reject.

			Max-eval test		Trace test			
p	T	p/T	raw	corrected	raw	corrected	J02	JHF05
2	20	0.10	8.59	6.60	8.69	6.61	9.28	6.74
2	100	0.02	6.09	5.65	5.55	5.25	5.65	5.31
5	20	0.25	29.98	9.34	36.72	8.85	17.13	7.24
5	100	0.05	7.67	6.06	7.95	5.83	6.85	5.87
10	25	0.40	95.92	14.35	99.73	19.42	51.40	16.23
10	50	0.20	32.60	9.16	51.59	8.40	12.68	7.62
10	100	0.10	14.04	7.50	19.17	6.69	8.40	6.61
20	45	0.44	100.0	16.87	100.0	28.52	89.05	60.13
20	60	0.33	98.36	9.64	100.0	12.04	28.77	12.62
20	150	0.13	27.55	7.84	59.47	6.50	8.22	6.08
20	500	0.04	9.66	6.33	13.05	5.38	5.78	5.27
50	105	0.48	100.0	20.47	100.0	44.23	100.0	100.0
50	150	0.33	100.0	8.97	100.0	12.09	31.96	29.93
50	200	0.25	99.73	8.27	100.0	8.21	10.32	8.85
50	500	0.10	34.98	6.75	94.13	6.02	5.59	4.86
50	1000	0.05	14.63	6.32	44.46	5.22	4.86	4.57
100	300	0.33	100.0	8.67	100.0	12.84	40.05	72.41
100	400	0.25	100.0	8.32	100.0	9.38	6.89	13.86
100	1000	0.10	64.72	7.49	100.0	6.52	4.05	4.32
100	2000	0.05	23.80	6.33	92.96	5.63	4.38	4.50

Table 3: Student's $t(3)$ version of Table 2.

Interestingly, such an over-rejection is much less pronounced for the trace test which uses our correction formula instead of that proposed by Johansen (2002). Moreover, the size of 'our' corrected test is closer to 5% than that of the 'J02-corrected' test for all (p,T)-pairs, except $(2,100)$ and $(100,400)$. It is also closer to 5% than that of the 'JHF05-corrected' test for all (p,T)-pairs, except $(2,100)$, $(5,20)$, $(10,25)$, $(10,50)$, and $(20,150)$. A further study of this phenomenon is left for future research.

Our final MC experiment assesses the robustness of the above findings to the deviations from the Gaussianity. We repeat all of the above MC experiments using the same settings, but drawing the i.i.d. components of η (see the data generating equation (1)), first, from the centered chi-squared distribution with one degree of freedom, $\chi^2(1)$, and then, from Student's $t(3)$ distribution. The centered $\chi^2(1)$ has zero mean, but is strongly skewed to the right, whereas Student's $t(3)$ has only two finite moments.

We find that the MC results reported in Table 1 and Figure 2 remain practically unchanged (see SM). However, the results reported in Table 2 change. Table 3 is the analogue of Table 2 for the $t(3)$ case. To save space, we report a similar table for the centered $\chi^2(1)$ in SM. Overall, the 'raw' tests over-reject even more, whereas the sizes of the corrected tests somewhat increase.

6 Empirical analysis

This section illustrates the real world relevance of our theoretical results with an empirical example inspired by Engel et al. (2015) (EMW15). That paper uses a few common factors extracted from a panel of 17 log exchange rates for prediction of the individual series. The factors are assumed to "soak up a common unit root component" in the log exchange rates, implying their cointegration. However, the paper does not attempt to test for cointegration, referring to the poor finite sample performance of the available tests in high dimensions.

There have been many previous studies of the cointegration among a smaller number of exchange rates. Baillie and Bollerslev (1989) find evidence for cointegration among seven exchange rates using daily data. Diebold et al. (1994) argue that including intercept in the VAR model changes the analysis so that there is no evidence for cointegration. More recently, Cheng and Phillips (2012) find no cointegration in a four-dimensional exchange rate system. However, they point out many previous studies that agree on the existence of the fractional cointegration between exchange rates.

