ASYMPTOTICS OF COINTEGRATION TESTS FOR HIGH-DIMENSIONAL VAR(k)

ANNA BYKHOVSKAYA AND VADIM GORIN

Abstract. The paper studies nonstationary high-dimensional vector autoregressions of order k, VAR(k). Additional deterministic terms such as trend or seasonality are allowed. The number of time periods, T, and the number of coordinates, N, are assumed to be large and of the same order. Under this regime the first-order asymptotics of the Johansen likelihood ratio (LR), Pillai–Bartlett, and Hotelling–Lawley tests for cointegration are derived: the test statistics converge to nonrandom integrals. For more refined analysis, the paper proposes and analyzes a modification of the Johansen test. The new test for the absence of cointegration converges to the partial sum of the Airy_1 point process. Supporting Monte Carlo simulations indicate that the same behavior persists universally in many situations beyond those considered in our theorems.

The paper presents empirical implementations of the approach for the analysis of S&P100 stocks and of cryptocurrencies. The latter example has a strong presence of multiple cointegrating relationships, while the results for the former are consistent with the null of no cointegration.

Date: November 26, 2023.

The authors would like to thank Bruce Hansen, Alexei Onatski, associate editor Zhipeng Liao, and three anonymous referees for valuable comments and suggestions. The authors are grateful to Victor Kleptsyn for his help with the proof of Lemma 24. Finally, the authors would like to thank Eszter Kiss for excellent research assistance. Gorin's work was supported by NSF grants DMS-1664619, DMS-1949820, and DMS-2246449, and BSF grant 2018248.

1. Introduction

Starting with the pioneering work of Sims [1980], vector autoregressions (VARs) became a workhorse model in macroeconomics and other fields. Many key time series in macroeconomics and finance (e.g., consumption and output) are nonstationary, and the properties of VARs can be very different depending on whether one is dealing with a stationary or nonstationary series. Moreover, there is a further subdivision to be accounted for in the case of nonstationary series: it is important to understand whether the data are cointegrated—that is, whether there exists a stationary nontrivial linear combination within the considered series (e.g., the log of consumption minus the log of output is stationary while the series themselves have unit roots).

Classical tools for testing cointegration (see, e.g., Johansen [1995], Maddala and Kim [1998], and Juselius [2006]) fail to achieve the desired finite sample performance when the number of time series, N, is large. Thus, they are not commonly used in such settings, and the design of proper tools to handle cointegration under a large N remained an open problem for years (see, e.g., Choi [2015, Sections 2.3.3, 2.4]). Recently Onatski and Wang [2018, 2019] and Bykhovskaya and Gorin [2022] have opened a new avenue based on the "T/N converging to a constant" asymptotic regime. However, the testing procedures of these texts cover only VAR(1), while, in practice, researchers rarely confine themselves to VARs of order 1, instead usually considering at least two lags. Indeed, as noted already in Pagan [1987], "most applications of Sims' methodology have put the number of lags between four and ten." Since Pagan [1987] the lengths of available time series and computing power have only increased, thus allowing researchers to work with even more complex models. Hence, it is important to generalize and extend the above papers to a VAR(k) setting, which is the main topic of our text.

Our paper analyses a family of tests for the absence of cointegration for nonstationary VAR(k), such as the Johansen likelihood ratio (LR) test (Johansen [1988, 1991]) and related Hotelling–Lawley and Pillai–Bartlett tests (see, e.g., Gonzalo and Pitarakis [1995] and references therein) as N and T jointly and proportionally go to infinity. The shared feature of these tests is that their statistics are based on the squared sample canonical correlations between certain transformations of current changes and past levels of the data. The main contribution of our paper is in the asymptotic analysis of these canonical correlations. First, we show that for VAR(k) with general k, under the null of no cointegration (and some additional technical conditions) the empirical distribution of the squared sample canonical correlations converges to the Wachter distribution. As a corollary, we deduce the first-order deterministic limits of the above test statistics. Second, we introduce a modification of the testing procedure and prove much more refined results in the modified setting. By computing

the exact asymptotic behavior of the probability distributions of individual canonical correlations after proper recentering and rescaling, we are able to compute the critical values for the test of no cointegration with correct asymptotic size as N and T jointly and proportionally go to infinity.

We remark that there is a wide scope of literature devoted to the corrections of Johansen's LR test and its relatives (originally developed based on fixed N, large T asymptotics) for large values of N (see, e.g., Reinsel and Ahn [1992], Johansen [2002], Swensen [2006], Cavaliere et al. [2012], and Onatski and Wang [2019]). The distinguishing feature of our work is that we are not trying to correct the finite N asymptotic statements, which stop working for large N, by introducing various empirical adjustments. Instead, we develop a theoretical framework for working with the large N case directly. One advantage is that our approach explains the general phenomenology and predicts the asymptotic behavior. As a result, the empirical or simulational adjustments for particular values of the parameters of the model are no longer needed.

To achieve the above, in our proofs we use the VAR(1) results of Bykhovskaya and Gorin [2022] as a cornerstone. The main technical work is devoted to producing recursive arguments, which reduce the VAR(k) behavior to that of VAR($k-1$) and eventually to VAR(1). The central role is played by highly nontrivial projections from the group of orthogonal $T\times T$ matrices to the smaller subgroup of orthogonal $(T-N)\times(T-N)$ matrices. While such projections have previously been used in asymptotic representation theory, to the authors' knowledge, this is their first appearance in the econometrics or statistics context. Thus, many new properties of those projections need to be developed in our framework.

The rest of the paper is organized as follows. Section 2 describes our setting and provides the first asymptotic results. Section 3 constructs our modified test and computes its asymptotics, while Section 4 presents supporting Monte Carlo simulations. Section 5 illustrates our theoretical findings on S&P100 data and on the prices of cryptocurrencies. Finally, Section 6 concludes. All proofs are in Sections 7.1–7.4. The accompanying R package is available at the Github `https://github.com/eszter-kiss/Largevars`.

2. First-order asymptotics of sample canonical correlations

We consider an N-dimensional vector autoregressive process of order k, VAR(k), based on a sequence of i.i.d. mean zero Gaussian[1] errors $\{\varepsilon_t\}$ with nondegenerate covariance matrix

[1]We expect that all our results continue to hold for non-normally distributed errors as long as they have enough moments (cf. such distribution-independence results in other high-dimensional models, as in Erdos and Yau [2012], Tao and Vu [2012], Han et al. [2018], and Yang [2022a]).

Λ. That is, written in the error correction form,

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

where $\Delta X_t := X_t - X_{t-1}$, D_t is a d_D-dimensional vector of deterministic terms, such as a constant, a trend or seasonality (extra explanatory variables are also allowed as long as they are observed), and $\Gamma_1, \dots, \Gamma_{k-1}$, Π, Φ are unknown parameters. The process is initialized at fixed $X_{1-k}, \dots, X_0$. We do not impose any restrictions on Λ; thus, we allow for arbitrary correlations across coordinates of X_t. In contrast, many previous approaches rely on specific properties of the covariance matrix Λ; see, e.g., Breitung and Pesaran [2008], Bai and Ng [2008, Section 7] and Zhang et al. [2018].

Remark 1. *Alternatively, the error correction form can be written as*

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \tilde{\Gamma}_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \qquad t = 1, \dots, T,$$

so that $\tilde{\Gamma}_i = \Gamma_i - \Pi$. *Whether we use the former (Eq.* (1)*) or the latter form does not affect our results. The testing procedures of our interest are based on the residuals from regressing* X_{t-k} *on* $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$*, which are the same as the residuals from regressing* $X_{t-1} = X_{t-k} + \Delta X_{t-1} + \ldots + \Delta X_{t-k+1}$ *on* $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$.

We are interested in the behavior of the squared sample canonical correlations between transformed past levels (lags) and changes (first differences) of the data X_t. As shown in [Johansen, 1988, 1991] (see also Anderson [1951]), the correlations are related to whether the process is cointegrated. To be more specific, they appear in the likelihood ratio test for the presence and rank of the cointegration. Let us formally define these correlations. Here and below * denotes matrix transposition.

Procedure 1 (Johansen [1991])**.** Let $Z_{0t} = \Delta X_t$, $Z_{1t} = (\Delta X^*_{t-1}, \dots, \Delta X^*_{t-k+1}, D^*_t)^*$, and $Z_{kt} = X_{t-k}$. We regress lags Z_{kt} and changes Z_{0t} on regressors Z_{1t} (lagged changes and deterministic terms) and define the residuals

$$R_{it} = Z_{it} - \left(\sum_{\tau=1}^{T} Z_{i\tau} Z^*_{1\tau} \right) \left(\sum_{\tau=1}^{T} Z_{1\tau} Z^*_{1\tau} \right)^{-1} Z_{1t}, \quad i = 0, k. \tag{2}$$

Define further $N \times N$ matrices $S_{ij} := \sum_{t=1}^{T} R_{it} R^*_{jt}$, $i, j = 0, k$ and finally set

$$\mathcal{C} = S^{-1}_{kk} S_{k0} S^{-1}_{00} S_{0k}.$$

The N eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ of $\mathcal{C}$ are squared sample canonical correlations of R_0 and R_k, where R_i is $N \times T$ matrix composed of columns R_{it}, $i = 0, k$.

Johansen's LR statistic for testing the hypothesis $\operatorname{rank}(\Pi) \le r_1$ (at most r_1 cointegrating relationships) versus the alternative $\operatorname{rank}(\Pi) \in (r_1, r_2]$ (between r_1 and r_2 cointegrating relationships) with $r_2 > r_1$ has the form[2]

$$\sum_{i=r_1+1}^{r_2} \ln(1-\lambda_i); \tag{3}$$

the Pillai–Bartlett statistic is

$$\sum_{i=r_1+1}^{r_2} \lambda_i; \tag{4}$$

and Hotelling–Lawley statistic is

$$\sum_{i=r_1+1}^{r_2} \frac{\lambda_i}{1-\lambda_i}. \tag{5}$$

See Gonzalo and Pitarakis [1995] for a discussion and many references about these statistics.

In Theorem 3 we show that the empirical measure of eigenvalues of $\mathcal{C}$ converges (weakly in probability) to the Wachter distribution. The theorem generalizes the results of Onatski and Wang [2018] from the VAR(1) to the VAR(k) setting.[3]

Definition 2. The Wachter distribution is a probability distribution on $[0,1]$ that depends on two parameters $\mathfrak{p} > 1$ and $\mathfrak{q} > 1$ and has density

$$\mu_{\mathfrak{p},\mathfrak{q}}(x) = \frac{\mathfrak{p}+\mathfrak{q}}{2\pi} \cdot \frac{\sqrt{(x-\lambda_-)(\lambda_+-x)}}{x(1-x)} \mathbf{1}_{[\lambda_-,\lambda_+]}, \tag{6}$$

where the support $[\lambda_-, \lambda_+] \subset (0,1)$ of the measure is defined via

$$\lambda_\pm = \frac{1}{(\mathfrak{p}+\mathfrak{q})^2} \left(\sqrt{\mathfrak{p}(\mathfrak{p}+\mathfrak{q}-1)} \pm \sqrt{\mathfrak{q}} \right)^2. \tag{7}$$

Theorem 3. *Let X_t follow Eq. (1). Suppose that k is fixed and, as $N \to \infty$,*

$$\lim_{N\to\infty} \frac{T}{N} = \tau > (k+1) \qquad \textit{and} \tag{8}$$

$$\lim_{N\to\infty} \frac{1}{N} \big(\operatorname{rank}(\Pi) + \operatorname{rank}(\Gamma_1) + \operatorname{rank}(\Gamma_2) + \cdots + \operatorname{rank}(\Gamma_{k-1}) + d_D \big) = 0. \tag{9}$$

Then, for each continuous function $f(x)$ on $x \in [0,1]$, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(\lambda_i) = \int_0^1 f(x) \mu_{2,\tau-k}(x)\, \mathrm{d}x, \qquad \textit{in probability.} \tag{10}$$

[2]We omit a usual scaling factor of T for the statistics (3), (4), and (5).

[3]While Onatski and Wang [2018, Theorem 1] allows the data to be VAR(k) under restriction (9), they construct the matrix $\mathcal{C}$ involved in the statistical testing procedures as if the data were VAR(1). That is, the formal procedure is based on a misspecified VAR(1) setting, while we use the true VAR(k) procedure. As illustrated in Section 4.3, using an underspecified VAR can lead to severe size distortions.

Equivalently, the empirical measure of eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ of $\mathcal{C}$ converges (weakly in probability) to the Wachter distribution of density $\mu_{2,\tau-k}$.

Imposing assumption (9) can be viewed as a dimension reduction (cf. sparsity assumption). Approximating data with low-rank matrices is a widely used and powerful technique in data science, in machine learning applications such as recommender systems (e.g., movie preference recognition), and in computational mathematics. We refer the reader to Udell and Townsend [2019] for theoretical explanations of the suitability of low-rank models and many references to situations in which they are very efficient. In our particular context, the number of unknown parameters in the VAR model (1) is proportional to N^2, and we have access to NT observations. Since N^2 and NT are of the same order in the asymptotic regime (8), the model (1) can overfit the data. We view the rank restriction (9) as a natural way to avoid overfitting[4] (cf. the discussion in Wang et al. [2022] and Wang and Tsay [2022]). Section 5 illustrates that the results obtained under this assumption are consistent with the behavior of large-dimensional financial datasets. Another setting in which we can expect (9) to be satisfied is when there are a few special coordinates in X_t, e.g., some macroeconomic indicators, that mostly drive the behavior of the entire vector X_t. This would correspond to the case where the columns of Γ_i corresponding to those indicators are nonzero while the other columns are zero.

Figure 1 illustrates Theorem 3 for independent standard normal errors and $k = 2$, $N = 150$, $T = 1500$. The parameters are $\Phi D_t = 1_N$, $\Gamma_1 = 0.95E_{12}$, $\Pi = -0.1E_{\cdot 1}$, where 1_N is an $N \times 1$-column matrix of ones, E_{12} is an $N \times N$ matrix with one at the intersection of the 1st row and the 2nd column and zeros everywhere else, and $E_{\cdot 1}$ is an $N \times N$ matrix with ones in the first column and zeros everywhere else. Thus, all the matrices have rank one. The parameters of the Wachter distribution are $\mathfrak{p} = 2$, $\mathfrak{q} = T/N - k = 8$. The single separated (rightmost) eigenvalue corresponds to $\operatorname{rank}(\Pi) = 1$, i.e., one cointegrating relationship. Generally, we expect that if there are r separated eigenvalues and the value of k in Procedure 1 is correctly specified, then there are at least r cointegrating relationships.

The proof of Theorem 3 is based on treating the setting of (1) as a small-rank perturbation of a more restrictive setting analyzed in Section 3. The small-rank assumption in (9) is crucial for the validity of the theorem, and we expect that the asymptotic behavior changes

[4]An alternative way to introduce a low-rank assumption into the VAR model is, instead of using the error correction form in (1), to rewrite the evolution as $X_t = \sum_{i=1}^{k} A_i X_{t-i} + \Phi D_t + \varepsilon_t$. Then, in the spirit of factor models, one can impose the low-rank assumption on A_i, $i = 1, \ldots, k$. Notice that $A_i = \Gamma_i - \Gamma_{i-1}$ for $i = 2, \ldots, k-1$, so that low-rank assumptions on higher-order lags in this and our setting are related. This alternative low-rank restriction complements ours via the rank of Π: in our setting, the number of cointegrating relationships grows sublinearly in N, while in the alternative setting, this number is close to N. While none of our theorems directly cover the factor setting, numeric simulations in Section 4.5 indicate that the tests that we develop remain useful.

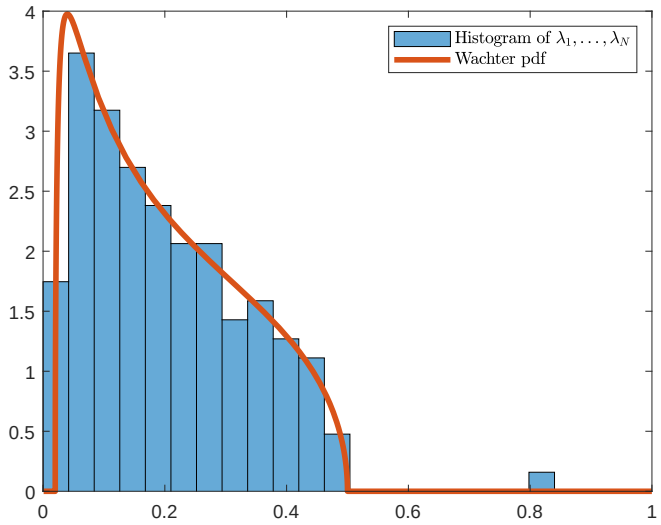


FIGURE 1. Illustration of Theorem 3: Eigenvalues and Wachter distribution. Data generating process: $\Delta X_t = 1_N + 0.95E_{12}\Delta X_{t-1} - 0.1E_{.1}X_{t-2} + \varepsilon_t$, $T = 1500$, $N = 150$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$.

in situations when (9) fails. This expectation is supported by the Monte Carlo experiment in Section 4.4. In contrast, the assumption that $T > (k+1)N$ can potentially be relaxed. When $T < (k+1)N$, the matrix $\mathcal{C}$ has deterministic eigenvalues equal to 1, which should be taken into account in $N \to \infty$ asymptotics. This case can be also addressed by our methods, but we do not continue in this direction.

An important corollary of Theorem 3 is that it provides the asymptotic behavior of various tests constructed from eigenvalues of $\mathcal{C}$.

Corollary 4. *Under the assumptions of Theorem 3, suppose that the ranks $r_1 = r_1(N)$ and $r_2 = r_2(N)$ are such that*

$$\lim_{N\to\infty} \frac{r_1}{N} = \rho_1, \quad \lim_{N\to\infty} \frac{r_2}{N} = \rho_2.$$

Let $F(x) = \int_x^1 \mu_{2,\tau-k}(z)\,\mathrm{d}z$. Then we have convergence in probability for the test statistics:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=r_1+1}^{r_2} \ln(1-\lambda_i) = \int_{F^{-1}(\rho_2)}^{F^{-1}(\rho_1)} \ln(1-x)\mu_{2,\tau-k}(x)\,\mathrm{d}x, \qquad \textit{if} \ \ \rho_1 > 0;$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=r_1+1}^{r_2} \lambda_i = \int_{F^{-1}(\rho_2)}^{F^{-1}(\rho_1)} x\mu_{2,\tau-k}(x)\,\mathrm{d}x;$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=r_1+1}^{r_2} \frac{\lambda_i}{1-\lambda_i} = \int_{F^{-1}(\rho_2)}^{F^{-1}(\rho_1)} \frac{x}{1-x}\mu_{2,\tau-k}(x)\,\mathrm{d}x, \qquad \textit{if} \ \ \rho_1 > 0;$$

and asymptotic inequalities: for each $\varepsilon > 0$,

$$\lim_{N\to\infty} \operatorname{Prob}\left(\frac{1}{N}\sum_{i=r_1+1}^{r_2} \ln(1-\lambda_i) \le \int_{F^{-1}(\rho_2)}^{1} \ln(1-x)\mu_{2,\tau-k}(x)\,\mathrm{d}x + \varepsilon\right) = 1, \qquad \textit{if } \rho_1 = 0;$$

$$\lim_{N\to\infty} \operatorname{Prob}\left(\frac{1}{N}\sum_{i=r_1+1}^{r_2} \frac{\lambda_i}{1-\lambda_i} \ge \int_{F^{-1}(\rho_2)}^{1} \frac{x}{1-x}\mu_{2,\tau-k}(x)\,\mathrm{d}x - \varepsilon\right) = 1, \qquad \textit{if } \rho_1 = 0.$$

Note that, when $\rho_1 = 0$, for the Johansen LR and Hotelling–Lawley (HW) statistics, we obtain inequalities rather than equalities. This is because of the singularity of $\ln(1-\lambda)$ and $\frac{\lambda}{1-\lambda}$ at $\lambda = 1$: while Eq. (10) controls average behavior, it does not control individual eigenvalues. Hence, the largest eigenvalue λ_1 can be arbitrarily close to 1, so that the LR and HW statistics reach large negative and positive values, respectively. However, for similar statistics in the modified setting of the next section the inequalities turn into equalities (as can be proven by combining Theorem 8 and Proposition 11).

For commonly used tests, one often takes $r_2 = N$ (i.e., $H_1 : \operatorname{rank}(\Pi) \le N$), in which case $F^{-1}(\rho_2) = 0$. We also remark that the integrals in Corollary 4 can be explicitly computed in many situations. For instance, the one appearing in the asymptotic of the Pillai–Bartlett statistic for $\rho_1 = 0$, $\rho_2 = 1$ is

$$\int_0^1 x\,\mu_{2,\tau-k}(x)\,\mathrm{d}x = \frac{2}{\tau+2-k}.$$

There are several applications of Theorem 3 and Corollary 4:

- They can be used for validation of the applicability of model (1) to a given dataset. Namely, if a VAR(k) model with low-rank matrices Γ_i and Π agrees with data, then irrespective of the true values of these parameters, we expect to see the Wachter distribution in the histogram of λ_i, $1 \le i \le N$. In Section 5 we perform such a validation on S&P100 and cryptocurrency data sets for VAR(k) with $1 \le k \le 4$ and observe a remarkable match. (VAR(1) for S&P100 is also reported in [Bykhovskaya and Gorin, 2022, Figure 7].)
- They can be used as a screening device for preliminary conclusions about the rank of Π: If the rank is finite, then for any r_1 and r_2 we should be in the ε-neighborhood of the limits in Corollary 4.
- As explained in Onatski and Wang [2018], such results can be used to explain over-rejection in some of the widely used tests for the rank of Π.

To draw further economical and statistical conclusions and to develop precise statistical tests and their critical values, one needs to go beyond the first-order asymptotic results of Theorem 3 and Corollary 4. In the next section we introduce relevant modifications and develop appropriate second-order asymptotics.

3. Cointegration test: Second-order asymptotics

In the regime of N and T growing simultaneously and proportionally, the first-order asymptotics of tests based on the squared sample canonical correlations are given in Corollary 4. To perform testing and be able to reject at a given significance level, we need to be more precise and find a centered limit, which would be a random variable rather than a constant. To do this, we need to impose additional conditions on Γ_i, D_t, and Φ in Eq. (1). Let us first describe the modified procedure and then state the asymptotic results.

