COINTEGRATION IN LARGE VARS

ANNA BYKHOVSKAYA AND VADIM GORIN

ABSTRACT. 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 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 at most $r > \rho$ 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 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%.

---

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

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

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

where 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.

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

---

[2]Accompanying paper Bykhovskaya and Gorin (2021) investigates 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.

---

[3]We 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.

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 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 integrals[4] leading to the identity of the law of the perturbed matrix with the Jacobi ensemble.

Put it otherwise, we discover an exactly-solvable[5] 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 VAR(k) context can be understood through (other) instances of the Jacobi ensemble. Justifications of this point of view are contained in Sections 7.2 ,7.3, and 9.5, and in Bykhovskaya and Gorin (2021).

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 the Appendix.

---

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

[5]A 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. Setting

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

$$(2) \qquad \Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-k} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where $\Gamma_1, \ldots, \Gamma_{k-1}$, Π, and 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 $X_{1-k}, \ldots, X_0$.

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 rank(Π) $= r_0$ 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.

In this paper we focus on VAR(1) setting:

$$(3) \qquad \Delta X_t = \Pi X_{t-1} + \mu + \varepsilon_t, \qquad t = 1, \ldots, T.$$

We discuss extensions to VAR(k), $k > 1$ in the accompanying paper Bykhovskaya and Gorin (2021).

2.2. Cointegration test. We test the null hypothesis H_0 of no cointegration, which is equivalent to rank(Π) $= 0$ or $\Pi \equiv 0$. We set rank(Π) $\leq r$, 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 rank(Π) $\equiv r_0$ is unknown one can use any r in order to reject H_0.[6] Formally, we test

$$(4) \qquad\qquad H_0 : \Pi \equiv 0 \quad \text{vs.} \quad H_1 : \text{rank}(\Pi) \leq r,$$

---

[6]In 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.

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-trend[8] the data and define

$$(5) \qquad \tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0).$$

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

$$(6) \qquad R_{0t} = \tilde{X}_t - \frac{1}{T}\sum_{\tau=1}^{T}\tilde{X}_\tau, \quad R_{1t} = \Delta X_t - \frac{1}{T}\sum_{\tau=1}^{T}\Delta X_\tau.$$

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

$$(7) \qquad S_{ij} = \sum_{t=1}^{T} R_{it}R_{jt}^*, \quad i,j = 0,1,$$

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

$$(8) \qquad \det(S_{10}S_{00}^{-1}S_{01} - \lambda S_{11}) = 0.$$

Step 4. Form the test statistic

$$(9) \qquad LR_{N,T}(r) = \sum_{i=1}^{r} \ln(1 - \lambda_i).$$

The subscript N, T in (9) 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).

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

---

[7]Johansen LR test is based on the maximization of the Gaussian likelihood.
[8]We discuss the role of de-trending in Section 7.2.

operator in a T–dimensional space defined by its matrix

$$(10) \qquad \mathcal{P} = \begin{pmatrix} 1 - \frac{1}{T} & -\frac{1}{T} & \cdots & -\frac{1}{T} \\ -\frac{1}{T} & 1 - \frac{1}{T} & \cdots & -\frac{1}{T} \\ & & \ddots & \\ -\frac{1}{T} & -\frac{1}{T} & \cdots & 1 - \frac{1}{T} \end{pmatrix}.$$

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

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

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

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

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

$$\mathfrak{a}_1 > \mathfrak{a}_2 > \mathfrak{a}_3 > \ldots$$

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 $\mu_{1;N} \geq \mu_{2;N} \geq \cdots \geq \mu_{N;N}$ be eigenvalues of $\frac{1}{2}(Y_N + Y_N^*)$. Then in the sense of convergence of finite-dimensional distributions*

$$(12) \qquad \lim_{N\to\infty}\left\{N^{1/6}\left(\mu_{i;N} - 2\sqrt{N}\right)\right\}_{i=1}^{N} = \{\mathfrak{a}_i\}_{i=1}^{\infty}.$$

The law of the first coordinate $\mathfrak{a}_1$ is known as the Tracy-Widom F_1 distribution; its distribution function can be written in terms of a solution of the Painleve II differential equation.

We remark that from the computational point of view (12) gives an efficient way to access the distribution of various functions of $\{\mathfrak{a}_i\}_{i=1}^{\infty}$.[9] From the theoretical perspective, one would

---

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

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]

Theorem 2. *Suppose that $T, N \to \infty$ 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 (3) with $\Pi = 0$. Then for each finite $r = 1, 2, \ldots$, we have convergence in distribution for the largest eigenvalues defined in Eq. (8):*

$$(13) \qquad \frac{\sum_{i=1}^{r} \ln(1 - \lambda_i) - r \cdot c_1(N, T)}{N^{-2/3} c_2(N, T)} \xrightarrow[T,N \to \infty]{d} \sum_{i=1}^{r} \mathfrak{a}_i,$$

where

$$(14) \quad c_1(N, T) = \ln(1 - \lambda_+), \qquad c_2(N, T) = -\frac{2^{2/3} \lambda_+^{2/3}}{(1 - \lambda_+)^{1/3}(\lambda_+ - \lambda_-)^{1/3}} (\mathfrak{p} + \mathfrak{q})^{-2/3} < 0,$$

$$(15) \qquad \mathfrak{p} = 2 - \frac{2}{N}, \qquad \mathfrak{q} = \frac{T}{N} - 1 - \frac{2}{N}, \qquad \lambda_\pm = \frac{1}{(\mathfrak{p} + \mathfrak{q})^2} \left[\sqrt{\mathfrak{p}(\mathfrak{p} + \mathfrak{q} - 1)} \pm \sqrt{\mathfrak{q}} \right]^2.$$

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

Quantiles of the distribution of $\sum_{i=1}^{r} \mathfrak{a}_i$ 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 $r = 1$ in Table 1.

r \ α	90%	95%	97.5%	99%
1	0.45	0.98	1.45	2.02

TABLE 1. Quantiles of $\mathfrak{a}_1$ from Bejan (2005, Table 2).

Note the shifts by $\frac{2}{N}$ in the definition (15) of $\mathfrak{p}$ and $\mathfrak{q}$. Theorem 2 remains valid without these shifts, i.e., for the simplified $\mathfrak{p} = 2$, $\mathfrak{q} = \frac{T}{N} - 1$. However, we found that for $T/N < 6$ the shifts improve the speed of convergence and, thus, the finite sample behavior of the test. We discuss the shifts in more detail in Sections 5 and 9.3.

We sketch the proof of Theorem 2 in Section 4 and give full details in the appendix. The two main ideas that we use are:

(I) We show that a small (vanishing as $N, T \to \infty$) perturbation of eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ leads to an explicit probability distribution known as the Jacobi

---

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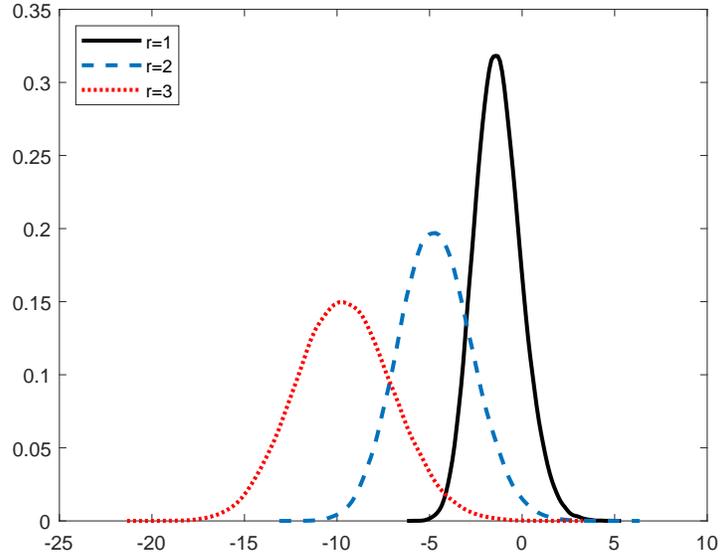
FIGURE 1. The probability density for the random variables $\sum_{i=1}^{r} \mathfrak{a}_i$. The histogram plots are obtained from 100,000 samples of 200×200 Gaussian matrices following Proposition 1.

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) 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 (13) 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]

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

$$(16) \qquad \sum_{i=1}^{\lfloor \rho N \rfloor} \ln(1 - \lambda_i) - N \cdot c_3(T/N, \rho) = o(N^\epsilon), \quad N \to \infty,$$

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 $N, T \to \infty$. (As far as the authors know, Eq. (13) in Theorem 2 is the very first statistic for which the computation of the distributional limit became possible in the $N, T \to \infty$, $T/N \to c$ 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 co-integrations 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 $N, T \to \infty$, $T/N \to c$, we have only finitely many observations per estimated parameter and consistent estimation is unlikely. The conclusion from this

---

[11]O'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.

[12]The 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.

dimension-based heuristics is that in the situation when the number of cointegration rela-
tionships 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 priory
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 coin-
tegrating 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 7.3.

3.2.2. *Estimation of rank and cointegration relationships.* In the small N cointegration anal-
ysis 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
(3) by subsequent rejections of the hypotheses of the rank being smaller than $2, 3, \ldots, r_0$.
Finally one wants to consistently estimate r_0 cointegrating relationships, closely related to
Π itself.

Switching to our limit regime $N, T \to \infty$, $T/N \to c$, 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 $\lambda_1, \ldots, \lambda_N$ and corresponding eigen-
vectors for the model (3) 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 (3),
we treat the $rank(\Pi) = r_0$ 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 $N \to \infty$ 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.,

$$
(17) \qquad \sum_{i=1}^{N} \ln(1 - \lambda_i)
$$

in the notations of Theorem 2. As we explain at the end of Section 3.1, the standard deviation of (17) is $O(1)$, i.e., after subtracting a constant as in Eq. (16), statistic (17) stays finite as $N, T \to \infty$. Hence, in contrast to (13) there should be no N–dependent rescaling in a test based on (17). This should lead to very different powers of tests based on statistics (13) and (17).

Indeed, we expect the main difference in the joint law of $(\lambda_1, \ldots, \lambda_N)$ 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., $\lambda_1 \to \lambda_1 + \nu$, is on the same scale as the standard deviation of (17), 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, (13) 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 (13) leads to a test of higher power than (17) in this situation.[13]

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 $\lambda_{r+1}, \ldots, \lambda_N$ agree with the predictions of our model: for any finite r we expect the empirical distribution function $\frac{1}{N-r} \sum_{i=r+1}^{N} \mathbf{1}_{\lambda_i \le x}$ to converge to the CDF of the Wachter distribution (under H_0 this can be deduced from our Theorems 4 and 14, 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.

---

[13]We do not analyze (17) 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.

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*

$$(18) \qquad \det(M)^{p-1} \det(I_N - M)^{q-1}, \quad 0 < M < I_N,$$

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 $N = 1$, (18) is the Beta distribution. For general N, the eigenvectors of random Jacobi-distributed M are uniformly distributed, while N eigenvalues $x_1 \geq \cdots \geq x_N$ admit an explicit density with respect to the Lebesgue measure given by

$$(19) \qquad \frac{1}{Z(N; p, q)} \prod_{1 \leq i < j \leq N} (x_i - x_j) \prod_{i=1}^{N} x_i^{p-1} (1 - x_i)^{q-1},$$

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 it[14], and as a result its asymptotic properties (usually as $N \to \infty$) 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 ith 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:

$$(20) \qquad \frac{\left(\sum_{i=1}^{T} X_i Y_i \right)^2}{\left(\sum_{i=1}^{T} (X_i)^2 \right) \left(\sum_{i=1}^{T} (Y_i)^2 \right)}.$$

A direct computation shows that if the elements of X and Y are i.i.d. mean zero Gaussians, then the law of (20) is given by Beta distribution for any $T = 1, 2, \ldots$.

---

[14]Just 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 generalization of (20) to $N, K \geq 1$ defines the *squared sample canonical correlations* $\{r_i\}$ as solutions to the equation

$$(21) \qquad \det\left(S_{XY}S_{YY}^{-1}S_{YX} - rS_{XX}\right) = 0,$$

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 (21) has $\min(N, K)$ non-zero solutions. They have a variational meaning. For instance, the maximal solution of Eq. (21) 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 $r_1 > r_2 > \cdots > r_N$ 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. (8) 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. (8) being a particular case) matches the one for the Jacobi ensemble (with shifted dimension parameters) in the limit as $N, T \to \infty$. 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 $N, T \to \infty$, 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 $T, N \to \infty$ in such a way that $T > 2N$ and the ratio T/N remains bounded. Under the hypothesis $\Pi = 0$ for (3), one can couple (i.e. define on the same probability space) the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ of the matrix $S_{10}S_{00}^{-1}S_{01}S_{11}^{-1}$ and*

eigenvalues $x_1 \geq \cdots \geq x_N$ 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

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

The proof of Theorem 4 is based on the following idea. Looking at Eq. (3) when $\Pi = 0$, one can notice that matrices entering into the test and given by Eq. (11) can be expressed in terms of the matrix of data X and the lag operator mapping $X_t \to X_{t-1}$. A computation shows that since we deal with the de-trended and de-meaned 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 the Appendix. 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 (9) 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 $N = 5, \ldots, 10$. 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.