Following EMW15, we consider logs of US dollar exchange rates for 17 OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Japan, Italy, Korea, Netherlands, Norway, Spain, Sweden, Switzerland, and the United Kingdom. We use two different datasets. The first one is EMW15's quarterly data from 1973:1 to 1998:4 (just before the introduction of the Euro). For these data, the ratio $c = p/T$ equals 0.16. Hence, the extreme canonical correlation may be relatively insensitive to stationary but persistent alternatives (see Figure 2), and the power of related cointegration tests may be low. Therefore, we also consider weekly data, where c equals 0.018. The weekly data are extracted from Federal Reserve, H10 dataset.[2]

[2] The dataset is daily from 13/04/1981 (the first available day for Korea) to 31/12/1998. We define the weekly series as the daily exchange rate on the last day of the week. Our results with weekly data are similar to those with daily data (not reported). We decided to use weekly rather than daily data because the value of c for the latter falls well outside the range studied in our Monte Carlo section.

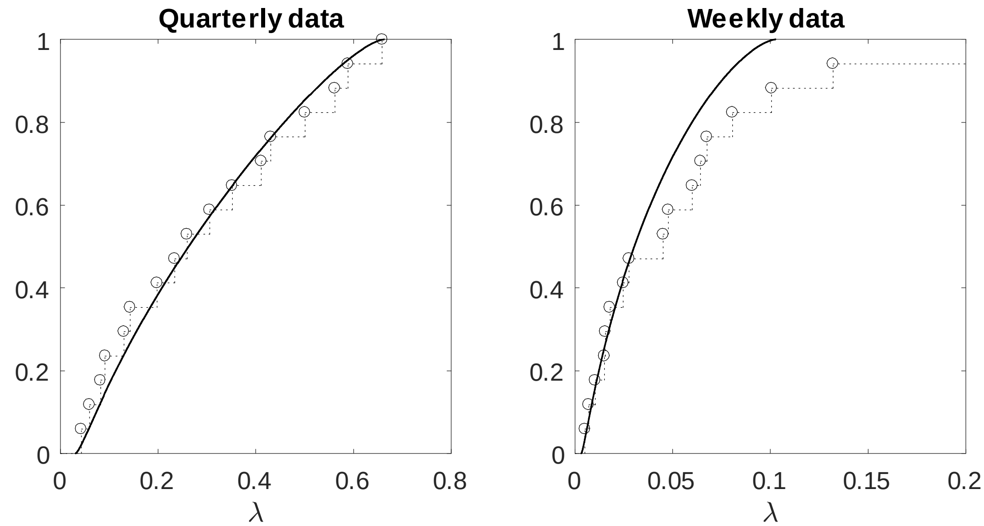

Figure 3: Cdf of W_c (solid line) superimposed with the cdf of the empirical distribution of λ_j, $j = 1, ..., 17$. The coordinates of the circle markers are: $x = \lambda_j, y = j/17$.

Let s_t be the 17-dimensional vector of log exchange rates at time t. We interpret the first observation as $t = 0$, and construct $X_t = s_t - s_0$ so that the initial value of X_t is zero. Next, we compute the squared sample canonical correlations λ_j, $j = 1, ..., 17$ between demeaned ΔX_t and demeaned $X_{t-1} - t\hat{\rho}_1$, where $\hat{\rho}_1 = X_{T+1}/(T+1)$. Here $T+1$ is the size of the sample with $T = 102$ for the quarterly data and $T = 924$ for the weekly data.

Figure 3 shows the cdf of the empirical distribution of λ_j, $j = 1, ..., 17$ superimposed on the cdf of W_c (solid lines). For the quarterly data, the fit between the two cdf's is tight and the extreme λ's almost coincide with the boundaries of the support of W_c. Hence, there is no obvious reason to quickly reject the no cointegration hypothesis.