3.1. Test. We restrict our attention to the case $D_t = 1$, i.e.,

$$\Delta X_t = \mu + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-k} + \varepsilon_t, \qquad t = 1, \dots, T. \tag{11}$$

The null hypothesis of no cointegration is $H_0 :\ \operatorname{rank}(\Pi) = 0$ or $\Pi \equiv 0$. The complement to H_0 is $\operatorname{rank}(\Pi) > 0$. However, to design our test we use an alternative hypothesis:

$$H(r) : \quad \operatorname{rank}(\Pi) \in [1, r].$$

As in Bykhovskaya and Gorin [2022], our test is based on a modification of the Johansen LR test. The Johansen LR test for the original H_0 (i.e., $\Pi \equiv 0$) versus $H(r)$ is

$$\sum_{i=1}^{r} \ln(1 - \lambda_i), \tag{12}$$

where $\lambda_1, \lambda_2, \dots$ are defined in Procedure 1. Let us describe how our modified test proceeds.

Procedure 2. Step 1. De-trend the data and define

$$\tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0). \tag{13}$$

Note that we do a time shift in line with the notation in Bykhovskaya and Gorin [2022].

Step 2. Define regressors and dependent variables: For any $a \in \mathbb{Z}$, set

$$a \mid T = a + kT, \quad \text{where } k \in \mathbb{Z} \text{ is such that } a + kT \in \{1, 2, \dots, T\}.$$

Define

$$\tilde{Z}_{0t} = \Delta X_{t|T} \equiv \Delta X_t, \quad \tilde{Z}_{kt} = \tilde{X}_{t-k+1|T}, \quad \tilde{Z}_{1t} = (\Delta X^*_{t-1|T}, \dots, \Delta X^*_{t-k+1|T}, 1)^*.$$

The main difference between $\tilde{Z}_{it}$ and Z_{it} from Procedure 1 is the usage of cyclic indices: values at $t = 0, -1, \dots$ are replaced by values at $t = T, T-1, \dots$.

Step 3. Calculate the residuals from regressions $\tilde{Z}_{0t}$ on $\tilde{Z}_{1t}$ and $\tilde{Z}_{kt}$ on $\tilde{Z}_{1t}$:

$$\tilde{R}_{it} = \tilde{Z}_{it} - \left(\sum_{\tau=1}^{T} \tilde{Z}_{i\tau} \tilde{Z}^*_{1\tau} \right) \left(\sum_{\tau=1}^{T} \tilde{Z}_{1\tau} \tilde{Z}^*_{1\tau} \right)^{-1} \tilde{Z}_{1t}, \quad i = 0, k. \tag{14}$$

Step 4. Calculate the squared sample canonical correlations between $\tilde{R}_0$ and $\tilde{R}_k$, where $\tilde{R}_i$ is an $N \times T$ matrix composed of columns $\tilde{R}_{it}$, $i = 0, k$. That is, define

$$\tilde{S}_{ij} = \sum_{t=1}^{T} \tilde{R}_{it}\tilde{R}^*_{jt}, \quad i, j = 0, k, \qquad \text{and} \tag{15}$$

$$\tilde{\mathcal{C}} = \tilde{S}_{k0}\tilde{S}_{00}^{-1}\tilde{S}_{0k}\tilde{S}_{kk}^{-1}. \tag{16}$$

Then, calculate N eigenvalues $\tilde{\lambda}_1 \geq \ldots \geq \tilde{\lambda}_N$ of the matrix $\tilde{\mathcal{C}}$. The eigenvalues solve the equation

$$\det(\tilde{S}_{k0}\tilde{S}_{00}^{-1}\tilde{S}_{0k} - \tilde{\lambda}\tilde{S}_{kk}) = 0. \tag{17}$$

Step 5. Form the test statistic

$$LR_{N,T}(r) = \sum_{i=1}^{r} \ln(1 - \tilde{\lambda}_i). \tag{18}$$

The subscript N, T in (18) indicates that we modify the 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 (see Theorem 9). Visually, rejections correspond to the case when the largest eigenvalues are separated from the rest (as in Figure 1).

One can also consider other functions of largest eigenvalues $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots$ such as Pillai–Barlett or Hotelling–Lawley statistics. The asymptotic behavior in those cases can be derived in the same way as we treat statistic (18) in Theorem 9.

An alternative way to write residuals $\tilde{R}_i$, $i = 0, k$ is via an orthogonal projector: Let $\mathcal{W}$ be a linear subspace of dimension $N(k-1)+1$ in T-dimensional vector space, spanned by vector $(1, 1, \ldots, 1)$ and all rows of matrices $(\Delta X)(L_c^i)^*$, $1 \leq i \leq (k-1)$, where L_c is a cyclic version of the conventional lag operator and L_c^i is its ith power, that is, the cyclic lag applied i times. The cyclic lag operator L_c maps a vector $(x_1, x_2, \ldots, x_T)$ to $(x_T, x_1, x_2, \ldots, x_{T-1})$. Let $P_{\perp\mathcal{W}}$ denote the projector on orthogonal complement to $\mathcal{W}$. Then,

$$\tilde{R}_0 = (\Delta X)P_{\perp\mathcal{W}}, \qquad \tilde{R}_k = \tilde{X}(L_c^{k-1})^* P_{\perp\mathcal{W}}. \tag{19}$$

3.2. Second-order asymptotics. In this section we show that, under additional restrictions, the eigenvalues $\tilde{\lambda}_i$, $i = 1, \ldots, N$ are very close (up to $N^{-1+\epsilon}$ for arbitrary $\epsilon > 0$) to a known random matrix distribution. From this result we deduce our main theorem (Theorem 9), which gives the large N, T limit of the test statistic $LR_{N,T}(r)$ in Eq. (18). Before we formally state the results, let us define the relevant random matrix distributions.

3.2.1. *Definitions.*

Definition 5. The (real) Jacobi ensemble $\mathbf{J}(N; p, q)$ is a distribution on $N \times N$ real symmetric matrices $\mathcal{M}$ of density proportional to

$$\det(\mathcal{M})^{p-1} \det(I_N - \mathcal{M})^{q-1} \, d\mathcal{M}, \qquad 0 < \mathcal{M} < I_N, \tag{20}$$

with respect to the Lebesgue measure, where $p, q > 0$ are two parameters, I_N is the $N \times N$ identity matrix, and $0 < \mathcal{M} < I_N$ means that both $\mathcal{M}$ and $I_N - \mathcal{M}$ are positive definite.

The Jacobi ensemble is a generalization of the Beta distribution to the space of square matrices (when $N = 1$, we obtain the Beta distribution). It plays a prominent role in statistics; e.g., it appears in canonical correlation analysis for independent data sets and in multivariate analysis of variance (see, e.g., Muirhead [2009]).

Definition 6. The Airy_1 point process is a random infinite sequence of reals

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

that can be defined through the following proposition.

Proposition 7 (Forrester [1993],Tracy and Widom [1996])**.** *Let X_N be an $N \times N$ matrix of i.i.d. $\mathcal{N}(0, 2)$ Gaussian random variables and let $\mu_{1;N} \geq \mu_{2;N} \geq \dots \mu_{N;N}$ be eigenvalues of $\frac{1}{2}(X_N + X_N^*)$. Then, in the sense of convergence of finite-dimensional distributions,*

$$\lim_{N\to\infty} \left\{ N^{1/6} \left(\mu_{i;N} - 2\sqrt{N} \right) \right\}_{i=1}^{N} = \{\mathfrak{a}_i\}_{i=1}^{\infty}. \tag{21}$$

The marginals of the Airy_1 point process can be calculated via various methods (see, e.g., Forrester [2010] for more details).

3.2.2. *Theorems.* The null H_0 for (11) is not a point hypothesis, as it does not specify Γ_i, $i = 1, \dots, k-1$. A simplifying procedure when we are faced with such a composite space of the maintained hypothesis is to assume some fixed values of the parameters as a proxy for the null hypothesis. Along these lines, for the next theorems we are going to introduce additional restrictions and specify the values of Γ_i, $i = 1, \dots, k-1$. Thus, our model is going to be fully specified (up to a constant μ, which will disappear in the testing procedure). We proceed to implement this approach in testing the hypothesis of no cointegration and introduce the restricted $\widehat{H}_0$[5]:

$$\widehat{H}_0 : \Pi = \Gamma_1 = \Gamma_2 = \dots = \Gamma_{k-1} = 0. \tag{22}$$

In other words, under $\widehat{H}_0$ the data generating process turns into

$$\Delta X_t = \mu + \varepsilon_t, \qquad t = 1, \dots, T, \tag{23}$$

where μ is an (unknown) N-dimensional vector.

[5]We discuss the consequences of using $\widehat{H}_0$ for testing the null H_0 after Theorem 9.

Theorem 8. *Fix $C > 0$, and suppose that $T, N \to \infty$ in such a way that $\frac{T}{N} \in [k+1+C^{-1}, C]$. For the data generating process (11) with restrictions $\widehat{H}_0$ given by (22), one can couple (i.e., define on the same probability space) the eigenvalues $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_N$ of the matrix $\tilde{S}_{k0}\tilde{S}_{00}^{-1}\tilde{S}_{0k}\tilde{S}_{kk}^{-1}$ and eigenvalues $x_1 \geq \cdots \geq x_N$ of the Jacobi ensemble $\mathbf{J}(N; \frac{N}{2}, \frac{T-(k+1)N}{2})$ in such a way that, for each $\epsilon > 0$, we have*[6]

$$\lim_{T,N\to\infty} \operatorname{Prob}\left(\max_{1\leq i\leq N} |\tilde{\lambda}_i - x_i| < \frac{1}{N^{1-\epsilon}}\right) = 1. \tag{24}$$

The proof of Theorem 8 relies on two steps. First, we modify our matrix $\tilde{\mathcal{C}}$ a bit, which leads to a surprising appearance of the Jacobi ensemble, as shown in Section 7.2. Second, in Section 7.3 we show that the distance between the original model and the modified one becomes small as $N \to \infty$.

Combining Theorem 8 with known asymptotic results for the Jacobi ensemble, which we recall in Proposition 11 in Section 7.1, we derive the asymptotics of (18) in the following theorem.

Theorem 9. *Fix $C > 0$, and suppose that $T, N \to \infty$ in such a way that $\frac{T}{N} \in [k+1+C^{-1}, C]$. For the data generating process (11) with restrictions $\widehat{H}_0$ given by (22), for each finite $r = 1, 2, \ldots,$ we have convergence in distribution for the largest eigenvalues defined in Eq. (17):*

$$\frac{\sum_{i=1}^{r} \ln(1-\tilde{\lambda}_i) - r \cdot c_1(N,T)}{N^{-2/3} c_2(N,T)} \xrightarrow[T,N\to\infty]{d} \sum_{i=1}^{r} \mathfrak{a}_i, \tag{25}$$

where

$$c_1(N,T) = \ln(1-\lambda_+), \qquad c_2(N,T) = -\frac{2^{2/3}\lambda_+^{2/3}}{(1-\lambda_+)^{1/3}(\lambda_+-\lambda_-)^{1/3}} (\mathfrak{p}+\mathfrak{q})^{-2/3} < 0, \tag{26}$$

$$\mathfrak{p} = 2, \qquad \mathfrak{q} = \frac{T}{N} - k, \qquad \lambda_\pm = \frac{1}{(\mathfrak{p}+\mathfrak{q})^2}\left[\sqrt{\mathfrak{p}(\mathfrak{p}+\mathfrak{q}-1)} \pm \sqrt{\mathfrak{q}}\right]^2. \tag{27}$$

Remark 10. *The condition $\frac{T}{N} \in [k+1+C^{-1}, C]$ is another way to require that T and N grow to infinity proportionally. For example, it is guaranteed by the joint limit (8). The role of C is only to make sure that T/N does not get too close to $k+1$ (if T/N approaches $k+1$, then λ_+ approaches 1 and c_1 explodes) or $+\infty$ (if T/N becomes large, then $\lambda_+ - \lambda_-$ and λ_+ tend to 0 at the same speed and c_2 vanishes).*

Theorem 9 gives us the basis of cointegration testing in the large N, T setting. Treating $\widehat{H}_0$ as a proxy for H_0, we can use our asymptotic results to test high-dimensional VARs for

[6]One can show that the probability in (24) is exponentially close to 1: there exists a constant $\delta > 0$, which depends on ϵ, C, and k, such that, for all N and T satisfying $\frac{T}{N} \in [k+1+C^{-1}, C]$, the probability under the limit in Eq. (24) is larger than $1 - \delta^{-1}\exp(\delta^{-1}N^{\delta})$. Analyzing the proof of Theorem 8, we can obtain this inequality by combining (66) with large deviations bounds for the smallest and largest eigenvalues of the Jacobi ensemble (see e.g., Anderson et al. [2010, Section 2.6.2] for the latter).

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

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

the presence of cointegration. Formally, to perform testing, one first needs to calculate the statistic $LR_{N,T}(r)$ following Procedure 2. We recommend using small[7] values of r, such as $r = 1, 2$, or 3. Then, one needs to calculate $\frac{LR_{N,T}(r) - r \cdot c_1(N,T)}{N^{-2/3} c_2(N,T)}$, as in Theorem 9, and compare the result with quantiles of the sum of Airy$_1$, $\sum_{i=1}^{r} \mathfrak{a}_i$. If the rescaled statistic is larger than the α quantile, we reject the null of no cointegration at the $(1 - \alpha)$ level. We report the quantiles for $r = 1, 2, 3$ in Table 1. See also Bykhovskaya et al. [2023] for more detailed tables for $r = 1, \ldots, 10$.

Note that, although the asymptotic result (25) is shown under the restrictions $\widehat{H}_0$, we believe that it extends well beyond $\widehat{H}_0$: the same asymptotic results and testing procedures continue to hold in many situations with nonzero Γ_i in Eq. (11). While we do not have a full rigorous proof, we expect the following to be true:

For the data generating process (11)*, assume that the ranks of all* Γ_i *are bounded, as are the norms of all the matrices and vectors involved in the specification of the process (see Section 8 for more details). Then conclusion* (25) *of Theorem 9 should continue to hold.*

We collect extensive evidence supporting this statement. In Section 4 we report results from Monte Carlo simulations consistent with it. Further, in Section 8.1 we present a precise mathematical conjecture in this direction and give a heuristic argument for its validity. The intuition is that generic small-rank matrices are negligible relative to the scale of the rest of the process and, thus, their addition does not change the asymptotics. For this intuition to hold, it is important to correctly specify the parameter k in the procedure to be equal to (or greater than) its true value. Otherwise (i.e., if we do not regress on the relevant ΔX_{t-i} in the procedure), the presence of Γ_i can have an effect similar to that of the presence of nonzero Π: it leads to the appearance of special highly correlated linear combinations of rows of $\tilde{R}_0$ and $\tilde{R}_k$, which changes the behavior of the largest canonical correlations $\tilde{\lambda}_i$; see also the simulations in Section 4.3.

[7]In Theorem 9 r is kept fixed as N and T grow. The role of this choice and the motivations for sticking to it are discussed in detail in [Bykhovskaya and Gorin, 2022, Section 3.2].

VAR(1)	VAR(2)	VAR(3)	VAR(4)
5.81%	5.92%	6.12%	6.95%

TABLE 2. Empirical size under no cointegration hypothesis (5% nominal level) based on VAR(k) tests, $k = 1, 2, 3, 4$. Data generating process: $\Delta X_{it} = \varepsilon_{it}$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0, 1)$, $T = 522$, $N = 92$, $MC = 1,000,000$ replications.

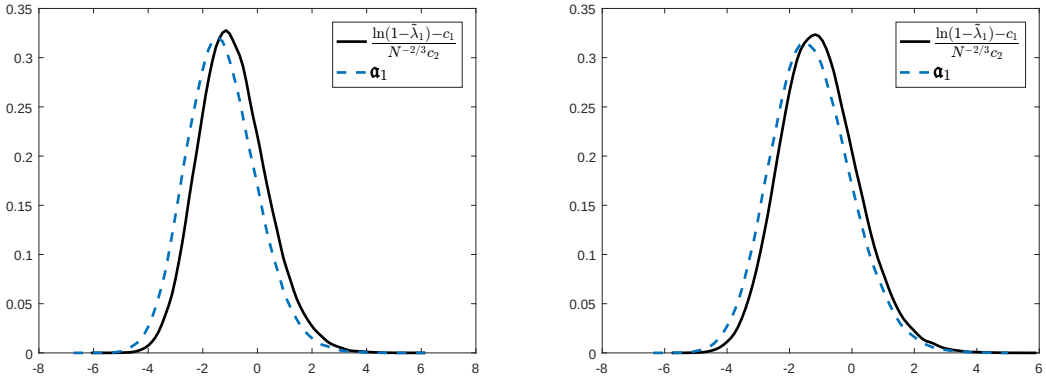
Theorem 9 means that under $\widehat{H}_0$ the largest eigenvalues $\tilde{\lambda}_i$ are close to λ_+, which is the right point of the support of the Wachter distribution in Eq. (6). The relevance of this theorem for cointegration testing stems from the fact that we expect some of the eigenvalues to be much larger than λ_+ when cointegrating relationships are present. As an illustration, see Figure 1, where Π of rank 1 leads to the largest eigenvalue being to the right of λ_+ and separated from the other eigenvalues. The separation is due to the small rank of Π. However, even if the rank of Π is large, we expect the largest eigenvalue to be significantly larger than λ_+, and, thus, the test remains relevant (see Section 4.5). Providing rigorous results on the consistency of the test is an important task for future research. We present the first result in this direction in Corollary 30 in Section 8.2, where we produce a lower bound on the power of the test against a particular "one cointegrating relationship" alternative and show that the power tends to 1 as T/N tends to infinity.

4. Monte Carlo simulations

4.1. Size. We refer to Bykhovskaya and Gorin [2022, Section 5] for the finite sample size performance of our test for $k = 1$. The results for VAR(k) are similar, and we do not show them in much detail here. For illustration purposes and to represent the comparative statics, Table 2 reports the empirical size for $T = 522$, $N = 92$ (those numbers correspond to our empirical example in Section 5.1) for tests based on VAR(k), $k = 1, 2, 3, 4$ procedures.[8] We can see that the numbers are close to the desired 5% and, for the same N and T, a lower order of VAR leads to slightly better results.

4.2. $\mathbf{H_0}$ vs. $\widehat{\mathbf{H}}_\mathbf{0}$. An important aspect of our analysis for $k > 1$ is the introduction of the additional restrictions $\widehat{H}_0$ maintained under the null. We would like to check whether Theorem 9 can hold under the less restrictive H_0 instead of $\widehat{H}_0$. Some theoretical results in this direction are provided in Section 8.1. Here we complement them with Monte Carlo simulations. For $N = 100$, $T = 500$ we simulate the data based on i.i.d. $\mathcal{N}(0, 1)$ errors ε_{it}, zero Π, and nonzero Γ_i (i.e., this corresponds to H_0 but not $\widehat{H}_0$). We then compare the density of the test based on the largest eigenvalue $\tilde{\lambda}_1$ ($r = 1$ case of Theorem 9 with statistic $\frac{\ln(1-\tilde{\lambda}_1)-c_1(N,T)}{N^{-2/3}c_2(N,T)}$), with the density of the first coordinate of the Airy$_1$ point process, $\mathfrak{a}_1$. If the

[8]Depending on the assumed order of autoregression, we have different numbers of regressors in Procedure 2.



(A) $\Gamma_1 = 0.95E_{11}$. (B) $\Gamma_1 = 0.1E_{\cdot 1} + 0.95E_{12}$.

FIGURE 2. Airy_1 and asymptotic distribution of the rescaled $\ln(1-\tilde{\lambda}_1)$ under H_0. Data generating process: $\Delta X_t = \Gamma_1 \Delta X_{t-1} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 100,000$ replications.

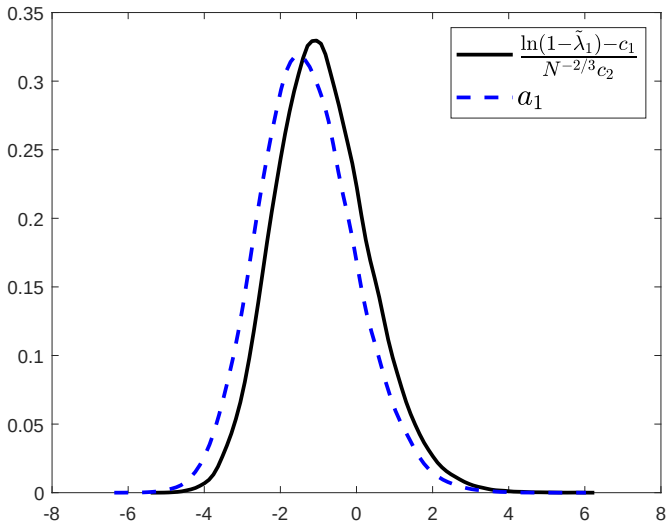


FIGURE 3. Airy_1 and asymptotic distribution of the rescaled $\ln(1-\tilde{\lambda}_1)$ under H_0. Data generating process: $\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \varepsilon_t$, $\Gamma_1 = E_{11}$, $\Gamma_2 = -\frac{2}{9}E_{11}$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 100,000$ replications.

densities coincide, then it means that we can still use the asymptotics from Theorem 9 to test the null of no cointegration.