To improve the finite-sample behavior for large c we suggest to ignore the $\frac{2}{N}$ correction in $\mathfrak{p}$ and $\mathfrak{q}$ in Theorem 2, Eq. (15), when T/N is large. That is, to use simplified formulas instead: $\mathfrak{p} = 2$, $\mathfrak{q} = T/N - 1$. We recalculate the empirical rejection rate under the simplified formulas for $\mathfrak{p}, \mathfrak{q}$ 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.
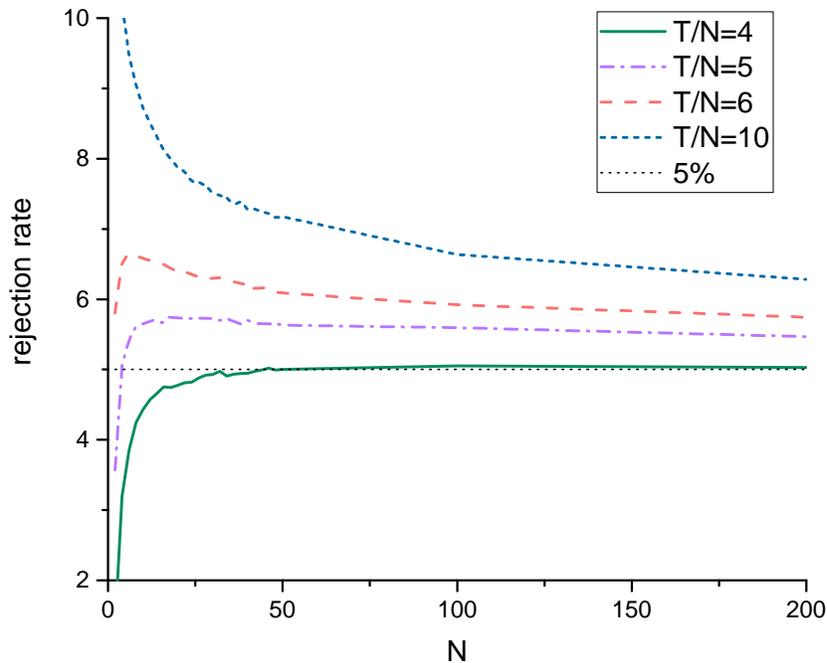
FIGURE 2. Empirical size when T/N is fixed and N increases.
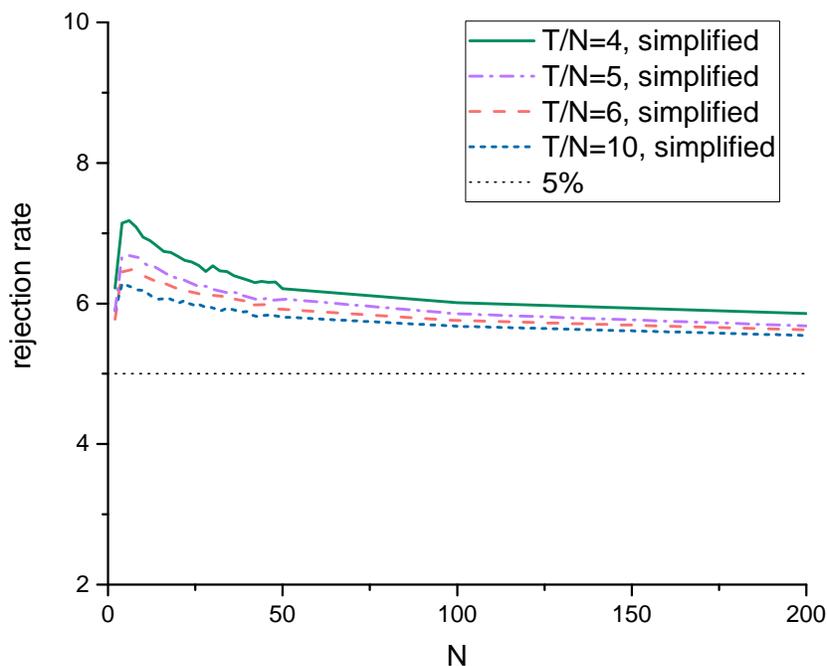
FIGURE 3. Empirical size when T/N is fixed and N increases. Test based on $\mathfrak{p} = 2$, $\mathfrak{q} = T/N - 1$.

An important feature of Figures 2 and 3 is that as $N, T \to \infty$ 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 to use in simulations. Hence, instead we proceed with various random alternatives.[15] Therefore, 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 $\mathfrak{p}$ and $\mathfrak{q}$ 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.

---

[15] Another 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.

[16] As N goes to infinity the non-zero eigenvalue is of order $N^{-1/2}$.
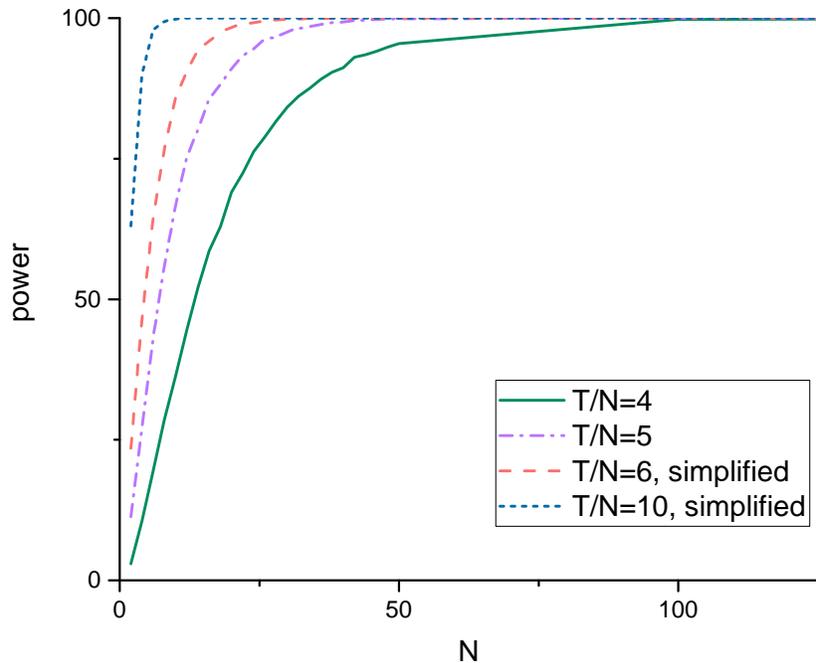
FIGURE 4. Power against random alternative of 1 cointegrating relationship when T/N is fixed and N increases.
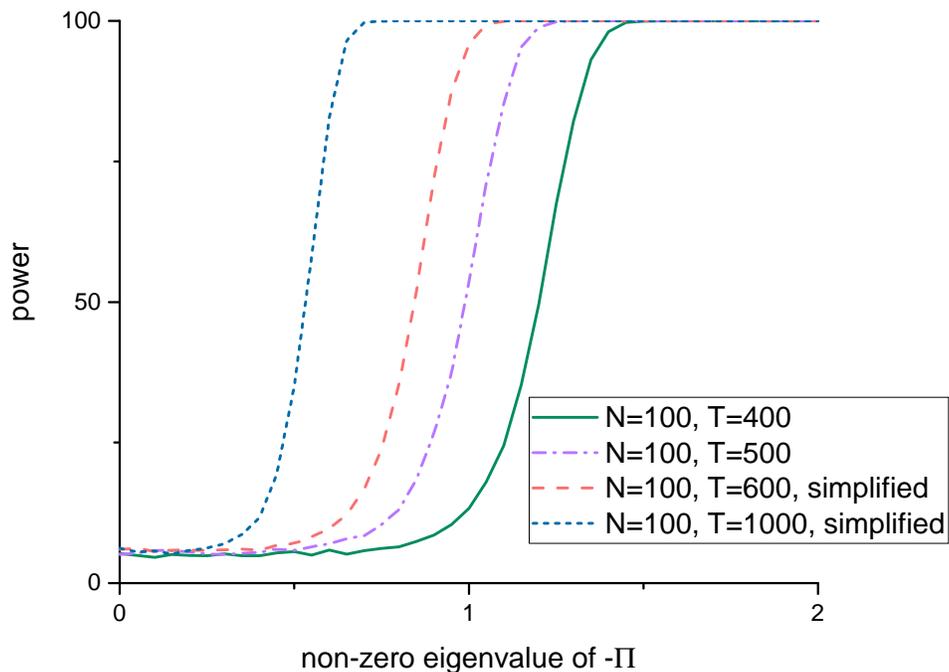

FIGURE 5. Power against random alternative of 1 cointegrating relationship when Π is symmetric. To ensure stationarity of ΔX a non-zero eigenvalue of $-\Pi$ lies between 0 and 2.
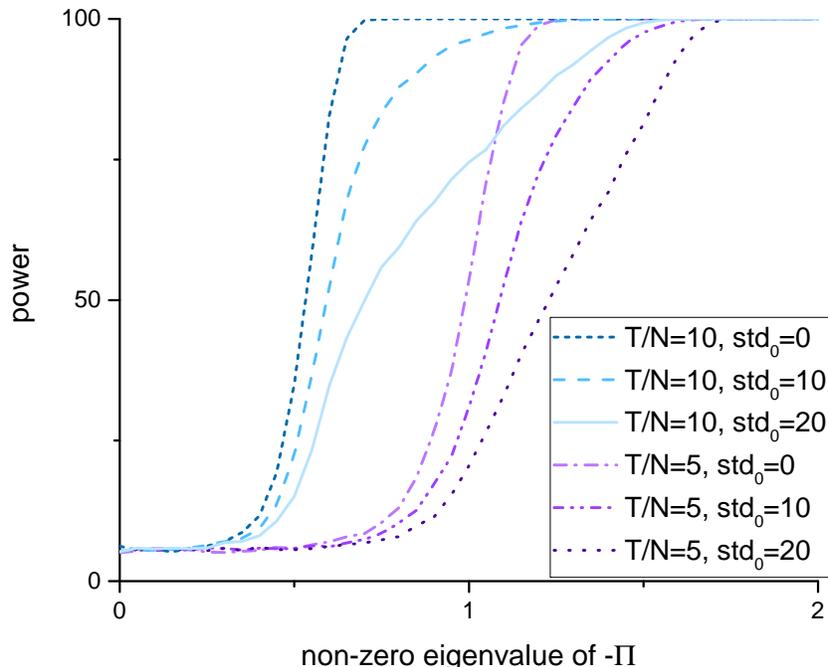
FIGURE 6. Power against random alternative of 1 cointegrating relationship when Π is symmetric. Initial condition is chosen as $X_0 = \text{std}_0 \cdot \mathcal{N}(0,1)$, $N = 100$.

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 $X_{i0} \sim$ 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.

---

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

6. Empirical illustration[18]

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 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 Appendix. Therefore, $N = 92$, $T = 521$ and $T/N \approx 5.66$.

Figure 7 shows the histogram of eigenvalues which solve Eq. 8 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 Appendix 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 (3) has $\Pi = 0$, then the theorem combined with the asymptotics of the Jacobi ensemble of Theorem 14 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.

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 78% percentile 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) 95% percentiles. Hence, we do not see evidence towards the presence of cointegration in S&P100 stock prices for the last 10 years.

---

[18]The first version of this paper posted in June of 2020 had different results in this section. However, later we discovered several suspicious points (four in one stock and one in another stock) in the weekly Yahoo Finance data that we were using. By comparing with the daily data we reached the conclusion that these are anomalies in the database rather than true spikes in the observations. The present version uses corrected values for these points.
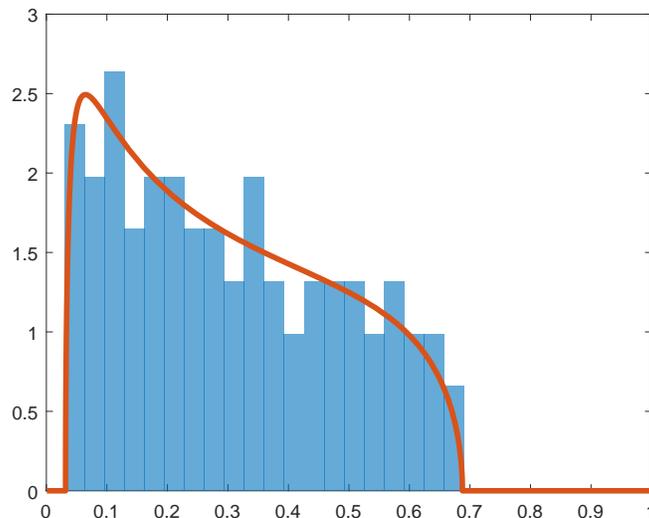
FIGURE 7. Eigenvalues from S&P data and Wachter distribution.