In contrast, for the weekly data, the fit is only good up to about the median. The largest squared sample canonical correlation (not shown on the graph to enhance visibility of the other λ's) equals 0.436, which is more than four times larger than the upper boundary of the support of W_c, $b_+ = 0.103$. This signals the existence of the cointegrating relationships.

For the quarterly data, the values of the maximum eigenvalue and trace statistics are 110.89 and 742.71, respectively. The corresponding asymptotic 5% critical values[3] equal 110.77 and 698.69. Therefore, the uncorrected tests reject the null

[3] We compute these values by simulation, using the response surface technique of MacKinnon et al (1999).

test	alternative	ρ=0	0.3	0.5	0.6	0.7	0.8	0.9	1
					'Quarterly data'				
max eigenvalue	(ii)	55.7	13.6	7.4	6.3	5.8	5.5	5.4	5.5
trace	(iii)	100	100	100	89.8	20.3	1.1	0.4	6.1
					'Weekly data'				
max eigenvalue	(ii)	100	100	100	100	100	99.1	19.2	5.1
trace	(iii)	100	100	100	100	100	100	100	5.4

Table 4: Power of the Bartlett-corrected maximum eigenvalue and trace tests. For the trace test 'our' correction formula is used. The last column of the table reports actual size of the test with the asymptotic size 5%.

of no cointegration. However, the Bartlett-corrected maximum eigenvalue statistic equals 99.16, overturning the conclusion of the uncorrected test. Similarly, the trace statistic, when corrected using 'our' formula, equals 672.08, and when corrected using Johansen's (2002) formula, equals 677.39. In both cases, we cannot reject the null of no cointegration.

For the weekly data, the values of the maximum eigenvalue and the trace statistics are 555.58 and 1290.59, respectively. The Bartlett-corrected maximum eigenvalue statistic equals 550.33. Therefore, the corrected maximum eigenvalue test still rejects the null. The trace statistic, when corrected using 'our' formula, equals 1278.54, and when corrected using Johansen's (2002) formula, equals 1279.68. In both cases, we do reject the null.

The difference between the conclusions of the tests for the quarterly and weekly data is likely due to the low power of the tests in the former case. To verify this conjecture, we simulate 17-dimensional Gaussian data with $T = 102$ (as in the quarterly data) and with $T = 924$ (as in the weekly data) under three different scenarios: (i) pure random walk; (ii) all but one of the 17 components of the data are random walks, whereas the special component is an AR(1) with the autoregressive coefficient ρ, and (iii) VAR(1) with the coefficient ρI_{17}. The values of ρ are 0, 0.3, 0.5, 0.6, 0.7, 0.8, and 0.9.

Table 4 reports the power of the Bartlett-corrected maximum eigenvalue test against alternative (ii), and that of the Bartlett-corrected (using 'our' formula) trace test against alternative (iii). In the 'quarterly data' case, the power against alternatives with $\rho \geq 0.7$ is extremely low, and is often even lower than the size, which is reported in the last column of the table.