Let E_{ij} be a matrix with 1 at the cell (i,j) and 0s everywhere else and let $E_{\cdot j}$ be a matrix with 1s filling the entire column j and 0s everywhere else. In the first two experiments we take $k = 2$. We set $\Gamma_1 = 0.95E_{11}$ in the first one, which guarantees stationarity of ΔX_t but allows for strong time correlations in the first coordinate via the 0.95 factor. In

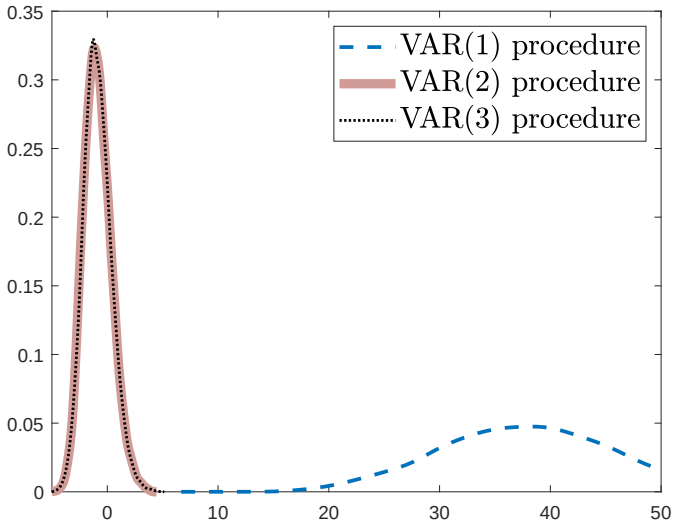


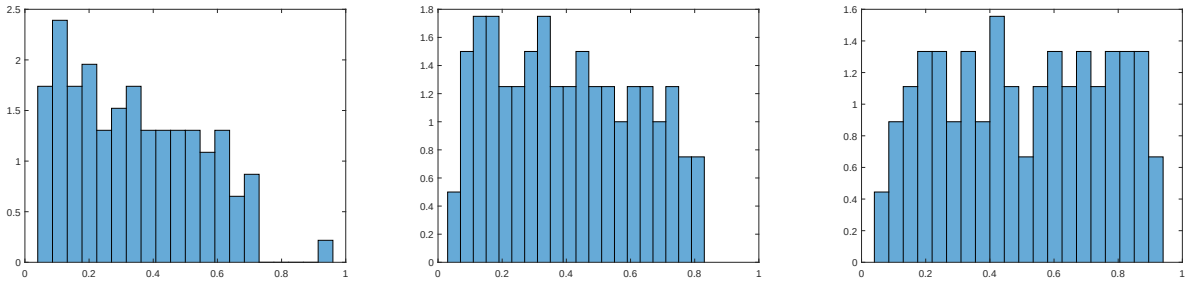
FIGURE 4. Density of rescaled $\ln(1-\tilde{\lambda}_1)$ obtained from various procedures. The correct procedure corresponds to VAR(2), $k=2$. Data generating process: $\Delta X_t = 0.95E_{11}\Delta X_{t-1} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 10,000$ replications.

this case the rank of Γ_1 is 1. In the second experiment we consider an asymmetric matrix $\Gamma_1 = 0.1E_{\cdot 1} + 0.95E_{12}$, which has a close to 1 singular value because of the 0.95 factor; the rank of Γ_1 is 2 in this case. The results are illustrated in Figure 2. In the third experiment, we take $k=3$, $\Gamma_1 = E_{11}$, and $\Gamma_2 = -\frac{2}{9}E_{11}$, so that both matrices are of rank 1. The value $-\frac{2}{9}$ guarantees stationarity of ΔX_t, since $1 - z + \frac{2}{9}z^2 = \left(1 - \frac{1}{3}z\right)\left(1 - \frac{2}{3}z\right)$. The result is shown in Figure 3. We interpret the outcomes of these three experiments as a strong argument toward the validity of an analogue of Theorem 9 well beyond the $\widehat{H}_0$ setting.[9]

4.3. Order of VAR. It is essential for the experiments in the last paragraph that the data generating process is VAR(2) and that the procedure we use also corresponds to VAR(2), i.e., $k=2$ in the notations of Sections 2 and 3. As illustrated in Figure 4, using a larger k would lead to similar results, while incorrectly using a VAR(1) procedure when the data generating process is VAR(2) would imply wrong centering and scaling. Moreover, one can spot in Figure 5 that underestimation of the order of the VAR can be misinterpreted as a presence of cointegration[10] (largest eigenvalue separated from the rest leading to the large value of the test statistic). However, as we increase the order, the largest eigenvalue

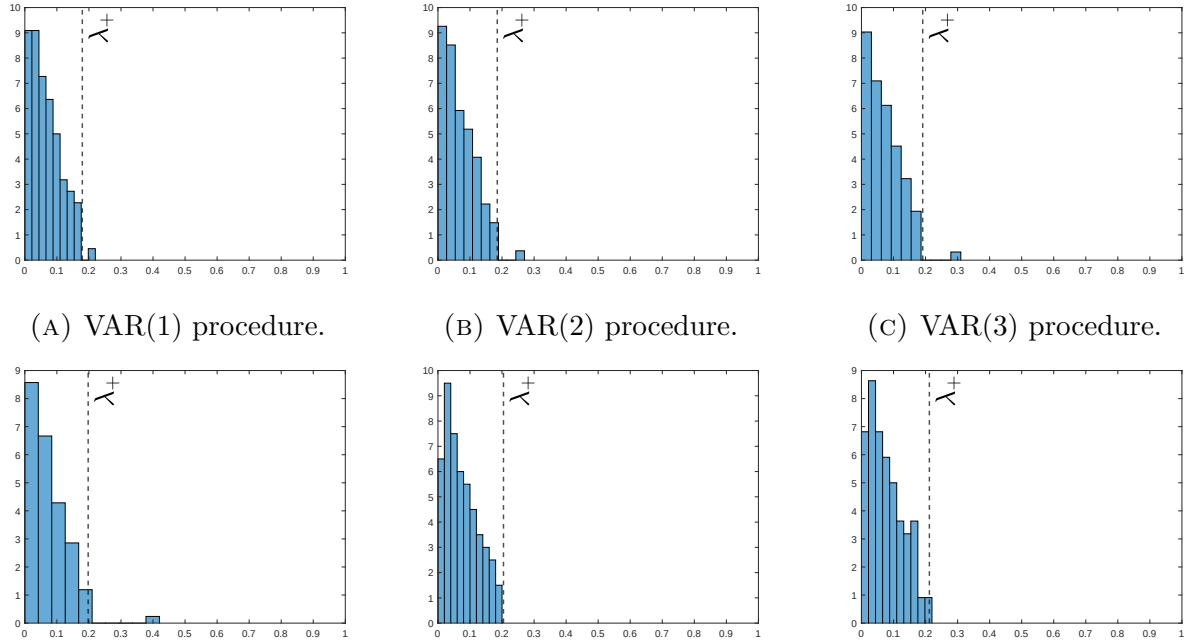
[9]The minor mismatches between densities as in Figures 2 and 3 should be expected even under $\widehat{H}_0$. Theorem 8 (after multiplication of the result by $N^{2/3}$, as in Eq. (25)) predicts errors of at least const $\cdot N^{-1/3}$ in the approximations under $\widehat{H}_0$.

[10]Related simulations for $\Gamma_1 = \theta E_{11}$ are also reported in [Bykhovskaya and Gorin, 2022, Section 7.3]: For small θ, such as $\theta = 0.5$, the VAR(1) procedure still performs well. However, as θ grows to 1, the performance quickly deteriorates ($\theta = 0.95$ in Figure 4).



(A) VAR(1) procedure. (B) VAR(2) procedure. (C) VAR(3) procedure.

FIGURE 5. Eigenvalues obtained from various procedures. The correct procedure corresponds to VAR(2), $k = 2$. Data generating process: $\Delta X_t = 0.95 E_{11} \Delta X_{t-1} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$.



(A) VAR(1) procedure. (B) VAR(2) procedure. (C) VAR(3) procedure.

(D) VAR(4) procedure. (E) VAR(5) procedure. (F) VAR(6) procedure.

FIGURE 6. Eigenvalues obtained from various procedures. The correct procedure corresponds to VAR(5), $k = 5$. Data generating process: $\Delta X_t = 0.9 E_{11} \Delta X_{t-4} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 3000$, $N = 100$.

becomes inseparable from the rest, and no sign of false cointegration remains present. Thus, practitioners are encouraged to experiment with the order of the VAR to make sure that they are detecting cointegration and not simply using the wrong model.

Note that one should be careful if using classical information criteria for estimating the order k of a VAR in our situation. They are known to be unreliable in high-dimensional settings and may underestimate k (see, e.g., the simulations in Gonzalo and Pitarakis [2002]).

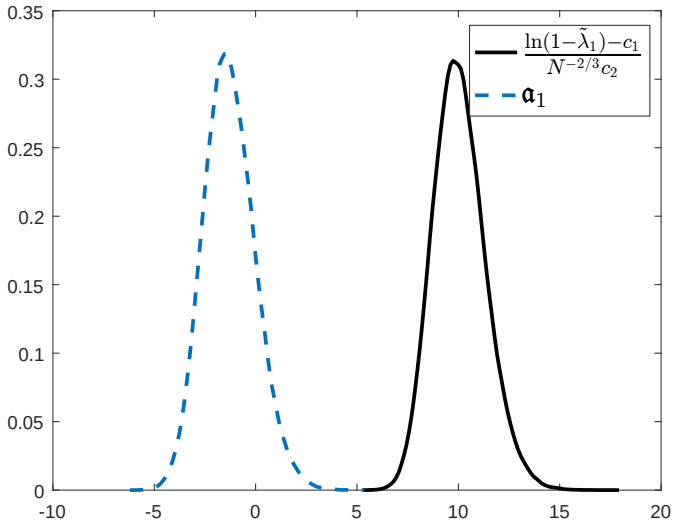


FIGURE 7. Airy_1 and asymptotic distribution of the rescaled $\ln(1-\tilde{\lambda}_1)$ under H_0 with Γ_1 of full rank. Data generating process: $\Delta X_{it} = 0.95\Delta X_{it-1} + \varepsilon_{it}$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 100,000$ replications.

A possible approach to choosing k is to look sequentially at histograms of eigenvalues at $k = 1, 2, \ldots$. If the outlier eigenvalues larger than λ_+ exist for the procedures with all k and perhaps move closer to λ_+ as k grows (corresponding to a decrease in the power of the test), then this is a strong indication of the presence of cointegration. On the other hand, if there is a sharp transition—i.e., outlier eigenvalues are present when we use the VAR(k') procedure for $k' < k$ and abruptly disappear at $k = k'$—then this is an indication that the true model is VAR(k) without cointegration (see Figure 6).

All of the above reinforces the importance of using the VAR(k) rather than the VAR(1) procedure.

4.4. Small ranks. To illustrate the importance of small ranks (e.g., (9) in Theorem 3), we also redo the same procedure for a matrix Γ_1 of full rank and set Γ_1 to be $0.95I_N$, where I_N is an $N \times N$ identity matrix. The result is shown in Figure 7. While the shape and the scale (corresponding to $N^{2/3}$ rescaling in Theorem 9) of the distribution remain similar, the location changes. Thus, the small-rank restriction of Eq. (9) is important not only in the context of Theorem 3 but also for correct centering in possible generalizations of Theorem 9.

4.5. Power. Finally, we simulate the process based on $\Pi \neq 0$ to assess the power of our cointegration testing procedure. We refer the reader to Bykhovskaya and Gorin [2022, Section 5.2] for many simulations in the $k = 1$ case and do not repeat similar experiments here.

For the first experiment, we use $k = 2$, $\Gamma_1 = 0$, and $\Pi = -0.95E_{11}$. Figure 8 shows the results of the simulation. Two curves are separated; moreover, the black straight line (test

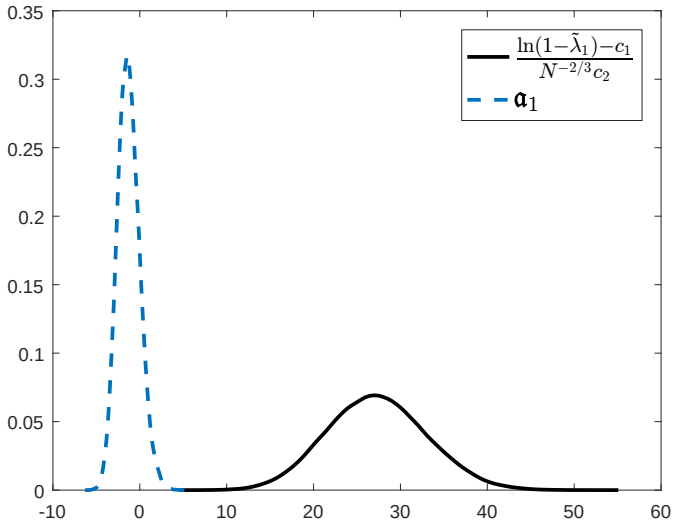
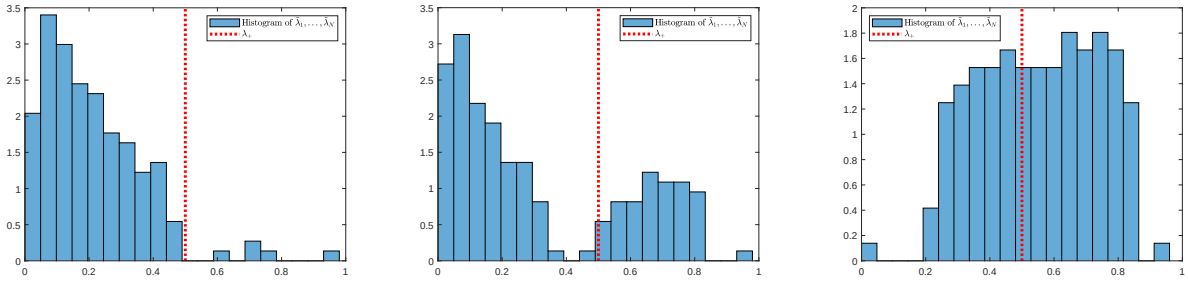


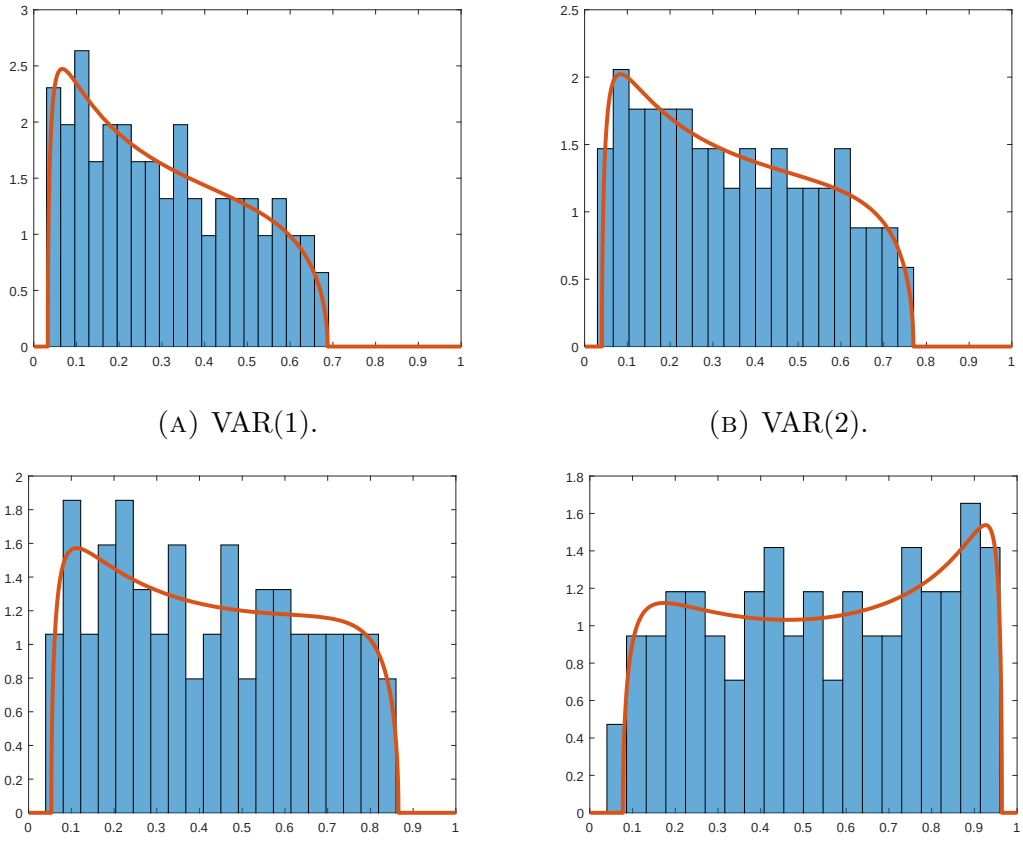
FIGURE 8. Airy$_1$ and asymptotic distribution of the rescaled $\ln(1-\tilde{\lambda}_1)$ under H_1. Data generating process: $\Delta X_t = -0.95E_{11}X_{t-2} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 100,000$ replications.



(A) $\rho = \text{rank}(\Pi) = 5$. (B) $\rho = \text{rank}(\Pi) = 50$. (C) $\rho = \text{rank}(\Pi) = 150$.

FIGURE 9. Separation of eigenvalues from λ_+ under various ranks of Π. Data generating process: $\Delta X_t = 1_N + 0.95E_{12}\Delta X_{t-1} - 0.8I_\rho X_{t-2} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 1500$, $N = 150$, I_ρ is a matrix with ones on the first ρ diagonal elements and zeros elsewhere.

distribution) is flatter than the blue dashed curve ($\mathfrak{a}_1$). First, the separation of the curves is in line with the usefulness of Theorem 9 in cointegration hypothesis testing, since the test statistic was designed to distinguish between $\Pi = 0$ and $\Pi \neq 0$. Second, the distinct variances are due to the fact that (as we expect from a comparison with results on spiked random matrices in the literature; see, e.g., Baik et al. [2005]) under the alternative ($\Pi \neq 0$) the test needs to be scaled differently: Instead of $N^{2/3}$ rescaling one should use $N^{1/2}$. This result is in line with the power analysis of our test in Section 8.2.



(A) VAR(1). (B) VAR(2).

(C) VAR(3). (D) VAR(4).

FIGURE 10. Eigenvalues from S&P data (blue histogram) and Wachter distribution (orange line) based on various VAR(k) settings.

For the second experiment, we use $k = 2$, $N = 150$, $T = 1500$, $\Gamma_1 = 0.95E_{12}$, and $\Pi = -0.8I_\rho X_{t-2}$, where I_ρ is a matrix with ones on the first ρ diagonal elements and zeros elsewhere. The histograms of eigenvalues $\tilde{\lambda}_1, \dots, \tilde{\lambda}_N$ for $\rho = 5, 50, 150$ are shown in Figure 9. As ρ becomes large, we are no longer in the framework of Theorem 3, and the result about the convergence to the Wachter distribution does not apply. Nevertheless, we observe that the largest eigenvalues are significantly larger than λ_+ of Theorem 9. Recall that our testing procedure is based on comparing the largest eigenvalues[11] with λ_+. Thus, this leads to the conclusion that our test is useful for all values of $\rho \in [1, N]$: the test rejects H_0 of no cointegration (i.e., the $\Pi = 0$ hypothesis) at a very high statistical significance level.

5. Empirical illustrations

5.1. S&P100. We illustrate our asymptotic theorems on the S&P100 data. We use logarithms of weekly prices of assets in the S&P100 over ten years (January 1, 2010, to January 1,

[11]More precisely, logarithms of 1 minus eigenvalues vs. $\log(1 - \lambda_+)$.

2020), which gives us 522 observations across time. More detailed description of the variables can be found in Bykhovskaya and Gorin [2022, Section 6].

For the S&P100 data set we use Procedure 2 to calculate $\tilde{\lambda}_1, \dots, \tilde{\lambda}_N$. We do this for various choices of k. The case $k = 1$ (VAR(1)) corresponds to Bykhovskaya and Gorin [2022]. The results are shown in Figure 10.

We see a striking match between the histograms and Wachter densities for all $k = 1, 2, 3, 4$, which is an indication that the setting of Theorem 3 is a proper modeling for the S&P data. We do not see any outliers in the largest eigenvalues, which would appear if there were cointegration. Indeed, our test statistics based on Theorem 9 are -0.28, -0.71, -1.07, -3.84 for $k = 1, 2, 3, 4$, respectively, while the 5% and 10% critical values are 0.97 and 0.44. Because the former numbers are smaller than the latter numbers, we do not reject the "no cointegration" hypothesis.

5.2. Cryptocurrencies. In this subsection we redo the calculations for cryptocurrencies instead of S&P stocks. We use the data from Keilbar and Zhang [2021] (25 series from the Github repository). Logarithms of daily prices for two years (from October 5, 2017, to October 4, 2019) are shown in Figure 11. The results of Procedure 2 used to calculate $\tilde{\lambda}_1, \dots, \tilde{\lambda}_N$ are shown in Figure 12.

Similarly to the S&P example in the previous subsection, we see a match between the eigenvalues and the Wachter distribution.[12] However, there is a major difference between Figures 10 and 12: The latter has around 3 eigenvalues to the right of the support of the orange curve (Wachter distribution). This is an indication of the presence of approximately 3 cointegrating relationships. This is reinforced by our test, which has p-values below 0.01 for all four choices of the order of VAR(k) ($k = 1, 2, 3, 4$).

The difference in results for traditional stocks and cryptocurrencies can be explained by the fact that the cryptocurrency market is still very inefficient and, thus, has numerous trading possibilities. The presence of cointegration can be one such inefficiency.

6. Conclusion

High-dimensional data are becoming increasingly widespread in economics and other sciences. Thus, appropriate machinery for handling such data is needed. We believe that the use of random matrix theory is inevitable for the development of the area: As soon as dimensions are high, random matrices start to contribute. Along these lines, in our paper the central role is played by random matrix objects: the Wachter distribution, Airy_1 point process, and Jacobi ensemble.

[12]The Wachter distribution depends on the order of the VAR, k, and on the ratio T/N. Thus, the orange curves in Figures 10 and 12 have different shapes and supports.

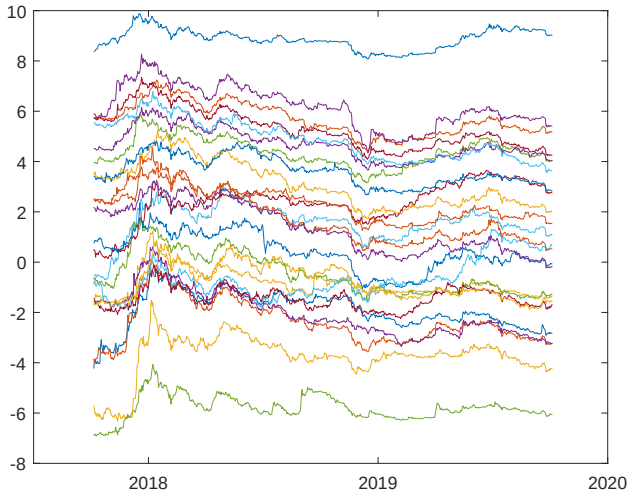
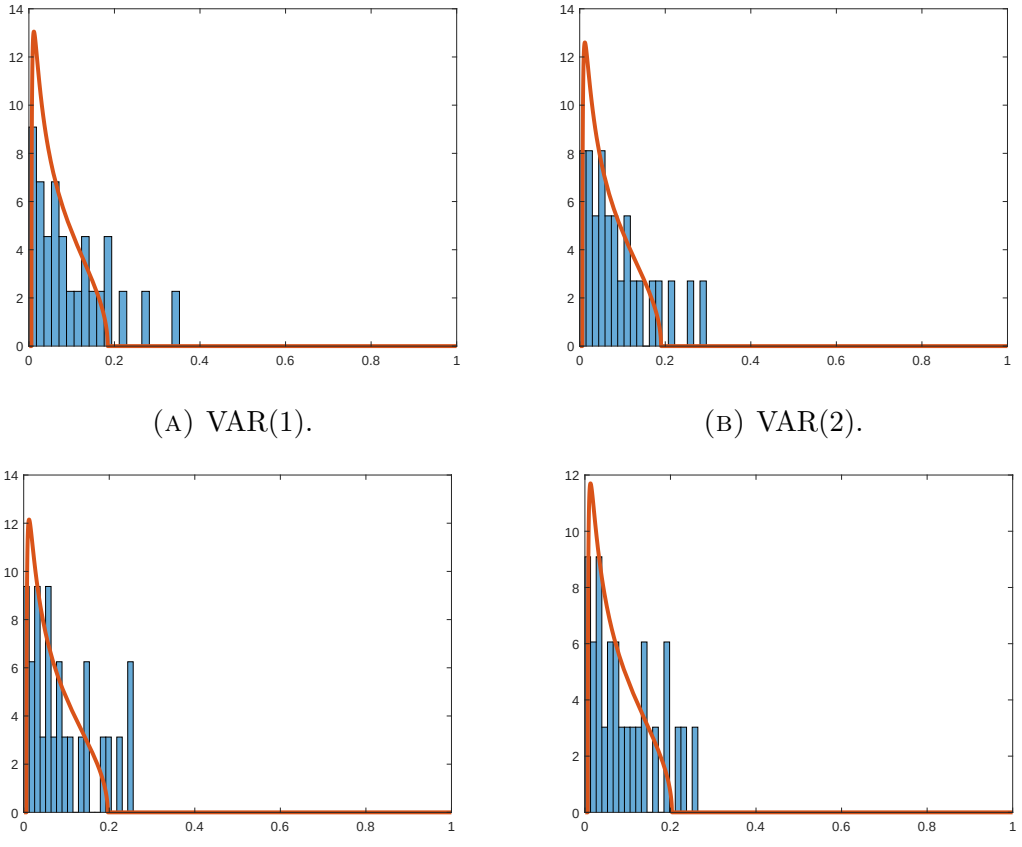


FIGURE 11. Time series of daily log prices for 25 cryptocurrencies.