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. In Subsection 7.3 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. (3) 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). 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. (3) 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:

$$(22) \qquad \Delta X_t = \Pi X_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T.$$

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 (22). If only de-trending is omitted, we get the Johansen test statistic corresponding to our original model (3). 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 Airy_1 point process 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
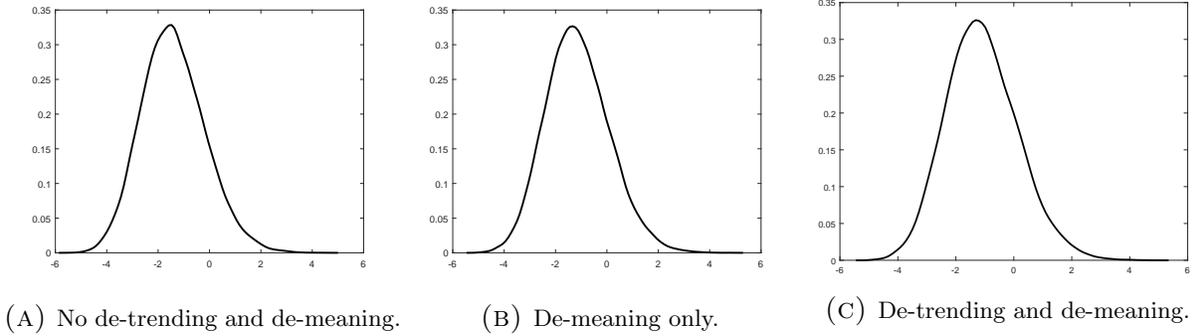
(A) No de-trending and de-meaning.   (B) De-meaning only.   (C) De-trending and de-meaning.

FIGURE 8. Asymptotic distribution of the rescaled largest eigenvalue under various testing procedures (Eq. (13) with $r = 1$). Data generating process: $\Delta X_{it} = \varepsilon_{it}$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 50,000$ replications.

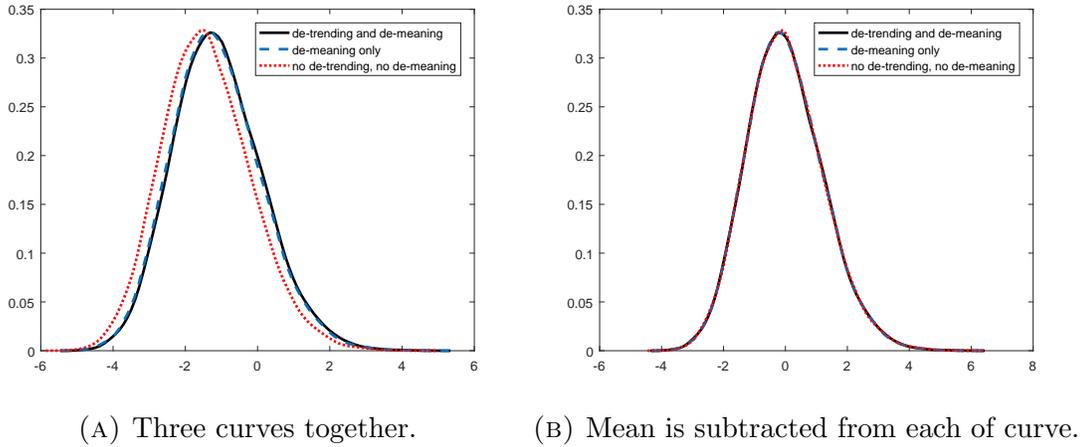(A) Three curves together.   (B) Mean is subtracted from each of curve.

FIGURE 9. Asymptotic distribution of the rescaled largest eigenvalue under various testing procedures with means normalized to zero (Eq. (13) with $r = 1$). Data generating process: $\Delta X_{it} = \varepsilon_{it}$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 50,000$ replications.

that some modifications of the constants in the asymptotic theorems are needed as soon as we start slightly adjusting the setting.

7.3. 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 (3). 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. (3)) becomes:

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

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. (10), we use the notation $\mathcal{P}$ for the de-meaning operator projecting on the hyperplane orthogonal to $(1, 1, \ldots, 1)$.

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:

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

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 $t \to t-1$ in X_t, as compared to Eq. (5). 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 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ as N roots to the equation

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

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 $T, N \to \infty$ 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 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ of the matrix $S_{10}^{w.n.}(S_{00}^{w.n.})^{-1} S_{01}^{w.n.}(S_{11}^{w.n.})^{-1}$ and eigenvalues $x_1 \geq \cdots \geq x_N$ 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

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

The proof of Theorem 5 follows a similar strategy as Theorem 4 and we refer to Section 9.5 in the Appendix 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. 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. 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 multivariave 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 (2020)). 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.

9. Appendix

9.1. New matrix models for the Jacobi ensemble. Recall that our testing procedure relies on the squared sample canonical correlations between two correlated data sets. As we later show in the proofs of Propositions 11 and 16, an equivalent point of view is that we deal with eigenvalues of a product of two orthogonal projectors P_1 and P_2, where P_1 is a projection on a random N-dimensional subspace $\mathcal{V}$ of a T–dimensional space and P_2 is a projection on $U\mathcal{V}$, where U is a certain deterministic linear operator.

We randomize this problem by replacing U with a *random* operator, whose spectrum is close to U. The randomized problem turns out to be exactly solvable — the eigenvalues of new P_1P_2 coincide with the classical Jacobi ensemble. The goal of this section is to prove this fact. The choices of U and $\mathcal{V}$ depend on the hypothesis that we are testing and, hence, we need several theorems. Throughout this section we are going to deal both with real and complex matrices. According to the customary random matrix theory notation, they are referred to as $\beta = 1$ and $\beta = 2$ cases, respectively.

In what follows I_n means the identity matrix in n–dimensional space. Sometimes we omit n and write simply I when the dimension is clear from the context. For a matrix X we let $[X]_{NN}$ to be its top-left $N \times N$ corner. The parameter $\mathcal{T}$ used in this section is related to T of the main text through $\mathcal{T} = T - 1$.

Throughout this section we repeatedly change the coordinates in various measures, which produces a factor given by the absolute value of the Jacobian of the transformation. We rely on the computation of the Jacobian of the multiplicative change of variables in the space of matrices. We need three forms of it, where in each of them Q stays for a $n \times n$ matrix:

- The map $Z \mapsto QZ$ on $n \times m$ matrices has the Jacobian

$$(26) \qquad \left| \frac{\partial(QZ)}{\partial Z} \right| = |\det Q|^{\beta m},$$

- The map $Z \mapsto QZQ^*$ from the space of $n \times n$ symmetric (Hermitian if $\beta = 2$) matrices to itself has the Jacobian

$$(27) \qquad \left| \frac{\partial(QZQ^*)}{\partial Z} \right| = |\det Q|^{\beta n + 2 - \beta} = \begin{cases} |\det Q|^{2n}, & \beta = 2, \\ |\det Q|^{n+1}, & \beta = 1. \end{cases}$$

- The map $Z \mapsto QZQ^*$ from the space of $n \times n$ skew-symmetric (skew-Hermitian if $\beta = 2$) matrices to itself has the Jacobian

$$(28) \qquad \left| \frac{\partial(QZQ^*)}{\partial Z} \right| = |\det Q|^{\beta n + \beta - 2} = \begin{cases} |\det Q|^{2n}, & \beta = 2, \\ |\det Q|^{n-1}, & \beta = 1. \end{cases}$$

The first identity (26) follows from the observation that each column of Z is transformed by linear map Q and there are m such columns. The second and third identities are similar and we refer to Forrester (2010, (1.35)) for details.

The first theorem of this section is relevant for the setting of Theorem 4 for VAR(1).

Theorem 6. *Assume $\mathcal{T} \geq 2N$. Let O be a random $\mathcal{T} \times \mathcal{T}$ matrix chosen from the uniform measure on the group of real orthogonal matrices of determinant 1 if $\beta = 1$ or of complex unitary matrices if $\beta = 2$. Define $U = (I_{\mathcal{T}} + O)^{-1}$. Then the matrix*

$$(29) \qquad M = [U]_{NN}([U^*U]_{NN})^{-1}[U^*]_{NN}$$

is distributed as the $N \times N$ real symmetric (if $\beta = 1$) or complex Hermitian (if $\beta = 2$) matrix of density proportional to

$$(30) \qquad \det(M)^{\frac{\beta}{2}N + \beta - 2} \det(I_N - M)^{\frac{\beta}{2}(\mathcal{T} - 2N + 1) - 1}, \quad 0 \leq M \leq I_N, \qquad \beta = 1, 2.$$

with respect to the Lebesgue measure on symmetric/Hermitian matrices.

Remark 1. *Define a $\mathcal{T} \times \mathcal{T}$–dimensional matrix $\mathcal{M}$ by putting M in $N \times N$ corner of $\mathcal{M}$ and filling the rest with zeros. Non-zero eigenvalues of M and $\mathcal{M}$ are the same. Simultaneously, $\mathcal{M}$ can be identified with $P_1 P_2 P_1$, where P_1 is the orthogonal projector on space $\mathcal{V}$ spanned by the first N coordinate vectors and P_2 is the projector on $U\mathcal{V}$.*

Remark 2. *If $\mathcal{T} < 2N$, then the spaces $\mathcal{V}$ and $U\mathcal{V}$ necessarily intersect, and therefore M has deterministic eigenvalues which equal 1. This should be taken into account when extending (30) to this case and we will not pursue this direction here.*

Proof of Theorem 6. We parameterize the orthogonal (or unitary if $\beta = 2$) group by the means of the Cayley transform. Namely, we set

$$\mathcal{R} = (I_{\mathcal{T}} - O)(I_{\mathcal{T}} + O)^{-1} = \frac{I_{\mathcal{T}} - O}{I_{\mathcal{T}} + O}, \quad \text{so that} \quad O = \frac{I_{\mathcal{T}} - \mathcal{R}}{I_{\mathcal{T}} + \mathcal{R}}.$$

Since $O^{-1} = O^*$, $\mathcal{R}$ is a skew-symmetric $\mathcal{T} \times \mathcal{T}$ matrix, i.e., $\mathcal{R}^* = -\mathcal{R}$. The uniform (Haar) measure on O leads to the following distribution on $\mathcal{R}$, in which we omitted the irrelevant for us normalization constant:

$$(31) \qquad \det(I_{\mathcal{T}} - \mathcal{R}^2)^{-\frac{\beta}{2}\mathcal{T} + \frac{2-\beta}{2}} d\mathcal{R}, \quad \beta = 1, 2,$$

see, e.g., Forrester (2010, (2.55)). Further, we have

$$U = \frac{I_{\mathcal{T}} + \mathcal{R}}{2},$$

so that

$$(32) \qquad M = [I_{\mathcal{T}} + \mathcal{R}]_{NN}([I_{\mathcal{T}} - \mathcal{R}^2]_{NN})^{-1}[I_{\mathcal{T}} - \mathcal{R}]_{NN}.$$

We partition $\mathcal{T} = N + (\mathcal{T} - N)$ and write $I_{\mathcal{T}} + \mathcal{R}$ in a block form according to this split:

$$I_{\mathcal{T}} + \mathcal{R} = \begin{pmatrix} I_N + A & B \\ -B^* & I_{\mathcal{T}-N} + C \end{pmatrix},$$

where A is a $N \times N$ skew-symmetric matrix, C is $(\mathcal{T} - N) \times (\mathcal{T} - N)$ skew-symmetric matrix and B is an arbitrary $N \times (\mathcal{T} - N)$ matrix.

