

COINTEGRATION IN LARGE VARS

BY ANNA BYKHOVSKAYA^{*} AND VADIM GORIN[†]

University of Wisconsin-Madison, ^{}anna.bykhovskaya@wisc.edu; [†]vadicgor@gmail.com*

The paper analyses cointegration in vector autoregressive processes (VARs) for the cases when both the number of coordinates, N , and the number of time periods, T , are large and of the same order. We propose a way to examine a VAR of order 1 for the presence of cointegration based on a modification of the Johansen likelihood ratio test. The advantage of our procedure over the original Johansen test and its finite sample corrections is that our test does not suffer from over-rejection. This is achieved through novel asymptotic theorems for eigenvalues of matrices in the test statistic in the regime of proportionally growing N and T . Our theoretical findings are supported by Monte Carlo simulations and an empirical illustration. Moreover, we find a surprising connection with multivariate analysis of variance (MANOVA) and explain why it emerges.

1. Introduction.

1.1. *Motivation.* The importance of cointegration in economics stems from the seminal papers [Granger \(1981\)](#) and [Engle and Granger \(1987\)](#). For example, as they show, monthly rates on 1-month and 20-year treasury bonds are cointegrated, which means that they are both non-stationary, but they have a stationary linear combination. A lot of other variables in macroeconomics and finance such as price level, consumption, output, trade flows, interest rates and so on are non-stationary, and, thus, are potentially subject to cointegration. When dealing with non-stationary time series, it is always a question whether one should work with levels or with differences. For multivariate settings such as vector autoregression (VAR), the choice of model would depend on whether the series is cointegrated or not.

There are several ways to test for the presence of cointegration (see, e.g., [Maddala and Kim \(1998\)](#) for the detailed description of various methods). One popular approach relies on checking whether the residuals from regressing one of the coordinates on the remaining ones are stationary. It is based on [Engle and Granger \(1987\)](#) and was later extended in [Phillips and Ouliaris \(1990\)](#). Another widely used technique is due to Søren Johansen ([Johansen, 1988, 1991](#))¹. This approach assumes VAR structure and relies on the likelihood ratio. It tests a null hypothesis of at most ρ cointegrating relationships versus an alternative of between ρ and $r > \rho$ cointegrating relationships. The Johansen test turns out to be related to the eigenvalues of some random matrix and has a non-standard asymptotic distribution.

Neither of the approaches is commonly used in the analysis of large systems. However, in many situations the data turns out to have both cross-sectional and time dimensions being large. Natural examples are given, e.g., by financial data (stock prices, exchange rates, etc.) or by monthly data on trade and investments between countries (where the number of pairs of countries is large). The main difficulty with applying the above cointegration testing approaches to high-dimensional settings is that both of them require N , the cross-sectional dimension of a time series, to be fixed and small. For the former approach, the larger is N , the more regressions we need to run and interpreting their results becomes ambiguous. For

Keywords and phrases: High-dimensional VAR, Cointegration, Johansen test, Jacobi ensemble.

¹A related approach was also proposed in [Stock and Watson \(1988\)](#).

the latter approach, it turns out, the asymptotic theory stops providing a good approximation for moderate values of T . The test starts to over-reject the null of fewer cointegrations in favor of the alternative (see, for example, [Ho and Sørensen \(1996\)](#), [Gonzalo and Pitarakis \(1999\)](#)). For instance, the simulations reported in Table 1 of [Gonzalo and Pitarakis \(1999\)](#) indicate that the empirical rejection rate based on the 95% asymptotic critical value for the no-cointegration hypothesis (constructed using the asymptotics of [Johansen \(1988, 1991\)](#)) is 20% at $N = 5$, $T = 30$. It eventually decreases to the desired 5% as T grows, but even at $T = 400$ the empirical size is still 6%.

After size distortions became clear econometricians developed various procedures for correcting over-rejection. Popular methods include finite sample correction (e.g., [Reinsel and Ahn \(1992\)](#), [Johansen \(2002\)](#)) and bootstrap (e.g., [Swensen \(2006\)](#), [Cavaliere et al. \(2012\)](#)). Such modifications help to restore correct size for moderate values of T (e.g., $T = 50$), when N is of a smaller order (e.g., $N = 5$). Yet, the question of larger N remained open.

A recent ground-breaking paper [Onatski and Wang \(2018\)](#) shows that the over-rejection when testing the null of no cointegration can be mathematically explained by considering an alternative asymptotics in which both N and T go to infinity jointly and proportionally. When N is large such asymptotic regime turns out to better suit the finite sample properties of the data. While [Onatski and Wang \(2018\)](#) point this out, they do not provide alternative testing procedures.

The observation that large VARs might have analyzable joint limits opens a new area of research. That is, one can try to develop various sophisticated joint asymptotics and derive, as a result, appropriate ways to test for cointegration in settings where both N and T are large. This, however, requires new tools rooted in the random matrix theory. In our paper we propose such cointegration tests and introduce asymptotic theorems which make testing possible. Let us describe our main results.

1.2. Results. By modifying the Johansen likelihood ratio (LR) test, we come up with a way to examine the presence of cointegration in a time series when the cross-sectional dimension, N , and the time dimension, T , are of the same order. We consider a vector autoregression (VAR) of order 1 in the error correction representation

$$(1) \quad \Delta X_t = \mu + \Pi X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T,$$

where $\Delta X_t := X_t - X_{t-1}$, errors $\{\varepsilon_t\}$ are Gaussian with $N \times N$ covariance matrix Λ , and Π , μ are unknown parameters.² We do not restrict Λ to be diagonal, which means that any cross-sectional correlation structure is allowed. Such cross-sectional heteroscedasticity assumption may be important in applications, where one expects variables, e.g., countries, to be correlated. In contrast, many previous approaches relied on specific forms of covariance Λ , see, e.g., [Breitung and Pesaran \(2008\)](#), [Bai and Ng \(2008, Section 7\)](#) for reviews and [Zhang et al. \(2018\)](#) for recent developments.

Our model allows for an arbitrary linear trend in X_t , which is a desirable feature when one goes to applications. Yet, it also encompasses as a special case a model without a trend ($\mu = 0$). The way we deal with the data turns out to be invariant to μ , and we do not propose any different procedure even if a researcher has a prior knowledge that there is no trend. (We discuss this in more detail in Section 7.2.)

We are interested in whether the N -dimensional non-stationary process X_t is cointegrated or not. That is, whether there is a non-zero vector β such that $\beta' X_t$ is trend stationary. Our main contributions lie in the construction of the appropriate test, analysis of its asymptotic

²Section 7.3 discusses possible generalizations of our approach to VAR(k) for general k .

distribution, and computation of critical values (some of them are reported in Table 1 of Section 3).

Our procedure for cointegration testing relies on the following steps. First, we de-trend X_t by subtracting $\frac{t}{T}(X_T - X_0)$ from X_t . Then we follow Johansen's procedure and regress both first differences, ΔX_t , and lagged de-trended X_t on a constant. That is, we de-mean the data. Then we calculate the squared sample canonical correlations between the residuals from those regressions. This corresponds to the eigenvalues from the modification of the Johansen LR statistics obtained by using the lagged de-trended version of X_t and not de-trended first differences ΔX_t .

The de-trending procedure can be interpreted in the following way: we take the first differences of the observed variables, de-mean them, and re-sum back. This is equivalent (up to a constant X_0 which disappears after we further de-mean the sums) to our de-trending. Such interpretation is related to Elliott et al. (1996) which analyzes unit root testing in the presence of a trend d_t . Under the null hypothesis of a unit root and when the trend d_t is linear, Elliott et al. (1996) suggests to de-mean the first differences.

As discussed in Xiao and Phillips (1999), de-trending plays an important role in improving the performance of cointegration tests. In our procedure there is an additional reason why de-trending matters. It leads to an unexpected connection with the Jacobi ensemble, a familiar object in random matrix theory, but a novel one in econometrics.³ Eventually, this connection is the main technical tool leading to our asymptotic results and construction of the test.

We show that after proper rescaling our test under the null hypothesis of no cointegration converges to the sum of the first r elements of the Airy_1 point process (we formally define that process in Section 3.1). Airy_1 process is a known object in the random matrix theory and its marginal distributions can be computed in various ways. We also present in Section 5 a simulation study of the speed of convergence, which supports the limiting results even for moderate values of N and T . In Table 2 of that section we demonstrate the significant improvement over the finite sample behavior of the Johansen test and its corrections reported in Gonzalo and Pitarakis (1999).

Along with the size we also report power simulations in Section 5. The power depends on the choice of the alternative and we perform several experiments with random rank 1 matrix Π and random initial condition X_0 of varying magnitudes. In many cases the power is very close to 100%. In particular, we found high power even for moderate N, T , e.g., $N \approx 30, T \geq 150$. In general, the results are encouraging and we see quite large power envelopes.

We also present a small empirical illustration of our testing procedure on weekly S&P100 log-prices over 10 years. This gives us approximately 500 observations across time and $T/N \approx 5$, corresponding to high power and close to 5% sample rejection rate in simulations. Log-prices are known to have a unit root, and we do not find any strong evidence that they are cointegrated.

1.3. Techniques. Let us briefly indicate technical aspects of our proofs and their mathematical novelty. The key observation that we make is that a small perturbation of the matrix arising in the modified Johansen test has an explicit distribution of its eigenvalues. This distribution is called Jacobi ensemble, and its usual appearances in statistics include sample canonical correlations for two sets of independent data (as opposed to highly dependent in our case, see the discussion in Section 4.1) and multivariate analysis of variance (MANOVA), see, e.g., Muirhead (2009). Our method has two main components which are new, as far as we are aware. First, our perturbation of the model is based on a replacement of a certain

³We remark that de-trending is implicitly present in the proofs of Onatski and Wang (2018, 2019), yet in our work it becomes a significant ingredient, rather than a technical detail.

permutation in the matrix formulation of the modified Johansen test by a uniformly random orthogonal matrix. The second ingredient is a challenging computation of matrix integrals⁴ leading to the identity of the law of the perturbed matrix with the Jacobi ensemble.

Put it otherwise, we discover an exactly-solvable⁵ model in a small neighborhood of the (random) matrix of the Johansen cointegration test. Let us center our attention on this exactly-solvable model. It can be used as an initial point for perturbative arguments leading to asymptotic theorems for our and possibly other modifications of the Johansen test. In addition, our exactly-solvable model is not isolated, but rather it is a representative of a whole class of similar cases. We expect that our approach works in several other situation in which various test statistics in the vector autoregression context can be understood through (other) instances of the Jacobi ensemble. Justifications of this point of view are contained in Sections 7.2, 7.4, and in Section 9.5 in Bykhovskaya and Gorin (2021, Supplementary Material).

1.4. Outline of the paper. Section 2 describes the setting and the main objects of interest. Section 3 presents asymptotic results, while Section 4 gives a sketch of their proofs. Section 5 shows supporting Monte Carlo simulations and Section 6 applies our test to S&P100. Additional results and extensions are presented in Section 7. Finally, Section 8 concludes. All proofs, unless otherwise noted, are in Bykhovskaya and Gorin (2021, Supplementary Material).

2. Setting.

2.1. Building block. We consider an N -dimensional vector autoregressive process of order 1, VAR(1), based on a sequence of i.i.d. mean zero Gaussian errors $\{\varepsilon_t\}$ with non-degenerate covariance matrix Λ . That is,

$$(2) \quad \Delta X_t = \Pi X_{t-1} + \mu + \varepsilon_t, \quad t = 1, \dots, T.$$

where $\Delta X_t := X_t - X_{t-1}$ and Π, μ are unknown parameters. We do not impose any restrictions on Λ , thus, we allow for arbitrary correlations across coordinates of X_t . The process is initialized at fixed X_0 . We outline possible extensions to autoregressions of higher order, VAR(k), in Section 7.3.