(A) VAR(1). (B) VAR(2).

(C) VAR(3). (D) VAR(4).

FIGURE 12. Eigenvalues from cryptocurrency data (blue histogram) and Wachter distribution (orange line) based on various VAR(k) settings.

The present paper focused on nonstationary high-dimensional VARs and presented the asymptotic limit of the Johansen LR test for cointegration and its modifications. Because the limit is nonrandom, the appropriate second-order statistic was derived, and a new test for the presence of cointegration was proposed. The new test builds upon the Johansen LR, while having some extra modifications. This new test is suitable for a vector autoregression of order k with an intercept.

The main focus of the present paper is the null of no cointegration. The next essential step is to be able to test whether the cointegration rank is r for $r > 0$, i.e., to find the true rank of cointegration. Heuristics for finding the correct value of r can already be seen in our simulations and data sets (cf. Figures 1, 10, and 12): When there are no cointegrations, all eigenvalues (squared canonical correlations) are to the left of the end-point of the support of the Wachter distribution. In contrast, we expect each cointegrating relationship to lead to an eigenvalue between the right end-point of the support of the Wachter distribution and 1. Identifying the exact conditions under which this heuristic is correct represents an important problem for future research.

7. Appendix 1: Proofs

This appendix contains the proofs of Theorems 3, 8, and 9 from the main text.

First, in Section 7.1 we collect known statements about the asymptotics of the Jacobi ensemble of Definition 5, which will be used in our subsequent proofs.

Second, in Theorem 12 of Section 7.2 we introduce a novel random matrix model for the Jacobi ensemble. Our proof of Theorem 12 proceeds through certain intricate inductive computations of large-dimensional matrix integrals.

Third, in Section 7.3 we connect the matrix model of Section 7.2 to the cointegration setting: for that we use the rotational symmetry of the Gaussian law to express the squared sample canonical correlations solving (17) under the hypothesis $\widehat{H}_0$ in terms of a certain deterministic orthogonal matrix. Replacing this deterministic matrix by a uniformly random one, we arrive at the Jacobi ensemble of Theorem 12. We proceed by bounding the error in this replacement, which relies on the rigidity estimate for orthogonal matrices (66), but needs special care due to various matrix inversions involved in our procedures. Eventually, we arrive at Theorem 8. This theorem is our main technical result. Combining Theorem 8 with Proposition 11 from Section 7.1, we finish the proof of Theorem 9 from the main text.

Finally, in Section 7.4 we prove Theorem 3 by combining Theorem 8 with Proposition 11 of Section 7.1 and general statements about small rank perturbations.

7.1. 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 5 as $N \to \infty$.

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

$$\frac{p}{N} = \frac{1}{2}(\mathfrak{p} - 1), \quad \mathfrak{p} \geq 1, \qquad \frac{q}{N} = \frac{1}{2}(\mathfrak{q} - 1), \quad \mathfrak{q} \geq 1, \tag{28}$$

where $\mathfrak{p}$ and $\mathfrak{q}$ are two parameters, which stay bounded away from 1 and from ∞ as $N \to \infty$.[13] We further define the *equilibrium measure* $\mu_{\mathfrak{p},\mathfrak{q}}$ of the Jacobi ensemble through:

$$\mu_{\mathfrak{p},\mathfrak{q}}(x)\,dx = \frac{\mathfrak{p}+\mathfrak{q}}{2\pi} \cdot \frac{\sqrt{(x-\lambda_-)(\lambda_+ - x)}}{x(1-x)} \mathbf{1}_{[\lambda_-,\lambda_+]}\,\mathrm{d}x, \tag{29}$$

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

$$\lambda_\pm = \frac{1}{(\mathfrak{p}+\mathfrak{q})^2}\left(\sqrt{\mathfrak{p}(\mathfrak{p}+\mathfrak{q}-1)} \pm \sqrt{\mathfrak{q}}\right)^2. \tag{30}$$

One can check that $0 < \lambda_- < \lambda_+ < 1$ for every $\mathfrak{p}, \mathfrak{q} > 1$. Further, define

$$c_\pm = \frac{(\mathfrak{p}+\mathfrak{q})}{2} \frac{\sqrt{\lambda_+ - \lambda_-}}{\lambda_\pm(1-\lambda_\pm)}, \tag{31}$$

and note that

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

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

Proposition 11 (See Johnstone [2008], Forrester [2010], and Han et al. [2016])**.** *Suppose that $N, p, q \to \infty$ in such a way that $\mathfrak{p} \geq 1$ and $\mathfrak{q} \geq 1$ in (28) 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*

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

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

$$\lim_{N\to\infty} \left| \frac{1}{N}\sum_{i=1}^{N} f(x_i) - \int_0^1 f(x)\mu_{\mathfrak{p},\mathfrak{q}}(x)\mathrm{d}x \right| = 0. \tag{32}$$

[13] Johnstone [2008, Theorem 1] suggests to use $\frac{p-1}{N}$ and $\frac{q-1}{N}$ instead of $\frac{p}{N}$ and $\frac{q}{N}$, respectively, in order to improve the speed of convergence. However, we found in Bykhovskaya and Gorin [2022] that for the tests in VAR(1) case the usefulness of this correction depends on the exact value of the ratio T/N and we are not going to pursue this direction here.

(2) *For* $\{\mathfrak{a}_i\}_{i=1}^{\infty}$ *as in Proposition 7, we have convergence in finite-dimensional distributions for the largest eigenvalues:*

$$\lim_{N\to\infty}\left\{N^{2/3}c_+^{2/3}\left(x_i-\lambda_+\right)\right\}_{i=1}^{\infty}\to\{\mathfrak{a}_i\}_{i=1}^{\infty}. \tag{33}$$

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

(3) *We also have convergence in distribution for the smallest eigenvalues*[14]

$$\lim_{N\to\infty}\left\{N^{2/3}c_-^{2/3}\left(\lambda_- - x_{N+1-i}\right)\right\}_{i=1}^{\infty}\to\{\mathfrak{a}_i\}_{i=1}^{\infty}. \tag{34}$$

7.2. A new model for the Jacobi ensemble. The Jacobi ensemble appearing in Theorem 8 originates in the following computation of exact distribution. In addition to real symmetric matrices ($\beta=1$ in the usual random matrix notations) it also covers the case of complex Hermitian matrices ($\beta=2$).

Theorem 12. *Fix* $k=1,2,\dots$ *and assume* $\mathcal{T}\geq(k+1)N$. *Let* $\mathcal{V}$ *be an* N*–dimensional subspace in the* $\mathcal{T}$*–dimensional space and let* O *be uniformly random orthogonal* $\mathcal{T}\times\mathcal{T}$ *matrix with determinant* 1 *if* $\beta=1$ *(*O *is a uniformly random unitary matrix if* $\beta=2$*). Let* P *be an orthogonal projector on the space orthogonal to* $O\mathcal{V}$, $O^2\mathcal{V},\dots,$ $O^{k-1}\mathcal{V}$. *Let* P_1 *be a projector on the subspace* $P\mathcal{V}$ *and* P_2 *be a projector on the subspace* $PO^{k-1}(I_{\mathcal{T}}+O)^{-1}\mathcal{V}$. *Then non-zero eigenvalues of* $P_1P_2P_1$ *coincide with those of the Jacobi ensemble of* $N\times N$ *real symmetric if* $\beta=1$ *(complex Hermitian if* $\beta=2$*) matrices of density proportional to*

$$\det(\mathcal{M})^{\frac{\beta}{2}N+\beta-2}\det(I_N-\mathcal{M})^{\frac{\beta}{2}(\mathcal{T}-(k+1)N+1)-1}\,d\mathcal{M},\quad 0\leq\mathcal{M}\leq I_N,\qquad \beta=1,2. \tag{35}$$

Remark 13. *We can replace* $PO^{k-1}(I_{\mathcal{T}}+O)^{-1}$ *in the definition of* P_2 *with* $PO^{k}(I_{\mathcal{T}}+O)^{-1}$. *Indeed, if* $k>1$, *then* $PO^{k}(I_{\mathcal{T}}+O)^{-1}\mathcal{V}+PO^{k-1}(I_{\mathcal{T}}+O)^{-1}\mathcal{V}=PO^{k-1}\mathcal{V}=0$. *If* $k=1$, *then* P *disappears (one can say that it becomes an identical operator) and the random operators* $(I_{\mathcal{T}}+O)^{-1}$ *and* $O(I_{\mathcal{T}}+O)^{-1}=(I_{\mathcal{T}}+O^{-1})^{-1}$ *has the same law, since the uniform measure on the orthogonal group is invariant under the inversion* $O\mapsto O^{-1}$.

By a similar argument applied inductively we can replace $PO^{k-1}(I_{\mathcal{T}}+O)^{-1}$ *with* $PO(I_{\mathcal{T}}+O)^{-1}$. *Note, however, that for* $k>1$ *we can not replace it with* $P(I_{\mathcal{T}}+O)^{-1}$.

The $k=1$ case of Theorem 12 is established in [Bykhovskaya and Gorin, 2022, Theorem 6 in Appendix]. The proof of Theorem 12 uses the following three auxiliary ingredients.

Lemma 14 (Block matrix inversion formula)**.** *For matrices* $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, *we have:*

$$\begin{pmatrix}\mathbf{A}&\mathbf{B}\\ \mathbf{C}&\mathbf{D}\end{pmatrix}^{-1}=\begin{pmatrix}\mathbf{Q}&-\mathbf{Q}\mathbf{B}\mathbf{D}^{-1}\\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{Q}&\mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}\mathbf{Q}\mathbf{B}\mathbf{D}^{-1}\end{pmatrix},\qquad \mathbf{Q}=(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}, \tag{36}$$

Proof. Direct computation. □

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

Lemma 15 (Cayley transform)**.** *Suppose that all eigenvalues of $N\times N$ matrix O are different from -1. Then O is an orthogonal matrix with determinant 1, if and only if the matrix $\mathcal{R}$ defined through*

$$\mathcal{R}=(I_N-O)(I_N+O)^{-1}=\frac{I_N-O}{I_N+O},\qquad \text{so that}\qquad O=\frac{I_N-\mathcal{R}}{I_N+\mathcal{R}} \tag{37}$$

is skew-symmetric, i.e., it satisfies $\mathcal{R}^*=-\mathcal{R}$.

Proof. The formulas (37) imply that $O^*=O^{-1}$ if and only if $\mathcal{R}^*=-\mathcal{R}$. On the other hand, for a skew-symmetric $\mathcal{R}$, we have $\det(I_N+\mathcal{R})=\det(I_N+\mathcal{R}^*)=\det(I_N-\mathcal{R})$. Hence, $\det(O)=\det\left(\frac{I_N-\mathcal{R}}{I_N+\mathcal{R}}\right)=1$. □

Lemma 16. *Choose two positive integers M, N and set $\mathcal{T}=M+N$. Let O be a uniformly random $\mathcal{T}\times\mathcal{T}$ orthogonal matrix with determinant 1. Write O in the block form according to $\mathcal{T}=M+N$ splitting:*

$$O=\begin{pmatrix}A & B\\ C & D\end{pmatrix}$$

Then $\tilde{O}:=A-B(I_N+D)^{-1}C$ is a $M\times M$ orthogonal matrix of determinant 1 uniformly distributed among all such matrices. In addition, the random matrices $\tilde{O}$ and $B(I_N+D)^{-1}$ are independent.

Remark 17. *The law of $\widehat{W}:=B(I_N+D)^{-1}$ is explicit. The computation (44) below implies that the density of $\widehat{W}$ is proportional to*

$$\det\bigl(I_N+\widehat{W}^*\widehat{W}\bigr)^{1/2-M-N/2}\,d\widehat{W}.$$

Remark 18. *The computation of the law of $\tilde{O}$ is mentioned in Olshanski [2003] and Neretin [2002] with its roots going back to Hua [1963]. However, we could not locate the statements concerning also $B(I_N+D)^{-1}$ in the literature.*

Proof of Lemma 16. First, note that the distribution of eigenvalues of D is absolutely continuous and, hence, I_N+D is almost surely invertible and the matrix $\tilde{O}$ is well-defined. Our next task is to show that $\tilde{O}$ is an orthogonal matrix with determinant 1. We use Cayley transform for that. Combining (37) with (36) we have

$$\begin{aligned}\mathcal{R}&=\frac{I_{\mathcal{T}}-O}{I_{\mathcal{T}}+O}=\frac{2}{I_{\mathcal{T}}+O}-I_{\mathcal{T}}\\ &=\begin{pmatrix}2Q-I_M & -2QB(I_N+D)^{-1}\\ -2(I_N+D)^{-1}CQ & 2(I_N+D)^{-1}+2(I_N+D)^{-1}CQB(I_N+D)^{-1}-I_N\end{pmatrix},\\ &\qquad\qquad Q=(I_M+A-B(I_N+D)^{-1}C)^{-1}.\end{aligned} \tag{38}$$

Since $\mathcal{R}$ is skew-symmetric, so is its top–left $M\times M$ corner $\tilde{\mathcal{R}}=2Q-I_M$. We claim that $\tilde{O}$ is the Cayley transform of $\tilde{\mathcal{R}}$, which would imply that $\tilde{O}$ is orthogonal of determinant 1.

Indeed,

$$\frac{I_M-\tilde{\mathcal{R}}}{I_M+\tilde{\mathcal{R}}}=\frac{2I_M-2Q}{2Q}=Q^{-1}-I_M=A-B(I_N+D)^{-1}C=\tilde{O}. \tag{39}$$

It remains to compute the distributions of $\tilde{O}$ and $B(I_N+D)^{-1}$ and show their independence. In terms of $\mathcal{R}$ the distribution of O (as a uniformly random orthogonal matrix of determinant 1) is given by the density proportional to

$$\det(I_{\mathcal{T}}-\mathcal{R}^2)^{-\frac{1}{2}\mathcal{T}+\frac{1}{2}}d\mathcal{R}=\det(I_{\mathcal{T}}-\mathcal{R})^{1-\mathcal{T}}d\mathcal{R}=\det(I_{\mathcal{T}}+\mathcal{R})^{1-\mathcal{T}}d\mathcal{R}, \tag{40}$$

see, e.g., Forrester [2010, (2.55)] and notice that $\det(I_{\mathcal{T}}-\mathcal{R})=\det(I_T+\mathcal{R})$ for the two equalities. We rewrite the block form (38) of $\mathcal{R}$ as

$$\mathcal{R}=\begin{pmatrix}\tilde{\mathcal{R}} & -W\\ W^* & \mathcal{R}_2\end{pmatrix},$$

where $\mathcal{R}_2$ is $N\times N$ skew-symmetric and W is an arbitrary $M\times N$ matrix. We further introduce the notation $\widehat{W}:=(I_M+\tilde{\mathcal{R}})^{-1}W$. Recalling that $\tilde{\mathcal{R}}=2Q-I_M$, we transform

$$\widehat{W}=2(I_M+\tilde{\mathcal{R}})^{-1}QB(I_N+D)^{-1}=B(I_N+D)^{-1}. \tag{41}$$

We also define

$$\widehat{\mathcal{R}}_2:=\big(I_N+\widehat{W}^*\widehat{W}\big)^{-1/2}\big(-\mathcal{R}_2+\widehat{W}^*\tilde{\mathcal{R}}\widehat{W}\big)\big(I_N+\widehat{W}^*\widehat{W}\big)^{-1/2}.$$

Note that $(\tilde{R},\widehat{W},\widehat{R}_2)$ is an alternative parameterization of $\mathcal{R}$, in which $\widehat{W}$ is an arbitrary $M\times N$ matrix and $\widehat{R}_2$ is an arbitrary $N\times N$ skew-symmetric matrix. Using the formula for the determinant of a block matrix

$$\det\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{pmatrix}=\det\mathbf{A}\cdot\det(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}), \tag{42}$$

we rewrite (40) as

$$\begin{aligned}\det(I_{\mathcal{T}}-\mathcal{R})^{1-\mathcal{T}}\,d\tilde{\mathcal{R}}\,dW\,d\mathcal{R}_2&=\det\begin{pmatrix}I_M-\tilde{\mathcal{R}} & W\\ -W^* & I_N-\mathcal{R}_2\end{pmatrix}^{1-\mathcal{T}}d\tilde{\mathcal{R}}\,dW\,d\mathcal{R}_2\\ &=\det(I_M-\tilde{\mathcal{R}})^{1-\mathcal{T}}\det(I_N-\mathcal{R}_2+W^*(I_M-\tilde{\mathcal{R}})^{-1}W)^{1-\mathcal{T}}\,d\tilde{\mathcal{R}}\,dW\,d\mathcal{R}_2\\ &=\det\big(I_M-\tilde{\mathcal{R}}\big)^{1-\mathcal{T}}\det\big(I_N-\mathcal{R}_2+\widehat{W}^*(I_M+\tilde{\mathcal{R}})\widehat{W}\big)^{1-\mathcal{T}}\,d\tilde{\mathcal{R}}\,dW\,d\mathcal{R}_2\end{aligned} \tag{43}$$

Further, notice that

$$\big(I_N+\widehat{W}^*\widehat{W}\big)^{1/2}\big(I_N+\widehat{\mathcal{R}}_2\big)\big(I_N+\widehat{W}^*\widehat{W}\big)^{1/2}=I_N-\mathcal{R}_2+\widehat{W}^*(I_M+\tilde{\mathcal{R}})\widehat{W}.$$

Hence, the last line of (43) is transformed into

$$\begin{aligned}(44)\quad &\det\bigl(I_M-\tilde{\mathcal{R}}\bigr)^{1-\mathcal{T}}\det\bigl(I_N+\widehat{W}^*\widehat{W}\bigr)^{1-\mathcal{T}}\det\bigl(I_N+\widehat{\mathcal{R}}_2\bigr)^{1-\mathcal{T}}\,d\tilde{\mathcal{R}}\,dW\,d\mathcal{R}_2\\ &\quad=\det\bigl(I_M+\tilde{\mathcal{R}}\bigr)^{1-M}\det\bigl(I_N+\widehat{W}^*\widehat{W}\bigr)^{1/2-M-N/2}\det\bigl(I_N+\widehat{\mathcal{R}}_2\bigr)^{1-M-N}\,d\tilde{\mathcal{R}}\,d\widehat{W}\,d\widehat{\mathcal{R}}_2,\end{aligned}$$

where in the last line we use $\mathcal{T}=M+N$ and change variables $dW\mapsto d\widehat{W}$ and $d\mathcal{R}_2\mapsto d\widehat{\mathcal{R}}_2$ using the general Jacobian computations:

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

$$\left|\frac{\partial(QZ)}{\partial Z}\right|=|\det Q|^m, \tag{45}$$

- The map $Z\mapsto QZQ^*$ from the space of $n\times n$ skew-symmetric matrices to itself has the Jacobian

$$\left|\frac{\partial(QZQ^*)}{\partial Z}\right|=|\det Q|^{n-1}. \tag{46}$$

The first identity (45) follows from the observation that each column of Z is transformed by linear map Q and there are m such columns. The second is similar and we refer to Forrester [2010, (1.35)] for details.

The key important feature of the last line of (44) is that it has a product form, which implies the joint independence of $\tilde{\mathcal{R}}$, $\widehat{W}$, and $\widehat{\mathcal{R}}_2$. Hence, the density of $\tilde{\mathcal{R}}$ is proportional to $\det\bigl(1+\tilde{\mathcal{R}}\bigr)^{1-M}d\tilde{\mathcal{R}}$. Comparing with (40) and noting that the dimension changed from $\mathcal{T}$ to M, we conclude that $\tilde{O}$ is a uniformly random $M\times M$ orthogonal matrix of determinant 1.

Next, recalling that $\tilde{O}$ is a deterministic function of $\tilde{\mathcal{R}}$ by (39), we conclude that $\tilde{O}$ is independent with $\widehat{W}$, which is precisely $B(I_N+D)^{-1}$ by (41). □

Proof of Theorem 12. We only give a proof for the real case $\beta=1$; the complex case can be proven by the same argument. The proof is induction in k with base case $k=1$ being [Bykhovskaya and Gorin, 2022, Theorem 6 in Appendix] and the induction step being based on Lemma 16.

Step 1. We first note that the particular choice of *deterministic* space $\mathcal{V}$ in the statement of the theorem is not important: any other deterministic choice of $\mathcal{V}$ can be achieved by a change of basis of the $\mathcal{T}$–dimensional space, which keeps the probability distribution of O and, hence, entire construction invariant. In particular, the probability distribution of $P_1P_2P_1$ is unchanged. However, we need to be more careful, if we would like to make $\mathcal{V}$ random, as correlations with O might cause issues.

Step 2. Take any $r\in\mathbb{Z}$. We claim that replacement of $\mathcal{V}$ with $O^r\mathcal{V}$ everywhere in the statement of Theorem 12 does not change the eigenvalues of $P_1P_2P_1$. Indeed, the only important feature of O^r here is that it is an orthogonal operator commuting with O. Hence, the change $\mathcal{V}\mapsto O^r\mathcal{V}$ leads to the image of the projector P being multiplied by O^r; in more

details, the transformation takes the form $P \mapsto O^r P O^{-r}$. Further, $P\mathcal{V}$ gets transformed to $O^r P\mathcal{V}$ and P_1 undergoes a similar transformation: $P_1 \mapsto O^r P_1 O^{-r}$. The same is true for P_2: it undergoes the transformation $P_2 \mapsto O^r P_2 O^{-r}$. We conclude that the product $P_1P_2P_1$ is transformed into $O^r P_1P_2P_1 O^{-r}$. Since conjugations do not change eigenvalues, we are done.