We make a change of variables by introducing $\tilde{B}$ so that

$$B = (I_N - A)\tilde{B}, \quad \text{and} \quad B^* = \tilde{B}^*(I_N + A).$$

Using (26) we compute the Jacobian of the transformation:

$$(33) \qquad \left|\frac{\partial B}{\partial \tilde{B}}\right| = |\det(I_N - A)|^{\beta(\mathcal{T}-N)}.$$

Note that

$$(34) \quad M = (I_N + A)(I_N - A^2 + BB^*)^{-1}(I_N - A)$$

$$= (I_N + A)\big((I_N - A)(I_N + \tilde{B}\tilde{B}^*)(I_N + A)\big)^{-1}(I_N - A) = \big(I_N + \tilde{B}\tilde{B}^*\big)^{-1}$$

Using the formula for the determinant of a block matrix

$$\det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det \mathbf{A} \cdot \det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}),$$

we also have

$$(35) \quad \det(I_{\mathcal{T}} + \mathcal{R}) = \det \begin{pmatrix} I_N + A & (I_N - A)\tilde{B} \\ -B^*(I_N + A) & I_{\mathcal{T}-N} + C \end{pmatrix}$$

$$= \det(I_N + A) \det \begin{pmatrix} I_N & (I_N - A)\tilde{B} \\ -\tilde{B}^* & I_{\mathcal{T}-N} + C \end{pmatrix} = \det(I_N + A) \det(I_{\mathcal{T}-N} + C + \tilde{B}^*(I_N - A)\tilde{B})$$

$$= \det(I_N + A) \det(I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B}) \det \left(I_{\mathcal{T}-N} + (I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B})^{-1/2}(C - \tilde{B}^*A\tilde{B})(I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B})^{-1/2} \right),$$

Next, we introduce

$$\tilde{C} = (I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B})^{-1/2}(C - \tilde{B}^*A\tilde{B})(I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B})^{-1/2}$$

and notice that the map $C \mapsto \tilde{C}$ preserves the space of all skew-symmetric $(\mathcal{T}-N) \times (\mathcal{T}-N)$ matrices. Using (28) the map has the Jacobian

$$(36) \qquad \frac{\partial C}{\partial \tilde{C}} = \det(I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B})^{\frac{\beta}{2}(\mathcal{T}-N) + \frac{\beta-2}{2}},$$

Combining (33), (35), and (36), we rewrite the measure (31) as follows:

$$(37) \quad \det(I_{\mathcal{T}} - \mathcal{R}^2)^{-\frac{\beta}{2}\mathcal{T} + \frac{2-\beta}{2}} \, dA \, dB \, dC = \left| \det(1_{\mathcal{T}} + \mathcal{R})^{-\beta\mathcal{T} + 2 - \beta} \right| \, dA \, dB \, dC$$

$$= \left| \det(I_N + A) \det(I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B}) \det(I_{\mathcal{T}-N} + \tilde{C}) \right|^{-\beta\mathcal{T} + 2 - \beta}$$

$$\times |\det(I_N - A)|^{\beta(\mathcal{T}-N)} \cdot \det(I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B})^{\frac{\beta}{2}(\mathcal{T}-N) + \frac{\beta-2}{2}} \, dA \, d\tilde{B} \, d\tilde{C}.$$

The key property of (37) is that the measure has factorized and projecting onto the $\tilde{B}-$ component is straightforward. We conclude that $\tilde{B}$ is distributed according to the measure

$$(38) \qquad \det(I_{\mathcal{T}-N} + \tilde{B}^*\tilde{B})^{-\frac{\beta}{2}(\mathcal{T}+N) + \frac{2-\beta}{2}} \, d\tilde{B} = \det(I_N + \tilde{B}\tilde{B}^*)^{-\frac{\beta}{2}(\mathcal{T}+N) + \frac{2-\beta}{2}} \, d\tilde{B},$$

where we used $\det(I + UW) = \det(I + WU)$ in the last equality. According to Forrester (2010, Exercise 3.2.q6), (38) implies that the symmetric (or Hermitian if $\beta = 2$) non-negative definite $N \times N$ matrix $F = \tilde{B}\tilde{B}^*$ has the law

$$(39) \qquad \det(F)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} \det(I_N + F)^{-\frac{\beta}{2}(\mathcal{T}+N) + \frac{2-\beta}{2}} \, dF.$$

We have by (34)

$$M = \frac{1}{I_N + F}, \quad F = \frac{1 - M}{M}, \quad dM = -\frac{1}{I_N + F} \, dF \, \frac{1}{I_N + F}.$$

Using (27) we have

$$(40) \qquad \left| \frac{\partial M}{\partial F} \right| = \left| \det \left(\frac{1}{I_N + F} \right) \right|^{\beta N - \beta + 2} = |\det M|^{\beta N - \beta + 2}.$$

Formulas (39) and (40) imply that the matrix M has the distribution

$$\det(I_N - M)^{\frac{\beta}{2}(\mathcal{T} - 2N + 1) - 1} \det(M)^{-\frac{\beta}{2}(\mathcal{T} - 2N + 1) + 1} \det(M)^{\frac{\beta}{2}(\mathcal{T} + N) - \frac{2 - \beta}{2}} \det(M)^{-\beta N + \beta - 2} dM$$

$$= \det(I_N - M)^{\frac{\beta}{2}(\mathcal{T} - 2N + 1) - 1} \det(M)^{\frac{\beta}{2}N + \beta - 2} dM, \quad 0 \le M \le I_N. \quad \square$$

Our next theorem gives another realisation of the Jacobi ensemble, which is relevant to testing $\Pi = -I_N$ in VAR(1) setting.

Theorem 7. *Assume $\mathcal{T} \ge 2N$. Let O be a random $\mathcal{T} \times \mathcal{T}$ real matrix chosen from the uniform measure on the group of orthogonal matrices with determinant 1 if $\beta = 1$ or of complex unitary matrices if $\beta = 2$. Let A be $N \times N$ top-left corner of O. Then the matrix*

$$M = (I_N + A)(2I_N + A + A^*)^{-1}(I_N + A^*)$$

is distributed as $N \times N$ real symmetric (if $\beta = 1$) or complex Hermitian (if $\beta = 2$) matrix of density proportional to

$$(41) \qquad \det(M)^{\frac{\beta}{2}(\mathcal{T} - N) + \beta - 2} \det(I_N - M)^{\frac{\beta}{2}(\mathcal{T} - 2N + 1) - 1}, \quad 0 \le M \le I_N, \qquad \beta = 1, 2.$$

with respect to the Lebesgue measure on symmetric/Hermitian matrices.

Proof. The computation of the law of A is well-known and Forrester (2010, (3.113)) provides a formula for the density with respect to the Lebesgue measure on all real/complex $N \times N$ matrices. It is

$$(42) \qquad \det(I_N - A^* A)^{\frac{\beta}{2}(\mathcal{T} - 2N + 1) - 1} dA, \quad 0 \le A^* A \le I_N,$$

where we omit here and below the normalization constant, which makes the total mass of measure equal to 1. The matrix M is a function of A and in the rest of the proof we transform the measure (42) to make the distribution of this function explicit.

Our first step is to rewrite (42) in terms of M. We claim that

$$(43) \qquad \det(I_N - A^* A) = \det(I_N - M) \det(2I_N + A + A^*)$$

Indeed, we first define

$$M_1 = (I_N + A^*)(I_N + A)(2I_N + A + A^*)^{-1}$$

and notice that

$$I_N - A^* A = 2I_N + A + A^* - M_1(2I_N + A + A^*) = (I_N - M_1)(2I_N + A + A^*).$$

Hence,

$$(44) \qquad \det(I_N - A^*A) = \det(I_N - M_1)\det(2I_N + A + A^*)$$

and (43) is obtained from (44) by using the identity $\det(I - CD) = \det(I - DC)$.

We conclude that the density (42) has the form proportional to

$$(45) \qquad \det(I_N - M)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1}\det(2I_N + A + A^*)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1}\,dA.$$

Note that the first factor has the desired form from the statement of the theorem.

We now split the Euclidean space of all $N \times N$ real (or complex) matrices into the symmetric (or Hermitian) and skew-symmetric (or skew-Hermitian) parts. For that we define

$$X = I_N + \frac{A + A^*}{2}, \quad Y = \frac{A - A^*}{2}.$$

Since A belongs to a unit matrix ball (because $AA^* + BB^* = I_N$, where B is the top-right $N \times (\mathcal{T} - N)$ corner of O), X is a positive-definite symmetric (or Hermitian) matrix, i.e., $X \geq 0$. The relation

$$(I_N + A)(2I_N + A + A^*)^{-1}(I_N + A^*) = M$$

is now rewritten as

$$(46) \qquad (X + Y)(2X)^{-1}(X - Y) = M \quad \text{or} \quad X = 2M + YX^{-1}Y.$$

Note that X^{-1} is positive definite, while Y is skew-Hermitian. This implies that $YX^{-1}Y \leq 0$, i.e., $YX^{-1}Y$ is a negative semi-definite Hermitian matrix. Hence, (46) implies

$$X \leq 2M,$$

which means that $2M - X$ is a non-negative definite matrix.

Using (46) we make a change of coordinates in the space of matrices $(X, Y) \to (M, Y)$. The Jacobian of this change of coordinates is

$$\det\begin{pmatrix} \frac{\partial M}{\partial X} & \frac{\partial Y}{\partial X} \\ \frac{\partial M}{\partial Y} & \frac{\partial Y}{\partial Y} \end{pmatrix} = \det\begin{pmatrix} \frac{\partial M}{\partial X} & 0 \\ \frac{\partial M}{\partial Y} & I \end{pmatrix} = \det\left(\frac{\partial M}{\partial X}\right)$$

Therefore, we need to compute the Jacobian of the map

$$X \mapsto M = \tfrac{1}{2}\left(X - YX^{-1}Y\right),$$

which maps the Euclidean space of symmetric (Hermitian if $\beta = 2$) matrices to itself. Differentiating, and using $dX^{-1} = X^{-1}dXX^{-1}$, we see the matrix identity

$$(47) \qquad 2dM = dX - YX^{-1}dXX^{-1}Y = dX + (YX^{-1})dX(YX^{-1})^*,$$

where we used $X = X^*$, $Y = -Y^*$ in the last identity. Hence, the Jacobian of the map $X \mapsto M$ is a function of YX^{-1} and we denote this function $J(YX^{-1})$. Therefore, (45)

becomes

$$(48) \qquad \sim \det(I_N - M)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} \det(X)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} J^{-1}(YX^{-1}) \, dM \, dY.$$

We now make another change of variables by setting

$$\tilde{Y} = M^{-1/2} Y M^{-1/2}, \qquad Y = M^{1/2} \tilde{Y} M^{1/2}.$$

The Jacobian of the map $(M, Y) \to (M, \tilde{Y})$ is the same as the Jacobian of the map $Y \to \tilde{Y}$, which is computed by (28) as

$$(49) \qquad \left| \frac{\partial \tilde{Y}}{\partial Y} \right| = |\det M|^{-\frac{\beta}{2}N + \frac{2-\beta}{2}} = \begin{cases} (\det M)^{-N}, & \beta = 2, \\ (\det M)^{-(N-1)/2}, & \beta = 1. \end{cases}$$

This converts (48) into

$$(50) \qquad \sim \det(I_N - M)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} (\det M)^{\frac{\beta}{2}N + \frac{\beta-2}{2}} \det(X)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} J^{-1}(YX^{-1}) \, dM \, d\tilde{Y}.$$

It remains to figure out the factors involving $\det X$ and YX^{-1} in the last formula. For that let us introduce the notation

$$\tilde{X} = M^{-1/2} X M^{-1/2},$$

and notice that (46) implies

$$(51) \qquad \tilde{X} = 2I_N + \tilde{Y} \tilde{X}^{-1} \tilde{Y}, \qquad 0 \leq \tilde{X} \leq 2I_N.$$

Solution to (51) gives us a function $\tilde{X} = \tilde{X}(\tilde{Y})$, such that $X = M^{1/2} \tilde{X} M^{1/2}$. Hence, we can transform

$$(52) \qquad \det(X)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} = \det(M)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} \det(\tilde{X})^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1}.$$

For $J(YX^{-1})$ we notice two properties. First,

$$YX^{-1} = M^{1/2} \tilde{Y} \tilde{X}^{-1} M^{-1/2}.$$

In addition, we have the following statement, which we prove later:

Claim. For the above Jacobian $J(\cdot)$, we have

$$(53) \qquad J(A) = J(B^{-1}AB), \qquad \text{whenever } B = B^*.$$

Thus, $J(YX^{-1}) = J(\tilde{Y}\tilde{X}^{-1}) = J(f(\tilde{Y}))$ for a certain function f, and using (52) and (53) we rewrite (50) as

$$(54) \qquad \det(I_N - M)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} (\det M)^{\frac{\beta}{2}N + \frac{\beta-2}{2}} \det(M)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} g(\tilde{Y}) \, dM \, d\tilde{Y}$$

for a certain function $g(\tilde{Y})$. Since the parts involving $\tilde{Y}$ and M are now decoupled, we can integrate out $\tilde{Y}$ arriving at the desired expression for the distribution of M:

$$(55) \qquad \det(I_N - M)^{\frac{\beta}{2}(\mathcal{T}-2N+1)-1} \det(M)^{\frac{\beta}{2}(\mathcal{T}-N)+\beta-2} \, dM.$$

It remains to prove the claim (53). By definition (47), $J(B^{-1}AB)$ is the Jacobian of the linear map

$$dX \mapsto \tfrac{1}{2}dX + \tfrac{1}{2}(B^{-1}AB)dX(BA^*B^{-1})$$

We split this map into a composition of three:

$$dX \mapsto B\,dX\,B \mapsto \tfrac{1}{2}(B\,dX\,B) + \tfrac{1}{2}A(B\,dX\,B)A^* \mapsto B^{-1}\left[\tfrac{1}{2}(B\,dX\,B) + \tfrac{1}{2}A(B\,dX\,B)A^*\right]B^{-1}.$$

$J(B^{-1}AB)$ is the product of Jacobians of these three maps. The first and the last maps are inverse to each other (above we used the explicit Jacobian of these maps, but this is not even needed) and their Jacobians cancel. The Jacobian of the middle map is the desired $J(A)$. $\qquad\square$

9.2. Proof of Theorem 4.

The proof of Theorem 4 is split into two steps. First, Proposition 11 uses rotation invariance of the Gaussian law to rewrite the matrix $S_{10}S_{00}^{-1}S_{01}S_{11}^{-1}$ as a product of corners of certain matrices, reminiscent of (29). Next, we show in Proposition 12 that replacement of a certain deterministic matrix (in that product) by its random perturbation leads precisely to (29) and simultaneously has controlled effect on the change of eigenvalues.