In contrast, the power in the 'weekly data' case is nearly 100% for all alter-

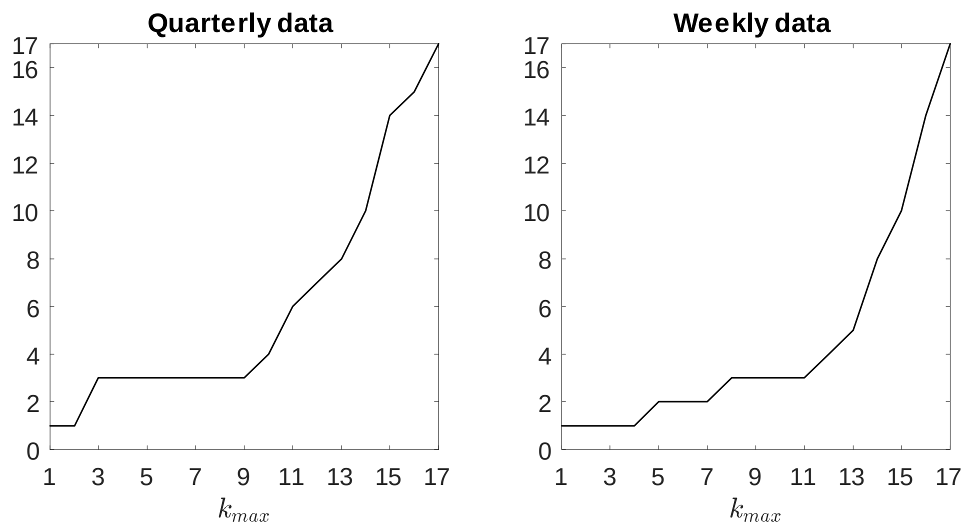

Figure 4: The number of estimated common factors in the log exchange rate data according to Bai's (2004) IPC_1 criterion.

natives, but the one with only one cointegrating relationship and $\rho = 0.9$. For that alternative, the power of the corrected maximum eigenvalue test is 19.2%. Overall, the table confirms our guess that the differences in the 'quarterly data' and 'weekly data' results are due to the low power of the tests in the 'quarterly data' case. The more informative 'weekly data' results favour the existence of cointegration between the 17 log exchange rates.

It is interesting to compare our analysis to factor-based approaches to cointegration. Bai (2004) considers non-stationary panels where the non-stationarity is entirely due to a few common factors. Using the criteria for the determination of the number of factors proposed in that paper (see its eq. (12)), one may consistently estimate the number of the cointegrating relationships by the dimensionality minus the estimate of the number of factors.

Figure 4 shows the number of factors estimated by Bai's (2004) IPC_1 criterion using our quarterly and weekly data. The number of estimated factors is a function of the user-supplied maximum number of factors $k_{\max}$. By construction, the criterion returns p factors if $k_{\max} = p$. Hence, strictly speaking, it cannot be used to test the null of no cointegration. Interestingly, for both datasets, when $k_{\max}$ is set to no more than half the dimensionality, the criterion estimates no more than three factors, which corresponds to no less than $17 - 3 = 14$ cointegrating relationships. Unfortunately, such an estimate may be unreliable because log exchange rates do not necessarily have stationary idiosyncratic components.

Bai and Ng (2004), as well as Barigozzi et al. (2018) and Banerjee et al. (2017) generalize the setup of Bai (2004) by allowing for non-stationary idiosyncratic dynamics. In such a framework, the presence of common factors does not necessarily imply cointegration in the data.

The focus of Barigozzi et al. (2018) and Banerjee et al. (2017) is on the impulse responses to the factor shocks interpreted as economy-wide disturbances. These papers pay much attention to the cointegration between the factors because it is consequential for the impulse responses. In contrast, the cointegration in the data, which may be due either to factors or to idiosyncratic relationships, is of little importance from the papers' perspective, and no systematic analysis of such cointegration is attempted.

Bai and Ng's (2004) PANIC methodology can, in principle, be used to test the hypothesis of no cointegration in the data. PANIC idea is, first, to estimate the factor and idiosyncratic components in the first-differenced data, then, cumulate these components, and finally, determine the cointegration rank of the cumulated factors and test for a unit root in each of the cumulated idiosyncratic terms. If the number of stochastic trends underlying the factors is smaller than the number of stationary idiosyncratic terms, then the data must be cointegrated.

We have applied PANIC to our data. Details of this analysis are reported in SM. Here, we only summarize the results. The number of factors was estimated at two. The MQ_c^τ and MQ_f^τ tests have not detected any cointegration between the two cumulated factors. The $\mathrm{ADF}_{\hat{e}}^\tau$ tests rejected the null of a unit root (at 5% significance level) in six cumulated idiosyncratic series (Australia, Japan, Korea, Sweden, Belgium, and Finland) when the quarterly data were used, but only in two such series (Korea and Norway) when the weekly data were used. Taken at face value, these results would imply the existence of cointegration in the quarterly data, but no evidence for cointegration in the weekly data.