We are interested in analyzing whether X_t is cointegrated. That is, whether there exists a non-zero $N \times r_0$ matrix β such that $\beta' X_t$ is (trend) stationary.

If there exists a $N \times r_0$ matrix β of rank r_0 such that $\beta' X_t$ is (trend) stationary, but there does not exist a $N \times (r_0 + 1)$ matrix $\tilde{\beta}$ of rank $r_0 + 1$ such that $\tilde{\beta}' X_t$ is (trend) stationary, then we say that X_t is cointegrated of order r_0 . As shown in Engle and Granger (1987, Granger representation theorem), Johansen (1995, Theorem 4.5) the necessary condition for X_t to be cointegrated of order r_0 is $\text{rank}(\Pi) = r_0$ and, thus, there exist two $N \times r_0$ matrices α and β of rank r_0 such that $\Pi = \alpha\beta'$. For the sufficiency one also needs to require an extra non-degeneracy condition involving α, β , and Γ_i . This additional requirement is used to rule out the $I(2)$ processes, i.e., such that their second differences are stationary, while the first differences are not.

⁴Our arguments have some similarities with proofs in Hua (1963).

⁵A random model is called exactly-solvable or integrable (see Borodin and Gorin (2016) for an overview) if there exist explicit formulas for the expectations of non-trivial random variables describing the system. Such formulas provide a basis for asymptotic analysis of the systems of interest. This is in contrast to generic models where explicit formulas are rarely available.

2.2. Cointegration test. We test the null hypothesis H_0 of no cointegration, which is equivalent to $\text{rank}(\Pi) = 0$ or $\Pi \equiv 0$. The complement to the null is $\text{rank}(\Pi) > 0$. Yet, in order to design our test we set $\text{rank}(\Pi) \in [1, r]$, where r is a fixed finite number, to be our alternative hypothesis H_1 . Thus, our alternative, in line with Granger representation theorem, can be interpreted as having at most r cointegrating relationships. While the test has different power depending on r and the true data-generating process, when working with real data for which the true value of $\text{rank}(\Pi) \equiv r_0$ is unknown one can use any r in order to reject H_0 .⁶ Formally, we test

$$(3) \quad H_0 : \Pi \equiv 0 \quad \text{vs.} \quad H_1 : \text{rank}(\Pi) \in [1, r],$$

Our testing procedure is a modification of the Johansen likelihood ratio (LR) test (Johansen, 1988, 1991).⁷ We focus on the small r regime, which differs from the classical Johansen LR test, where $r = N$ is more commonly used as an alternative. In that sense, our approach is close to the maximum eigenvalue ($\lambda_{\max}$) test. That turns out to be important in the large N, T setting, see Section 3.2. Our approach consists of several steps:

Step 1. De-trend⁸ the data and define

$$(4) \quad \tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0).$$

Note that we also do a time shift. This shift is in line with Johansen test, where lags and first differences are regressed on the observables.

Step 2. De-mean the data. That is, regress de-trended lags, $\tilde{X}_t$, and first differences ΔX_t on a constant. Define the residuals from those regressions as

$$(5) \quad R_{0t} = \Delta X_t - \frac{1}{T} \sum_{\tau=1}^T \Delta X_\tau, \quad R_{1t} = \tilde{X}_t - \frac{1}{T} \sum_{\tau=1}^T \tilde{X}_\tau.$$

Step 3. Calculate the squared sample canonical correlations between R_0 and R_1 . That is, define

$$(6) \quad S_{ij} = \sum_{t=1}^T R_{it} R_{jt}^*, \quad i, j = 0, 1,$$

where here and below the notation X^* means transpose of the matrix X (transpose-conjugate whenever complex matrices appear). Then calculate N eigenvalues $\lambda_1 \geq \dots \geq \lambda_N$ of the matrix $S_{10} S_{00}^{-1} S_{01} S_{11}^{-1}$. The eigenvalues solve the equation

$$(7) \quad \det(S_{10} S_{00}^{-1} S_{01} - \lambda S_{11}) = 0.$$

Step 4. Form the test statistic

$$(8) \quad LR_{N,T}(r) = \sum_{i=1}^r \ln(1 - \lambda_i).$$

The subscript N, T in (8) indicates that we modify Johansen LR test to develop the large N, T asymptotics. This statistic after centering and rescaling will be compared with appropriate critical values to decide whether one can reject H_0 or not (see Theorem 2).

⁶In practice we recommend using reasonably small values of r , since our asymptotic theorems are valid in the regime of fixed r and large N, T .

⁷Johansen LR test is based on the maximization of the Gaussian likelihood.

⁸We discuss the role of de-trending in Section 7.2.

Throughout the proofs and extensions, we will be using an alternative way to write the residuals and matrices S_{ij} . For this let us define the *demeaning* operator $\mathcal{P}$. It is a linear operator in a T -dimensional space defined by its matrix

$$(9) \quad \mathcal{P} = \begin{pmatrix} 1 - \frac{1}{T} & -\frac{1}{T} & \cdots & -\frac{1}{T} \\ -\frac{1}{T} & 1 - \frac{1}{T} & \cdots & -\frac{1}{T} \\ & & \ddots & \\ -\frac{1}{T} & -\frac{1}{T} & \cdots & 1 - \frac{1}{T} \end{pmatrix}.$$

$\mathcal{P}$ is an orthogonal projector on the space orthogonal to the vector $(1, 1, \dots, 1)$. By definition

$$R_{0it} = [\Delta X \mathcal{P}]_{it} = \Delta X_{it} - \frac{1}{T} \sum_{\tau=1}^T \Delta X_{i\tau}, \quad R_{1it} = [\tilde{X} \mathcal{P}]_{it} = \tilde{X}_{it} - \frac{1}{T} \sum_{\tau=1}^T \tilde{X}_{i\tau}.$$

Using the fact that $\mathcal{P}^2 = \mathcal{P}$,

$$(10) \quad S_{00} = \Delta X \mathcal{P} \Delta X^*, \quad S_{01} = \Delta X \mathcal{P} \tilde{X}^*, \quad S_{10} = S_{01}^* = \tilde{X} \mathcal{P} \Delta X^*, \quad S_{11} = \tilde{X} \mathcal{P} \tilde{X}^*.$$

Let us emphasize that our test differs from the original Johansen test in the fact that we use $\tilde{X}_t$ instead of X_{t-1} . Note that $\tilde{X}_t$ can be viewed as a rank 1 perturbation of X_{t-1} . Hence, our test statistic is a finite rank perturbation of the original Johansen procedure.

3. Asymptotic results. In this section we formulate our main asymptotic results, explain and discuss the role of the precise setting we chose, and indicate directions for generalisations. We provide a sketch of the proof of our asymptotic theorem in Section 4.

3.1. Large (N, T) limit of the test. Our results use the Airy_1 point process. Thus, let us introduce it before we formulate the main theorems. The Airy_1 point process is a random infinite sequence of reals

$$\alpha_1 > \alpha_2 > \alpha_3 > \dots$$

which can be defined through the following proposition.

PROPOSITION 1 ([Forrester \(1993\)](#), [Tracy and Widom \(1996\)](#)). *Let Y_N be $N \times N$ matrix of i.i.d. $\mathcal{N}(0, 2)$ Gaussian random variables and let $\mu_{1;N} \geq \mu_{2;N} \geq \dots \geq \mu_{N;N}$ be eigenvalues of $\frac{1}{2}(Y_N + Y_N^*)$. Then in the sense of convergence of finite-dimensional distributions*

$$(11) \quad \lim_{N \rightarrow \infty} \left\{ N^{1/6} \left(\mu_{i;N} - 2\sqrt{N} \right) \right\}_{i=1}^N = \{\alpha_i\}_{i=1}^\infty.$$

The law of the first coordinate α_1 is known as the Tracy-Widom F_1 distribution; its distribution function can be written in terms of a solution of the Painlevé *II* differential equation.

We remark that from the computational point of view (11) gives an efficient way to access the distribution of various functions of $\{\alpha_i\}_{i=1}^\infty$.⁹ From the theoretical perspective, one would like to have a more structural definition, which can be used for the analysis. Such definitions exist, yet, unfortunately, none of them is particularly simple.¹⁰

⁹An even faster way uses tridiagonal matrix models ([Dumitriu and Edelman, 2002](#)).

¹⁰ There are several equivalent ways to define Airy_1 point process: through Pfaffian formulas for the correlation functions ([Forrester, 1993](#); [Tracy and Widom, 1996](#)), through combinatorial formulas for the Laplace transform ([Sodin, 2014](#)), through eigenvalues of the Stochastic Airy Operator ([Ramirez et al., 2011](#)).

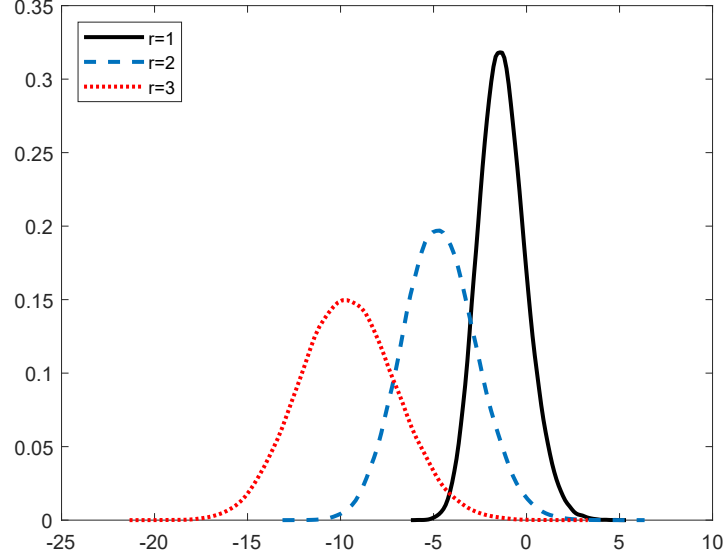


Figure 1: The probability density for the random variables $\sum_{i=1}^r \mathbf{a}_i$. The histogram plots are obtained from 100,000 samples of 200×200 Gaussian matrices following Proposition 1.

THEOREM 2. Suppose that $T, N \rightarrow \infty$ in such a way that the ratio T/N belongs to $[2 + \gamma_1, \gamma_2]$ for some $\gamma_1, \gamma_2 > 0$. Suppose that H_0 holds, that is, we have (2) with $\Pi = 0$. Then for each finite $r = 1, 2, \dots$, we have convergence in distribution for the largest eigenvalues defined in Eq. (7):

$$(12) \quad \frac{\sum_{i=1}^r \ln(1 - \lambda_i) - r \cdot c_1(N, T)}{N^{-2/3} c_2(N, T)} \xrightarrow[T, N \rightarrow \infty]{d} \sum_{i=1}^r \mathbf{a}_i,$$

where

$$(13) \quad c_1(N, T) = \ln(1 - \lambda_+), \quad c_2(N, T) = -\frac{2^{2/3} \lambda_+^{2/3}}{(1 - \lambda_+)^{1/3} (\lambda_+ - \lambda_-)^{1/3}} (\mathbf{p} + \mathbf{q})^{-2/3} < 0,$$

$$(14) \quad \mathbf{p} = 2 - \frac{2}{N}, \quad \mathbf{q} = \frac{T}{N} - 1 - \frac{2}{N}, \quad \lambda_{\pm} = \frac{1}{(\mathbf{p} + \mathbf{q})^2} \left[\sqrt{\mathbf{p}(\mathbf{p} + \mathbf{q} - 1)} \pm \sqrt{\mathbf{q}} \right]^2.$$

REMARK 1. The condition $T/N > 2$ is important: The procedure for constructing λ_i involves computing squared sample canonical correlations between two N -dimensional subspaces in T -dimensional space. If $T/N < 2$, then two subspaces necessarily intersect, leading to $\lambda_1 = 1$. Hence, Eq. (12) would need an adjustment in such case.