We need to introduce some deterministic $T \times T$ matrices. The summation matrix Φ has 1's below the diagonal and 0's on the diagonal and everywhere above the diagonal:

$$\Phi = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ & & \ddots & & \\ 1 & 1 & \ldots & 1 & 0 \end{pmatrix}.$$

Definition 8. *V is the $(T-1)$–dimensional hyperplane orthogonal to $(1,1,\ldots,1)$.*

We let $\mathcal{P}$ be the orthogonal projector on V (see Eq. (10)). Finally, we set

$$\tilde{\Phi} = \mathcal{P}\Phi\mathcal{P}.$$

We also need the cyclic version of the lead operator F mapping $(x_1, x_2, \ldots, x_T)$ to $(x_2, x_3, x_4, \ldots, x_T, x_1)$. V is an invariant subspace for F and we denote through F_V the restriction of F onto V.

Lemma 9. *The operator $\tilde{\Phi}$ preserves the space V. Its restriction onto V coincides with $-(I_V - F_V)^{-1}$, where I_V is the identical operator acting in V.*

Proof of Lemma 9. Let e_i, $i = 1, \ldots, T$ denote the ith coordinate vector. Then $(e_i - e_{i+1})_{i=1}^{T-1}$ gives a linear basis of V. Since $\Phi e_i = e_{i+1} + e_{i+2} + \cdots + e_T$, we have

$$\tilde{\Phi}(e_i - e_{i+1}) = \mathcal{P}\Phi(e_i - e_{i+1}) = \mathcal{P}e_{i+1} = e_{i+1} - \frac{1}{T}\sum_{j=1}^{T} e_j.$$

Applying $I - F$ to the last vector and noticing that $Fe_{i+1} = e_i$, we conclude that

$$(I - F)\tilde{\Phi}(e_i - e_{i+1}) = e_{i+1} - e_i. \quad \square$$

Let us introduce one more notation. Choose some orthonormal basis $\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_{T-1}$ of V, e.g., this can be an ortogonalization of $e_1 - e_2$, $e_2 - e_3$, $\ldots$, $e_{T-1} - e_T$ (but the exact choice is irrelevant for the following theorems).

Definition 10. *For an operator A acting in V we set $[A]_{NN}^V$ to be the $N \times N$ corner of the operator A, taken in the basis $\{\tilde{e}_i\}$.*

Next, we take a uniformly-random orthogonal (or unitary if $\beta = 2$) operator $\tilde{O}$ acting in $(T-1)$–dimensional space V and define an operator $\tilde{U}$ acting in V:

$$\tilde{U} = (I_V - \tilde{O}F_V\tilde{O}^*)^{-1}.$$

Let us introduce a symmetric (or Hermitian if $\beta = 2$) $N \times N$ matrix $\tilde{M}$,

(56) $$\tilde{M} = [\tilde{U}]_{NN}^V([\tilde{U}^*\tilde{U}]_{NN}^V)^{-1}[\tilde{U}^*]_{NN}^V,$$

Proposition 11. *Choose an arbitrary positive definite covariance matrix Λ. Let ε be $N \times T$ matrix of random variables (real if $\beta = 1$ and complex if $\beta = 2$), such that T columns of ε are i.i.d., and each of them is an N-dimensional mean zero Gaussian vector with covariance Λ. Fix arbitrary N–dimensional vectors X_0 and μ. Define the $N \times T$ matrix $X = (X_1, X_2, \ldots, X_T)$ as in Eq. (3) via recurrence*

$$X_t = X_{t-1} + \mu + \varepsilon_t, \quad t = 1, \ldots, T.$$

Further set $\Delta X_t = X_t - X_{t-1}$ and $\tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0)$, $t = 1, \ldots, T$. Define $N \times N$ matrices:

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

Then the eigenvalues of the matrix $S_{10}S_{00}^{-1}S_{01}S_{11}^{-1}$ have the same distribution as those of $\tilde{M}$ in (56).

Proof. We start by expressing the matrices S_{ij} via ε. Clearly, $\Delta X_t = \mu + \varepsilon_t$. Further,

$$\Delta X_t\mathcal{P} = \mu + \varepsilon_t - \frac{1}{T}\sum_{i=1}^{T}(\mu + \varepsilon_i), \qquad \Delta X\mathcal{P} = \varepsilon\mathcal{P}.$$

Hence, $S_{00} = \varepsilon \mathcal{P} \varepsilon^*$. Next,

$$\tilde{X}_t = X_0 + (t-1)\mu + \sum_{i=1}^{t-1} \varepsilon_i - \frac{t-1}{T}\left(T\mu + \sum_{i=1}^{T} \varepsilon_i\right) = X_0 + \sum_{i=1}^{t-1} \varepsilon_i - \frac{t-1}{T}\sum_{i=1}^{T} \varepsilon_i$$

We claim that $\tilde{X}\mathcal{P} = \varepsilon\tilde{\Phi}^*$. Indeed, $\tilde{X}\mathcal{P}$ coincides with $\tilde{\tilde{X}}\mathcal{P}$, where

$$\tilde{\tilde{X}}_t = \tilde{X}_t - X_0 = \sum_{i=1}^{t-1} \varepsilon_i - \frac{t-1}{T}\sum_{i=1}^{T} \varepsilon_i = \sum_{i=1}^{t-1}\left(\varepsilon_i - \frac{1}{T}\sum_{j=1}^{T} \varepsilon_j\right).$$

Since Φ is the summation operator, we have $\tilde{\tilde{X}} = (\Phi(\varepsilon\mathcal{P})^*)^* = \varepsilon\mathcal{P}\Phi^*$ and the claim is proven as $\tilde{\Phi}^* = \mathcal{P}\Phi^*\mathcal{P}$.

We conclude that

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

Next, note that we can (and will) assume without loss of generality that the covariance matrix Λ is identical, i.e., all matrix elements of ε are i.i.d. random variables $\mathcal{N}(0,1)$ if $\beta = 1$ and $\mathcal{N}(0,1) + \mathbf{i}\mathcal{N}(0,1)$ if $\beta = 2$. Indeed, using (57), if we take any non-degenerate $N \times N$ matrix A and replace ε by $A\varepsilon$, then the matrix $S_{10}S_{00}^{-1}S_{01}S_{11}^{-1}$ is conjugated by A, which keeps eigenvalues unchanged. By choosing appropriate A, we can guarantee that the covariance of columns of $A\varepsilon$ is identical, and then rename $A\varepsilon$ as ε.

In the remaining proof we explicitly couple ε with $\tilde{O}$ (arising in the definition of $\tilde{M}$) by constructing $\tilde{O}$ using randomness coming from ε.

Since for any two rectangular matrices A and B of the same sizes, AB^* and B^*A have the same non-zero eigenvalues, the eigenvalues of $S_{10}S_{00}^{-1}S_{01}S_{11}^{-1}$ can be identified (using also $\mathcal{P}\tilde{\Phi} = \tilde{\Phi}$) with those of

(58) $$\left(\mathcal{P}\varepsilon^* S_{00}^{-1} \varepsilon\mathcal{P}\right)\left(\tilde{\Phi}\varepsilon^* S_{11}^{-1} \varepsilon\tilde{\Phi}^*\right) = P_1 P_2,$$

where P_1 is the orthogonal projector onto the space spanned by N columns of $\mathcal{P}\varepsilon^*$ and P_2 is the orthogonal projector onto the N columns of $\tilde{\Phi}\varepsilon^*$. Since, the eigenvalues of $P_1 P_2$ are the same as those of $P_1 P_2 P_1$, and the latter operator acts as 0 on the orthogonal complement of V, we can restrict all the operators to V. Due to invariance of the Gaussian law with respect to rotations by orthogonal (or unitary if $\beta = 2$) matrices, the columns of $\mathcal{P}\varepsilon^*$ span a rotationally-invariant N–dimensional subspace in V. The columns of $\tilde{\Phi}\varepsilon^* = \tilde{\Phi}\mathcal{P}\varepsilon^*$ then span the $\tilde{\Phi}$–image of this subspace.

The eigenvalues of $P_1 P_2 P_1$ are preserved when we conjugate each of the projectors with an orthogonal transformation O, i.e. replace $P_1 P_2 P_1$ with $OP_1O^*OP_2O^*OP_1O^* = OP_1P_2P_1O^*$. In this transformation the spaces to where P_1 and P_2 are projecting are replaced by their O–images. We take O to be a random operator satisfying two conditions:

- O maps the span of the columns of $\mathcal{P}\varepsilon^*$ to the subspace spanned by the first N basis vectors, $\tilde{e}_1, \ldots, \tilde{e}_N$ in V;
- O is distributed as uniformly random orthogonal operator acting in V, i.e., $O \overset{d}{=} \tilde{O}$.

The existence of such O follows from the rotational invariance of the Gaussian law.[19] Hence, OP_1O^* is a projector onto $\langle \tilde{e}_1, \ldots, \tilde{e}_N \rangle$, and we conclude that the eigenvalues of $P_1P_2P_1$ coincide with those of

$$[OP_2O^*]_{NN}^V.$$

Since P_2 is the projector on the span of columns of $\tilde{\Phi}\varepsilon^* = O^*(O\tilde{\Phi}O^*)O\varepsilon^*$, OP_2O^* is the projector on the O–image of this span, i.e., OP_2O^* is the projector onto the columns of $(O\tilde{\Phi}O^*)O\varepsilon^*$. Since columns of $O\varepsilon^*$ span $\langle \tilde{e}_1, \ldots, \tilde{e}_N \rangle$ and $O\tilde{\Phi}O^* = O(-(I_V - F_V)^{-1})O^* = -\tilde{U}$ on V by Lemma 9, we reach the formula for $\tilde{M}$ as the $N \times N$ corner of the projector on the first N columns of $\tilde{U}$. □

Set $\mathcal{T} = T - 1$ and let O be uniformly random real orthogonal of determinant 1 (if $\beta = 1$) or complex unitary (if $\beta = 2$) $\mathcal{T} \times \mathcal{T}$ matrix. Define $U = (I_{\mathcal{T}} + O)^{-1}$. Then we set, as in (29),

$$(59) \qquad\qquad M = [U]_{NN}([U^*U]_{NN})^{-1}[U^*]_{NN}$$

For a matrix A with real spectrum, we set $\lambda_i(A)$ to be the ith largest eigenvalue of A.

Proposition 12. *Assume $\mathcal{T} = T - 1 \geq 2N$. One can couple M from (59) with $\tilde{M}$ from (56) in such a way that for each $\epsilon > 0$ we have*

$$\mathrm{Prob}\left(\max_{1 \leq i \leq N} |\lambda_i(M) - \lambda_i(\tilde{M})| < \frac{1}{N^{1-\epsilon}} \right) \to 1,$$

as $T, N \to \infty$ in such a way that the ratio T/N remains bounded.

Remark 3. *We believe that the condition $\mathcal{T} \geq 2N$ can be weakened and replaced by T/N bounded away from 0, but we stick to it, since that's the situation when Theorem 6 applies.*

Let us compare the definitions of M and $\tilde{M}$. There is only one difference between them: the eigenvalues of $-\tilde{O}F_V\tilde{O}^*$ in the definition of $\tilde{M}$ are deterministic, while the eigenvalues of O are random. Note that both matrices $-\tilde{O}F_V\tilde{O}^*$ and O have uniformly-random eigenvectors. Then the idea of the proof is to rely on the phenomenon of *rigidity* for random matrix eigenvalues, which says that as $T \to \infty$, the eigenvalues of O are very close to their deterministic expected positions, which, in turn, essentially coincide with eigenvalues of $-F_V$. One technical difficulty is that we need to deal with $(I + O)^{-1}$, whose eigenvalues can be large (the absolute value of the largest eigenvalue grows linearly in T), however, as we will

---

[19]One way to see it is by noticing that the span of the columns of $\mathcal{P}\varepsilon^*$ and the span of N columns of O^{-1} (for uniformly random O) have the same distribution given by the uniform measure on all N–dimensional subspace of T–dimensional space.

see, one can study $(I + O)$ instead of $(I + O)^{-1}$, and the problem of unbounded operators disappears. We proceed to the detailed proof.