The arguments of Steps 1 and 2 might give a feeling that we can actually replace $\mathcal{V}$ by any random space. However, this is not the case. Repeating the same arguments, we see that replacement $\mathcal{V} \mapsto A\mathcal{V}$ leads to the same eigenvalues of the projector $P_1P_2P_1$ as if we replaced $O \mapsto A^*OA$. In both Steps 1 and 2 O had the same distribution as A^*OA, hence, the eigenvalues were unchanged. But in general, if A is correlated with O in a non-trivial way, then the distribution might change.[15]

Step 3. We now transform the statement of Theorem 12 by replacing $\mathcal{V}$ with $O^{1-k}\mathcal{V}$ and further replacing O by O^{-1} everywhere. Since the uniform (Haar) measure on the orthogonal matrices is invariant under inversion, the law of eigenvalues of $P_1P_2P_1$ is unchanged and the ingredients of Theorem 12 are now as follows:

- O is a uniformly random $\mathcal{T} \times \mathcal{T}$ orthogonal matrix with determinant 1 and $\mathcal{V}$ is an arbitrary (deterministic) N–dimensional subspace of the $\mathcal{T}$–dimensional space, whose choice is irrelevant for the statement.
- P is the projector on the orthogonal complement of $O^{k-2}\mathcal{V}$, $O^{k-3}\mathcal{V}$, …, $O\mathcal{V}$, $\mathcal{V}$.
- P_1 is the projector on the subspace $PO^{k-1}\mathcal{V}$ and P_2 is the projector on the subspace $PO(I_T + O)^{-1}\mathcal{V}$. (The latter can be replaced by $P(I_T + O)^{-1}\mathcal{V}$ without changing the outcome. Indeed, for that we need start from $PO^k(I_T + O)^{-1}$ instead of $PO^{k-1}(I_T + O)^{-1}$, which is possible by Remark 13).
- The claim is that the eigenvalues of $P_1P_2P_1$ are distributed as (35).

We are going to prove this last statement by induction in k. For that we choose $\mathcal{V}$ to be the span of the last N coordinate vectors, split $\mathcal{T} = M + N$ with $M = \mathcal{T} - N$ and project everything on the first M coordinate vectors (which are orthogonal complement to $\mathcal{V}$). We rely on Lemma 16 and use A, B, C, D and $\tilde{O}$ notation from that lemma.

Step 4. We claim that the subspace in M–dimensional space spanned by the first M coordinates of $O^{k-2}\mathcal{V}$, $O^{k-3}\mathcal{V}$, …, $O\mathcal{V}$ (since $\mathcal{V}$ has zero projection on the first M coordinates, we do not need it here) is the same as the subspace spanned by $\tilde{O}^{k-3}\langle B\rangle$, $\tilde{O}^{k-4}\langle B\rangle$, …, $\langle B\rangle$, where $\langle B\rangle$ is N–dimensional space spanned by columns of the $M \times N$ matrix B.

Indeed, the first M coordinates of $O\mathcal{V}$ are $\langle B\rangle$ by definition of the block structure in Lemma 16. Further, to go from powers of O to powers of $\tilde{O}$ we make the following observation: take a vector w in $\mathcal{T}$–dimensional space and write it as $w = \binom{w_2}{w_1}$, where w_1 is N-dimensional vector (one can think of w_1 being in $\mathcal{V}$) and w_2 is M–dimensional vector (one can think of

[15]For instance, if $\mathcal{V}$ is spanned by eigenvectors of O, then the spaces $O\mathcal{V}$, $O^2\mathcal{V}$,…,$O^{k-1}\mathcal{V}$ all coincide, which is a very different behavior from the case of deterministic $\mathcal{V}$.

w_2 being in the orthogonal complement of $\mathcal{V}$) and write

$$O\begin{pmatrix} w_2 \\ w_1 \end{pmatrix} = \begin{pmatrix} u_2 \\ u_1 \end{pmatrix}.$$

Then the M–dimensional vector u_2 takes the form

$$u_2 = Aw_2 + Bw_1 = \tilde{O}w_2 + B((I_N + D)^{-1}Cw_2 + w_1).$$

Since we only care about the linear span of columns and $\langle B\rangle$ already belongs to the desired linear span, the last term can be ignored and we arrive at $\tilde{O}w_2$, which then implies the claim.

Step 5. Next, consider the projection of $O^{k-1}\mathcal{V}$ on the orthogonal complement to $O^{k-2}\mathcal{V}$, $O^{k-3}\mathcal{V}$, …, $O\mathcal{V}$, $\mathcal{V}$. This is the same as the the projection of the first M coordinates of $O^{k-1}\mathcal{V}$ on the orthogonal complement (in M–dimensional space) to first M coordinates of $O^{k-2}\mathcal{V}$, $O^{k-3}\mathcal{V}$, …, $O\mathcal{V}$. Hence, combining with the argument of Step 4, this is the same as the projection of $\tilde{O}^{k-2}\langle B\rangle$ on the orthogonal complement of $\tilde{O}^{k-3}\langle B\rangle$, $\tilde{O}^{k-4}\langle B\rangle$, …, $\langle B\rangle$. It is convenient to note that $\langle B\rangle = \langle B(I_N + D)^{-1}\rangle$.

Step 6. Finally, consider the projection of $(I_T + O)^{-1}O\mathcal{V}$ on the orthogonal complement of $O^{k-2}\mathcal{V}$, $O^{k-3}\mathcal{V}$, …, $O\mathcal{V}$, $\mathcal{V}$. By Steps 4 and 5 this is the same as the projection of the first M coordinates of $(I_T + O)^{-1}O\mathcal{V}$ on the orthogonal complement of $\tilde{O}^{k-3}\langle B(I_N + D)^{-1}\rangle$, $\tilde{O}^{k-4}\langle B(I_N + D)^{-1}\rangle$, …, $\langle B(I_N + D)^{-1}\rangle$. Representing $(I_T + O)^{-1}O\mathcal{V}$ in the block form, the first M coordinates of $(I_T + O)^{-1}O\mathcal{V}$ are the span of the columns of the sum of the top–left corner of $(I_T + O)^{-1}$ multiplied by B plus the top-right corner of $(I_T + O)^{-1}$ multiplied by D. Using Lemma 14, we get the span of the columns of

$$(I_M + A - B(I_N + D)^{-1}C)^{-1}B - (I_M + A - B(I_N + D)^{-1}C)^{-1}B(I_N + D)^{-1}D$$
$$= (I_M + \tilde{O})^{-1}B(I_N + D)^{-1},$$

which is the same as $(I_M + \tilde{O})^{-1}\langle B(I_N + D)^{-1}\rangle$.

Step 7. Combining the results of Steps 6 and 7 with Lemma 16, we identify the eigenvalues of $P_1P_2P_1$ with the eigenvalues of $\tilde{P}_1\tilde{P}_2\tilde{P}_1$ obtained by the following procedure:

- $\tilde{O}$ is a uniformly random $M \times M$ orthogonal matrix with determinant 1, where $M = \mathcal{T} - N$.
- $\tilde{P}$ is the projector on orthogonal complement of $\tilde{O}^{k-3}\langle B(I_N + D)^{-1}\rangle$, $\tilde{O}^{k-4}\langle B(I_N + D)^{-1}\rangle$, …, $\langle B(I_N + D)^{-1}\rangle$.
- $\tilde{P}_1$ is the projector on the subspace $\tilde{P}\tilde{O}^{k-2}\langle B(I_N + D)^{-1}\rangle$ and $\tilde{P}_2$ is the projector on the subspace $\tilde{P}(I_M + \tilde{O})^{-1}\langle B(I_N + D)^{-1}\rangle$.

Since $\langle B(I_N + D)^{-1}\rangle$ is independent from $\tilde{O}$ by Lemma 16, this is the same form as the one at the end of Step 3, but with k decreased by 1, $\mathcal{T}$ decreased by N, and $\mathcal{V}$ replaced by

$\langle B(I_N + D)^{-1}\rangle$. Decreasing k by 1 and $\mathcal{T}$ by N leaves the formula (35) unchanged, hence, we can invoke the induction assumption, thus, finishing the proof. □

7.3. A perturbation of the Jacobi ensemble. In this section we use Theorem 12 to prove Theorem 8.

Recall the cyclic shift[16] operator L_c acting in T–dimensional space. Let V be the $(T-1)$–dimensional space orthogonal to the vector $(1, 1, \dots, 1)$, i.e., $V = \{(x_1, \dots, x_T) \mid x_1 + \dots + x_T = 0\}$. Note that V is an invariant space for L_c and let L_V denote the restriction of L_c on the subspace V.

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

$$\tilde{L} = -\tilde{O} L_V \tilde{O}^*.$$

Proposition 19. *Assume $T > (k+1)N$ and let $\tilde{L}$ be as above. Take an arbitrary N–dimensional subspace $\mathcal{U}$ in $(T-1)$–dimensional space V. Let P be the orthogonal projector on the space orthogonal to $\tilde{L}\mathcal{U}$, $\tilde{L}^2\mathcal{U}$,..., $\tilde{L}^{k-1}\mathcal{U}$. Let P_1 be the projector on the subspace $P\mathcal{U}$ and P_2 be the projector on the subspace $P\tilde{L}^k(I_V + \tilde{L})^{-1}\mathcal{U}$. Then the distributions of non-zero eigenvalues of $P_1P_2P_1$ coincides with that of the squared sample canonical correlations solving* (17) *under the hypothesis $\widehat{H}_0$.*

Comparing Proposition 19 with Theorem 12 and Remark 13 one notices that the differences are in restricting on the subspace V (hence, decreasing the dimension by 1) and in replacement $O \leftrightarrow \tilde{L}$.

Proof of Proposition 19.

Step 1. We start by transforming the Gaussian noise ε_t. Let ε be $N \times T$ matrix, whose t-th column is ε_t. Take any non-degenerate $N \times N$ matrix A and transform $\varepsilon \mapsto A\varepsilon$. Thus, we leave X_0 unchanged and recalculate X_t, $1 \le t \le T$. We claim that the canonical correlations solving Eq. (17) are unchanged. Indeed, the linear subspace $\mathcal{W}$ stays the same and so does the projector $P_{\perp\mathcal{W}}$. For each $t = 1, 2, \dots, T$, the vector ΔX_t is transformed by $\Delta X_t \mapsto A\Delta X_t + (I_N - A)\mu$ and $\tilde{X}_t$ is transformed by $\tilde{X}_t \mapsto A\tilde{X}_t + (I_N - A)X_0$. Recall that the space $\mathcal{W}$ includes vector $(1, \dots, 1)$, which leads to the projector $P_{\perp\mathcal{W}}$ canceling the additional terms $(I_N - A)\mu$ and $(I_N - A)X_0$ in the last two formulas. Hence, the matrices $\tilde{R}_0$ and $\tilde{R}_k$ are transformed by $\tilde{R}_0 \mapsto A\tilde{R}_0$ and $\tilde{R}_k \mapsto A\tilde{R}_k$. Therefore,

$$\tilde{S}_{k0}\tilde{S}_{00}^{-1}\tilde{S}_{0k} \mapsto A\tilde{S}_{k0}\tilde{S}_{00}^{-1}\tilde{S}_{0k}A^*, \qquad \tilde{S}_{kk} \mapsto A\tilde{S}_{kk}A^*.$$

We conclude that Eq. (17) is multiplied by $\det(A)\det(A^*)$ and, hence, its roots are preserved.

[16]Note that in Bykhovskaya and Gorin [2022] we expressed all the operators in terms of $F = L_c^{-1}$ rather than L_c.

By choosing an $A = \Lambda^{-1/2}$ the covariance matrix Λ becomes identical. Hence, for the rest of the proof we assume without loss of generality that Λ is identical, which means that the matrix elements of ε are i.i.d. standard Gaussians.

Step 2. Let us now reduce the canonical correlations solving Eq. (17) to eigenvalues for a product of projectors. By definition the canonical correlations are eigenvalues of $N \times N$ matrix

$$\tilde{S}_{k0}\tilde{S}_{00}^{-1}\tilde{S}_{0k}\tilde{S}_{kk}^{-1} = \tilde{R}_k\tilde{R}_0^*(\tilde{R}_0\tilde{R}_0^*)^{-1}\tilde{R}_0\tilde{R}_k^*(\tilde{R}_k\tilde{R}_k^*)^{-1}$$

Note that for any two rectangular matrices A and B of the same sizes the non-zero eigenvalues of AB^* and of B^*A coincide. Hence, the desired canonical correlations are also eigenvalues of $T \times T$ matrix

$$\big[\tilde{R}_0^*(\tilde{R}_0\tilde{R}_0^*)^{-1}\tilde{R}_0\big] \cdot \big[\tilde{R}_k^*(\tilde{R}_k\tilde{R}_k^*)^{-1}\tilde{R}_k\big].$$

The last matrix is a product of two projectors:[17] the first one projects on the space spanned by columns of $\tilde{R}_0^*$ and the second one projects on columns of $\tilde{R}_k^*$.

Step 3. The next step is to express via ε various matrices involved in constructing $\tilde{R}_0$ and $\tilde{R}_k$. Let $\mathcal{P}$ be the orthogonal projector on the subspace V. Under $\widehat{H}_0$ we have $\Delta X_t = \mu + \varepsilon_t$. Also

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

Further, we define the $T \times T$ summation matrix Φ. It has 1's below the diagonal and 0's on the diagonal and everywhere above the diagonal:

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

We set

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

By a straightforward linear algebra (see [Bykhovskaya and Gorin, 2022, Section 9.2] for some details) one shows that the linear operator $\tilde{\Phi}$ preserves the space V (orthogonal to $(1, 1, \dots, 1)$). In addition, its restriction on the subspace V coincides with $L_V(I_V - L_V)^{-1}$, where I_V is the identical operator acting in V.

[17] For a closer match to the proposition that we are proving, note also that if P_1 and P_2 are projectors, then eigenvalues of P_1P_2 and $P_1P_2P_1$ are the same.

We can write

$$\tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0) = X_0 + (t-1)\mu + \sum_{\tau=1}^{t-1}\varepsilon_\tau - \frac{t-1}{T}\left(T\mu + \sum_{\tau=1}^{T}\varepsilon_\tau\right) = X_0 + \sum_{\tau=1}^{t-1}\varepsilon_\tau - \frac{t-1}{T}\sum_{\tau=1}^{T}\varepsilon_\tau. \tag{47}$$

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

$$\tilde{\tilde{X}}_t = \tilde{X}_t - X_0 = \sum_{\tau=1}^{t-1}\varepsilon_\tau - \frac{t-1}{T}\sum_{\tau=1}^{T}\varepsilon_\tau = \sum_{\tau=1}^{t-1}\left(\varepsilon_\tau - \frac{1}{T}\sum_{s=1}^{T}\varepsilon_s\right).$$

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

Step 4. Previous steps yield the following expressions for $\tilde{R}_0$ and $\tilde{R}_k$. Take the N–dimensional space $\tilde{\mathcal{U}}$ (belonging to $(T-1)$-dimensional space V) spanned by the columns of $\mathcal{P}\varepsilon^*$. Let $\tilde{P}$ be the orthogonal projector on the space orthogonal to $L_c\tilde{\mathcal{U}}$, $L_c^2\tilde{\mathcal{U}}$, …, $L_c^{k-1}\tilde{\mathcal{U}}$. (Note that L_c can be replaced by L_V in the last definition without changing $\tilde{P}$). Then the space spanned by N columns of $\tilde{R}_0^*$ is $\tilde{P}\tilde{\mathcal{U}}$. On the other hand, the space spanned by N columns of $\tilde{R}_k^*$ is $\tilde{P}L_c^{k-1}\tilde{\Phi}\tilde{\mathcal{U}} = \tilde{P}L_V^k(I_V - L_V)^{-1}\tilde{\mathcal{U}}$. At this point, we see strong similarities with objects in the statement of Proposition 19 with main difference being in the assignment of randomness: L_c is deterministic and $\tilde{\mathcal{U}}$ is random, but $\tilde{L}$ is random and $\mathcal{U}$ is deterministic. Thus, it remains to relocate the random part.

For that we notice that due to the rotational invariance of the Gaussian law (here it is important that we made the covariance matrix Λ identical on the first step), the space $\tilde{\mathcal{U}}$ spanned by the columns of $\mathcal{P}\varepsilon^*$ has the same law as $\tilde{O}^*\mathcal{U}$. The reason is that both laws give uniformly random N–dimensional subspace of $(T-1)$–dimensional space V.

Since everything was previously expressed through the span of columns of $\mathcal{P}\varepsilon^*$, denote $\tilde{\mathcal{U}}$, we now simply replace those by the columns of $\tilde{O}^*\mathcal{U}$. Then the space orthogonal to $L_c\tilde{\mathcal{U}}$, $L_c^2\tilde{\mathcal{U}}$, …, $L_c^{k-1}\tilde{\mathcal{U}}$ becomes the space orthogonal to $L_c\tilde{O}^*\mathcal{U}$, $L_c^2\tilde{O}^*\mathcal{U}$, …, $L_c^{k-1}\tilde{O}^*\mathcal{U}$. Equivalently, this is the space orthogonal to $\tilde{O}^*\tilde{L}\mathcal{U}$, $\tilde{O}^*\tilde{L}^2\mathcal{U}$, …, $\tilde{O}^*\tilde{L}^{k-1}\mathcal{U}$. $\tilde{P}$ is the projector on this space. We conclude that the law of canonical correlations (17) is the same as the law of non-zero eigenvalues of the product of two projectors: the first one projects on the subspace $\tilde{P}\tilde{O}^*\mathcal{U}$ and the second one projects on the subspace $\tilde{P}L_c^k(I_V - L_c)^{-1}\tilde{O}^*\mathcal{U} = \tilde{P}\tilde{O}^*\tilde{L}^k(I_V + \tilde{L})^{-1}\mathcal{U}$. Up to a change of basis (by matrix $\tilde{O}$), which does not change the eigenvalues, we have arrived precisely at the expression from the statement of the proposition. □

The next proposition explains the effect of the replacement $O \leftrightarrow \tilde{L}$ on the eigenvalues of the product of projectors in Theorem 12 and Proposition 19. We need to introduce some additional notations.

Choose positive integers k, N, and $\mathcal{T}$, such that $\mathcal{T} \geq (k+1)N$ and an arbitrary N–dimensional subspace $\mathcal{V}$ in $\mathcal{T}$–dimensional space. Let

$$f^{k,N,\mathcal{T};\mathcal{V}} : SO(\mathcal{T}) \to \{0 \leq x_1 \leq x_2 \leq \cdots \leq x_N \leq 1\}$$

be a map from the group $SO(\mathcal{T})$ of orthogonal $\mathcal{T} \times \mathcal{T}$ matrices of determinant 1 to N–tuples of reals on $[0,1]$ interval, defined by the following procedure: Take $O \in SO(\mathcal{T})$. Let P be the orthogonal projector on the space orthogonal to $O\mathcal{V}$, $O^2\mathcal{V}$,..., $O^{k-1}\mathcal{V}$. Let P_1 be the projector on the subspace $P\mathcal{V}$ and P_2 be the projector on the subspace $PO^k(I_{\mathcal{T}} + O)^{-1}\mathcal{V}$. Then $f^{k,N,\mathcal{T};\mathcal{V}}$ maps O to N largest eigenvalues of $P_1P_2P_1$.

We also need three norms:

(1) $\|v\|_2$ is the L_2 norm of a vector $v = (v_1, v_2, \ldots, v_N)$, defined as $\|v\|_2 = \sqrt{\sum_{i=1}^N v_i^2}$.
(2) $\|v\|_\infty$ is the supremum norm of a vector $v = (v_1, \ldots, v_N)$, defined as $\|v\|_\infty = \max_i |v_i|$.
(3) $\|A\|_2$ is the spectral norm of a matrix A, defined as the square root of the largest eigenvalue of AA^*. Equivalently, $\|A\|_2 = \max_v \frac{\|Av\|_2}{\|v\|_2}$.

Proposition 20. *Suppose that k is fixed, while N is growing and $\mathcal{T}$ depends on N in such a way that $\frac{\mathcal{T}}{N} \in [k+1+C_1, C_2]$ for some $C_1, C_2 > 0$. Let O_1 and O_2 be two $\mathcal{T} \times \mathcal{T}$ random matrices, such that:*

- *O_1 is a uniformly random $\mathcal{T} \times \mathcal{T}$ orthogonal matrix with determinant 1.*
- *The eigenvalues of O_2 are almost surely different from -1.*
- *For each $\varepsilon > 0$ we have*

$$\lim_{N\to\infty} \operatorname{Prob}\left(\|O_1 - O_2\|_2 < \frac{1}{N^{1-\varepsilon}}\right) = 1. \tag{48}$$

Then for each $\varepsilon > 0$ we have

$$\lim_{N\to\infty} \operatorname{Prob}\left(\|f^{k,N,\mathcal{T};\mathcal{V}}(O_1) - f^{k,N,\mathcal{T};\mathcal{V}}(O_2)\|_\infty < \frac{1}{N^{1-\varepsilon}}\right) = 1. \tag{49}$$

Proposition 20 claims a continuity of map $f^{k,N,\mathcal{T};\mathcal{V}}$. The proof needs care because of the inversions in the definition of the map $f^{k,N,\mathcal{T};\mathcal{V}}$.

Remark 21. *Proposition 20 has a version for complex numbers, in which all orthogonal matrices are replaced by unitary matrices. The proof of the complex version is the same.*

The proof of Proposition 20 relies on three lemmas which we prove later in this section. For these lemmas we write matrices O_1 and O_2 of Proposition 20 in the block forms according to the splitting $\mathcal{T} = (\mathcal{T} - N) + N$:

$$O_1 = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \qquad O_2 = \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix}. \tag{50}$$

Lemma 22. *Let O be a $\mathcal{T}\times T$ orthogonal matrix of determinant 1 written in the block form $O=\begin{pmatrix}A & B\\ C & D\end{pmatrix}$ according to the splitting $\mathcal{T}=(\mathcal{T}-N)+N$. If all eigenvalues of O are different from -1, then so are the eigenvalues of D and of $A-B(I_N+D)^{-1}C$.*

Lemma 23. *Under the assumptions of Proposition 20 we have*

$$\lim_{N\to\infty}\operatorname{Prob}\left(\|(A_1-B_1(I_N+D_1)^{-1}C_1)-(A_2-B_2(I_N+D_2)^{-1}C_2)\|_2<\frac{1}{N^{1-\varepsilon}}\right)=1. \tag{51}$$

Lemma 24. *Under the assumptions of Proposition 20, let $\tilde{\mathcal{V}}_0$ be the N–dimensional subspace of $(\mathcal{T}-N)$–dimensional space spanned by the last N coordinate vectors. There exists an $(\mathcal{T}-N)\times(\mathcal{T}-N)$ orthogonal matrix U_1, depending only on $B_1(I_N+D_1)^{-1}$ and an $(\mathcal{T}-N)\times(\mathcal{T}-N)$ orthogonal matrix, U_2, depending both on $B_1(I_N+D_1)^{-1}$ and on $B_2(I_N+D_2)^{-1}$, such that $U_1\tilde{\mathcal{V}}_0=\langle B_1\rangle$, $U_2\tilde{\mathcal{V}}_0=\langle B_2\rangle$ and*

$$\lim_{N\to\infty}\operatorname{Prob}\left(\|U_1-U_2\|_2<\frac{1}{N^{1-\varepsilon}}\right)=1. \tag{52}$$

Proof of Proposition 20. The proof is induction in k with the base case $k=1$ proven in Bykhovskaya and Gorin [2022], see the continuity of $M(Z)$ in the proof of Proposition 13 there. For the induction step we recycle the ideas in the proof of Theorem 12.