Figure 1 shows densities for the random variables $\sum_{i=1}^r \mathbf{a}_i$ for $r = 1, 2, 3$. As r grows, the skewness of $\sum_{i=1}^r \mathbf{a}_i$ decreases, the expectations of $\sum_{i=1}^r \mathbf{a}_i$ tend to $-\infty$ and the variances tend to $+\infty$.

Quantiles of the distribution of $\sum_{i=1}^r \mathbf{a}_i$ serve as critical values for testing the hypothesis H_0 of no cointegrations against the alternative of at most r cointegrations. We report the quantiles for $r = 1, 2, 3$ in Table 1.

$\alpha \backslash r$	0.9	0.95	0.975	0.99
1	0.44	0.97	1.45	2.01
2	-1.88	-1.09	-0.40	0.41
3	-5.91	-4.91	-4.03	-2.99

TABLE 1

Quantiles of $\sum_{i=1}^r \alpha_i$ for $r = 1, 2, 3$ (based on 10^6 Monte Carlo simulations of $10^8 \times 10^8$ tridiagonal matrices of Dumitriu and Edelman (2002)).

Note the shifts by $\frac{2}{N}$ in the definition (14) of $\mathbf{p}$ and $\mathbf{q}$. Such finite sample correction is inspired by Johnstone (2008, Discussion before Theorem 1). Theorem 2 remains valid without these shifts, i.e., for the simplified $\mathbf{p} = 2$, $\mathbf{q} = \frac{T}{N} - 1$. However, our simulations show that for $T/N < 6$ the shifts improve the speed of convergence and, thus, the finite sample behavior of the test, see Section 5 and Section 9.3 in Bykhovskaya and Gorin (2021, Supplementary Material) for details.

We sketch the proof of Theorem 2 in Section 4 and give full details in Bykhovskaya and Gorin (2021, Supplementary Material). The two main ideas that we use are:

- (I) We show that a small (vanishing as $N, T \rightarrow \infty$) perturbation of eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ leads to an explicit probability distribution known as the Jacobi ensemble (see Definition 3 below). While this distribution appears in many problems of random matrix theory and multivariate statistics, its connection to the law of the Johansen test statistic remained previously unknown.
- (II) Further, we rely on the universality phenomenon from random matrix theory, which says that the particular choice $(\frac{1}{2}(Y_N + Y_N^*))$ for the law of random matrix made in Proposition 1 is not of central importance. Instead, the Airy_1 point process is a universal scaling limit for the largest eigenvalues in various ensembles of symmetric random matrices of growing sizes, see, e.g., Erdos and Yau (2012) and Tao and Vu (2012). In particular, an asymptotic statement similar to Proposition 1 is known for the Jacobi ensemble (see Section 9.3 in Bykhovskaya and Gorin (2021, Supplementary Material)) and, by combining it with the first idea, we eventually arrive at Theorem 2.

It is natural to ask about an extension of Theorem 2 for the case of r growing together with N . We distinguish two subcases here:

- Slow growth: $r = \lfloor N^\theta \rfloor$ for some $0 < \theta < 1$.
- Linear growth: $r = \lfloor \rho N \rfloor$ for some $0 < \rho \leq 1$.

In both cases we expect that (after proper adjustment of c_1 and c_2) the limit in (12) becomes Gaussian. Although we are not going to pursue this direction here, for the slow growth case the proof of asymptotic normality can probably be obtained by the same methods as we use in Theorem 2: we can again use the Jacobi ensemble for the computation. For the Jacobi ensemble individual eigenvalues λ_i (in the regime of growing i and $N - i$) become asymptotically Gaussian, hence, also their sums.¹¹

¹¹O'Rourke (2010) proves asymptotic Gaussianity for eigenvalues of Gaussian random matrices and we expect the same proof to work for the Jacobi ensemble case, see also Bao et al. (2013) for more details in the complex setting.

For the linear growth case the situation is more complicated. Our present tools only allow us to prove an asymptotic upper-bound of the form

$$(15) \quad \sum_{i=1}^{\lfloor \rho N \rfloor} \ln(1 - \lambda_i) - N \cdot c_3(T/N, \rho) = o(N^\epsilon), \quad N \rightarrow \infty,$$

for any $\epsilon > 0$ (for $\epsilon = 1$ this also follows from the results of [Onatski and Wang \(2018, 2019\)](#)),¹² while we expect the expression to be $O(1)$. Yet, there exist very general statements on asymptotic Gaussianity of linear statistics of functions of random matrices (see, e.g., [Guionnet and Novak \(2015\)](#), [Mingo and Popa \(2013\)](#)), and, therefore, there is little doubt in the fact that asymptotic distribution is Gaussian. Note, however, that these methods usually give very limited information about asymptotic variance and it is unclear at this moment how to find a reasonably simple explicit formula for it.

3.2. Theorem 2 and the role of finite r : Discussion. An important feature of our setting and our asymptotic results is that r is kept finite as $N, T \rightarrow \infty$. (As far as the authors know, Eq. (12) in Theorem 2 is the very first statistic for which the computation of the distributional limit became possible in the $N, T \rightarrow \infty, T/N \in [2 + \gamma_1, \gamma_2]$ regime.) In practice, when the dimensions are given to us in the data, this means that r should be much smaller than N and T . Notice that for the purpose of rejecting H_0 any statistic with known asymptotic distribution can be useful. It could be tempting to instead choose $r = N$ and test H_0 of no cointegrations against the alternative of arbitrary number of cointegrations, yet, we believe that in many situations small r allows to infer much more and is a good choice by the following reasons:

3.2.1. What can be estimated and when? The ultimate goal of the cointegration analysis is to find the cointegrating relationships. Turns out, the feasibility of that task crucially depends on their true number. Imagine for a second that the true number of cointegrating relationships is large, say $N/2$, then they span an $N/2$ -dimensional subspace of $\mathbb{R}^N$. Thus, the number of unknown parameters which need to be estimated is N^2 up to a multiplicative constant. On the other hand, the total number of observations that we have is NT . Hence, in our asymptotic regime $N, T \rightarrow \infty, T/N \in [2 + \gamma_1, \gamma_2]$, we have only finitely many observations per estimated parameter and consistent estimation is unlikely. The conclusion from this dimension-based heuristics is that in the situation when the number of cointegration relationships is large, our data does not contain enough information for the consistent estimation of all cointegrating relationships and, in a sense, the problem is hopeless. On the other hand, the same heuristics shows that we are in a much better scenario, if we have some a priori reasons to expect that either the number of cointegrating relationships is small, or if this number is in a small neighbourhood of N (in the latter case instead of the space of cointegrating relationships we can estimate its low-dimensional orthogonal complement). The former situation precisely corresponds to our choice of small r in Theorem 2; the first step towards the asymptotic analysis in the latter situation is discussed in Section 7.4.

3.2.2. Estimation of rank and cointegration relationships. In the small N cointegration analysis described in detail in [Johansen \(1988, 1991, 1995\)](#), rejection of the hypothesis of no cointegration is the first step. The next step is the estimation of the true rank r_0 of Π in (2) by subsequent rejections of the hypotheses of the rank being smaller than $2, 3, \dots, r_0$. Finally one wants to consistently estimate r_0 cointegrating relationships, closely related to Π itself.

¹²The value of the constant c_3 can be computed through certain integrals involving the equilibrium measure of the Jacobi ensemble, see (72) and (75) for more details.

Switching to our limit regime $N, T \rightarrow \infty$, $T/N \in [2 + \gamma_1, \gamma_2]$, there are two separate cases. If r_0 also grows linearly in N , then the full success of the above program is unlikely: as outlined in Section 3.2.1, we are trying to consistently estimate too many parameters from too few data points.

On the other hand, if true r_0 is small, then the situation is different (another potentially feasible case is that of small $N - r_0$, which we do not discuss here). We believe that our present results — Theorem 2 and the new approach underlying it — open a path to the full realization of the cointegration analysis for growing N . Let us outline the reasons.

From the technical point of view a key ingredient which needs to be developed is the asymptotics of squared sample canonical correlations $\lambda_1, \dots, \lambda_N$ and corresponding eigenvectors for the model (2) with Π of a fixed and not growing rank r_0 . In particular Theorem 2 should be extended from $H_0 : r_0 = 0$ to arbitrary finite r_0 . A natural approach here is by further developing the link to random matrices established in our proof of Theorem 2 via the Jacobi ensemble. This leads to the theory of spiked random matrices. A basic question of this theory is to infer the information on a large-dimensional small rank matrix B from observing $A + B$, where A is a matrix of pure noise (e.g., one can use $A = \frac{1}{2}(Y_N + Y_N^*)$ in the notations of Proposition 1). This is a well-developed theory, see e.g., Johnstone (2001), Baik et al. (2005) for seminal contributions and Bao et al. (2019), Johnstone and Onatski (2020) for the most recent progress. In order to obtain a connection with the VAR-framework (2), we treat the $\text{rank}(\Pi) = r_0$ case as a finite rank deformation of the “pure noise case” $\Pi = 0$, which we consider in this text. Hence, by combining our present results with the spiked random matrix theory, we expect to establish the complete asymptotic theory for any finite as $N \rightarrow \infty$ value of r_0 . (This will require further technical and conceptual efforts and, therefore, is left for future work.)

3.2.3. Power. Suppose that the true number of cointegrating relationships is small, say, $r_0 = 1$, yet we are trying to reject the null of no cointegrations using the LR-statistic with $r = N$, i.e.,

$$(16) \quad \sum_{i=1}^N \ln(1 - \lambda_i)$$

in the notations of Theorem 2. As we explain at the end of Section 3.1, the standard deviation of (16) is $O(1)$, i.e., after subtracting a constant as in Eq. (15), statistic (16) stays finite as $N, T \rightarrow \infty$. Hence, in contrast to (12) there should be no N -dependent rescaling in a test based on (16). This should lead to very different powers of tests based on statistics (12) and (16).

Indeed, we expect the main difference in the joint law of $(\lambda_1, \dots, \lambda_N)$ under the H_0 and under the alternative of one cointegrating relationship to be in the behavior of λ_1 (which is in line with $\ln(1 - \lambda_1)$ being the likelihood ratio test statistic in such setting). Change of λ_1 by $\nu \in \mathbb{R}$, i.e., $\lambda_1 \rightarrow \lambda_1 + \nu$, is on the same scale as the standard deviation of (16), and ν needs to be large in order for the test statistic to detect this change and to reject H_0 . On the other hand, (12) is changing in the same situation by a number of order $N^{2/3}\nu$, which is much larger, and, hence, rejection of H_0 is more likely. Therefore, one could anticipate that (12) leads to a test of higher power than (16) in this situation.¹³

¹³We do not analyze (16) in this text and, thus, we do not provide any rigorous justification of the above. We remark that the existing in the literature similar comparisons, e.g., Lütkepohl et al. (2001), Paruolo (2001), do not apply in our situation, since they only deal with the small N case.

3.2.4. Checking model validity. The fact that we assume r is finite and only use first r largest eigenvalues allows us to use the remaining ones to verify the plausibility of our model specification. That is, we can check whether the statistical properties of $\lambda_{r+1}, \dots, \lambda_N$ agree with the predictions of our model: for any finite r we expect the empirical distribution function $\frac{1}{N-r} \sum_{i=r+1}^N \mathbf{1}_{\lambda_i \leq x}$ to converge to the CDF of the Wachter distribution (under H_0 this can be deduced from our Theorem 4 and Theorem 14 in [Bykhovskaya and Gorin \(2021, Supplementary Material\)](#), while more general result under H_1 with finite r follows from ([Onatski and Wang, 2018, Theorem 1](#))). Our empirical example in Section 6, indeed, shows the match of the S&P100 data with the Wachter distribution (see Figure 7).

4. Outline of the proofs. The proof of Theorem 2 rests on the notion of the Jacobi ensemble. Let us first define it and then provide a sketch of the proof of Theorem 2.