Proof of Proposition 12. We start by explicitly constructing the desired coupling. The eigenvalues of F_V are all roots of unity of order T different from 1. If $\beta = 2$, then we can diagonalize F_V to turn it into $\mathcal{T} \times \mathcal{T} = (T-1) \times (T-1)$ diagonal matrix with the roots of unity on the diagonal. If $\beta = 1$, the matrix F_V should be block-diagonalized (with blocks of size 2 and one additional block of size 1 corresponding to eigenvalue -1 if $\mathcal{T}$ is even): the pair of complex conjugate roots of unity ω and $\bar{\omega}$ gives rise to the 2×2 matrix of rotation by the angle $|\arg(\omega)|$. Let us denote by D the resulting (block) diagonal matrix multiplied by -1. In order to avoid ambiguity about the order of eigenvalues, we assume that the blocks correspond to the increasing order of $|\arg(-\omega)|$, i.e. the top-left 2×2 corner of D corresponds to the pair of the closest to 1 eigenvalues of D.

The eigenvalues of O also lie on the unit circle and if $\beta = 1$ they come in complex-conjugate pairs. Hence, O can be similarly block-diagonalized (we do not need to multiply by -1 this time) and we denote through D^{rand} the result. The distinction with F_V is that the eigenvalues are *random* and so is D^{rand}. The law of the eigenvalues of O is explicitly known in the random-matrix literature. Both for $\beta = 1$ and $\beta = 2$ they form a determinantal point process on the unit circle with explicit kernel. The *repulsion* between the eigenvalues leads to them being very close to evenly spaced as $T \to \infty$. We summarize this property in the following statement (which is a manifestation of much more general rigidity of eigenvalues, see, e.g., Erdos and Yau (2012)), whose proof can be found in Meckes and Meckes (2013, Lemma 10, $m = 1$, $u = T^\delta$ case, and Section 5).

Claim. There exist constants $c_1(\beta), c_2(\beta) > 0$, such that for $\beta = 1, 2$, every $\delta > 0$, there exists $\mathcal{T}_0(\delta)$ and for every $\mathcal{T} > \mathcal{T}_0(\delta)$ we have [20]

$$(60) \qquad \mathrm{Prob}\left(\max_{1 \le i,j < \mathcal{T}} |D - D^{\mathrm{rand}}|_{ij} > \frac{1}{\mathcal{T}^{1-\delta}}\right) < c_1(\beta) \cdot \mathcal{T} \cdot \exp\left(-c_2(\beta)\frac{\mathcal{T}^{2\delta}}{\log \mathcal{T}}\right).$$

We remark that since D and D^{rand} are block-diagonal, the bound on the maximum matrix element of their difference is equivalent to a similar bound for any other norm, e.g., for the maximum absolute value among the eigenvalues of $D - D^{\mathrm{rand}}$.

We now choose another $\mathcal{T} \times \mathcal{T}$ uniformly-random orthogonal (or unitary if $\beta = 2$) matrix O_2 (independent from the rest), replace $-\tilde{O}F_V\tilde{O}^*$ with $O_2 D O_2^*$ and replace O with $O_2 D^{\mathrm{rand}} O_2^*$. The invariance of the uniform measure on the orthogonal group $SO(N)$ (or on the unitary group $U(N)$ if $\beta = 2$) with respect to right/left multiplications, implies the distributional identities:

$$-\tilde{O}F_V\tilde{O}^* \overset{d}{=} O_2 D O_2^*, \qquad O \overset{d}{=} O_2 D^{\mathrm{rand}} O_2^*.$$

---

[20] All the constants can be made explicit, following Meckes and Meckes (2013).

The right-hand sides of the identities provide the desired coupling and (60) implies that these two random matrices are close to each other as $\mathcal{T} \to \infty$. Both matrices M and $\tilde{M}$ are obtained from the above two matrices by the following mechanism:

Map $M(Z)$: Given an orthogonal (or unitary if $\beta = 2$) matrix Z in $\mathcal{T}$–dimensional space, we define $U(Z) = (I_\mathcal{T} + Z)^{-1}$ and an $N \times N$ symmetric (or hermitian) matrix $M(Z)$ through

$$(61) \qquad M(Z) = [U(Z)]_{NN}([U^*(Z)U(Z)]_{NN})^{-1}[U^*(Z)]_{NN}.$$

Clearly, under our coupling

$$M = M(O_2 D^{\mathrm{rand}} O_2^*), \qquad \tilde{M} = M(O_2 D O_2^*),$$

and it remains to prove a form of the uniform continuity of the map $Z \mapsto M(Z)$ as $N, \mathcal{T} \to \infty$.

We note that

$$2U(Z) - I_\mathcal{T} = \frac{I_\mathcal{T} - Z}{I_\mathcal{T} + Z}$$

is a skew-symmetric matrix and we use it to write $2U(Z)$ in the block form according to the splitting $\mathcal{T} = N + (\mathcal{T} - N)$.

$$2U(Z) = \begin{pmatrix} I_N + A(Z) & B(Z) \\ -B^*(Z) & I_{\mathcal{T}-N} + C(Z) \end{pmatrix}, \qquad A^*(Z) = -A(Z), \quad C^*(Z) = -C(Z).$$

Then, as in the proof of Theorem 6, we have

$$(62) \qquad M(Z) = \left(I_N + \tilde{B}(Z)\tilde{B}^*(Z)\right)^{-1}, \qquad \tilde{B}(Z) = \left(I_N - A(Z)\right)^{-1}B(Z).$$

and our statement boils down to the uniform continuity of the map $Z \mapsto \tilde{B}(Z)$. We remark that in our limit regime the matrix $B(Z)B^*(Z)$ might have large eigenvalues (in fact, of order $\mathcal{T}$) and therefore, $B(Z)$ is potentially an exploding factor in the definition of $\tilde{B}(Z)$. However, the exploding parts would precisely cancel out with similarly growing parts in $(I_N - A(Z))$. In order to see that, we use the formula for the inverse of the block matrix, which reads

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{Q}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Q} \\ -\mathbf{Q}\mathbf{C}\mathbf{A}^{-1} & \mathbf{Q} \end{pmatrix}, \qquad \mathbf{Q} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$$

The advantage of this formula is that when used for $2U(Z)$, the ratio $-\mathbf{C}\mathbf{A}^{-1} = B^*(z)(1 + A(z))^{-1} = \tilde{B}^*(Z)$ gets expressed as $\mathbf{Q}^{-1}(-\mathbf{Q}\mathbf{C}\mathbf{A}^{-1})$, which is the ratio of two blocks in $U(Z)^{-1} = 1 + Z$.[21] Thus, if we write $1 + Z$ itself in the block form according to $\mathcal{T} = N + (\mathcal{T} - N)$ splitting:

$$I_\mathcal{T} + Z = \begin{pmatrix} I_N + Z_{11} & Z_{12} \\ Z_{21} & I_{\mathcal{T}-N} + Z_{22} \end{pmatrix},$$

---

[21]The same observation can be used to link Theorems 6 and 7 to each other.

then
$$\tilde{B}^*(Z) = \mathbf{Q}^{-1}(-\mathbf{QCA}^{-1}) = (I_{\mathcal{T}-N} + Z_{22})^{-1} Z_{21},$$
$$M(Z) = \left(I_N + Z_{21}^* \big((I_{\mathcal{T}-N} + Z_{22})(I_{\mathcal{T}-N} + Z_{22}^*)\big)^{-1} Z_{21} \right)^{-1}$$

Hence, the matrix $M(Z)$ has the same (other from 1) eigenvalues as

$$M'(Z) = \left(I_{\mathcal{T}-N} + \big((I_{\mathcal{T}-N} + Z_{22})(I_{\mathcal{T}-N} + Z_{22}^*)\big)^{-1} Z_{21} Z_{21}^* \right)^{-1}$$

The latter can be simplified, since Z is orthogonal, implying $Z_{21} Z_{21}^* + Z_{22} Z_{22}^* = I_{\mathcal{T}-N}$:

$$(63) \quad M'(Z) = \left(I_{\mathcal{T}-N} + \big((I_{\mathcal{T}-N} + Z_{22} + Z_{22}^* + Z_{22}Z_{22}^*)\big)^{-1}(I_{\mathcal{T}-N} - Z_{22}Z_{22}^*) \right)^{-1}$$
$$= \left(\big((I_{\mathcal{T}-N} + Z_{22} + Z_{22}^* + Z_{22}Z_{22}^*)\big)^{-1}(2 \cdot I_{\mathcal{T}-N} + Z_{22} + Z_{22}^*) \right)^{-1}$$
$$= (2 \cdot I_{\mathcal{T}-N} + Z_{22} + Z_{22}^*)^{-1}\big(I_{\mathcal{T}-N} + Z_{22} + Z_{22}^* + Z_{22}Z_{22}^*\big)$$

At this point we rely on the lemma, which is proven below:

Lemma 13. *Fix any small $\upsilon > 0$. If Z is a uniformly random element of the group $SO(\mathcal{T})$ of orthogonal determinant 1 matrices (or $U(\mathcal{T})$ if $\beta = 2$), then there exists $r > 0$ such that with probability tending to 1 (uniformly) as $\mathcal{T}, N \to \infty$ in such a way that $1 + \upsilon < \mathcal{T}/N < \upsilon^{-1}$, the smallest eigenvalue of $2 I_{\mathcal{T}-N} + Z_{22} + Z_{22}^*$ is larger than r.*

Now we are ready to compare $M'(Z)$ with $M'(\tilde{Z})$, where $Z = O_2 D^{\mathrm{rand}} O_2^*$, $\tilde{Z} = O_2 D O_2^*$. Fix $\delta > 0$. By (60), with probability tending to 1 as $N, \mathcal{T} \to \infty$, $\|Z - \tilde{Z}\| < \mathcal{T}^{\delta-1}$, where $\|\cdot\|$ is the spectral norm of the matrix (i.e. its largest singular value). Hence, since cutting corners of the matrix only decreases the spectral norm, with probability tending to 1,

$$(64) \qquad \left\| \big(I_{\mathcal{T}-N} + Z_{22} + Z_{22}^* + Z_{22}Z_{22}^*\big) - \big(I_{\mathcal{T}-N} + \tilde{Z}_{22} + \tilde{Z}_{22}^* + \tilde{Z}_{22}\tilde{Z}_{22}^*\big) \right\| < 4\mathcal{T}^{\delta-1}.$$

Simultaneously, by Lemma 13, the spectral norm of $(2 I_{\mathcal{T}-N} + Z_{22} + Z_{22}^*)^{-1}$ is bounded from above by r^{-1} (and, hence, a similar bound holds for Z replaced with $\tilde{Z}$). We conclude that the part of the increment $M'(Z) - M'(\tilde{Z})$ due to the change of the numerator in the right-hand side of (63) is bounded from above (in spectral norm) by

$$(65) \qquad\qquad\qquad \mathrm{const} \cdot \mathcal{T}^{\delta-1}.$$

For the denominator in (63), we use the matrix identity

$$(66) \qquad\qquad (G + \Delta)^{-1} - G^{-1} = -(G + \Delta)^{-1}\Delta\, G^{-1}$$

with $G = 2 I_{\mathcal{T}-N} + Z_{22} + Z_{22}^*$, $\Delta = \tilde{Z}_{22} + \tilde{Z}_{22}^* - Z_{22} - Z_{22}^*$. By (60), $\|\Delta\| < 2\mathcal{T}^{\delta-1}$ with probability tending to 1. Hence, using Lemma 13 we upper bound the spectral norm of (66) by $\mathrm{const} \cdot r^{-2} \mathcal{T}^{\delta-1}$. Since the spectral norm of the numerator in the right-hand side of

(63) is at most 4, we conclude that the part of the increment $M'(Z) - M'(\tilde{Z})$ due to the change of the denominator also admits a bound (65). Summing up, we have shown that with probability tending to 1,

$$\|M'(O_2 D^{\mathrm{rand}} O_2^*) - M'(O_2 D O_2^*)\| < \mathrm{const} \cdot \mathcal{T}^{\delta-1}.$$

This finishes the proof of Proposition 12. $\qquad\square$

Proof of Lemma 13. We would like to show that for some deterministic $r > 0$ the following inequality on the norm holds with probability tending to 1 as $N, \mathcal{T} \to \infty$:

$$\text{(67)} \qquad \left\| \frac{Z_{22} + Z_{22}^*}{2} \right\| < 1 - \frac{r}{2}.$$

For that we denote $H = \frac{Z+Z^*}{2}$ and notice that H is a Hermitian matrix with spectrum between -1 and 1. In fact, Forrester (2010, Section 3.7) explains that $\frac{1+H}{2}$ is again a Jacobi ensemble. The law of large numbers for the latter (reviewed in Theorem 14) implies that as $\mathcal{T} \to \infty$, the empirical measure of the eigenvalues $\frac{1}{\mathcal{T}} \sum_{i=1}^{\mathcal{T}} \delta_{\lambda_i(H)}$ converges (weakly in probability[22]) to a certain explicit deterministic measure on $[-1, 1]$ with a density $p(x)$.