First, recall a property of function f, which we established in Steps 1 and 2 of Theorem 12: if U is a $\mathcal{T}\times\mathcal{T}$ orthogonal matrix, then

$$f^{k,N,\mathcal{T};U\mathcal{V}}(O)=f^{k,N,\mathcal{T};\mathcal{V}}(U^*OU). \tag{53}$$

Note that conjugations (by the same orthogonal matrix for O_1 and O_2) leave the three conditions of Proposition 20 unchanged, hence, the (53) implies that statement of proposition remains the same for any choice $\mathcal{V}$.

Second, we make the replacements of Steps 2 and 3 of Theorem 12 individually for $f^{k,N,\mathcal{T};\mathcal{V}}(O_1)$ and $f^{k,N,\mathcal{T};\mathcal{V}}(O_2)$. The replacement $\mathcal{V}\mapsto O^{k-1}\mathcal{V}$ does not change the eigenvalues of $P_1P_2P_1$, while inversion of O_1 and O_2 keeps the conditions of Proposition 20 unchanged. Summing up, we replace the map $O\mapsto f^{k,N,\mathcal{T};\mathcal{V}}(O)$ in Proposition 20 by a new map $O\mapsto\tilde{f}^{k,N,\mathcal{T};\mathcal{V}_0}(O)$ defined through: Let $\mathcal{V}_0$ be the subspace spanned by the last N coordinate vectors in $\mathcal{T}$–dimensional space. Let P be the orthogonal projector on the space orthogonal to $\mathcal{V}, O\mathcal{V}$, $O^2\mathcal{V},\dots,O^{k-2}\mathcal{V}$. Let P_1 be the projector on the subspace $PO^{k-1}\mathcal{V}$ and P_2 be the projector on the subspace $P(I_{\mathcal{T}}+O)\mathcal{V}$ (or, equivalently, on $PO(I_{\mathcal{T}}+O)\mathcal{V}$). Then $\tilde{f}^{k,N,\mathcal{T};\mathcal{V}_0}(O)$ is N largest eigenvalues of $P_1P_2P_1$.

Using the block notations (50), steps 4-7 in the proof of Theorem 12 imply the following almost sure identities:

$$\tilde{f}^{k,N,\mathcal{T};\mathcal{V}_0}(O_1)=\tilde{f}^{k-1,N,\mathcal{T}-N;\langle B_1\rangle}(A_1-B_1(I_N+D_1)^{-1}C_1), \tag{54}$$

$$\tilde{f}^{k,N,\mathcal{T};\mathcal{V}_0}(O_2) = \tilde{f}^{k-1,N,\mathcal{T}-N;\langle B_2\rangle}(A_2 - B_2(I_N + D_2)^{-1}C_2). \tag{55}$$

We would like to check that the right-hand sides of (54) and (55) are close by using the induction assumption.

Using (53) and Lemma 24, we rewrite the right-hand sides of (54) and (55) as:

(56)
$$\tilde{f}^{k-1,N,\mathcal{T}-N;\tilde{\mathcal{V}}_0}\big(U_1^*(A_1 - B_1(I_N+D_1)^{-1}C_1)U_1\big); \qquad \tilde{f}^{k-1,N,\mathcal{T}-N;\tilde{\mathcal{V}}_0}\big(U_2^*(A_2 - B_2(I_N+D_2)^{-1}C_2)U_2\big).$$

Let us check that we can apply the induction assumption to deduce that the expressions of (56) are close to each other:

- By Lemma 16, $A_1 - B_1(I_N + D_1)^{-1}C_1$ is a uniformly random $(\mathcal{T} - N) \times (\mathcal{T} - N)$ orthogonal matrix. By Lemma 24, U_1 is a function of $B_1(I_N + D_1)^{-1}$. Hence, using Lemma 16 again, we conclude that U_1 is independent from $A_1 - B_1(I_N + D_1)^{-1}C_1$. Therefore, $U_1^*(A_1 - B_1(I_N + D_1)^{-1}C_1)U_1$ is a uniformly random orthogonal matrix, as desired.
- By Lemma 22, $(I_N+D_2)^{-1}$ is well-defined and no eigenvalues of $A_2-B_2(I_N+D_2)^{-1}C_2$ are equal to -1. Hence, the eigenvalues of $U_2^*(A_2 - B_2(I_N + D_2)^{-1}C_2)U_2$ are almost surely different from -1.
- Combining Lemmas 23 and 24 we conclude that

$$\lim_{N\to\infty} \operatorname{Prob}\left(\|U_1^*(A_1 - B_1(I_N + D_1)^{-1}C_1)U_1 - U_2^*(A_2 - B_2(I_N + D_2)^{-1}C_2)U_2\|_2 < \frac{1}{N^{1-\varepsilon}} \right) = 1.$$

Hence, using the $(k-1)$ statement, the expressions in (56) are close to each other as $N \to \infty$ and, therefore, (54) is close to (55). □

Proof of Lemma 22. Let us show that D has no eigenvalues -1. We argue by contradiction and assume that there exists an N–dimensional vector v of length 1 such that $Dv = -v$. Note that $B^*B + D^*D = I_N$ by orthogonality of O. Hence, using the notation $\langle \cdot, \cdot \rangle$ for the scalar product, we have

$$\langle Bv, Bv\rangle = \langle B^*Bv, v\rangle = \langle (I_N - D^*D)v, v\rangle = \langle v, v\rangle - \langle Dv, Dv\rangle = 1 - 1 = 0.$$

Therefore, $Bv = 0$, which readily implies that the $\mathcal{T}$–dimensional vector $\binom{0}{v}$ is an eigenvector of O with eigenvalue -1. Contradiction.

Next, for the matrix $A-B(I_N+D)^{-1}C$, let us use its representation as a Cayley transform developed in (39):

$$A - B(I_N + D)^{-1}C = \frac{I_{\mathcal{T}-N} - \mathcal{R}}{I_{\mathcal{T}-N} + \mathcal{R}},$$

where $\mathcal{R}$ is a $(\mathcal{T}-N)\times(\mathcal{T}-N)$ skew-symmetric matrix. If v was an eigenvector of $A - B(I_N+D)^{-1}C$ with eigenvalue -1, then we would have

$$(I_{\mathcal{T}-N}-\mathcal{R})v = -(I_{\mathcal{T}-N}+\mathcal{R})v,$$

which is impossible for non-zero v. □

Proof of Lemma 23. Note that whenever X is a submatrix of Y, we have $\|X\|_2 \le \|Y\|_2$. Hence, the spectral norms of the differences A_1-A_2, B_1-B_2, C_1-C_2, D_1-D_2 are all small with probability tending to 1 as $N\to\infty$. Addition, multiplication, and inversion of matrices are all Lipschitz operations as long as factors are bounded for the multiplication and singular values are bounded away from 0 for the inversion. Therefore, it remains to show that the norms of the factors B_1, $(I_N+D_1)^{-1}$, and C_1 are uniformly bounded (since B_1, $(I_N+D_1)^{-1}$, and C_1 are close to B_2, $(I_N+D_2)^{-1}$, and C_2, respectively, the norms of the latter are then going to be bounded as well). For B_1 and C_1 the bound on the norm is straightforward, as they are submatrices of O_1, whose norm is 1. Hence, $\|B_1\|_2\le 1$ and $\|C_1\|_2\le 1$.

In order to deal with $(I_N+D_1)^{-1}$ we rely on the fact that the distribution of the symmetric $N\times N$ matrix $Y = D_1^*D_1$ is explicit. It has density (see, e.g., [Forrester, 2010, (3.113) and the formula immediately after]) proportional to:

$$\det Y^{-1/2}\det(I_N-Y)^{\frac{\mathcal{T}}{2}-N-1/2}\,dY, \qquad 0<Y<I_N. \tag{57}$$

This is a particular case of the Jacobi ensemble of Definition 5 and we can use the large N asymptotic of the latter recorded in Proposition 11. Therefore, there exists a constant $0<c<1$, such that all the eigenvalues of Y are smaller than c with probability tending to 1 as $N\to\infty$. Hence, by the triangular inequality

$$\min_{\|v\|_2=1}\|(I_N+D_1)v\|_2 \ge 1-\sqrt{c},$$

with probability tending to 1 as $N\to\infty$. We conclude that

$$\lim_{N\to\infty}\operatorname{Prob}\left(\|(I_N+D_1)^{-1}\| < \frac{1}{1-\sqrt{c}}\right) = 1. \quad \square.$$

Proof of Lemma 24. We will be proving a slightly different statement, in which $\tilde{\mathcal{V}}_0$ is the span of the first (rather than last) coordinate vectors. The desired statement of the theorem is then obtained by replacing $U_1\mapsto U_1\cdot\mathfrak{S}$, and $U_2\mapsto U_2\cdot\mathfrak{S}$, where $\mathfrak{S}$ is the (orthogonal matrix) which swaps ith and $(\mathcal{T}-N+1-i)$th basis vectors for $i=1,2,\dots,\mathcal{T}-N$.

We know that the matrices B_1 and B_2 are close to each other and our aim is to show that the orthonormal bases of $\langle B_1\rangle = \langle B_1(I_N+D_1)^{-1}\rangle$ and its orthogonal complement, and $\langle B_2\rangle = \langle B_2(I_N+D_2)^{-1}\rangle$ and its orthogonal complement can be chosen to also be close to each other. For that we need to produce some formulas for these bases, which is what we do in the rest of the proof. The delicacy of this argument stems from the fact that given a

space, in general, there might be no continuous way to produce an orthogonal matrix, such that the space is spanned by its first columns. (For instance, by the hairy ball theorem one can not continuously complement a unit vector in 3–dimensional space to an orthonormal basis.) Hence, we need to be more careful.

We start by replacing B_1 with

$$X_1 := B_1(I_N + D_1)^{-1}\big((I_N + D_1^*)^{-1}B_1^*B_1(I_N + D_1)^{-1}\big)^{-1/2}$$

and replacing B_2 with

$$X_2 := B_2(I_N + D_2)^{-1}\big((I_N + D_2^*)^{-1}B_2^*B_2(I_N + D_1)^{-1}\big)^{-1/2}.$$

Clearly, $\langle B_1\rangle = \langle X_1\rangle$ and $\langle B_2\rangle = \langle X_2\rangle$. The advantage of X_1 and X_2 is that their columns are orthonormal. Indeed,

$$X_1^*X_1 = \big((I_N + D_1^*)^{-1}B_1^*B_1(I_N + D_1)^{-1}\big)^{-1/2}(I_N + D_1^*)^{-1}B_1^*B_1(I_N + D_1)^{-1} \\ \times \big((I_N + D_1^*)^{-1}B_1^*B_1(I_N + D_1)^{-1}\big)^{-1/2} = I_N$$

and similarly for X_2.

Claim. X_1 and X_2 are asymptotically close to each other:

$$\lim_{N\to\infty} \operatorname{Prob}\left(\|X_1 - X_2\|_2 < \frac{1}{N^{1-\varepsilon}}\right) = 1. \tag{58}$$

Note that X_1 and X_2 are built out of O_1 and O_2 with operations of addition, multiplication, inversion, and square root. The first one is Lipschitz in spectral norm, the second one is Lipschitz as long as the factors are uniformly bounded, and for the last two we additionally need the singular values of the factors to be uniformly bounded away from 0 uniformly[18]. We already explained in the proof of Lemma 23 that B_1 has spectral norm at most 1 and that $(I_N + D_1)$ (and hence also its inverse and its transpose) has singular values bounded away from 0 and ∞. Hence, it remains only to deal with $B_1^*B_1$ in the definition of X_1. Since B_1 is a $(\mathcal{T} - N) \times N$ submatrix of uniformly random $\mathcal{T} \times \mathcal{T}$ matrix, the law of $\Lambda = B_1^*B_1$ is explicit. It has density (see, e.g., [Forrester, 2010, (3.113) and the formula immediately after]) proportional to:

$$\det \Lambda^{\frac{\mathcal{T}}{2}-N-1/2} \det(I_N - \Lambda)^{-1/2}\, d\Lambda, \qquad 0 < \Lambda < I_N. \tag{59}$$

This is a particular case of the Jacobi ensemble of Definition 5 and we can use the large N asymptotic of the latter recorded in Proposition 11, which implies that the eigenvalues of Λ are bounded away from 0 as $N \to \infty$. The claim is proven.

[18]For the square root operation on positive-definite matrices $x \mapsto \sqrt{x}$ we can first rescale x so that its spectrum belongs to $[c_0, 1]$ segment for some $c_0 > 0$ and then use Taylor series expansion of the square root: $\sqrt{x} = \sqrt{1 + (x-1)} = 1 + \frac{x-1}{2} - \frac{1}{4}(x-1)^2 + \dots$ to deduce the Lipschitz property.

Next, we produce the desired orthogonal matrix U_1 by the Gramm-Schmidt orthogonalization procedure: letting e_k be the k–th coordinate vector in $(\mathcal{T}-N)$–dimensional space, and X_1^k be the k–th column of X_1, we start from $(\mathcal{T}-N)$ vectors

$$X_1^1, X_1^2, \dots, X_1^N,\ e_{N+1}, e_{N+2}, \dots, e_{\mathcal{T}-N}$$

and orthogonalize them. This is a valid procedure, since the Gramm matrix of the above vectors is almost surely non-degenerate (this is equivalent to the non-degeneracy of the top $N\times N$ corner of X_1, which is true due to absolute continuity of the distribution of this corner with respect to the Lebesgue measure on $N\times N$ matrices that can be deduced from Remark 17).

We set the columns of U_1 to be the vectors from the orthogonalization procedure. Since the vectors are orthonormal, U_1 is orthogonal. Note that since the columns of X_1 are orthonormal, the first N steps of the orthogonalization procedure are trivial and the first N columns of U_1 are $X_1^1, X_1^2, \dots, X_1^N$. In particular, these N columns span $\langle B_1\rangle$, as desired.

We proceed to the construction of U_2. It is tempting to do exactly the same procedure (with all indices 1 replaced by indices 2), but that is not going to work: the problem is that while the top $N\times N$ corner of X_1 was almost surely non-degenerate, but it can have singular values arbitrary close to 0. Eventually, this leads to unstability of the orthogonalization procedure and, hence, there is no way to guarantee that the results of orthogonalization for X_1 and X_2 are close to each other.

Therefore, we proceed in a different way. Set $F := U_1^{-1}X_2$. Because the first N columns of U_1 are X_1, we have

$$U_1^{-1}X_1 = \begin{pmatrix} I_N \\ 0_{(\mathcal{T}-2N)\times N}\end{pmatrix},$$

where $0_{(\mathcal{T}-2N)\times N}$ stays for the $(\mathcal{T}-2N)\times N$ filled with 0 matrix elements. Hence, since X_1 and X_2 were close, we have

$$\lim_{N\to\infty} \operatorname{Prob}\left(\left\| F - \begin{pmatrix} I_N \\ 0_{(\mathcal{T}-2N)\times N}\end{pmatrix}\right\|_2 < \frac{1}{N^{1-\varepsilon}}\right) = 1. \tag{60}$$

Let F^k, $k=1,\dots,N$, denote the columns of F and consider $\mathcal{T}-N$ vectors

$$F^1, F^2, \dots, F^N,\ e_{N+1}, e_{N+2}, \dots, e_{\mathcal{T}-N}.$$

We are going to orthogonalize these vectors. The advantage over the procedure we used for X_1 is that now the top $N\times N$ submatrix of F is close to identity, which is going to make the orthogonalization procedure well-behaved. In order to make the orthogonalization procedure explicit, we are going to use a block version of the Cholesky decomposition.

For that set $M = \mathcal{T} - 2N$ and write F in the block form according to the splitting $\mathcal{T} - N = N + M$:

$$F = \begin{pmatrix} Y_2 \\ Z_2 \end{pmatrix}.$$

Let W_2 denote the $(\mathcal{T} - N) \times (\mathcal{T} - N)$ matrix written in the $N + M$ block form as

$$W_2 = \begin{pmatrix} Y_2 & 0 \\ Z_2 & I_M \end{pmatrix}.$$

We would like to perform orthogonalization of the columns of W_2. For that we first compute

$$W_2^* W_2 = \begin{pmatrix} Y_2^* Y_2 + Z_2^* Z_2 & Z_2^* \\ Z_2 & I_M \end{pmatrix}. \tag{61}$$

We further would like to represent $W_2^* W_2$ as

$$W_2^* W_2 = \begin{pmatrix} I_N & 0 \\ Q_2 & I_M \end{pmatrix} \begin{pmatrix} G_2 & 0 \\ 0 & H_2 \end{pmatrix} \begin{pmatrix} I_N & Q_2^* \\ 0 & I_M \end{pmatrix} = \begin{pmatrix} G_2 & G_2 Q_2^* \\ Q_2 G_2 & Q_2 G_2 Q_2^* + H_2 \end{pmatrix}. \tag{62}$$

Comparing with (61) we conclude that

$$G_2 = Y_2^* Y_2 + Z_2^* Z_2 = F^* F, \quad Q_2 = Z_2 (F^* F)^{-1}, \quad H_2 = I_M - Z_2 (F^* F)^{-1} Z_2^*. \tag{63}$$

Since the spectral norm of a submartix is at most the spectral norm of the matrix, (60) implies that

$$\lim_{N \to \infty} \operatorname{Prob}\left(\|Y_2 - I_N\|_2 < \frac{1}{N^{1-\varepsilon}} \right) = 1, \qquad \lim_{N \to \infty} \operatorname{Prob}\left(\|Z_2 - 0_{M \times N}\|_2 < \frac{1}{N^{1-\varepsilon}} \right) = 1. \tag{64}$$

Therefore, W_2 is close to I_{M+N}, G_2 is close to I_N, Q_2 is close to $0_{N \times M}$, H_2 is close to I_M.

Note that G_2 and H_2 are positive-definite symmetric matrices, hence, they have well-defined square roots. In addition,

$$\begin{pmatrix} I_N & Q_2^* \\ 0 & I_M \end{pmatrix}^{-1} = \begin{pmatrix} I_N & -Q_2^* \\ 0 & I_M \end{pmatrix}.$$

The goal of all these manipulations with matrices is to define

$$\tilde{U}_2 = W_2 \cdot \begin{pmatrix} I_N & -Q_2^* \\ 0 & I_M \end{pmatrix} \cdot \begin{pmatrix} G_2^{-1/2} & 0 \\ 0 & H_2^{-1/2} \end{pmatrix}. \tag{65}$$

The two key properties of $\tilde{U}_2$ are:

- The span of the first N columns of $\tilde{U}_2$ coincides with $\langle F \rangle$.
- $\tilde{U}_2$ is orthogonal. Indeed, using (62) we have

$$\tilde{U}_2 \tilde{U}_2^* = W_2 \begin{pmatrix} I_N & -Q_2^* \\ 0 & I_M \end{pmatrix} \begin{pmatrix} G_2^{-1} & 0 \\ 0 & H_2^{-1} \end{pmatrix} \begin{pmatrix} I_N & 0 \\ -Q_2 & I_M \end{pmatrix} W_2^* = W_2 (W_2^* W_2)^{-1} W_2^* = I_{\mathcal{T}-N}.$$

Hence, we can finally set

$$U_2 := U_1 \cdot \tilde{U}_2,$$

We have

$$U_2 \tilde{\mathcal{V}}_0 = \langle U_1 F \rangle = \langle X_2 \rangle = \langle B_2 \rangle,$$

as desired. It remains to show that the matrix $\tilde{U}_2$ is very close to identity, as this would imply that U_2 is close to U_1. For that we consider each factor in (65) and see that they are close to identical matrices by (64). Hence,

$$\lim_{N\to\infty} \mathrm{Prob}\left(\left\| \tilde{U}_2 - I_{N+M} \right\|_2 < \frac{1}{N^{1-\varepsilon}} \right) = 1,$$

as desired. □

Proof of Theorem 8. We start by explicitly constructing the desired coupling. For the Jacobi ensemble we use the realization of Theorem 12 and for the matrix of the Johansen test we use the realization of Proposition 19. We set $\mathcal{T} = T - 1$ to match the notations and it remains to couple O of Theorem 12 with $\tilde{L} = -\tilde{O} L_V \tilde{O}^*$ of Proposition 19.

The eigenvalues of L_V are all roots of unity of order T different from 1. In the complex case $\beta = 2$ we can diagonalize L_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. In the real case $\beta = 1$, the matrix L_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 $\beta = 1$, then 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 L_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 a 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 [19]

$$\text{Prob}\left(\max_{1\le i,j<\mathcal{T}} \left|D - D^{\text{rand}}\right|_{ij} > \frac{1}{\mathcal{T}^{1-\delta}}\right) < c_1(\beta)\cdot\mathcal{T}\cdot\exp\left(-c_2(\beta)\frac{\mathcal{T}^{2\delta}}{\log\mathcal{T}}\right). \tag{66}$$

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 spectral norm, which we used in Proposition 20.

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 $-\tilde{O}L_V\tilde{O}^*$ with $O_2DO_2^*$ and replace O with $O_2D^{\text{rand}}O_2^*$. 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:

$$-\tilde{O}L_V\tilde{O}^* \stackrel{d}{=} O_2DO_2^*, \qquad O \stackrel{d}{=} O_2D^{\text{rand}}O_2^*.$$

The right-hand sides of the identities provide the desired coupling and (66) implies that these two random matrices are close to each other as $\mathcal{T}\to\infty$.