4.1. Jacobi ensemble.

DEFINITION 3. A (real) Jacobi ensemble $\mathbf{J}(N; p, q)$ is a distribution on $N \times N$ symmetric matrices M of density proportional to

$$(17) \quad \det(M)^{p-1} \det(I_N - M)^{q-1}, \quad 0 < M < I_N,$$

with respect to the Lebesgue measure, where $p, q > 0$ are two parameters and $0 < M < I_N$ means that both M and $I_N - M$ are positive definite.

When $N = 1$, (17) is the Beta distribution. For general N , the eigenvectors of random Jacobi-distributed M are uniformly distributed, while N eigenvalues $x_1 \geq \dots \geq x_N$ admit an explicit density with respect to the Lebesgue measure given by

$$(18) \quad \frac{1}{Z(N; p, q)} \prod_{1 \leq i < j \leq N} (x_i - x_j) \prod_{i=1}^N x_i^{p-1} (1 - x_i)^{q-1},$$

where $Z(N; p, q)$ is an (explicitly known) normalization constant.

The Jacobi ensemble is widely used in statistics ([Muirhead, 2009](#)), theoretical physics ([Mehta, 2004](#)), and random matrix theory ([Forrester, 2010](#)). There are numerous tools for studying it¹⁴, and as a result its asymptotic properties (usually as $N \rightarrow \infty$) are known in great detail.

Two particularly important appearances of the Jacobi ensemble in statistics are in the multivariate analysis of variance (MANOVA) and in the canonical-correlation analysis, see [Johnstone \(2008, Sections 1 and 2\)](#) for more detailed discussions and more examples. Let us explain the latter setting, as it has some resemblance with the Johansen test. We start with two rectangular matrices of data: X of size $N \times T$ and Y of size $K \times T$. One can interpret X_{it} (and Y_{it}) as the i th variable at time t .

If $N = K = 1$, then one can think about data sets X and Y as T observations of two variables. Then one can measure the (linear) dependence between these variables by computing the (squared) sample correlation coefficient:

$$(19) \quad \frac{\left(\sum_{t=1}^T X_t Y_t \right)^2}{\sum_{t=1}^T (X_t)^2 \sum_{t=1}^T (Y_t)^2}.$$

¹⁴Just to mention some: Pfaffian point processes ([Mehta, 2004](#)), combinatorics of moments ([Bai and Silverstein, 2010](#)), variational problems for log-gases ([Ben Arous and Guionnet, 1997](#)), Kadell integrals ([Kadell, 1997](#)), Schwinger-Dyson/Loop equations ([Johansson, 1998](#)), tridiagonal models ([Killip and Nenciu, 2004](#)), Macdonald processes ([Borodin and Gorin, 2015](#)).

A direct computation shows that if the elements of X and Y are i.i.d. mean zero Gaussians, then the law of (19) is given by Beta distribution for any $T = 1, 2, \dots$

A generalization of (19) to $N, K \geq 1$ defines the *squared sample canonical correlations* $\{r_i\}$ as solutions to the equation

$$(20) \quad \det(S_{XY}S_{YY}^{-1}S_{YX} - rS_{XX}) = 0,$$

where $S_{XY} = XY^*$, $S_{YY} = YY^*$, $S_{YX} = YX^*$, $S_{XX} = XX^*$ (by X^* we mean transpose of X when X is a real matrix and conjugate-transpose of X when X is complex). Generically the equation (20) has $\min(N, K)$ non-zero solutions. They have a variational meaning. For instance, the maximal solution of Eq. (20) is the maximal squared correlation coefficient between a^*X and b^*Y , where the maximization goes over all N -dimensional vectors a and K -dimensional vectors b .

Now suppose that the columns of X are i.i.d. N -dimensional mean 0 Gaussians with (non-degenerate) covariance matrix Λ and the columns of Y are i.i.d. K -dimensional mean 0 Gaussians with (non-degenerate) covariance matrix Λ' . Assume for simplicity that $N \leq K$ and $N + K \leq T$. If X and Y are independent, then it can be shown that the (squared) sample canonical correlations $r_1 > r_2 > \dots > r_N$ have the law of the eigenvalues of the Jacobi ensemble $\mathbf{J}(N; \frac{K-N+1}{2}, \frac{T-N-K+1}{2})$.

At this point we observe both a similarity and a difference between the above instance of the Jacobi ensemble and the matrix appearing in the Johansen test. On one hand, the latter also deals with sample canonical correlations of two data sets. On the other hand, the data sets are no longer independent, instead one is obtained from another by a deterministic linear transformation. So are the roots of Eq. (7) related to the Jacobi ensemble? There is evidence in both directions. First, computer simulations quickly reveal that in one-dimensional case the distribution of the single eigenvalue in the Johansen test is *not* the Beta distribution. Yet, second, recent results of Onatski and Wang (2018, 2019) show that the Law of Large Numbers for the empirical distribution of the eigenvalues appearing in the Johansen test (with roots of Eq. (7) being a particular case) matches the one for the Jacobi ensemble (with shifted dimension parameters) in the limit as $N, T \rightarrow \infty$. Those articles were asking for an explanation.

4.2. Sketch of the proof of Theorem 2. To prove Theorem 2, we first need to establish the following central statement: a small perturbation of the Johansen test matrix, obtained by replacing the deterministic summation matrix in its definition by a random analogue, exactly matches the Jacobi ensemble. We further show that the perturbation vanishes in the limit as $N, T \rightarrow \infty$, thus, allowing us to obtain the asymptotics of the variants of the Johansen test from the known asymptotic results for the Jacobi ensemble.

THEOREM 4. *Suppose that $T, N \rightarrow \infty$ in such a way that $T > 2N$ and the ratio T/N remains bounded. Under the hypothesis $\Pi = 0$ for (2), one can couple (i.e. define on the same probability space) the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ of the matrix $S_{10}S_{00}^{-1}S_{01}S_{11}^{-1}$ and eigenvalues $x_1 \geq \dots \geq x_N$ of the Jacobi ensemble $\mathbf{J}(N; \frac{N}{2}, \frac{T-2N}{2})$ in such a way that for each $\epsilon > 0$ we have*

$$\lim_{T, N \rightarrow \infty} \text{Prob} \left(\max_{1 \leq i \leq N} |\lambda_i - x_i| < \frac{1}{N^{1-\epsilon}} \right) = 1.$$

The proof of Theorem 4 is based on the following idea. Looking at Eq. (2) when $\Pi = 0$, one can notice that matrices entering into the test and given by Eq. (10) can be expressed in terms of the matrix of data X and the lag operator mapping $X_t \rightarrow X_{t-1}$. A computation shows that since we deal with the de-trended and de-measured data, we can replace the lag operator with

its *cyclic version*, which maps X_1 to X_T rather than X_0 . The latter is an orthogonal operator whose eigenvalues are roots of unity of order T . Then, the idea is to replace this operator by *uniformly random* orthogonal operator. From that we proceed in two steps:

- (I) We show that when the lag operator is replaced by its random counterpart, the eigenvalues of $S_{10}S_{00}^{-1}S_{01}S_{11}^{-1}$ have distribution $\mathbf{J}(N; \frac{N}{2}, \frac{T-2N}{2})$. We remark that this is a new appearance of the Jacobi ensemble, which was not present in the literature before.
- (II) We show that replacement of the lag operator by its random counterpart introduces an error, which can be upper-bounded by $N^{\epsilon-1}$ for any $\epsilon > 0$. This part is based on the rigidity results from random matrix theory, which say that eigenvalues of a uniformly random orthogonal matrix can be very closely approximated by equally spaced roots of unity.

The full proof of Theorem 4 is given in [Bykhovskaya and Gorin \(2021, Supplementary Material\)](#). In addition to the real case, we also simultaneously prove a similar statement for complex matrices, encountering Jacobi ensemble of Hermitian matrices. Generally complex settings are rare guests in economics. Yet, they play a major role in the spectral analysis of time series data and in other areas, such as high energy physics.

By combining Theorem 4 with known asymptotic results for the Jacobi ensemble we can obtain the asymptotics of test statistic (8) in various regimes.

5. Monte Carlo simulations. In this section we illustrate the performance of our test via Monte Carlo simulations. We consider both size (rejection rate) and power.

5.1. Rejection rate. First, we compare the finite sample performance of our approach versus Johansen’s LR test and one of its corrected versions for (reasonably small) $T = 30$ and $N = 5, \dots, 10$. The finite sample correction takes the form $\frac{T-N}{T}$ and was suggested in [Reinsel and Ahn \(1992\)](#). (Let us note that there are several more advanced empirical finite sample correction procedures for Johansen’s LR test and its variations, we refer to [Onatski and Wang \(2019, Table 2\)](#) for a recent comparison of some of those for a wide range of values of N and T .) Table 2 summarizes the results (numbers closer to 5 mean better performance). The $LR_{N,T}$ column is our test and the last two columns are from [Gonzalo and Pitarakis \(1999\)](#). They correspond to the same null of no cointegration, but use a different from ours alternative hypothesis. Our alternative is at most r cointegrating relationships (in Table 2 we use $r = 1$, which means we look at the largest eigenvalue). The LR and RALR tests consider “at most N ” cointegrations as H_1 , which means that they use the sum of all eigenvalues. We can see that in finite samples when both N and T are of a similar magnitude, our approach significantly outperforms the alternatives based on small N , large T asymptotics.

N	$LR_{N,T}, r = 1$	LR	RALR
5	6.60	20.75	3.59
6	5.45	31.66	2.68
7	4.52	47.44	1.98
8	3.80	67.42	2.00
9	3.16	85.00	1.32
10	2.60	96.69	0.96

TABLE 2

Empirical size under no cointegration hypothesis (5% nominal level). Data generating process: $\Delta X_{it} = \varepsilon_{it}$, $\varepsilon_{it} \sim i.i.d. \mathcal{N}(0, 1)$, $T = 30$, $MC = 1,000,000$ replications for $LR_{N,T}$ and $MC = 10,000$ for LR and RALR.

To illustrate the performance of our test as the sample size increases, we fix the ratio $T/N \equiv c$ and plot the empirical size as a function of N . This is shown in Figure 2, where the target is 5% rejection rate. The picture suggests that the test has rejection rate close to 5%. The green solid line corresponding to $c = 4$ achieves 5% rejection rate very fast. Other three curves overshoot 5% by couple percents (e.g., both $c = 5$ and $c = 6$ are always below 7%). Moreover, the larger is c , the higher is the corresponding rejection curve. E.g., $c = 10$ curve (blue, short dash) is strictly above other curves.

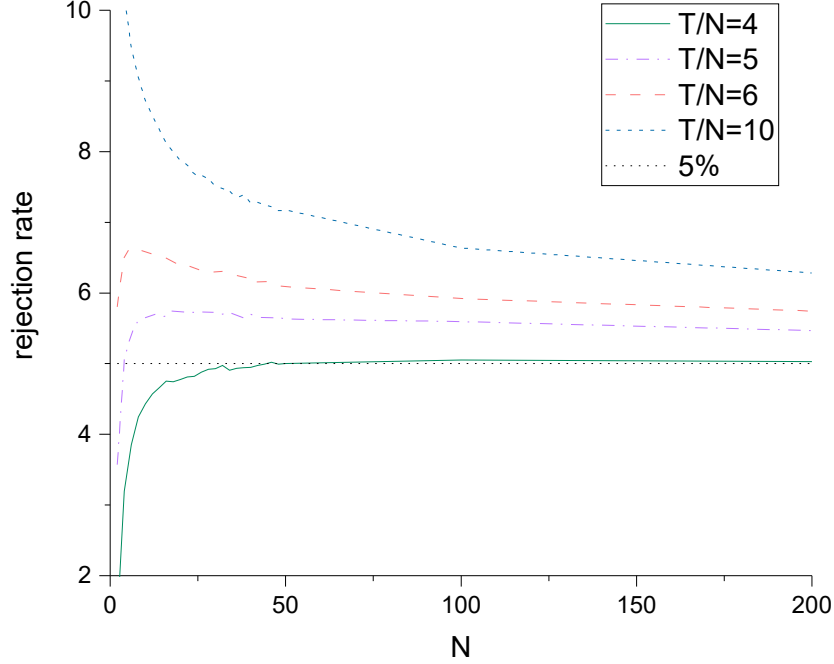


Figure 2: Empirical size when T/N is fixed and N increases.