Note however, that PANIC does not analyze cointegration between the non-stationary idiosyncratic terms. Therefore, its not finding a sufficient number of the stationary idiosyncratic series in the weekly data does not necessarily contradict the evidence for cointegration contained in Figure 3 and supported by the Bartlett-corrected maximum eigenvalue and trace tests.

As to the cointegration in the quarterly data found by PANIC, unfortunately, this result may be spurious. To show this, we simulated 17-dimensional random walk data with $T = 102$, driven by innovations η_t that have covariance matrix Σ equal to the sample covariance matrix of the actual first-differenced quarterly data. In contrast to the methods based on the squared sample canonical correlations, PANIC is not invariant with respect to the choice of Σ, so the calibration is important. We repeated the simulation 10,000 times, each time applying PANIC analysis assuming the existence of two factors. Even though the data are not cointegrated by construction, PANIC rejected the null of a unit root for three or more idiosyncratic components (which is sufficient for the existence of the cointegration) in 7,092 cases. It found 6 or more stationary components in 1,864 cases.

In conclusion of this section, we would like to remind the reader that in this paper we do not propose any new methods for determining the number of cointegrating relationships. Looking at Figure 3, we are tempted to "estimate" it by the number of λ's that 'deviate' from the cdf of W_c. For the weekly data, the number of the 'deviating' λ's is, perhaps, 9. However, we have nothing to say about the properties or the quality of this estimator. Its study is left for future research.

7 Conclusion and discussion

This paper establishes the a.s. convergence of the largest and the smallest eigenvalues of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ to the upper and lower boundaries of the support of the Wachter distribution W_{c_0}. This complements OW18's result on the a.s. weak convergence of the empirical distribution of the eigenvalues to W_{c_0}. The strategy of our proofs is similar to that of the proof of the convergence of the extreme eigenvalues of the sample covariance matrix in BS98. However, the fact that we have to deal with the product of four dependent stochastic matrices, S_{01}, S_{11}^{-1}, S_{10}, and S_{00}^{-1}, presents substantial challenges that we overcome.

Eigenvalues of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ can be interpreted as squared sample canonical correlations between demeaned innovations of a high-dimensional random walk and detrended and demeaned levels of this random walk. Such eigenvalues form the basis for the trace and maximum eigenvalue tests of no cointegration in high-dimensional VAR(1). Both the LR_{trace} and $LR_{\max}$ statistics have singularities at unity, and OW18's result cannot be used to establish their a.s. convergence.

The result of this paper shows that the singularity can be ignored because none of the eigenvalues of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ are close to unity asymptotically. Thus, our Corollary 1 establishes the a.s. limit of the scaled trace and maximum eigenvalue statistics. We use this result to obtain previously unknown analytic formulae for Bartlett-type correction coefficients for the trace and maximum eigenvalue tests.

We establish Theorem 1 under Gaussianity of the errors η_t of model (1). We need the Gaussianity for two reasons. First, it allows us to reduce the analysis of $S_{01}S_{11}^{-1}S_{10}S_{00}^{-1}$ to that of $C'D^{-1}CA^{-1}$, where C, D, and A have form $\varepsilon' W\varepsilon$ with block-diagonal W, and ε has i.i.d. elements. Second, we use it to derive bounds on the expected value of the inverse of the smallest eigenvalue of A (in SM). In principle, the first reason can be circumvented by simply assuming that the matrix ε of the discrete Fourier transforms of η has i.i.d. (but not necessarily Gaussian) elements. This still leaves the second reason intact. Unfortunately even a seemingly innocuous assumption that the elements of ε are i.i.d. Bernoulli random variables leads to non-invertibility of A with small but positive probability, and hence, to nonexistence of the expected value of the inverse of the smallest eigenvalue of A. We leave removing the Gaussianity assumption as an important topic for future research.

The Monte Carlo analysis in Section 5 shows that the extreme squared sample canonical correlations are well approximated by the boundaries of the support of the Wachter distribution in finite samples and without Gaussianity. Our Bartlett-correction formulae deliver well-sized tests, especially in situations where p/T is relatively small. Interestingly, our Bartlett correction outperforms the correction proposed in Johansen (2002) for most of the Monte Carlo settings.