It now remains to apply Proposition 20 (see also Remark 21) with the first matrix being $O_2D^{\text{rand}}O_2^*$ and the second matrix being $O_2DO_2^*$. □

7.4. Small rank perturbations. In this section we prove Theorem 3 by combining Theorem 8 with Proposition 11 and general statements about small rank perturbations. The key step of the proof is the following observation:

Theorem 25. *Let R_i, $i = 0, k$ be as in Section 2 for X_t solving Eq. (1) and let $\tilde{R}_i$, $i = 0, k$ be as in Section 3 under $\widehat{H}_0$ for X_t solving Eq. (23). Suppose that X_0 and the noises ε_t used in the constructions of R_i and $\tilde{R}_i$ are the same. Introduce $T\times T$ projection matrices:*

$$P_0 = R_0^*(R_0R_0^*)^{-1}R_0, \quad P_k = R_k^*(R_kR_k^*)^{-1}R_k, \quad \tilde{P}_0 = \tilde{R}_0^*(\tilde{R}_0\tilde{R}_0^*)^{-1}\tilde{R}_0, \quad \tilde{P}_k = \tilde{R}_k^*(\tilde{R}_k\tilde{R}_k^*)^{-1}\tilde{R}_k. \tag{67}$$

Then under the assumptions (8), (9) of Theorem 3 we have

$$\lim_{N\to\infty}\frac{1}{N}\text{rank}\left(P_0P_kP_0 - \tilde{P}_0\tilde{P}_k\tilde{P}_0\right) = 0. \tag{68}$$

Proof. Throughout the proof we assume that the matrices $R_iR_i^*$ and $\tilde{R}_i\tilde{R}_i^*$ are invertible. In principle, invertibility might fail for some N: in such situation we can still use Moore–Penrose inverse in order for the statements to make sense, and we are not going to detail this.

In the following argument we use various properties of ranks:

- If a matrix A differs from a matrix B only in $\mathfrak{r}$ columns, then $\text{rank}(A-B)\le\mathfrak{r}$;
- $\text{rank}(CA-CB)\le\text{rank}(A-B)$;

[19] All the constants can be made explicit, following Meckes and Meckes [2013].

- $\operatorname{rank}(A+B) \le \operatorname{rank}(A) + \operatorname{rank}(B)$;
- If matrices A and B are invertible, then $\operatorname{rank}(A^{-1} - B^{-1}) = \operatorname{rank}(A - B)$.

We refer to the time series defined by (1) as X_t and to the time series defined by (23) as $\mathcal{X}_t$. We form two $N \times T$ matrix X and $\mathcal{X}$ with columns X_t and $\mathcal{X}_t$, $t = 1, \dots, T$, respectively. Our first task is to show that $\frac{1}{N}\operatorname{rank}(X - \mathcal{X}) \to 0$ as $N \to \infty$.

For that we subtract (23) from (1) to get:

$$\Delta X_t - \Delta \mathcal{X}_t = \Pi X_{t-k} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t - \mu, \quad t = 1, 2, \dots, T. \tag{69}$$

In the matrix form, (69) represents the $N \times T$ matrix $\Delta(X - \mathcal{X})$ with columns $\Delta X_t - \Delta \mathcal{X}_t$ as a sum of $k+2$ low rank matrices, with the total rank (coming from the right-hand side of (69)) at most

$$\operatorname{rank}(\Pi) + \sum_{i=1}^{k-1} \operatorname{rank}(\Gamma_i) + d_D + 1. \tag{70}$$

The matrix $X - \mathcal{X}$ is obtained from $\Delta(X - \mathcal{X})$ by multiplication by the summation matrix

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

Hence, the rank of $X - \mathcal{X}$ is at most (70) and $\frac{1}{N}\operatorname{rank}(X - \mathcal{X}) \to 0$ by assumption (9) of Theorem 3.

Next, we should take into account that the procedures for constructing R_0, R_k from X and $\tilde{R}_0$, $\tilde{R}_k$ from $\mathcal{X}$ are slightly different. Namely, the latter involves cyclic shifts of indices, rather than usual shifts, involves regressing over only constants, rather than d_D deterministic terms D_t, and finally involves detrending (13). However, cyclic shifts only affect the first $k-1$ indices t and, hence, lead to bounded difference in ranks, and similarly for detrending. Regressing on d_D terms leads to another $O(d_D)$ difference in ranks, which is negligible after division by N in the limit $N \to \infty$ by (9). The conclusion is that R_0, R_k from one side and $\tilde{R}_0$, $\tilde{R}_k$ on the the other side are constructed from two finite sets of matrices, which differ by small rank perturbations, by finitely many of operations of addition, multiplication, and inversion. Each of these operations preserves the smallness of the rank of perturbations and, hence,

$$\lim_{N\to\infty} \frac{1}{N}\operatorname{rank}(R_0 - \tilde{R}_0) = 0, \quad \lim_{N\to\infty} \frac{1}{N}\operatorname{rank}(R_k - \tilde{R}_k) = 0. \tag{71}$$

Since $P_0P_kP_0$ and $\tilde{P}_0\tilde{P}_k\tilde{P}_0$ are obtained from R_0, R_k and $\tilde{R}_0$, $\tilde{R}_k$, respectively, by the same algebraic operations, (68) follows from (71). □

Proof of Theorem 3. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ denote the eigenvalues of $\mathcal{C} = S_{kk}^{-1}S_{k0}S_{00}^{-1}S_{0k}$ and let $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_N$ denote the eigenvalues of $\tilde{\mathcal{C}} = \tilde{S}_{kk}^{-1}\tilde{S}_{k0}\tilde{S}_{00}^{-1}\tilde{S}_{0k}$. Combining Theorem 8 with Proposition 11 we conclude that the empirical measure of $\tilde{\lambda}_i$ converges:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta_{\tilde{\lambda}_i} = \mu_{2,\tau-k}. \tag{72}$$

We would like to show that $\tilde{\lambda}_i$ can be replaced by λ_i in (72). Note that although the spectra of matrices $\mathcal{C}$ and $\tilde{\mathcal{C}}$ are real, but these matrices are not symmetric. Similarly to (68), $\frac{1}{N}\mathrm{rank}(\mathcal{C}-\tilde{\mathcal{C}}) \to 0$, however, in general, for non-symmetric matrices even rank 1 perturbations can lead to significant changes in the spectrum. Hence, we need to be more careful and symmetrize $\mathcal{C}$ and $\tilde{\mathcal{C}}$ by using projectors as in (68).

Note that for any two $K \times M$ matrices A and B, the non-zero eigenvalues of AB^* and of B^*A coincide. Recalling that $S_{ij} = R_iR_j^*$ and using the notations (67), we conclude that the eigenvalues of $\mathcal{C}$ are the same as N largest eigenvalues of P_kP_0. Since $P_0^2 = P_0$, they are also the same as N largest eigenvalues of $P_kP_0P_0$ and the same as those of $P_0P_kP_0$. Similarly, the eigenvalues of $\tilde{\mathcal{C}}$ are the same as N largest eigenvalues of $\tilde{P}_0\tilde{P}_k\tilde{P}_0$.

Denote $\mathfrak{r} = \mathrm{rank}(P_0P_kP_0 - \tilde{P}_0\tilde{P}_k\tilde{P}_0)$. All the involved matrices are symmetric and we can use classical inequalities between eigenvalues of a Hermitian matrix A and Hermitian matrix $A+B$, where B has rank $\mathfrak{r}$, see, e.g., Horn and Johnson [2013, Corollary 4.3.5]. In our situation the inequalities read

$$\lambda_{m-\mathfrak{r}} \geq \tilde{\lambda}_m \geq \lambda_{m+\mathfrak{r}}, \quad 1 \leq m-\mathfrak{r} \leq m+\mathfrak{r} \leq N. \tag{73}$$

Therefore, for any points $0 < a < b < 1$,

$$\Big|\#\{1 \leq i \leq N \mid \lambda_i \in [a,b]\} - \#\{1 \leq i \leq N \mid \tilde{\lambda}_i \in [a,b]\}\Big| \leq 2\mathfrak{r}.$$

Hence, (72) implies

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i} = \mu_{2,\tau-k}. \tag{74}$$

□

8. Appendix 2. Discussion of asymptotics under H_0 and H_1

The goal of this section is to discuss the asymptotics of the test statistic $LR_{N,T}(r)$ of (18) under various data generating processes (11) generalizing $\widehat{H}_0$ of (22) and Theorem 9.

8.1. Beyond $\widehat{H}_0$. We start by working under a slightly more restrictive assumption than (9) of Theorem 3. Let $\|A\|_2$ be the spectral norm of a matrix A.

Conjecture 26. *Fix some $k \in \mathbb{N}$, $C > 0$. Suppose that the data generating process is*

$$\Delta X_t = \mu + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t, \qquad t = 1, \dots, T, \qquad \text{where} \tag{75}$$

(1) $\varepsilon_t \sim$ *i.i.d.* $\mathcal{N}(0, \Lambda)$ *and the covariance matrix* Λ *satisfies* $\|\Lambda\|_2 < C$ *and* $\|\Lambda^{-1}\|_2 < C$*;*

(2) $\|\Gamma_i\|_2 < C$ *and* $\operatorname{rank}(\Gamma_i) < C$ *for all* $1 \le i \le k-1$*;*

(3) *All roots of the following characteristic equation* (76) *satisfy*[20] $|z| > 1 + C^{-1}$*:*

$$\det\left(I_N - \sum_{i=1}^{k-1} \Gamma_i z^i\right) = 0; \tag{76}$$

(4) $\|\Gamma_j \Delta X_{1-i}\|_2 \le C$ *and* $\|\Gamma_j \mu\| \le C$ *for all* $1 \le i, j \le k-1$.

Then as $T, N \to \infty$ in such a way that $\frac{T}{N} \in [k + 1 + C^{-1}, C]$, the conclusion of Theorem 9 continues to hold with the same $c_1(N,T)$ and $c_2(N,T)$:

$$\frac{\sum_{i=1}^{r} \ln(1 - \tilde{\lambda}_i) - r \cdot c_1(N,T)}{N^{-2/3} c_2(N,T)} \xrightarrow[T,N\to\infty]{d} \sum_{i=1}^{r} \mathfrak{a}_i. \tag{77}$$

We do not expect the conditions in Conjecture 26 to be optimal. For instance, the Gaussianity assumption can likely be relaxed, as the simulations of Bykhovskaya and Gorin [2022, Section 7.1] indicate, and it is plausible that $\operatorname{rank}(\Gamma_i) < C$ condition can be replaced with slow growth of $\operatorname{rank}(\Gamma_i)$, as in Theorem 3. Nevertheless, we wanted to record Conjecture 26 in the present form, as a precise statement to be addressed in the future work. We are not giving a proof of Conjecture 26 here: the required mathematical apparatus does not exist so far. Instead, we are going to provide a heuristic argument for its validity based on our recent results in Bykhovskaya and Gorin [2023] in a related, yet different setting.

Bykhovskaya and Gorin [2023] studied the following general setting: let $\mathbf{U}$ and $\mathbf{V}$ be two random linear subspaces in S–dimensional space with $\dim(\mathbf{U}) = K$, $\dim(\mathbf{V}) = M$ and all three numbers K, M, S assumed to be growing to infinity. In addition, suppose that there are $\mathfrak{q}$ special vectors $\mathbf{u}_1, \dots, \mathbf{u}_\mathfrak{q}$ inside $\mathbf{U}$ and other $\mathfrak{q}$ special vectors $\mathbf{v}_1, \dots, \mathbf{v}_\mathfrak{q}$ inside $\mathbf{V}$, where $\mathfrak{q}$ is assumed to stay finite as other parameters grow. We directly observe $\mathbf{U}$ and $\mathbf{V}$, but not $\mathbf{u}_1, \dots, \mathbf{u}_\mathfrak{q}$ or $\mathbf{v}_1, \dots, \mathbf{v}_\mathfrak{q}$. Can we reconstruct $\mathbf{u}_1, \dots, \mathbf{u}_\mathfrak{q}$, $\mathbf{v}_1, \dots, \mathbf{v}_\mathfrak{q}$, or at least identify their presence by looking at the squared sample canonical correlations between $\mathbf{U}$ and $\mathbf{V}$ and corresponding vectors?

The connection to our cointegration tests comes from taking as $\mathbf{U}$ the space spanned by the N rows of $\tilde{R}_0$, as defined after (14), and as $\mathbf{V}$ the space spanned by the rows of $\tilde{R}_k$. The value $\mathfrak{q}$ corresponds to the cointegration rank and $\mathbf{v}_1, \dots, \mathbf{v}_\mathfrak{q}$ correspond to the cointegrating relationships.

[20] This guarantees that ΔX_t is $I(0)$ process, which is a standard assumption in the cointegration literature.

While any finite $\mathbb{q}$ can be analyzed in a similar fashion, let us stick to $\mathbb{q} = 1$ case for simplicity, so that we have a single vector $\mathbf{u} \in \mathbf{U}$ and another vector $\mathbf{v} \in \mathbf{V}$. The most important quantity is the sample squared correlation coefficient r^2 between vectors $\mathbf{u}$ and $\mathbf{v}$. It turns out that if r^2 is large (i.e., close to 1, because $0 \le r^2 \le 1$), then the largest canonical correlation between $\mathbf{U}$ and $\mathbf{V}$ is clearly separated from the rest (reminiscent of Figure 1) and the corresponding eigenvectors can be used to extract information on $\mathbf{u}$ and $\mathbf{v}$. On the other hand, if r^2 is small, then the histogram of the canonical correlations does not have such a spiked eigenvalue and all the information about $\mathbf{u}$ and $\mathbf{v}$ is washed out. Bykhovskaya and Gorin [2023, Theorem 2.5, Theorem 3.2, Theorem 3.3, Theorem 3.4] proved the existence of $r^2_{\text{critical}} \in (0,1)$ separating the above two regimes for a variety of settings for the data generating process for $\mathbf{U}$, $\mathbf{V}$, $\mathbf{u}$, and $\mathbf{v}$, see also Bao et al. [2019], Yang [2022b]. However, the results of Bykhovskaya and Gorin [2023] do not address the setting relevant to cointegration and further new ideas would be necessary to find the value of r^2_{critical} for cointegration or rigorously prove its existence. Nevertheless, because the cointegration testing is also based on canonical correlations, one expects that the same phenomenology is true for it and, therefore, there should be the following dichotomy:

(1) If the linear subspace (in T–dimensional space) spanned by the N rows of $\tilde{R}_0$ (as defined after (14)) has a special vector $\mathbf{u}$ and the linear subspace spanned by rows of $\tilde{R}_k$ has a special vector $\mathbf{v}$, such that the sample squared correlation coefficient between $\mathbf{u}$ and $\mathbf{v}$ is atypically large compared to correlation coefficients of other vectors (e.g., if it is close to 1), then the histogram of all squared canonical correlations would have a spike as in Figure 1 and we should be able to reject the null of no cointegration.
(2) Otherwise, there would be no spikes (e.g., as in Figure 10) and we expect validity of asymptotics as in (25) and (77) consistent with the hypothesis of no cointegration.

We now present heuristics in favor of Conjecture 26 based on this dichotomy. Some of the technical details are omitted as we try to express the key ideas instead.

Heuristics for Conjecture 26. For simplicity of the presentation we stick to the case $k = 2$, take the covariance matrix Λ to be identical, set $\mu = 0$, and let Γ_1 to be a matrix, which has θ in the upper-left corner and 0 everywhere else. Clearly, $\operatorname{rank}(\Gamma_1) = 1$ and the only root of (76) is $1/\theta$, hence, the third condition in the statement of Conjecture 26 turns into $|\theta| < 1$.

Note that if we look only at the last $(N-1)$ out of N coordinates of X_t, then we are in the setting of Theorem 9 and asymptotics (25) holds. In particular, the largest canonical correlation is not separated from the rest. Hence, we only need to investigate how the addition of the special first row changes the situation. We will rely on the above dichotomy for our assessment. There are two ways how the addition of the first coordinate changes the setting compared to the situation when it did not exist (and N was smaller by 1):

(1) The matrices $\tilde{R}_0$ and $\tilde{R}_k$ have a new first row each. Hence, we should check how large is the correlation between these first rows, if they are viewed as the special vectors $\mathbf{u}$ and $\mathbf{v}$.
(2) We projected the data orthogonally to $\tilde{Z}_{1t}$ in Step 3 of the procedure, see (14). The $\tilde{Z}_{1t}$ matrix also has a new first row, hence, we are now decreasing the dimension by 1 via projecting orthogonally to an additional vector.

Let y_t, $t = 1, 2, \dots, T$ denote the first coordinate of X_t. It solves the scalar recurrence

$$\Delta y_t = \theta \Delta y_{t-1} + \xi_t, \tag{78}$$

where ξ_t is the first coordinate of ε_t, and therefore a Gaussian $\mathcal{N}(0,1)$ random variable, i.i.d. in time t. Iterating (78), we get

$$\Delta y_t = \theta^t (y_0 - y_{-1}) + \sum_{\tau=1}^{t} \theta^{t-\tau} \xi_\tau, \qquad y_t = y_0 + \frac{\theta - \theta^{t+1}}{1-\theta}(y_0 - y_{-1}) + \sum_{\tau=1}^{t} \frac{1-\theta^{t+1-\tau}}{1-\theta}\xi_\tau. \tag{79}$$

Next, we make the detrending of Step 1 in Procedure 2 of Section 3. As in (13), we define

$$\begin{aligned} \tilde{y}_t &= y_{t-1} - \frac{t-1}{T}(y_T - y_0) \\ &= y_0 + \frac{\theta - \theta^t}{1-\theta}(y_0 - y_{-1}) + \sum_{\tau=1}^{t-1} \frac{1-\theta^{t-\tau}}{1-\theta}\xi_\tau - \frac{t-1}{T}\left(\frac{\theta - \theta^{T+1}}{1-\theta}(y_0 - y_{-1}) + \sum_{\tau=1}^{T} \frac{1-\theta^{T+1-\tau}}{1-\theta}\xi_\tau \right). \end{aligned} \tag{80}$$

Recalling cyclic shifts of Step 2 in Procedure 2, we set

$$\tilde{z}_t = \tilde{y}_{t-1}, \quad 2 \le t \le T, \qquad \tilde{z}_1 = \tilde{y}_T.$$

The vector $\tilde{z}_t$, $1 \le t \le T$, is the first row of $\tilde{Z}_{2t}$, $1 \le t \le T$, viewed as an $N \times T$ matrix. Simultaneously, the vector Δy_t, $1 \le t \le T$, is the first row of $\tilde{Z}_{0t}$. Recalling Step 3 in Procedure 2 and its restatement in terms of projectors at the end of Section 3.1, we analyze the sample correlation coefficients between the vectors $\tilde{z}_t$ and Δy_t projected orthogonally to the constant vector and $\tilde{Z}_{1t}$. The first row of $\tilde{Z}_{1t}$ is Δy_{t-1} (with cyclic shift of index, so that the $t = 1$ coordinate is actually Δy_T). Hence, we are allowed to subtract multiples of the constant vector and multiples of Δy_{t-1} from either of $\tilde{z}_t$ or Δy_t without changing the desired sample correlation coefficient. Therefore, subtracting multiples of constants, we replace $\tilde{z}_t$ with a vector whose t-th coordinate for $2 \le t \le T$ is

$$\frac{-\theta^{t-1} - \frac{t-T/2}{T}(1-\theta^{T+1})}{1-\theta}(y_0 - y_{-1}) + \sum_{\tau=1}^{t-2} \frac{1-\theta^{t-\tau-1}}{1-\theta}\xi_\tau - \frac{t-2}{T}\left(\sum_{\tau=1}^{T} \frac{1-\theta^{T+1-\tau}}{1-\theta}\xi_\tau \right), \tag{81}$$

and the first coordinate is given by a similar expression, which we omit. By subtracting $\theta \Delta y_{t-1}$, we replace Δy_t with a vector whose t-th coordinate for $2 \le t \le T$ is simply ξ_t

Claim. The squared sample correlation coefficient between the vectors (81) and ξ_t tends to 0 as $T \to \infty$.

The claim follows from three computations:

A) The scalar product between these two vectors grows as $O(T)$.
B) The scalar product of (81) with itself is of order T^2.
C) The scalar product of the vector ξ_t, $1 \le t \le T$, with itself is of order T.

Each computation is a straightforward application of the Law of Large Numbers and Central Limit Theorem for i.i.d. random variables and we leave the details to the reader. We only note that the condition $|\theta| < 1$ and boundness of $y_0 - y_{-1}$ are both used here. Together, these computations imply that the sample correlation coefficient is of order $O(T^{-1})$, thus proving the claim.

In order to further pass from the two vectors in the claim to the first rows of the two matrices $\tilde{R}_0$ and $\tilde{R}_k$, we need to project orthogonally to the constant vector, to the vector Δy_{t-1} (which is the first row of $\tilde{Z}_{1t}$), and to the remaining $(N-1)$ rows of $N \times T$ matrix $\tilde{Z}_{1t}$, $1 \le t \le T$. One can check that projecting orthogonally to the first two vectors does not change the conclusion of the claim — this is simply because these two vectors are very close to being orthogonal to the vectors of the claim. Showing that projecting orthogonally to the last $N-1$ rows of $\tilde{Z}_{1t}$ preserves the conclusion of the claim is a more challenging computation, which we record in the following abstract lemma, which is proven later.

Lemma 27. *Suppose that as $T \to \infty$ we are given a T–dimensional space and the following random data inside it: two vectors $\mathbf{a}$ and $\mathbf{b}$, such that the angle[21] between them tends to $\pi/2$ as $T \to \infty$, and a linear subspace $\mathcal{V}$ of dimension M. We assume that the ratio M/T tends to a number α such that $0 < \alpha < 1$ and that $\mathcal{V}$ is uniformly distributed among all subspaces of dimension M and is independent of $\mathbf{a}$ and $\mathbf{b}$. Then the angle between orthogonal projections of $\mathbf{a}$ and $\mathbf{b}$ onto $\mathcal{V}$ tends to $\pi/2$ as $T \to \infty$.*

The lemma is applicable in our situation, because the last $N-1$ rows of $\tilde{Z}_{1t}$, $1 \le t \le T$, are formed by $(N-1)T$ i.i.d. $\mathcal{N}(0,1)$ random variables, independent from y_t. Because of the invariance of the Gaussian law with identical covariance matrix under orthogonal transformations, the distribution of the space spanned by these $N-1$ rows is invariant under orthogonal transformations, which is the same as being uniformly distributed.

The overall conclusion from the discussion is that the sample correlation coefficient between the new first rows of the matrices $\tilde{R}_0$ and $\tilde{R}_k$ tends to 0 as $N, T \to \infty$. Hence, by the dichotomy, these rows can not be special vectors which cause the appearance of a spike in the histogram of eigenvalues. Therefore, we expect that (77) holds. □

[21]Note that the cosine of the angle between $\mathbf{a}$ and $\mathbf{b}$ matches the sample correlation coefficient $\frac{\langle \mathbf{a},\mathbf{b}\rangle}{\sqrt{\langle \mathbf{a},\mathbf{a}\rangle \langle \mathbf{b},\mathbf{b}\rangle}}$.