To improve the finite-sample behavior for large c we suggest to ignore the $\frac{2}{N}$ correction in p and q in Theorem 2, Eq. (14), when T/N is large. That is, to use simplified formulas instead: $p = 2$, $q = T/N - 1$. We recalculate the empirical rejection rate under the simplified formulas for p, q in Figure 3. Under the simplified formulas, the larger is c , the smaller is over-rejection and the closer the curve is to 5% line. As can be seen by comparing Figures 2 and 3, for small c the formula with the correction leads to better rejection rates, while for large c it is the opposite, and we do not gain from finite sample correction. The conclusion is to use the simplified formula when $c = T/N$ is at least 6.

An important feature of Figures 2 and 3 is that as $N, T \rightarrow \infty$ with fixed T/N the results improve towards the perfect match to 5%. This is in contrast to various finite sample corrections of Johansen's LR test used earlier. In particular, examination of Table 2 in Onatski and Wang (2019) reveals that each procedure has its own pairs (N, T) where the results are close to perfect (empirical size of the test is very close to the desired 5%) and some others where the results are much worse, yet rejection rates in general do not improve as the sample size increases keeping T/N fixed.

5.2. Power. When one analyzes the power of a test, the question is which data generating process (dgp) to use under an alternative H_1 . In our setting the space of alternative dgps has growing with N dimension. Thus, there is no clear choice of the proxy alternative hypothesis

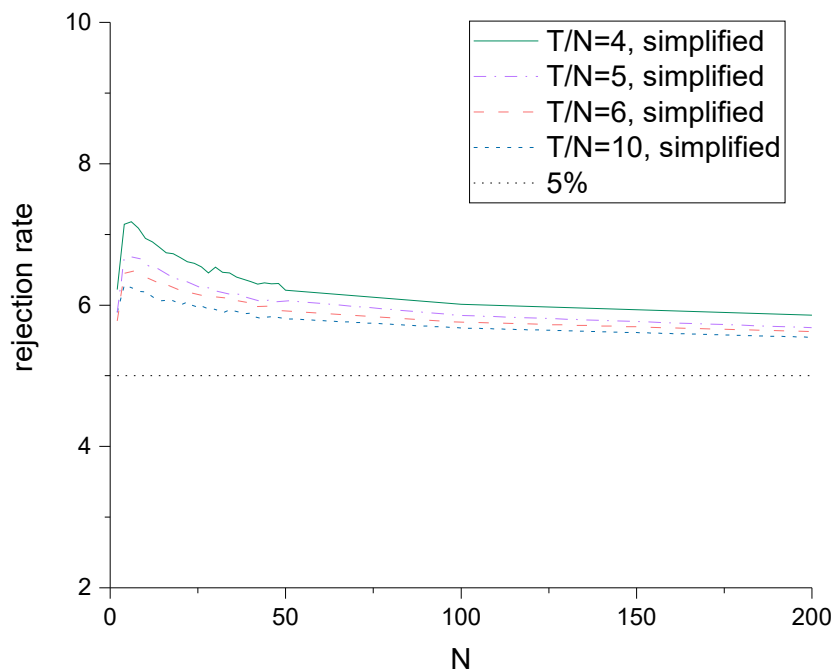


Figure 3: Empirical size when T/N is fixed and N increases. Test based on $p = 2$, $q = T/N - 1$.

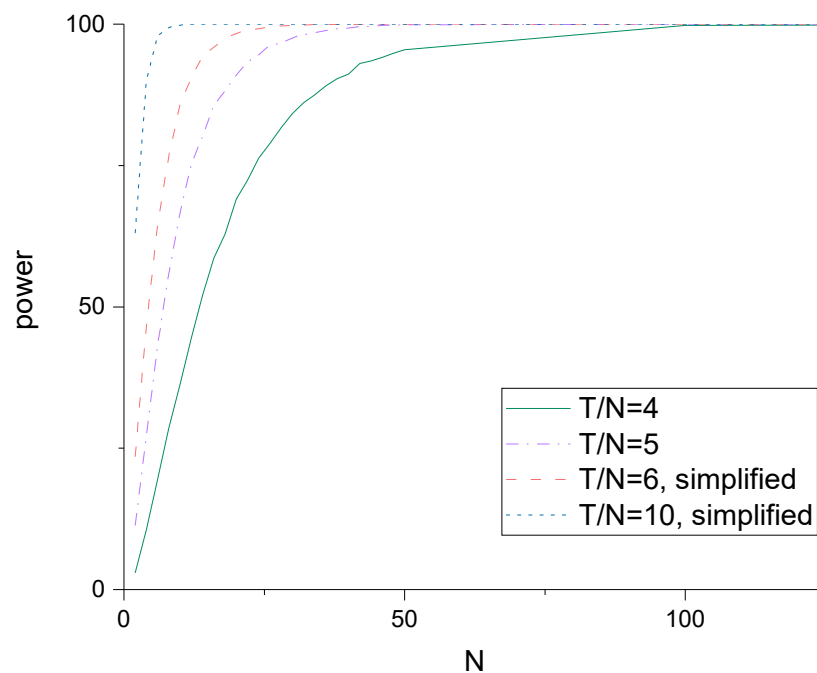


Figure 4: Power against random alternative of 1 cointegrating relationship when T/N is fixed and N increases.

to use in simulations. Hence, instead we proceed with various random alternatives.¹⁵ There-

¹⁵Another approach would be to stick to some ad hoc matrix Π . Yet, we do not expect that to affect our simulations in a qualitatively significant way.

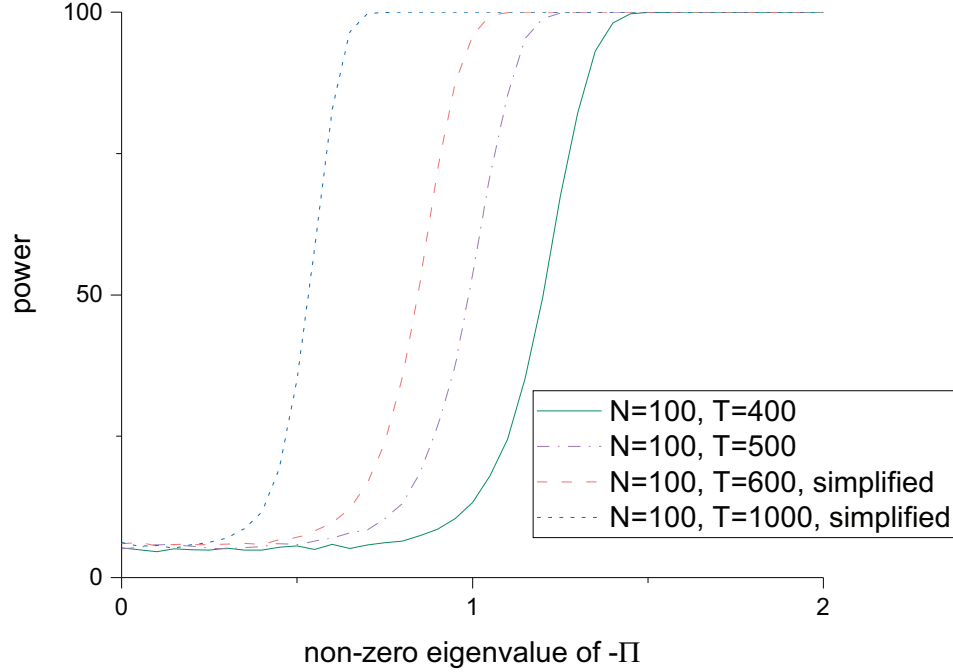


Figure 5: Power against random alternative of 1 cointegrating relationship when Π is symmetric. To ensure stationarity of ΔX a non-zero eigenvalue of $-\Pi$ lies between 0 and 2.

fore, the corresponding power is also random. For illustrational convenience, we only report averages of those powers.

To analyze the power of our test we conduct several experiments. In all of them the errors ε_{it} are generated as i.i.d. $\mathcal{N}(0, 1)$. In the first experiment we randomly sample a matrix Π of rank 1. We do this by generating two uniformly random N -dimensional unit vectors, u, v , such that the non-zero eigenvalue of uv^* is negative. Then we set $\Pi = uv^*$. By construction, Π has rank 1, singular value 1, and one eigenvalue between -1 and 0 (others are zero),¹⁶ so that X_t is non-stationary, while ΔX_t is stationary. The average power constructed from such random alternative is shown in Figure 4. As in the previous subsection we fix the ratio T/N and plot the average power as a function of N . Following our recommendation, we use simplified formulas for p and q when $T/N \geq 6$. We can see that the average power quickly reaches 100% for all ratios of T/N . The larger is that ratio, the faster we reach 100%. This is due to the fact that higher ratio means higher time span. This gives the process more chances to accumulate the effects of the presence of cointegration.

In the second experiment, we randomly generate a symmetric matrix Π of rank 1. We do this by generating a uniformly random N -dimensional unit vector v . Then we set $\Pi = -\lambda vv^*$, where λ goes from 0 to 2. The coefficient λ equals to the non-zero eigenvalue of $-\Pi$. The fact that it lies between 0 and 2 guarantees that X_t is non-stationary, while ΔX_t is stationary. Figure 5 shows the power as a function of λ for $N = 100$ and various values of T corresponding to $T/N = \{4, 5, 6, 10\}$ as in the previous experiments. The larger is $-\lambda$, the larger is power, which eventually reaches 100%. When $\lambda = 0$, the dgp corresponds to the null H_0 . Thus, all curves start from $\approx 5\%$, which corresponds to the empirical size of the test. We can see that again the larger is T/N , the faster we reach 100%. The reason is the same as in the previous simulation.

¹⁶As N goes to infinity the non-zero eigenvalue is of order $N^{-1/2}$.

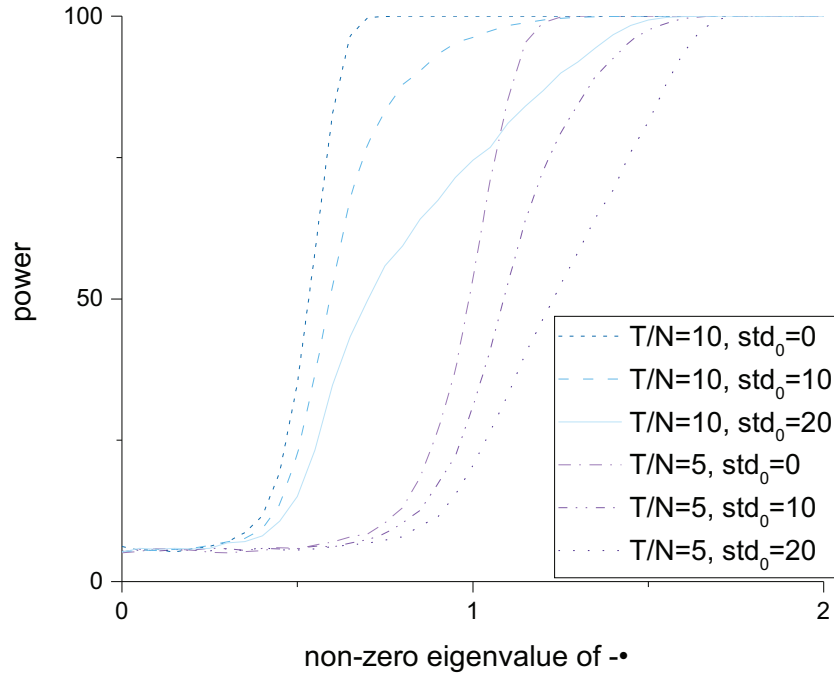


Figure 6: Power against random alternative of 1 cointegrating relationship when Π is symmetric. Initial condition is chosen as $X_0 = \text{std}_0 \cdot \mathcal{N}(0, 1)$, $N = 100$.