The matrix $\frac{Z_{22} + Z_{22}^*}{2}$ is a $(\mathcal{T} - N) \times (\mathcal{T} - N)$ corner of H. The largest eigenvalue of this matrix has the following representation:

$$\text{(68)} \qquad \lambda_1 \left(\frac{Z_{22} + Z_{22}^*}{2} \right) = \max_{\substack{u \in \langle e_{N+1}, \dots, e_{\mathcal{T}} \rangle \\ |u|=1}} \left(\sum_{i=1}^{\mathcal{T}} \lambda_i(H) \left| (u, v_i(H)) \right|^2 \right),$$

where $\lambda_i(H)$ is the ith largest eigenvalue of H, $v_i(H)$ is the corresponding eigenvector and $(u, v_i(H))$ is the scalar product with this eigenvector. Since the vectors v_i are orthonormal, $\sum_{i=1}^{N} \left| (u, v_i(H)) \right|^2 = 1$. On the other hand, all $\lambda_i(H)$ are not greater than 1 and only few of them are close to 1. Hence, (67) would follow, if we manage to show that the sum in (68) is not concentrated on few largest $\lambda_i(H)$. To show that, it suffices to upper-bound $\left| (u, v_i(H)) \right|^2$, which we now do.

Let us fix k with $k < \min(N, \mathcal{T} - N)$ and upper bound (68) by

$$\text{(69)} \qquad \max_{\substack{u \in \langle e_{N+1}, \dots, e_{\mathcal{T}} \rangle \\ |u|=1}} \left(\sum_{i=1}^{k} \left| (u, v_i(H)) \right|^2 + \lambda_{k+1}(H) \sum_{i=k+1}^{\mathcal{T}} \left| (u, v_i(H)) \right|^2 \right)$$

$$= \lambda_{k+1}(H) + (1 - \lambda_{k+1}(H)) \max_{\substack{u \in \langle e_{N+1}, \dots, e_{\mathcal{T}} \rangle \\ |u|=1}} \left(\sum_{i=1}^{k} \left| (u, v_i(H)) \right|^2 \right).$$

---

[22]A sequence of *random* probability measures μ_n converges weakly in probability to a measure μ if for each bounded continuous function $f(x)$, the random variables $\int f(x)\mu_n(dx)$ converge in probability to $\int f(x)\mu(dx)$.

Since the law of the matrix H is invariant under conjugations with orthogonal (or unitary) matrices, the eigenvectors $v_1(H), v_2(H), \ldots, v_{\mathcal{T}}(H)$ are uniformly-distributed, i.e. the $\mathcal{T} \times \mathcal{T}$ matrix formed by them is a uniformly-random orthogonal (or unitary) matrix. Hence,

$$\max_{\substack{u \in \langle e_{N+1}, \ldots, e_{\mathcal{T}-N} \rangle \\ |u|=1}} \left(\sum_{i=1}^{k} |(u, v_i(H))|^2 \right).$$

is a maximum eigenvalue of the $(\mathcal{T} - N) \times (\mathcal{T} - N)$ corner of the projector in $\mathcal{T}$–dimensional space on random k–dimensional subspace chosen uniformly at random. If we denote through P_k and $P_{\mathcal{T}-N}$ the projectors on the first k and $\mathcal{T} - N$ basis vectors, respectively, and take a uniformly random orthogonal (or unitary if $\beta = 2$) $\mathcal{T} \times \mathcal{T}$ matrix O, then we deal with the maximal eigenvalue of $P_{\mathcal{T}-N} O P_k O^* P_{\mathcal{T}-N}$. The maximal eigenvalue of this product is the same as the one of $O P_k O^* P_{\mathcal{T}-N}$, or of $P_k O^* P_{\mathcal{T}-N} O$, or of $P_k O^* P_{\mathcal{T}-N} O P_k$. Equivalently, this is the maximal eigenvalue of $k \times k$ corner Y of a projector on uniformly-random $(\mathcal{T} - N)$–dimensional space. Such $k \times k$ matrix Y is distributed as Jacobi ensemble with density proportional to

(70) $$\det(Y)^{\frac{\beta}{2}(\mathcal{T}-N-k+1)-1} \det(1 - Y)^{\frac{\beta}{2}(N-k+1)-1} dY,$$

see Forrester (2010, (3.113) and the following formula). We now choose $k = N/2$. Then the largest eigenvalue of the ensemble (70) converges as $N, \mathcal{T} \to \infty$ in probability (see e.g., Johnstone (2008), or Anderson et al. (2010, Section 2.6.2), or Holcomb and Moreno Flores (2012)) to a deterministic number $0 < c < 1$, depending on the ratio $\mathcal{T}/N$.

Simultaneously, $\lambda_{N/2+1}(H)$ converges as $\mathcal{T}, N \to \infty$ to another deterministic number $0 < c' < 1$, due to the aforementioned convergence of the empirical measure of H. Hence, (69) implies that the largest eigenvalue of $\frac{Z_{22} + Z_{22}^*}{2}$ is bounded away from 1. $\qquad \square$

Remark 4. *Although we are not pursuing this direction, it is possible to upgrade the above argument to the case when $N, \mathcal{T} \to \infty$ in such a way that $\mathcal{T}/N \to \infty$. Then r is tending to 0, but we can find the speed of convergence (and then propagate it to the precise bound in Proposition 12). For that we would need to analyze the largest eigenvalue of (70) more precisely (when $\mathcal{T}/N \to \infty$, the Jacobi ensemble concentrates near 1 and degenerates into the Wishart ensemble after proper rescaling), and also use exact asymptotics of $\lambda_{k+1}(H)$ (which can be obtained from the local law of the Jacobi ensemble for H near 1).*

We now have all the ingredients for Theorem 4 (as well as for its $\beta = 2$ version).

Proof of Theorem 4. By Theorem 6, the eigenvalues $x_1 \geq \cdots \geq x_N$ of Jacobi ensemble $\mathbf{J}(N; \frac{N}{2}, \frac{T-2N}{2})$ have the same distribution as those of M defined in (59). On the other hand, by Proposition 11 the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ of the matrix $S_{10} S_{00}^{-1} S_{01} S_{11}^{-1}$ have the

same distribution as those of $\tilde{M}$ defined in (56). Hence, Proposition 12 gives the desired statement. □

9.3. Asymptotic of Jacobi ensemble.

In this section we review the asymptotic results for the Jacobi ensemble $\mathbf{J}(N; p, q)$ introduced in the Definition 3 as $N \to \infty$.

We assume that as $N \to \infty$, also $p, q \to \infty$, in such a way that

$$(71) \qquad \frac{p-1}{N} = \frac{1}{2}(\mathfrak{p} - 1), \quad \mathfrak{p} \geq 1, \qquad \frac{q-1}{N} = \frac{1}{2}(\mathfrak{q} - 1), \quad \mathfrak{q} \geq 1,$$

where $\mathfrak{p}$ and $\mathfrak{q}$ are two parameters, which stay bounded away from 1 and from ∞ as $N \to \infty$.[23] We further define the *equilibrium measure* $\mu_{\mathfrak{p},\mathfrak{q}}$ of the Jacobi ensemble through:

$$(72) \qquad \mu_{\mathfrak{p},\mathfrak{q}}(x)\, dx = \frac{\mathfrak{p} + \mathfrak{q}}{2\pi} \cdot \frac{\sqrt{(x - \lambda_-)(\lambda_+ - x)}}{x(1 - x)} \mathbf{1}_{[\lambda_-, \lambda_+]}\, dx,$$

where the support $[\lambda_-, \lambda_+]$ of the measure is defined via

$$(73) \qquad \lambda_\pm = \frac{1}{(\mathfrak{p} + \mathfrak{q})^2} \left(\sqrt{\mathfrak{p}(\mathfrak{p} + \mathfrak{q} - 1)} \pm \sqrt{\mathfrak{q}} \right)^2.$$

One can check that $0 < \lambda_- < \lambda_+ < 1$ for every $\mathfrak{p}, \mathfrak{q} > 1$. The law (72) is known as Wachter distribution. Further, define

$$(74) \qquad c_\pm = \frac{(\mathfrak{p} + \mathfrak{q})}{2} \frac{\sqrt{\lambda_+ - \lambda_-}}{\lambda_\pm (1 - \lambda_\pm)},$$

and note that

$$\mu_{\mathfrak{p},\mathfrak{q}}(x) \approx \frac{c_\pm}{\pi} \sqrt{|x - \lambda_\pm|}, \text{ as } x \to \lambda_\pm \quad \text{inside} \quad [\lambda_-, \lambda_+],$$

where the normalization $\frac{1}{\pi}\sqrt{|x - \lambda_\pm|}$ was chosen to match the behavior of the Wigner semi-circle law $\frac{1}{2\pi}\sqrt{4 - x^2}$ near edges ± 2.

Theorem 14. *Suppose that $N, p, q \to \infty$ in such a way that $\mathfrak{p} \geq 1$ and $\mathfrak{q} \geq 1$ in (71) stay bounded. For the second conclusions we additionally assume that $\mathfrak{q}$ is bounded away from 1 and for the third conclusion we additionally require $\mathfrak{p}$ to be bounded away from 1. Let $x_1 \geq x_2 \geq \cdots \geq x_N$ be N random eigenvalues of Jacobi ensemble $\mathbf{J}(N; p, q)$. Then*

$$(\mathrm{I}) \quad \lim_{N \to \infty} \left| \frac{1}{N} \sum_{i=1}^N \delta_{x_i} - \mu_{\mathfrak{p},\mathfrak{q}} \right| = 0, \text{ weakly in probability.}$$

---

[23]The use of $p-1$ and $q-1$ instead of p and q is related to the choice of notations in (18). The following theorems would work without this shift as well, however, the shift leads to a faster speed of convergence, cf. Johnstone (2008, Discussion before Theorem 1).

> This means that for any continuous function $f(x)$ we have convergence in probability:

$$(75) \qquad \lim_{N \to \infty} \left| \frac{1}{N} \sum_{i=1}^{N} f(x_i) - \int_0^1 f(x) \mu_{\mathfrak{p},\mathfrak{q}}(x) dx \right| = 0.$$

(II) *For $\{\mathfrak{a}_i\}_{i=1}^{\infty}$ as in Proposition 1, we have convergence in finite-dimensional distributions for the largest eigenvalues:*

$$(76) \qquad \lim_{N \to \infty} \left\{ N^{2/3} c_+^{2/3} (x_i - \lambda_+) \right\}_{i=1}^{N} \to \{\mathfrak{a}_i\}_{i=1}^{\infty}.$$

> *In particular, $N^{2/3} c_+^{2/3} (x_1 - \lambda_+)$ converges to the Tracy-Widom distribution F_1.*

(III) *We also have convergence in distribution for the smallest eigenvalues[24]*

$$(77) \qquad \lim_{N \to \infty} \left\{ N^{2/3} c_-^{2/3} (\lambda_- - x_{N+1-i}) \right\}_{i=1}^{N} \to \{\mathfrak{a}_i\}_{i=1}^{\infty}.$$

The first proof of (75) appeared in Wachter (1980), for other proofs see Bai and Silverstein (2010) and Dumitriu and Paquette (2012); we follow the notations of the last reference. The asymptotics (76), (77) can be found in Johnstone (2008), and it is a manifestation of the universality for the distributions of largest/smallest eigenvalues of random matrices holding in much wider generality, see, e.g., Deift and Gioev (2007), Erdos and Yau (2012), Tao and Vu (2012).

9.4. Proof of Theorem 2.

As an intermediate step we prove the following theorem.