Remark 28. *One additional effect which we have not examined in the above heuristics is that the last $N-1$ rows of $\tilde{Z}_0$ and $\tilde{Z}_2$ matrices should also be projected orthogonally to Δy_{t-1} (in addition to projecting orthogonally to $\tilde{Z}_1$ covered by the setting of Theorem 9). Because Δy_{t-1} is independent from the rest and is close to being orthogonal to every other vector entering into the procedure, we do not expect this effect to significantly change the asymptotics of the canonical correlations.*

We now come back to Lemma 27.

Proof of Lemma 27. Let $\mathcal{W}$ denote the two-dimensional space spanned by the vectors $\mathbf{a}$ and $\mathbf{b}$. Let us introduce canonical bases of spaces $\mathcal{W}$ and $\mathcal{V}$, see Anderson [2003, Chapter 12] or Muirhead [2009, Section 11.3] for the general introduction to canonical correlations and corresponding variables. Thus, we choose an orthonormal basis of $\mathcal{W}$, $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$ and an orthonormal basis of $\mathcal{V}$, $\mathbf{v}_1, \dots, \mathbf{v}_M \in \mathcal{V}$, such that $\langle \mathbf{w}_1, \mathbf{v}_1 \rangle = c_1$, $\langle \mathbf{w}_2, \mathbf{v}_2 \rangle = c_2$ and all other scalar products $\langle \mathbf{w}_i, \mathbf{v}_j \rangle$ are zeros. We can assume without loss of generality that $1 \geq c_1 \geq c_2 \geq 0$. These numbers are canonical correlations between spaces $\mathcal{W}$ and $\mathcal{V}$. Because the space $\mathcal{W}$ is uniformly distributed along all M–dimensional subspaces, the distribution of the squared correlations (c_1^2, c_2^2) is explicit, it equals the distribution of eigenvalues of the Jacobi ensemble $\mathbf{J}(2; \frac{M-1}{2}, \frac{T-M-1}{2})$, see Muirhead [2009, Corollary 11.3.3], Johnstone [2008, Sections 2.1.1, 2.1.2], and references therein. This means that the joint density of (c_1^2, c_2^2) denoted $\rho(x, y)$ is proportional to:

$$\rho(x,y) \sim (x-y)\, x^{\frac{M-3}{2}} (1-x)^{\frac{T-M-3}{2}}\, y^{\frac{M-3}{2}} (1-y)^{\frac{T-M-3}{2}}. \tag{82}$$

As $T, M \to \infty$, the density $\rho(x, y)$ is sharply concentrated around its maximum. Hence, directly computing the asymptotics of $\rho(x, y)$ we find that

$$\lim_{T\to\infty} (c_1^2, c_2^2) = \lim_{T\to\infty} \left(\frac{M}{T}, \frac{M}{T} \right) = (\alpha, \alpha). \tag{83}$$

Let us expand $\mathbf{a}$ and $\mathbf{b}$ in $(\mathbf{w}_1, \mathbf{w}_2)$ basis:

$$\mathbf{a} = a_1 \mathbf{w}_1 + a_2 \mathbf{w}_2, \qquad \mathbf{b} = b_1 \mathbf{w}_1 + b_2 \mathbf{w}_2.$$

The squared cosine of the angle between $\mathbf{a}$ and $\mathbf{b}$ is then computed as

$$\frac{(a_1 b_1 + a_2 b_2)^2}{(a_1^2 + a_2^2)(b_1^2 + b_2^2)}. \tag{84}$$

The orthogonal projections of $\mathbf{a}$ and $\mathbf{b}$ onto $\mathcal{V}$ are

$$\mathrm{proj}(\mathbf{a}) = c_1 a_1 \mathbf{v}_1 + c_2 a_2 \mathbf{v}_2, \qquad \mathrm{proj}(\mathbf{b}) = c_1 b_1 \mathbf{v}_1 + c_2 b_2 \mathbf{v}_2.$$

Hence, the squared cosine of the angle between two projections is

$$\frac{(c_1^2 a_1 b_1 + c_2^2 a_2 b_2)^2}{(c_1^2 a_1^2 + c_2^2 a_2^2)(c_1^2 b_1^2 + c_2^2 b_2^2)}. \tag{85}$$

Using (83), it becomes clear that (85) tends to 0 as $T \to \infty$ whenever (84) does. $\square$

8.2. Power. We proceed to our next computation, supplementing Conjecture 26. This time we would like to explain what changes in the asymptotics, if the data generating process satisfies the alternative H_1, rather than the null-hypothesis H_0. For the clarity of the exposition, we only concentrate on one particular instance of $k = 1$ case here and deal with the N-dimensional data generating process

$$\Delta X_t = \theta E_{11} X_{t-1} + \varepsilon_t, \qquad t = 1, \dots, T, \qquad \text{where} \tag{86}$$

$\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \Lambda)$, E_{11} is the matrix with 1 in top-left corner and 0s everywhere else. θ is a real parameter, we set $\beta = 1 + \theta$, which implies that the first coordinate of X_t is a scalar process y_t solving

$$y_t = \beta y_{t-1} + \xi_t, \qquad \xi_t \text{ is the first coordinate of } \varepsilon_t. \tag{87}$$

Note that ξ_t are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables for some constant σ^2. Because in the notations of (11) the matrix Π now has rank 1, one hopes that our test statistic of Section 3.1 under (86) behaves significantly differently than under H_0 of Conjecture 4. This would imply that the no-cointegration test based on Theorem 9 has high power against the rank one alternative (86). Let us prove that this is indeed true for large values of the ratio T/N.

Proposition 29. *In the notations of* (86)–(87) *assume that* $|\beta| < 1$ *and let* σ^2 *be the variance of* ξ_t. *Let* $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_N$ *be eigenvalues of the matrix* $\tilde{\mathcal{C}}$ *from Section 3.1 constructed using the* $k = 1$ *procedure. For each* $\epsilon > 0$, *we have*

$$\lim_{T \to \infty} \operatorname{Prob}\left(\tilde{\lambda}_1 > \frac{1}{\frac{2}{1-\beta} + \frac{1+\beta}{6\sigma^2} \left(y_T - y_0\right)^2} - \epsilon \right) = 1, \tag{88}$$

where N *can depend on* T *in* (88) *in an arbitrary way.*

As a corollary, we deduce that our cointegration test has a significant power against rank one stationary alternative, and this power tends to 1 as $T/N \to \infty$. Here is a precise statement:

Corollary 30. *Suppose that* $T, N \to \infty$ *in such a way that* $\lim_{T,N\to\infty} \frac{T}{N} = \tau$. *Fix a confidence level* $0 < \alpha < 1$, y_0, σ^2, *and* $\beta = 1 + \theta$ *such that* $|\beta| < 1$. *Let* H_1 *be the data generating process* (86)–(87). *Then the* $k = 1$ *cointegration test based on Theorem 9 has asymptotic power against* H_1 *at least* $p(\alpha, \tau)$ *as* $T, N \to \infty$. *Here* $p(\alpha, \tau)$ *is a non-negative function, such that for each* α *we have* $\lim_{\tau\to\infty} p(\alpha, \tau) = 1$.

Our approach to the proof of Corollary 30 gives a lower bound on $p(\alpha, \tau)$. Finding exact formulas for the power under H_1 of this corollary and under other alternatives remains an important open problem for the future research.

Proof of Proposition 29. By definition, $\hat{\lambda}_1$ is the largest sample canonical correlations between matrices $\tilde{R}_0$ and $\tilde{R}_k$. The variational interpretation for $\hat{\lambda}_1$ (see, e.g., Anderson [2003, Chapter 12]) as the maximal sample correlation coefficient between vectors in linear spans of $\tilde{R}_0$ and $\tilde{R}_1$, implies that $\hat{\lambda}_1$ is larger or equal than the correlation coefficient between the first rows of $\tilde{R}_0$ and $\tilde{R}_k$. In the rest of the proof we estimate this correlation coefficient and show it is satisfies the asymptotic inequality (88).

Let us compute these first rows by following the procedure of Section 3.1. From Eq. (87) we obtain

$$y_t = \beta^t y_0 + \sum_{i=1}^{t} \beta^{t-i}\xi_i, \qquad \Delta y_t = (\beta-1)\beta^{t-1}y_0 + (\beta-1)\sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i + \xi_t.$$

Then

$$\tilde{y}_t = y_{t-1} - \frac{t-1}{T}(y_T - y_0) = \beta^{t-1}y_0 + \sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i - \frac{t-1}{T}(y_T - y_0).$$

After regressing on a constant we get residuals (here $\tilde{R}_{0t,1}$ is the first element of the column $\tilde{R}_{0t}$ and similarly for $\tilde{R}_{kt,1}$)

(89)

$$\begin{aligned}\tilde{R}_{0t,1} &= \Delta y_t - \frac{1}{T}\sum_{\tau=1}^{T}\Delta y_\tau \\ &= y_0\left((\beta-1)\beta^{t-1} + \frac{1-\beta^T}{T}\right) + \xi_t - \frac{1}{T}\sum_{i=1}^{T}\xi_i + (\beta-1)\sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i + \frac{1}{T}\sum_{i=1}^{T}(1-\beta^{T-i})\xi_i,\end{aligned}$$

$$\begin{aligned}\tilde{R}_{kt,1} = \tilde{y}_t - \frac{1}{T}\sum_{\tau=1}^{T}\tilde{y}_\tau &= y_0\left(\beta^{t-1} - \frac{1-\beta^T}{T(1-\beta)}\right) + \sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i - \frac{1}{T}\sum_{i=1}^{T}\frac{1-\beta^{T-i}}{1-\beta}\xi_i \\ &\quad - \left(\frac{2t-1}{2T} - \frac{1}{2}\right)(y_T - y_0).\end{aligned} \tag{90}$$

In order to compute the sample correlation coefficient, we analyze three sums representing sample variances and covariance: $\frac{1}{T}\sum_{t=1}^{T}\tilde{R}^2_{0t,1}$, $\frac{1}{T}\sum_{t=1}^{T}\tilde{R}^2_{kt,1}$, $\frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{0t,1}\tilde{R}_{kt,1}$. Let us analyze the

sums sequentially. Summing the geometric series and using the law of large numbers, we get

(91)
$$\frac{y_0^2}{T}\sum_{t=1}^{T}\left((\beta-1)\beta^{t-1}+\frac{1-\beta^T}{T}\right)^2 \xrightarrow[T\to\infty]{} 0, \qquad \frac{1}{T}\sum_{t=1}^{T}\xi_t^2 \xrightarrow[T\to\infty]{P} \sigma^2, \qquad \left(\frac{1}{T}\sum_{i=1}^{T}\xi_i\right)^2 \xrightarrow[T\to\infty]{P} 0,$$
$$\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i\right)^2 = \frac{1}{T}\sum_{i=1}^{T-1}\xi_i^2\frac{1-\beta^{2(T-i)}}{1-\beta^2}+\frac{2}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-2}\sum_{j=i+1}^{t-1}\beta^{2(t-1)-i-j}\xi_i\xi_j \xrightarrow[T\to\infty]{P} \frac{\sigma^2}{1-\beta^2},$$
$$\left(\frac{1}{T}\sum_{i=1}^{T}(1-\beta^{T-i})\xi_i\right)^2 = \frac{1}{T^2}\sum_{i=1}^{T}(1-\beta^{T-i})^2\xi_i^2+\frac{1}{T^2}\sum_{i\neq j}(1-\beta^{T-i})(1-\beta^{T-j})\xi_i\xi_j \xrightarrow[T\to\infty]{P} 0.$$

We also have

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\xi_t\left(\sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i\right)\right]^2 = \frac{1}{T^2}\sum_{t=1}^{T}\sigma^2\sum_{i=1}^{t-1}\beta^{2(t-1-i)}\sigma^2 \xrightarrow[T\to\infty]{} 0, \tag{92}$$

which implies that the expression under expectation tends to 0. Using formulas (91),(92) and Cauchy-Schwarz inequality to show that the remaining averages of cross-products of terms in Eq. (89) converge to 0, we get

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{0t,1}^2 \xrightarrow[T\to\infty]{P} \sigma^2+(\beta-1)^2\frac{\sigma^2}{1-\beta^2}=\frac{2\sigma^2}{1+\beta}. \tag{93}$$

To analyze $\frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{kt,1}^2$, we again sum geometric series and use the law of large numbers:

(94)
$$\frac{y_0^2}{T}\sum_{t=1}^{T}\left(\beta^{t-1}-\frac{1-\beta^T}{T(1-\beta)}\right)^2 \xrightarrow[T\to\infty]{} 0, \qquad \frac{(y_T-y_0)^2}{T}\sum_{t=1}^{T}\left(\frac{2t-1}{2T}-\frac{1}{2}\right)^2 \approx_{T\to\infty} \frac{(y_T-y_0)^2}{12},$$

where the $\approx_{T\to\infty}$ sign means that the ratio of the left-hand side and the right-hand side tends to 1 in probability. It is also straightforward to show that

$$\frac{y_T-y_0}{T}\sum_{t=1}^{T}\left[\left(\frac{2t-1}{2T}-\frac{1}{2}\right)\sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i\right] \xrightarrow[T\to\infty]{P} 0. \tag{95}$$

Using formulas (94), (95), second and third lines of formulas (91), and Cauchy-Schwarz inequality to show that the remaining averages of cross-products of terms in Eq. (90) converge to 0, we get

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{kt,1}^2 \approx_{T\to\infty} \frac{\sigma^2}{1-\beta^2}+\frac{1}{12}\left(y_T-y_0\right)^2. \tag{96}$$

We are left with the analysis of the covariance $\frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{0t,1}\tilde{R}_{kt,1}$, which relies on similar computations as for the two variances. The only asymptotically non-vanishing term is given by

the computation of the second line in (91): we multiply $(\beta-1)\sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i$ from (89) by $\sum_{i=1}^{t-1}\beta^{t-1-i}\xi_i$ and sum over t. Thus,

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{0t,1}\tilde{R}_{kt,1}\xrightarrow[T\to\infty]{}(\beta-1)\frac{\sigma^2}{1-\beta^2}=-\frac{\sigma^2}{1+\beta}. \tag{97}$$

Combining (93), (96), (97) together we get

$$\left(\widehat{corr}(\tilde{R}_{0,1},\tilde{R}_{k,1})\right)^2=\frac{\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{0t,1}\tilde{R}_{kt,1}\right)^2}{\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{R}_{0t,1}^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}R_{kt,1}^2\right)}\approx_{T\to\infty}\frac{\frac{\sigma^4}{(1+\beta)^2}}{\frac{2\sigma^2}{1+\beta}\left[\frac{\sigma^2}{1-\beta^2}+\frac{1}{12}\left(y_T-y_0\right)^2\right]} \tag{98}$$

$$=\frac{1}{\frac{2}{1-\beta}+\frac{1+\beta}{6\sigma^2}\left(y_T-y_0\right)^2}. \qquad \square$$

Proof of Corollary 30. Cointegration test based on Theorem 9 has the form: reject H_0, if

$$\frac{\sum_{i=1}^{r}\ln(1-\tilde{\lambda}_i)-r\cdot c_1(N,T)}{N^{-2/3}c_2(N,T)}\geq\kappa, \tag{99}$$

where κ is a constant depending on r and the confidence level α (κ is found from the equation $\mathrm{Prob}(\sum_{i=1}^{r}\mathfrak{a}_i\leq\kappa)=\alpha$). In order to prove Corollary 30, we need to find the probability of the event (99) under H_1 given by (86)–(87), and show that this probability tends to 1 in the double limit in which we first send $T,N\to\infty$ with $\lim\frac{T}{N}=\tau$ and then send τ to infinity.

Recall that $c_2(N,T)$ is negative, as given in (26). Hence, using deterministic inequalities $\ln(1-\tilde{\lambda}_i)\leq 0$ and $\ln(1-\tilde{\lambda}_1)\leq-\tilde{\lambda}_1$, we conclude that the probability of the event (99) is larger than the probability of a simpler event

$$\tilde{\lambda}_1\geq -r\cdot c_1(N,T)-\kappa N^{-2/3}c_2(N,T). \tag{100}$$

Note that both terms in the right-hand side of (100) are positive and the second one vanishes as $N,T\to\infty$. As for the first one, using (26), we see that it converges to a positive constant as $N,T\to\infty$ with $\lim\frac{T}{N}=\tau$, and this constant further tends to 0 as $\tau\to\infty$. The conclusion is that the right-hand side of (100) tends to 0 in our double limit.

On the other hand, under H_1 by Proposition 29, for any $\epsilon>0$, with probability tending to 1 as $T\to\infty$, we have

$$\tilde{\lambda}_1>\frac{1}{\frac{2}{1-\beta}+\frac{1+\beta}{6\sigma^2}\left(y_T-y_0\right)^2}-\epsilon. \tag{101}$$

Note that y_0 is assumed to be bounded. Simultenously, we assumed $|\beta| < 1$, and therefore, y_T, which due to (87) can be expressed as

$$y_T = \beta^T y_0 + \sum_{t=1}^{T} \beta^{T-t} \xi_t,$$

has uniformly bounded second moment. Therefore, the denominator in (101) does not explode. Hence, (101) implies that (100) holds with probability tending to 1 in the double limit. □

References

G. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Cambridge university press, 2010.

T. W. Anderson. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, 22(3):327–351, 1951.

T. W. Anderson. *Introduction to multivariate statistical analysis, 3rd edition*. John Wiley & Sons, 2003.

J. Bai and S. Ng. Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3(2):89–163, 2008.

J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

Z. Bao, J. Hu, G. Pan, and W. Zhou. Canonical correlation coefficients of high-dimensional gaussian vectors: Finite rank case. *Annals of Statistics*, 47(1):612–640, 2019.

J. Breitung and M. H. Pesaran. Unit roots and cointegration in panels. In *Mátyá L., Sevestre P. (eds) The Econometrics of Panel Data. Advanced Studies in Theoretical and Applied Econometrics, vol. 46*, pages 279–322. Springer, Berlin, Heidelberg, 2008.

A. Bykhovskaya and V. Gorin. Cointegration in large vars. *Annals of Statistics*, 2022.

A. Bykhovskaya and V. Gorin. High-dimensional canonical correlation analysis. *arXiv preprint arXiv:2306.16393*, 2023.

A. Bykhovskaya, V. Gorin, and E. Kiss. Largevars: an R package for testing large VARs for the presence of cointegration. 2023. `https://github.com/eszter-kiss/Largevars`.

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

I. Choi. Panel cointegration. In *Baltagi B.H. (eds) The Oxford handbook of panel data*. Oxford University Press, 2015.

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

L. Erdos and H. T. Yau. Universality of local spectral statistics of random matrices. *Bulletin of the American Mathematical Society*, 49(3):377—-414, 2012.

P. J. Forrester. The spectrum edge of random matrix ensembles. *Nuclear Physics B*, 402(3): 709–728, 1993.

P. J. Forrester. *Log-gases and random matrices*. Princeton University Press, 2010.

J. Gonzalo and J. Y. Pitarakis. Comovements in large systems. *Statistics and Econometrics Series, Vol. 10. Working Paper 95-38, Universidad Carlos III de Madrid*, 1995.

J. Gonzalo and J. Y. Pitarakis. Lag length estimation in large dimensional systems. *Journal of Time Series Analysis*, 23(4):401–423, 2002.

C. Han, G. Pan, and Q. Yang. A unified matrix model including both cca and f matrices in multivariate analysis: The largest eigenvalue and its applications. *Bernoulli*, 24(4B): 3447–3468, 2018.

X. Han, G. M. Pan, and B. Zhang. The tracy-widom law for the largest eigenvalue of f type matrix. *The Annals of Statistics*, 44(4):1564–1592, 2016.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2013.

L. Hua. *Harmonic analysis of functions of several complex variables in the classical domains*, volume 6. American Mathematical Soc., 1963.

S. Johansen. Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control*, 12(2–3):231–254, 1988.

S. Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59:1551–1580, 1991.

S. Johansen. *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press, 1995.

S. Johansen. A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrics*, 70(5):1929–1961, 2002.

I. Johnstone. Multivariate analysis and jacobi ensembles: largest eigenvalue, tracy-widom limits and rates of convergence. *Annals of statistics*, 36(6):2638–2716, 2008.

K. Juselius. *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, 2006.

G. Keilbar and Y. Zhang. On cointegration and cryptocurrency dynamics. *Digital Finance*, 3(1):1–23, 2021. `https://github.com/QuantLet/CryptoDynamics/blob/master/CryptoDynamics_Series/price.csv`.

G. S. Maddala and I.-M. Kim. *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press, 1998.

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

R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, 2009.

Y. Neretin. Hua-type integrals over unitary groups and over projective limits of unitary groups. *Duke Mathematical Journal*, 114(2):239–266, 2002.

G. Olshanski. The problem of harmonic analysis on the infinite-dimensional unitary group. *Journal of Functional Analysis*, 205(2):464–524, 2003.

A. Onatski and C. Wang. Alternative asymptotics for cointegration tests in large vars. *Econometrica*, 86(4):1465–1478, 2018.

A. Onatski and C. Wang. Extreme canonical correlations and high-dimensional cointegration analysis. *Journal of Econometrics*, 2019.

A. Pagan. Three econometric methodologies: A critical appraisal. *Journal of Economic surveys*, 1(1):3–23, 1987.

G. C. Reinsel 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*, 13(4):353–375, 1992.

C. A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980.

A. R. Swensen. Bootstrap algorithms for testing and determining the cointegration rank in var models. *Econometrica*, 74(6):1699–1714, 2006.

T. Tao and V. Vu. Random matrices: the universality phenomenon for wigner ensembles. *Modern aspects of random matrix theory*, 72:121—-172, 2012.

C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754, 1996.

M. Udell and A. Townsend. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, 2019.

D. Wang and R. S. Tsay. Rate-optimal robust estimation of high-dimensional vector autoregressive models. *arXiv preprint arXiv:2107.11002*, 2022.

D. Wang, Y. Zheng, H. Lian, and G. Li. High-dimensional vector autoregressive time series modeling via tensor decomposition. *Journal of the American Statistical Association*, 117 (539):1338–1356, 2022.

F. Yang. Sample canonical correlation coefficients of high-dimensional random vectors: Local law and tracy–widom limit. *Random Matrices: Theory and Applications*, 11(1):2250007, 2022a.

F. Yang. Limiting distribution of the sample canonical correlation coefficients of high-dimensional random vectors. *Electronic Journal of Probability*, 27:1–71, 2022b.

B. Zhang, G. M. Pan, and J. T. Gao. Clt for largest eigenvalues and unit root testing for high-dimensional nonstationary time series. *The Annals of Statistics*, 46(5):2186–2215, 2018.

(Anna Bykhovskaya) Duke University
Email address: anna.bykhovskaya@duke.edu

(Vadim Gorin) University of California at Berkeley
Email address: vadicgor@gmail.com