Drawing the intuition from the unit root testing literature (e.g., Müller and Elliott (2003), Harvey et al. (2009)), we also analyze the sensitivity of power to the choice of initial condition, X_0 .¹⁷ In two previous experiments we set $X_0 = 0$. Figure 6 shows how the power against random alternative of 1 cointegrating relationship for symmetric Π changes for various magnitudes of X_0 . To be more specific, we redo the same simulations as in the previous paragraph, but start with $X_{i0} \sim \text{i.i.d. } \text{std}_0 \cdot \mathcal{N}(0, 1)$ for each Monte Carlo draw. We consider $T/N = 5$ and $T/N = 10$. Curves with $\text{std}_0 = 0$ are the same as on Figure 5 (they are also represented with the same style and color on both pictures). We can see that the larger is the magnitude of X_0 , as measured by its standard deviation std_0 , the slower the power reaches 100%.

In contrast to the previous paragraph, the power against random alternative of 1 cointegrating relationship when Π is asymmetric (as in Figure 4) does not exhibit any substantial changes with respect to the magnitude of X_0 . Hence, we do not redraw the analogue of Figure 4 for non-zero X_0 .

Overall, the simulations suggests the good asymptotic performance of our test procedure both under H_0 and H_1 . The theoretical analysis of the power remains an open and challenging question.

6. Empirical illustration. In this section we illustrate our testing strategy by analyzing cointegration in log prices of various stocks. The search for cointegrations is a part of a stock market strategy called pairs trading, see, e.g., Krauss (2017) and references therein. For us this is a convenient testing ground, as both N and T are large.

We use logarithms of weekly S&P100 prices over ten years: 01.01.2010 – 01.01.2020, which gives us 522 observations across time. The time range is chosen so that we do not

¹⁷Note that X_0 does not affect rejection rates in Section 5.1, because under H_0 it cancels out.

need to worry about potential structural breaks due to financial crisis of 2007 – 2008 and due to COVID-19. S&P100 includes 101 (because one of its component companies, Google, has 2 classes of stock) leading U.S. stocks with exchange-listed options. We use 92 of those stocks (those which were available for the whole ten years span and only one of two Google's stocks). More details on those stocks are in Section 9.6 in [Bykhovskaya and Gorin \(2021, Supplementary Material\)](#). Therefore, $N = 92$, $T = 521$ and $T/N \approx 5.66$.

Figure 7 shows the histogram of eigenvalues which solve Eq. 7 for the S&P100 data. The key feature of this histogram is that it closely resembles the Wachter distribution, which density is shown by thick orange line in Figure 7. This distribution governs the asymptotics of the eigenvalues of the Jacobi ensemble (see Section 9.3 in [Bykhovskaya and Gorin \(2021, Supplementary Material\)](#) for the details). In particular, we see a precise match between supports of the histogram and of the Wachter distribution. The resemblance is in line with our Theorem 4. Indeed, if we believe that the true data-generating process (2) has $\Pi = 0$, then the theorem combined with the asymptotics of the Jacobi ensemble of Theorem 14 in [Bykhovskaya and Gorin \(2021, Supplementary Material\)](#) predicts convergence to the Wachter distribution. The conclusion is that Figure 7 shows a close match between theoretical predictions and real data. Simultaneously, this figure is consistent with H_0 of no cointegration.

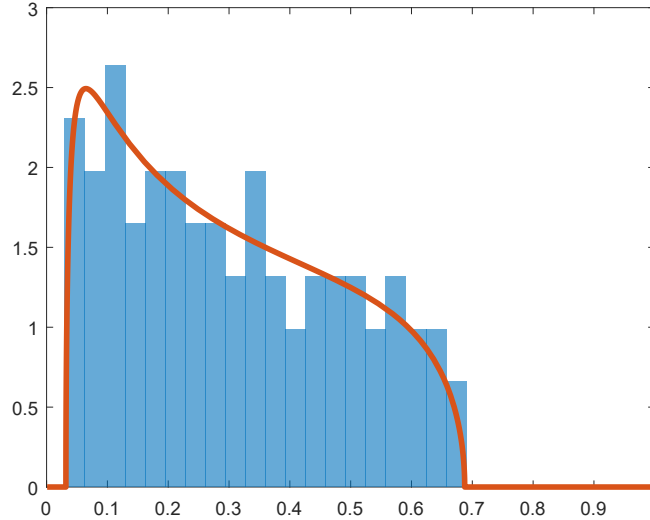


Figure 7: Eigenvalues from S&P data and Wachter distribution.

We compute our test statistic for $r = 1, 2, 3$ using the S&P100 data. In neither of the cases the value of the statistic is large enough for a statistically significant rejection of H_0 . If $r = 1$, then the value is -0.27 , which corresponds to approximately the 0.78 quantile of the asymptotic distribution shown in Figure 1. For $r = 2, 3$ the values are closer to the right tails of the distribution, but still below the (one-sided) 0.95 quantiles. Hence, we do not see evidence towards the presence of cointegration in S&P100 stock prices for the last 10 years.

7. Extensions. Let us describe possible extensions and modifications of our results. In Subsection 7.1 we consider non-Gaussian errors ε_t . Subsection 7.2 looks at the model without an intercept μ (linear trend in X_t) and considers the effect of de-trending vs. no de-trending in such setting. Subsection 7.3 investigates the performance of our test when the true process follows higher order of autoregression. Finally, in Subsection 7.4 we discuss testing the hypothesis $\Pi = -I_N$ using the same approach as for $\Pi = 0$.

7.1. Non-gaussian errors. The result of Theorem 2 is obtained under the assumption that the errors ε_t in Eq. (2) are Gaussian. However, we believe that it should be possible to remove this restriction and it is reasonable to expect that the very same statement should hold for any (independent and identical across time t) distribution of ε_t , as long as it has sufficiently many moments. The underlying reason for this belief is the so-called universality phenomenon in the random matrix theory: asymptotic local spectral characteristics of a random matrix are almost independent of the distributions of the matrix elements, see, e.g., Erdos and Yau (2012), Tao and Vu (2012) for general reviews and Han et al. (2016), Han et al. (2018) for the recent work in contexts of multivariate analysis of variance and canonical correlations. In particular, for the Wigner matrices ($Y + Y^*$, where Y is a square matrix with real i.i.d. entries), it is known that the asymptotic behavior of the largest eigenvalues depends only on the first two moments (expectation and variance) of the distribution of an individual matrix element. Note, however, that our asymptotic result in Theorem 2 holds for any choice of the first two moment of Gaussian noise ε_t : in Eq. (2) the covariance matrix Λ is arbitrary and any shift in expectation can be absorbed into the parameter μ . Hence, we conjecture that the result of Theorem 2 would hold for any distribution of ε_t , as long as it is sufficiently well-behaved.

In order to test this conjecture we made simulations for the case when elements of ε_t are non-gaussian, but i.i.d. across both i and t (corresponding to a diagonal covariance matrix Λ). We ran Monte-Carlo simulations for three different distributions for ε_{it} : uniform on the interval $[0, 1]$, uniform on 3 points $\{1, 2, 3\}$, and the product of two independent $\mathcal{N}(0, 1)$ random variables. In each case for $T = 900$, $N = 300$, and small values of r , we do not see any significant changes in the distribution of $\sum_{i=1}^r \ln(1 - \lambda_i)$ from the limit in Theorem 2. However, things go differently when the distribution has heavy tails. In the forth experiment we chose ε_{it} to be Cauchy-distributed, and then the distribution of $\sum_{i=1}^r \ln(1 - \lambda_i)$ dramatically changed from what we saw in the Gaussian case. Hence, we conclude, that the existence of at least some number of moments of the errors $\{\varepsilon_t\}$ should be necessary for the validity of the conjecture. Rigorous proof of the conjecture remains an important problem for the future research.

7.2. Model without trend. One of the important steps in our testing procedure is de-trending the data. Moreover, the exact form of the de-trending that we use (we subtract the slope of the line which connects X_0 at $t = 0$ and X_T at $t = T$) is an ingredient substantially used in the proof of Theorem 2. Yet, we expect that this is just a technical artifact and statements similar to Theorem 2 should also be true with other forms of de-trending or in models where this step is not needed at all. Let us provide some evidence.

Our model allows for any linear trend and works even if the true value of μ is zero. Yet, when one has such prior knowledge it may seem natural to ignore the de-trending and de-meaning steps (Steps 1 and 2 in Section 2.2) and proceed without them. In this section we compare tests with and without de-trending for the model without μ :

$$(21) \quad \Delta X_t = \Pi X_{t-1} + \varepsilon_t, \quad t = 1, \dots, T.$$

If both de-trending and de-meaning steps are omitted, we get (the small r version of) the classical Johansen test statistic for the model (21). If only de-trending is omitted, we get the Johansen test statistic corresponding to our original model (2). When both de-trending and de-meaning are implemented, we get our procedure described in Section 2.2. To compare the asymptotic properties of those procedures, we perform a Monte Carlo simulation. Results are reported in Figure 8.

As Figure 8 suggests, the densities have almost identical shape. Figure 9 plots all three of them together as well as their versions with subtracted mean. As we can see, after subtracting the mean, all three densities are identical. This suggests robustness of the Airy₁ point process

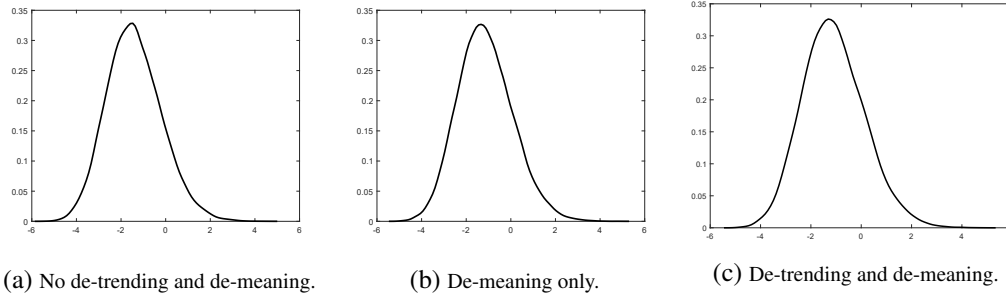


Figure 8: Asymptotic distribution of the rescaled $\ln(1 - \lambda_1)$ under various testing procedures (Eq. (12) with $r = 1$). Data generating process: $\Delta X_{it} = \varepsilon_{it}$, $\varepsilon_{it} \sim \text{i.i.d. } \mathcal{N}(0, 1)$, $T = 500$, $N = 100$, $MC = 50,000$ replications.