Theorem 15. *Suppose that $T, N \to \infty$ in such a way that $T/N > 2$ remains bounded away from 2 and from ∞. Let $\lambda_1 \geq \cdots \geq \lambda_N$ be eigenvalues of the matrix $S_{10} S_{00}^{-1} S_{01} S_{11}^{-1}$ defined as in Theorem 2 and Proposition 11. Then for $\{\mathfrak{a}_i\}_{i=1}^{\infty}$ as in Proposition 1, we have convergence in finite-dimensional distributions for the largest eigenvalues:*

$$(78) \qquad \lim_{N \to \infty} \left\{ N^{2/3} c_+^{2/3} (\lambda_i - \lambda_+) \right\}_{i=1}^{\infty} \to \{\mathfrak{a}_i\}_{i=1}^{\infty},$$

where

$$(79) \qquad \lambda_{\pm} = \frac{1}{(\mathfrak{p} + \mathfrak{q})^2} \left[\sqrt{\mathfrak{p}(\mathfrak{p} + \mathfrak{q} - 1)} \pm \sqrt{\mathfrak{q}} \right]^2, \qquad c_+ = (\mathfrak{p} + \mathfrak{q}) \frac{\sqrt{\lambda_+ - \lambda_-}}{2 \cdot \lambda_+ \cdot (1 - \lambda_+)},$$

$$\mathfrak{p} = 2 - \frac{2}{N}, \qquad \mathfrak{q} = \frac{T}{N} - 1 - \frac{2}{N}.$$

Proof. Theorem 4 implies that the asymptotics of largest eigenvalues $x_1, x_2, \dots$ is the same as the one for the largest eigenvalues of $\mathbf{J}(N; \frac{N}{2}, \frac{T-2N}{2})$. For the latter we use Theorem 14 with $\mathfrak{p} = 2 - \frac{2}{N}$, $\mathfrak{q} = \frac{T}{N} - 1 - \frac{2}{N}$. $\qquad \square$

---

[24]The limiting processes $\{\mathfrak{a}_i\}_{i=1}^{\infty}$ arising for the largest and smallest eigenvalues are independent.

Proof of Theorem 2. We use Theorem 15 and the fact that for small x

$$\ln\big(1 - (\lambda_+ + x)\big) = \ln(1 - \lambda_+) - \frac{1}{1 - \lambda_+} x + o(x). \quad \square$$

9.5. Asymptotics under white noise assumption. In this section we explain how the proof of Theorem 5 is obtained. We follow the notations of Section 9.2 for the deterministic operators and spaces. We take a uniformly-random orthogonal (or unitary if $\beta = 2$) operator $\tilde{O}$ acting in $(T - 1)$–dimensional space V and define an operator $\tilde{W}$ acting in V:

$$\tilde{W} = I_V - \tilde{O}(F_V)^{-1}\tilde{O}^*,$$

Let us introduce a symmetric (or Hermitian if $\beta = 2$) $N \times N$ matrix $\widehat{M}$ through

$$(80) \qquad \widehat{M} = [\tilde{W}]_{NN}^V([\tilde{W}^*\tilde{W}]_{NN}^V)^{-1}[\tilde{W}^*]_{NN}^V.$$

Proposition 16. *Choose an arbitrary positive definite covariance matrix Λ. Let ε be $N \times T$ matrix of random variables (real if $\beta = 1$ and complex if $\beta = 2$), such that T columns of ε are i.i.d., and each of them is an N-dimensional mean zero Gaussian vector with covariance Λ. Fix an arbitrary N–dimensional vector μ. Define the $N \times T$ matrix $X = (X_1, X_2, \ldots, X_T)$ as in Eq. (23) via*

$$X_t = \mu + \varepsilon_t, \quad t = 1, \ldots, T.$$

Further set $\Delta^c X_t = X_{t+1} - X_t$, $t = 1, \ldots, T - 1$ and $\Delta^c X_T = X_1 - X_t$. Define $N \times N$ matrices:

$$(81) \quad S_{00}^{w.n.} = \Delta^c X \mathcal{P}(\Delta^c X)^*, \quad S_{01}^{w.n.} = \Delta^c X \mathcal{P} X^*, \quad S_{10}^{w.n.} = X \mathcal{P}(\Delta^c X)^*, \quad S_{11}^{w.n.} = X \mathcal{P} X^*,$$

Then the eigenvalues of the matrix $S_{10}^{w.n.}(S_{00}^{w.n.})^{-1}S_{01}^{w.n.}(S_{11}^{w.n.})^{-1}$ have the same distribution as those of $\widehat{M}$ in (80).

Proof. The proof is similar to that of Proposition 11 and we omit many details. First, note that μ cancels out both in $X\mathcal{P}$ and in $\Delta^c X$. Hence, without loss of generality we can assume $\mu = 0$.

Next, we define

$$G = \mathcal{P}(F^{-1} - I_T)\mathcal{P}.$$

Note that $\mathcal{P}$ and $(F^{-1} - I_T)$ commute with each other. Using additionally $\mathcal{P}^2 = \mathcal{P}$, we get:

$$S_{00}^{w.n.} = \varepsilon G^*G\varepsilon^*, \quad S_{01}^{w.n.} = \varepsilon G^*\varepsilon^*, \quad S_{10}^{w.n.} = \varepsilon G\varepsilon^*, \quad S_{11}^{w.n.} = \varepsilon \mathcal{P}\varepsilon^*.$$

At this point, arguing as in the proof of Proposition 11, we can assume that the covariance matrix Λ is identical, since any other covariance matrix can be obtained at the cost of conjugation of $S_{10}^{w.n.}(S_{00}^{w.n.})^{-1}S_{01}^{w.n.}(S_{11}^{w.n.})^{-1}$, which leaves the eigenvalues unchanged.

We can further identify non-zero eigenvalues of $S_{10}^{w.n.}(S_{00}^{w.n.})^{-1}S_{01}^{w.n.}(S_{11}^{w.n.})^{-1}$ with those of the product of two projectors $P_2 P_1 P_2$ (similarly to (58)), where P_1 is the orthogonal projector

onto the space spanned by N columns of $G\varepsilon^*$ and P_2 is the orthogonal projector onto the space spanned by N columns of $\mathcal{P}\varepsilon^*$. Next, we rotate the space V, so that the span of N columns of $\mathcal{P}\varepsilon^*$ turns into the span of the first N basis vectors $\tilde{e}_1, \ldots, \tilde{e}_N$ in V. At this point we arrive at (80) with $\tilde{W}$ replaced by $-\tilde{W}$. It remains to notice that the introduction of prefactor -1 in front of $\tilde{W}$ does not change the matrix (80). $\qquad\square$

Let $\mathcal{T} = T - 1$ and let O be uniformly random real orthogonal of determinant 1 (if $\beta = 1$) or complex unitary (if $\beta = 2$) $\mathcal{T} \times \mathcal{T}$ matrix. Define $W = I_{\mathcal{T}} + O$. Then we set

$$(82) \qquad M = [W]_{NN}([W^*W]_{NN})^{-1}[W^*]_{NN}.$$

Recall that for a matrix A with real spectrum, $\lambda_i(A)$ is the ith largest eigenvalue of A.

Proposition 17. *Assume $\mathcal{T} = T - 1 \geq 2N$. One can couple M from (82) with $\widehat{M}$ from (80) in such a way that for each $\epsilon > 0$ we have*

$$\mathrm{Prob}\left(\max_{1 \leq i \leq N} |\lambda_i(M) - \lambda_i(\widehat{M})| < \frac{1}{N^{1-\epsilon}}\right) \to 1,$$

as $T, N \to \infty$ in such a way that the ratio T/N remains bounded.

The proof is almost identical to that of Proposition 12 and we omit it: the key idea is to notice that the only difference between M and $\widehat{M}$ is in the replacement of O by $-\tilde{O}(F_V)^{-1}\tilde{O}^*$; however, by the results of Meckes and Meckes (2013) the latter two matrices can be coupled so that their eigenvectors are the same, while eigenvalues are very close to each other.

Combining Proposition 16, Proposition 17, and Theorem 7, we arrive at the statement of Theorem 5.

9.6. Data. As of May 12, 2020, S&P100 consists of the following companies: Apple Inc., AbbVie Inc., Abbott Laboratories Accenture, Adobe Inc., American International Group Allstate, Amgen Inc., American Tower, Amazon.com, American Express Boeing Co., Bank of America Corp, Biogen, The Bank of New York Mellon, Booking Holdings, BlackRock Inc, Bristol-Myers Squibb, Berkshire Hathaway, Citigroup Inc, Caterpillar Inc., Charter Communications, Colgate-Palmolive, Comcast Corp., Capital One Financial Corp., ConocoPhillips, Costco Wholesale Corp., salesforce.com, Cisco Systems, CVS Health, Chevron Corporation, DuPont de Nemours Inc., Danaher Corporation, The Walt Disney Company, Dow Inc., Duke Energy, Emerson Electric Co., Exelon, Ford Motor Company, Facebook Inc., FedEx, General Dynamics, General Electric, Gilead Sciences, General Motors, Alphabet Inc. (Class C), Alphabet Inc. (Class A), Goldman Sachs, Home Depot, Honeywell, International Business Machines, Intel Corp., Johnson & Johnson, JPMorgan Chase & Co., Kraft Heinz, Kinder Morgan, The Coca-Cola Company, Eli Lilly and Company, Lockheed Martin, Lowe's, MasterCard Inc, McDonald's Corp, Mondelēz International, Medtronic plc, MetLife Inc., 3M

Company, Altria Group, Merck & Co., Morgan Stanley, Microsoft, NextEra Energy, Netflix, Nike Inc., NVIDIA Corp., Oracle Corporation, Occidental Petroleum Corp., PepsiCo, Pfizer Inc, Procter & Gamble Co, Philip Morris International, PayPal Holdings, Qualcomm Inc., Raytheon Technologies, Starbucks Corp., Schlumberger, Southern Company, Simon Property Group, Inc., AT&T Inc, Target Corporation, Thermo Fisher Scientific, Texas Instruments, UnitedHealth Group, Union Pacific Corporation, United Parcel Service, U.S. Bancorp, Visa Inc., Verizon Communications, Walgreens Boots Alliance, Wells Fargo, Walmart, Exxon Mobil Corp.

Eight of the above companies are not available for the entire period under consideration $01.01.2010 - 01.01.2020$. Those companies are AbbVie Inc. (founded in 2013), Charter Communications (was bancrupt in 2009 and got released on NASDAQ in the middle of 2010), Dow Inc. (was spun off of DowDuPont on April 1, 2019), Facebook Inc. (went on IPO on February 1, 2012), General Motors (approached bankruptcy in 2009 and returned to NASDAQ as a new company at the end of 2010), Kraft Heinz (Kraft Foods and H.J. Heinz merged into Kraft Heinz in 2015), Kinder Morgan (was taken in a buyout and began trading again on the NYSE on February 11, 2011), PayPal Holdings (was part of eBay until 2015).

References

Anderson, G., A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, 2010.

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.

Bejan, A., "Largest eigenvalues and sample covariance matrices. Tracy-Widom and painlevé ii: computational aspects and realization in s-plus with applications," *Preprint: http://www. vitrum. md/andrew/MScWrwck/TWinSplus.pdf*, 2005.

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 $\beta-$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.

Bykhovskaya, A., "Time Series Approach to the Evolution of Networks: Prediction and Estimation," *Working Paper*, 2020.

Bykhovskaya, A. and V. Gorin, "Jacobi Ensemble and Cointegration in VAR(k)," *Working Paper*, 2021.

Cavaliere, G., A. Rahbek, and A.R. Taylor, "Bootstrap determination of the cointegration rank in vector autoregressive models," *Econometrica*, 2012, *80* (4), 1721–1740.

Deift, P. and D. Gioev, "Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices," *Communications on Pure and Applied Mathematics*, 2007, *69* (6), 867—-910.

Dumitriu, I. and A. Edelman, "Matrix models for beta ensembles," *Journal of Mathematical Physics*, 2002, *43* (11), 5830–5847.

Dumitriu, I. and E. Paquette, "Global fluctuations for linear statistics of β-Jacobi ensembles," *Random Matrices: Theory and Applications*, 2012, *1* (04), 1250013.

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.

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.

Holcomb, D. and G.R. Moreno Flores, "Edge scaling of the β-Jacobi ensemble," *Journal of Statistical Physics*, 2012, *149* (9), 1136–1160.

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," *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.

Meckes, E. and M. Meckes, "Spectral measures of powers of random matrices," *Electronic communications in probability*, 2013, *18*.

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.

Wachter, K.W., "The limiting empirical measure of multiple discriminant ratios," *Annals of Statistics*, 1980, *8* (5), 937–957.

Xiao, Z. and P.C.B. Phillips, "Efficient detrending in cointegrating regression," *Econometric Theory*, 1999, *15* (4), 519–548.

University of Wisconsin-Madison

E-mail address, A.B.: `anna.bykhovskaya@wisc.edu`

E-mail address, V.G.: `vadicgor@gmail.com`