Guided by the theoretical results of this paper and of OW18, we propose a quick graphical method for cointegration detection in a high-dimensional VAR(1). The empirical cdf of the squared sample canonical correlations is superimposed on the cdf of the Wachter distribution $W_{p/T}$. A visible mismatch signals the presence of cointegration. We view this technique as complementary to more formal analysis based on statistical tests.

As an empirical application of the proposed graphical method and of the obtained Bartlett correction formulae, we study cointegration between log exchange rates of 17 OECD countries. We find a clear evidence of the cointegration in the weekly data, but not in the quarterly data. This is consistent with our Monte Carlo results that find low power of the cointegration detection procedures in situations where p/T is relatively large, which is a feature of our quarterly data.

In our opinion, there are two most important directions for future research. First, the results of this paper and of OW18 should be generalized to the VAR(k) setting with $k > 1$. A preliminary Monte Carlo analysis shows that the empirical distribution of the squared sample canonical correlations would still converge to a Wachter distribution, but parameters of that Wachter would depend on the parameters of VAR(k) under the null. Second, it would be interesting to study the asymptotic fluctuations of the empirical distribution of the squared sample canonical correlations around the Wachter limit. This would allow one to directly derive the simultaneous large-p and large-T asymptotics for the cointegration tests, instead of relying on the Bartlett-type correction of the tests based on the sequential asymptotics. We are following these two research directions in separate projects.

Acknowledgements: We are very grateful to Frank Diebold and two anonymous referees for many suggestions that substantially improved the paper.

References

[1] Bai, J., 2004. "Estimating cross-section common stochastic trends in nonstationary panel data," *Journal of Econometrics* 122, 137–183.

[2] Bai, J. and S. Ng, 2004. "A PANIC attack on unit roots and cointegration," *Econometrica* 72, 1127–1177.

[3] Bai, Z. and Silverstein, J., 1998. "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large-dimensional Sample Covariance Matrices," *The Annals of Probability* 26, 316-345.

[4] Baillie, R. T. and T. Bollerslev, 1989. "Common Stochastic Trends in a System of Exchange Rates," *Journal of Finance* 44, 167–181.

[5] 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.

[6] Banerjee, A., M. Marcellino, and I. Masten, 2017. "Structural FECM: Cointegration in large-scale structural FAVAR models," *Journal of Apllied Econometrics* 32, 1069–1086.

[7] Barigozzi, M., M. Lippi, and M. Luciani, 2018. "Non-Stationary Dynamic Factor Models for Large Datasets," *arXiv 1602.02398v3*

[8] Brillinger, D.R., 1981. *Time series: data analysis and theory.* Holden Day, Inc., San Francisco.

[9] Capitaine, M. and Donati-Martin, C., 2016. "Spectrum of deformed random matrices and free probability," *arXiv 1607.05560v*

[10] Cheng, X. and P.C.B. Phillips, 2012. "Cointegrating Rank Selection in Models with Time-varying Variance," *Journal of Econometrics* 169, 155–165.

[11] 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.

[12] Diebold, F. X., J. Gardeazabal, and K. Yilmaz, 1994. "On Cointegration and Exchange Rate Dynamics," *The Journal of Finance* 49, 727–735.

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

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

[15] 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.

[16] 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, pp. 212-229

[17] 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.

[18] Horn, R.A. and C.R. Johnson, 1985. *Matrix Analysis*, Cambridge University press.

[19] 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.

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

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

[22] Johansen, S., 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press Inc., New York.

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

[24] 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/.

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

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

[27] 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.

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

[29] Onatski, A. and C. Wang, 2018. "Alternative asymptotics for cointegration tests in large VARs," forthcoming in *Econometrica*.

[30] Pedroni, P.L., Vogelsang, T.J., Wagner, M., and Westerlund, J., 2015. "Nonparametric rank tests for non-stationary panels," *Journal of Econometrics* 185, 378-391.