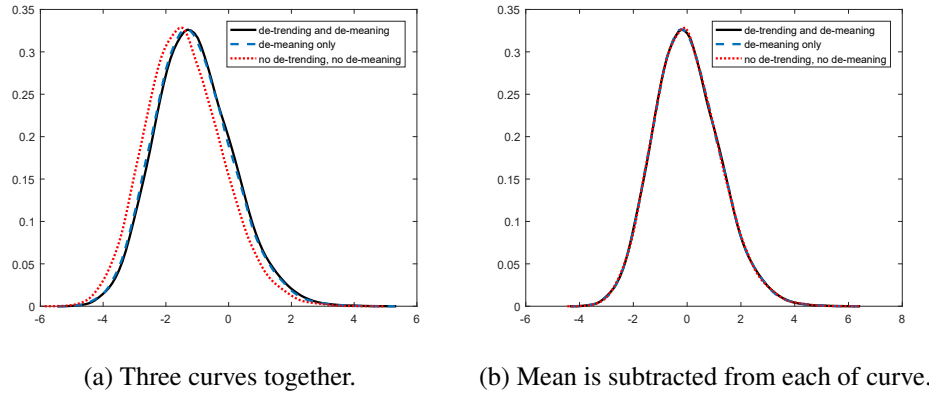


Figure 9: Asymptotic distribution of the rescaled $\ln(1 - \lambda_1)$ under various testing procedures with means normalized to zero (Eq. (12) with $r = 1$). Data generating process: $\Delta X_{it} = \varepsilon_{it}$, $\varepsilon_{it} \sim \text{i.i.d. } \mathcal{N}(0, 1)$, $T = 500$, $N = 100$, $MC = 50,000$ replications.

in our asymptotic results of Theorem 2 (which correspond to Figure 8c) and predicts that some modifications, which preserve the limiting distribution, are possible.

Let us emphasize that our limit theorems currently only apply to Figure 8c, but not to the settings of Figures 8a and 8b. The mismatch of the means in Figure 9a makes one suspect that some modifications of the constants in the asymptotic theorems are needed as soon as we start slightly adjusting the setting.

7.3. Higher order of VAR. In this subsection we discuss the performance of our test when the true data generating process (dgp) is $\text{VAR}(k)$, $k > 1$. That is

$$(22) \quad \Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-k} + \mu + \varepsilon_t, \quad t = 1, \dots, T,$$

and the no cointegration situation corresponds to $\Pi \equiv 0$.

Even if $k > 1$, we expect to see the Tracy-Widom distribution and marginals of Airy₁ process (as in Theorem 2) in the asymptotic behavior of the squared sample canonical correlations from Section 2.2 under mild restrictions on Γ_i . The belief is based on the universality intuition of the random matrix theory. However, we do not expect the scalings (such as the

coefficients c_1 and c_2 in Theorem 2) to remain the same. The most plain analogy is the dependence of centering and scaling in the classical Central Limit Theorem on the underlying process. Closer to our context is the asymptotic behavior of sample covariance matrices: when the data is i.i.d., the empirical distribution of the eigenvalues of the sample covariance matrix converges to the Marchenko-Pastur law (and the largest eigenvalues concentrate near the right edge of this distribution), while data with general covariance structure leads to much richer limits, see [Bai and Silverstein \(2010, Chapter 4\)](#) and references therein. Figuring out (even heuristically) any formulas for the scaling coefficients c_1 and c_2 as functions of Γ_i is a challenging open problem for the future research.

There is an important family of cases where one can hope that the formulas for c_1 and c_2 from Theorem 2 remain valid (perhaps, with minor modifications). This is when Γ_i are small and evolution of X_t given by (22) can be treated as a small perturbation of the VAR(1) process. One way to formalize the “smallness” of Γ_i is by requiring them to be of small rank (cf. discussion of rank in Section 3.2) and of small norm.

In order to investigate the above conjecture, we run Monte Carlo simulations for $k = 2$ and Γ_1 of rank 1. We consider the null $\Pi \equiv 0$ and calculate our test statistic based on $r = 1$ under VAR(2) and VAR(1) data generating processes and compare their asymptotic distribution. Under VAR(1) $\Gamma_1 \equiv 0$, while for VAR(2) we use $\Gamma_1 = 0.5E_{11}$ and $\Gamma_1 = 0.5E_{12}$, where E_{ij} is a matrix which has 1 on the intersection of row i and column j and 0 everywhere else; the components of the noise ε_{it} are i.i.d. $\mathcal{N}(0, 1)$. The asymptotic distribution of our test statistic with $r = 1$ for various dgps is shown in Figure 10. We see that the distributions on each panel of Figure 10 are close to each other. Hence, testing based on Theorem 2 in such a VAR(2) setting remains valid. Yet, if we consider more general situation of $\Gamma_1 = \theta E_{11}$ or $\Gamma_1 = \theta E_{12}$, then we observe in simulations (not shown) that the quality of approximations significantly deteriorates as θ grows to 1.

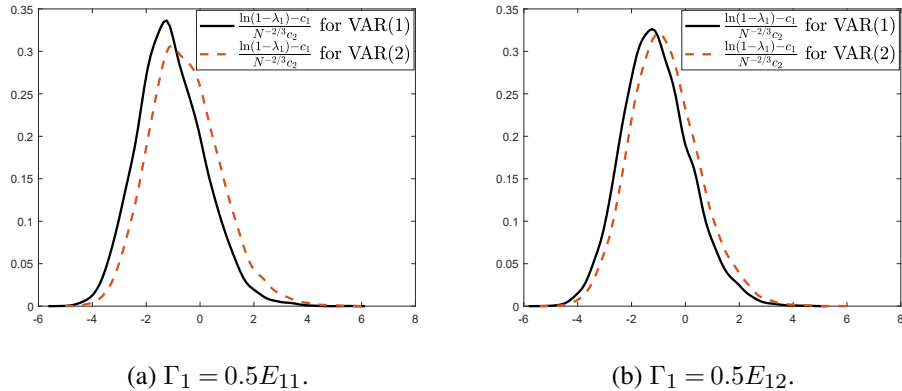


Figure 10: Asymptotic distribution of the rescaled $\ln(1 - \lambda_1)$ (Eq. (12) with $r = 1$) under VAR(1) with $\Pi = 0$ and VAR(2) with $\Pi = 0$ and Γ_1 of rank 1.

Data generating process: $\varepsilon_{it} \sim \text{i.i.d. } \mathcal{N}(0, 1)$, $T = 500$, $N = 100$, $MC = 10,000$ replications.

7.4. Testing for white noise hypothesis in VAR(1) setting. The main result of Theorem 2 is a development of a test for the hypothesis $\Pi = 0$ in the VAR(1) model (2). One could also try to understand for which other Π can the testing be possible. The asymptotic distribution depends on the choice of Π , and it is not possible to estimate Π consistently in our regime of T and N growing to infinity proportionally. Thus, for general Π the problem seems infeasible

at this point. However, there is another particular choice of Π for which an approach very similar to Theorem 4 still works: $\Pi = -I_N$. Denoting this hypothesis $H_0^{w.n.}$, where $w.n.$ stays for the white noise, the data generating process (Eq. (2)) becomes:

$$(23) \quad H_0^{w.n.} : \quad X_t = \mu + \varepsilon_t, \quad t = 1, \dots, T.$$

In other words, we are now testing the hypothesis that the time series X_t is independent across time t against various VAR(1) alternatives. Here is one setup where such testing can be relevant. Suppose that we want to forecast some variable Y and we chose some model for it. After estimating parameters of the model we obtain residuals X_t . If we know that X_t are independent, then they are unforecastable and we cannot further improve our forecasting model. To check the above we can take the residuals X_t and then apply our white noise hypothesis testing procedure.

Let us introduce an adaptation of the Johansen statistic to $H_0^{w.n.}$. As in Eq. (9), we use the notation $\mathcal{P}$ for the de-meaning operator projecting on the hyperplane orthogonal to $(1, 1, \dots, 1)$.

Following Johansen (1988, 1991) we are going to use the de-meaned data $X_t \mathcal{P}$. For the increments ΔX_t in addition to the conventional de-meaning we use an extra modification: we deal with *cyclic increments* $\Delta^c X_t$ defined as:

$$\Delta^c X_t = \begin{cases} X_{t+1} - X_t, & t = 1, 2, \dots, T-1, \\ X_1 - X_T, & t = T. \end{cases}$$

While it might seem bizarre to subtract the last observation from the first one, if we recall that our current hypothesis of interest (23) ignores the time ordering, then this becomes less controversial. Note also a shift of index by 1, as compared to the conventional ΔX_t , which is compensated by the lack of shift $t \rightarrow t-1$ in X_t , as compared to Eq. (4). Our choice of definition of Δ^c is important for the following precise asymptotic results. As in Section 2, the conventional Johansen statistic should be thought of as a finite rank perturbation of the modified version that we now introduce.

$$(24) \quad S_{00}^{w.n.} = \Delta^c X \mathcal{P} (\Delta^c X)^*, \quad S_{01}^{w.n.} = \Delta^c X \mathcal{P} X^*, \quad S_{10}^{w.n.} = X \mathcal{P} (\Delta^c X)^*, \quad S_{11}^{w.n.} = X \mathcal{P} X^*,$$

We further define N numbers $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ as N roots to the equation

$$(25) \quad \det(S_{10}^{w.n.} (S_{00}^{w.n.})^{-1} S_{01}^{w.n.} - \lambda S_{11}^{w.n.}) = 0.$$

Equivalently, $\{\lambda_i\}$ are eigenvalues of $S_{10}^{w.n.} (S_{00}^{w.n.})^{-1} S_{01}^{w.n.} (S_{11}^{w.n.})^{-1}$.

THEOREM 5. *Suppose that $T, N \rightarrow \infty$ in such a way that $T > 2N$ and the ratio T/N remains bounded. Under the hypothesis $H_0^{w.n.}$ one can couple (i.e. define on the same probability space) the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ of the matrix $S_{10}^{w.n.} (S_{00}^{w.n.})^{-1} S_{01}^{w.n.} (S_{11}^{w.n.})^{-1}$ and eigenvalues $x_1 \geq \dots \geq x_N$ of Jacobi ensemble $\mathbf{J}(N; \frac{T-N-1}{2}, \frac{T-2N}{2})$ in such a way that for each $\epsilon > 0$ we have*

$$\lim_{T, N \rightarrow \infty} \text{Prob} \left(\max_{1 \leq i \leq N} |\lambda_i - x_i| < \frac{1}{N^{1-\epsilon}} \right) = 1.$$

The proof of Theorem 5 follows a similar strategy as Theorem 4 and we refer to Section 9.5 in Bykhovskaya and Gorin (2021, Supplementary Material) for details; in particular, the proofs rely on yet another novel appearance of the Jacobi ensemble.

The remaining straightforward step to obtain the asymptotics of various statistics built on the eigenvalues $\{\lambda_i\}$ is to combine Theorem 5 with asymptotic results for the Jacobi ensemble presented in Section 9.3 in Bykhovskaya and Gorin (2021, Supplementary Material). This is in the spirit of Theorem 2.

Note that the hypothesis $H_0^{w.n.}$ implies the maximal amount of cointegrating relationships: each of the N components of X_t is already stationary. A reasonable alternative hypothesis H_1 is the presence of $N - r$ cointegrating relationships. For simplicity, let us concentrate on the case $r = 1$. Then the alternative can be also interpreted as a presence of a single growing factor. In this situation we expect the *smallest* eigenvalue λ_N to be a good test statistic. We see in numerical simulations that λ_N is bounded away from 0 under $H_0^{w.n.}$. It can also be formally proved by combining Theorem 5 with asymptotics of the Jacobi ensemble from Section 9.3 in [Bykhovskaya and Gorin \(2021, Supplementary Material\)](#). Thus, if λ_N is close to 0, then we are able to reject $H_0^{w.n.}$. The same simulations indicate that λ_N starts to be close to 0 when there are at most $N - 1$ cointegrating relationships. Hence, the test based on λ_N should have a good asymptotic power. We leave rigorous justifications of this observation till future research, and for now only mention the following heuristics: the stationary linear combinations of X_t are strongly correlated with the same linear combinations of $\Delta^c X_t$; on the other hand, the growing linear combinations of X_t have very weak correlation with the same linear combinations of $\Delta^c X_t$ (cf. correlations of a one dimensional random walk with its increments). Hence, if the latter are present, the smallest canonical correlations of $X\mathcal{P}$ and $\Delta^c X\mathcal{P}$, which coincide with the eigenvalues of the matrix $S_{10}^{w.n.}(S_{00}^{w.n.})^{-1}S_{01}^{w.n.}(S_{11}^{w.n.})^{-1}$, should become small leading to close to 0 value of λ_N .

8. Conclusion. The paper presents a cointegration test which has desirable empirical size when N and T are of the same magnitude. To our knowledge, this is a first paper which constructs and analyzes asymptotic properties of a test that does not suffer from significant distortions (such as over-rejection) for comparable N and T . The test is based on the Johansen LR test and incorporates some additional steps. First, our procedure reinforces the importance of de-trending in cointegration testing. It turns out, that de-trending is crucial for deriving desirable asymptotic properties. (E.g., only after de-trending one can rewrite the lagged process as a linear function of its first differences.) Second, our asymptotic results reveal and explain an unexpected connection between the Johansen cointegration test and the Jacobi ensemble — a classical ensemble of the random matrix theory whose previous appearances in statistics include multivariate analysis of variance (MANOVA) and sample canonical correlations for independent sets of data.

On the theoretical side the next step would be to go from null hypothesis of zero cointegration to analyzing the behaviour of our test under r cointegrations. This will allow us to calculate the power of the test, reinforcing our simulational findings in Section 5.2, as well as to perform tests of r versus $r + 1$ cointegrations.

On the empirical side it would be interesting to apply our test to other data sets beyond what is presented in Section 6. Annual cross-country data provides a natural example of our setting where the number of years and countries is comparable. Another example arises if one considers network-type settings which evolve over time (e.g., as in [Bykhovskaya \(2021\)](#)). Data on trade or on foreign direct investment can potentially be non-stationary, especially if we focus on largest and the most active countries. Moreover, although such monthly data is available, for many countries it only covers 20 years. Thus, we have $T \approx 200$. If we look at directed pairs across 10 largest countries, this gives us $N = 90$ cross-section units, which fits ideally in our setting. Classical cointegration tests are known to perform poorly in the above settings. However, the asymptotic results of our paper open up a possibility of detecting the presence of cointegration in such time-series data.

SUPPLEMENTARY MATERIAL

See [Bykhovskaya and Gorin \(2021, Supplementary Material\)](#) for proofs and data description.

Acknowledgements. The authors are grateful to Donald Andrews, Alexei Borodin, Giuseppe Cavaliere, Alice Guionnet, Bruce Hansen, Søren Johansen, Grigori Olshanski, and anonymous referees for valuable suggestions. V.G. acknowledges support from the NSF Grants DMS-1664619 and DMS-1949820.

REFERENCES

- Bai, J. and S. Ng**, “Large Dimensional Factor Analysis,” *Foundations and Trends in Econometrics*, 2008, 3 (2), 890163.
- Bai, Z. and J.W. Silverstein**, *Spectral analysis of large dimensional random matrices*, Vol. 20, New York: Springer, 2010.
- Baik, J., G. Ben Arous, and S. Péché**, “Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices,” *The Annals of Probability*, 2005, 33 (5), 1643–1697.
- Bao, Z., G. Pan, and W. Zhou**, “Central limit theorem for partial linear eigenvalue statistics of Wigner matrices,” *Journal of Statistical Physics*, 2013, 150, 88–129.
- Bao, Z., J. Hu, G. Pan, and W. Zhou**, “Canonical correlation coefficients of high-dimensional Gaussian vectors: finite rank case,” *The Annals of Statistics*, 2019, 47 (1), 612–640.
- Ben Arous, G. and A. Guionnet**, “Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy,” *Probability theory and related fields*, 1997, 108 (4), 517–542.
- Borodin, A. and V. Gorin**, “General β -Jacobi Corners Process and the Gaussian Free Field,” *Communications on Pure and Applied Mathematics*, 2015, 68 (10), 1774–1844.
- Borodin, A. and V. Gorin**, “Lectures on integrable probability,” in “Probability and statistical physics in St. Petersburg,” Vol. 91, American Mathematical Soc., 2016, pp. 155–214.
- Breitung, J. and M.H. Pesaran**, “Unit Roots and Cointegration in Panels,” in “Mátyás L., Sevestre P. (eds) The Econometrics of Panel Data. Advanced Studies in Theoretical and Applied Econometrics, vol. 46,” Berlin, Heidelberg: Springer, 2008, pp. 279–322.
- Bykhovskaya, A.**, “Time Series Approach to the Evolution of Networks: Prediction and Estimation,” *Journal of Business & Economic Statistics*, 2021.
- Bykhovskaya, A. and V. Gorin**, “Supplement to ‘Cointegration in large VARs’,” 2021.
- Cavaliere, G., A. Rahbek, and A.R. Taylor**, “Bootstrap determination of the co-integration rank in vector autoregressive models,” *Econometrica*, 2012, 80 (4), 1721–1740.
- Dumitriu, I. and A. Edelman**, “Matrix models for beta ensembles,” *Journal of Mathematical Physics*, 2002, 43 (11), 5830–5847.
- Elliott, G., T.J. Rothenberg, and J.H. Stock**, “Efficient tests for an autoregressive unit root,” *Econometrica*, 1996, 64 (4), 813–836.
- Engle, R.F. and C.W. Granger**, “Co-integration and error correction: representation, estimation, and testing,” *Econometrica*, 1987, 55 (2), 251–276.
- Erdos, L. and H.T. Yau**, “Universality of local spectral statistics of random matrices,” *Bulletin of the American Mathematical Society*, 2012, 49 (3), 377–414.
- Forrester, P.J.**, “The spectrum edge of random matrix ensembles,” *Nuclear Physics B*, 1993, 402 (3), 709–728.
- Forrester, P.J.**, *Log-gases and random matrices*, Princeton University Press, 2010.
- Gonzalo, J. and J.Y. Pitarakis**, “Dimensionality effect in cointegration analysis,” in “Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive WJ Granger,” Oxford: Oxford University Press, 1999, chapter 9, pp. 212–229.
- Granger, C.W.**, “Some properties of time series data and their use in econometric model specification,” *Journal of Econometrics*, 1981, 16 (1), 121–130.
- Guionnet, A. and J. Novak**, “Asymptotics of unitary multimatrix models: The Schwinger–Dyson lattice and topological recursion,” *The Schwinger–Dyson lattice and topological recursion*, 2015, 268 (10), 2851–2905.
- Han, X., G. M. Pan, and B. Zhang**, “The Tracy–Widom law for the Largest Eigenvalue of F Type Matrix,” *The Annals of Statistics*, 2016, 44 (4), 1564–1592.
- Han, X., G.M. Pan, and Q. Yang**, “A unified matrix model including both CCA and F matrices in multivariate analysis: the largest eigenvalue and its applications,” *Bernoulli*, 2018, 24 (4B), 3447–3468.
- Harvey, D.I., S.J. Leybourne, and A.R. Taylor**, “Unit root testing in practice: dealing with uncertainty over the trend and initial condition,” *Econometric theory*, 2009, 25 (3), 587–636.
- Ho, M.S. and B.E. Sørensen**, “Finding cointegration rank in high dimensional systems using the Johansen test: an illustration using data based Monte Carlo simulations,” *The Review of Economics and Statistics*, 1996, 78 (4), 726–732.
- Hua, L.**, *Harmonic analysis of functions of several complex variables in the classical domains*, Vol. 6, American Mathematical Soc., 1963.

- Johansen, S.**, “Statistical Analysis of Cointegrating Vectors,” *Journal of Economic Dynamics and Control*, 1988, 12 (2–3), 231–254.
- Johansen, S.**, “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models,” *Econometrica*, 1991, 59, 1551–1580.
- Johansen, S.**, *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford University Press, 1995.
- Johansen, S.**, “A small sample correction for the test of cointegrating rank in the vector autoregressive model,” *Econometrics*, 2002, 70 (5), 1929–1961.
- Johansson, K.**, “On fluctuations of eigenvalues of random Hermitian matrices,” *Duke mathematical journal*, 1998, 91 (1), 151–204.
- Johnstone, I.**, “Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence,” *The Annals of Statistics*, 2008, 36 (6), 2638–2716.
- Johnstone, I.**, “On the distribution of the largest eigenvalue in principal components analysis,” *The Annals of Statistics*, 2001, 29 (2), 295–327.
- Johnstone, I. and A. Onatski**, “Testing in high-dimensional spiked models,” *The Annals of Statistics*, 2020, 48 (3), 1231–1254.
- Kadell, K.W.**, “The Selberg–Jack symmetric functions,” *Advances in mathematics*, 1997, 130 (1), 33–102.
- Killip, R. and I. Nenciu**, “Matrix models for circular ensembles,” *International Mathematics Research Notices*, 2004, 2004 (50), 2665–2701.
- Krauss, C.**, “Statistical arbitrage pairs trading strategies: Review and outlook,” *Journal of Economic Surveys*, 2017, 31 (2), 513–545.
- Lütkepohl, H., P. Saikkonen, and C. Trenkler**, “Maximum eigenvalue versus trace tests for the cointegrating rank of a VAR process,” *The Econometrics Journal*, 2001, 4 (2), 287–310.
- Maddala, G.S. and In-Moo Kim**, *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, 1998.
- Mehta, M.L.**, *Random matrices*, Elsevier, 2004.
- Mingo, J.A. and M. Popa**, “Real second order freeness and Haar orthogonal matrices,” *Journal of Mathematical Physics*, 2013, 54 (5), 051701.
- Muirhead, R.J.**, *Aspects of multivariate statistical theory*, John Wiley & Sons, 2009.
- Müller, U.K. and G. Elliott**, “Tests for unit roots and the initial condition,” *Econometrica*, 2003, 71 (4), 1269–1286.
- Onatski, A. and C. Wang**, “Alternative asymptotics for cointegration tests in large vars,” *Econometrica*, 2018, 86 (4), 1465–1478.
- Onatski, A. and C. Wang**, “Extreme canonical correlations and high-dimensional cointegration analysis,” *Journal of Econometrics*, 2019.
- O’Rourke, S.**, “Gaussian Fluctuations of Eigenvalues in Wigner Random Matrices,” *Journal of Statistical Physics*, 2010, 138 (6), 1045–1066.
- Paruolo, P.**, “The power of lambda max,” *Oxford Bulletin of Economics and Statistics*, 2001, 63 (3), 395–403.
- Phillips, P.C.B. and S. Ouliaris**, “Asymptotic properties of residual based tests for cointegration,” *Econometrica*, 1990, 58 (1), 165–193.
- Ramirez, J., B. Rider, and B. Virág**, “Beta ensembles, stochastic Airy spectrum, and a diffusion,” *Journal of the American Mathematical Society*, 2011, 24 (4), 919–944.
- Reinsel, G.C. and S.K. Ahn**, “Vector autoregressive models with unit roots and reduced rank structure: Estimation. Likelihood ratio test, and forecasting,” *Journal of time series analysis*, 1992, 13 (4), 353–375.
- Sodin, S.**, “Several applications of the moment method in random matrix theory,” *Proceedings of the International Congress of Mathematicians*, 2014.
- Stock, J.H. and M.W. Watson**, “Testing for common trends,” *Journal of the American statistical Association*, 1988, 83 (404), 1097–1107.
- Swensen, A.R.**, “Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models,” *Econometrica*, 2006, 74 (6), 1699–1714.
- Tao, T. and V. Vu**, “Random matrices: the universality phenomenon for Wigner ensembles,” *Modern aspects of random matrix theory*, 2012, 72, 121–172.
- Tracy, C.A. and H. Widom**, “On orthogonal and symplectic matrix ensembles,” *Communications in Mathematical Physics*, 1996, 177 (3), 727–754.
- Xiao, Z. and P.C.B. Phillips**, “Efficient detrending in cointegrating regression,” *Econometric Theory*, 1999, 15 (4), 519–548.
- Zhang, B., G. M. Pan, and J. T. Gao**, “CLT for largest eigenvalues and unit root testing for high-dimensional nonstationary time series,” *The Annals of Statistics*, 2018, 46 (5), 2186–2215.