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

ANNA BYKHOVSKAYA AND VADIM GORIN

ABSTRACT. The paper studies non-stationary 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 such regime the first-order asymptotics of the Johansen likelihood ratio (LR), Pillai-Barlett, and Hotelling-Lawley tests for cointegration is derived: Test statistics converge to non-random 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 our theorems.

The paper presents an empirical implementation of the approach to the analysis of stocks in S&P100 and of cryptocurrencies. The latter example has strong presence of multiple cointegrating relationships, while the former is consistent with the null of no cointegration.

---

1. Introduction

Starting from the pioneering work Sims (1980), vector autoregressions (VARs) became a workhorse model in macroeconomics and other fields. A lot of data in macroeconomics and finance is non-stationary (e.g., consumption and output), while properties of VARs can be very different depending on whether one deals with stationary or non-stationary time series. There is a further subdivision in the non-stationary case: It is important to understand if the data is cointegrated or not. That is, whether there exists a stationary non-trivial linear combination of considered series (e.g., log of consumption minus log of output is stationary, while the series per se are unit roots).

Classical tools for testing cointegration (see, e.g., Johansen (1995), Maddala and Kim (1998), Juselius (2006)) failed to achieve desired finite sample performance when the number of time series, N, is large. Thus, they are not commonly used in such settings and designing proper tools to handle the cointegration for 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 "T/N converging to a constant" asymptotic regime. Yet, the testing procedures of these texts covered only VAR(1), while in practice researchers rarely stick to VARs of order 1, instead usually at least two lags are considered. Indeed, as noted already in Pagan (1987), "most applications of Sims' methodology have put number of lags between four and ten." Since Pagan (1987) the lengths of available time-series and computing power only increased, thus, allowing 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 non-stationary 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 Johanssen'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), 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, so that 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 non-trivial 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 were previously used in the asymptotic representation theory, to the authors' knowledge, this is their first appearance in 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 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 7–10.

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 non-degenerate covariance matrix Λ. That is, written in the error correction form,

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

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

---

[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), or Han et al. (2018), Yang (2022).

are observed), and $\Gamma_1, \ldots, \Gamma_{k-1}$, Π, Φ are unknown parameters. The process is initialized at fixed $X_{1-k}, \ldots, X_0$. We do not impose any restrictions on Λ, thus, we allow for arbitrary correlations across coordinates of X_t. In contrast, many previous approaches relied 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 error correction form can be written as*

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

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 residuals from regressing X_{t-k} on $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}$, which are the same as residuals from regressing $X_{t-1} = X_{t-k} + \Delta X_{t-1} + \ldots + \Delta X_{t-k+1}$ on $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}$.

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), they are related to whether the process is cointegrated or not. 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 $*$ means matrix transposition.

Procedure 1 (Johansen (1991))**.** Let $Z_{0t} = \Delta X_t$, $Z_{1t} = (\Delta X_{t-1}^*, \ldots, \Delta X_{t-k+1}^*, D_t^*)^*$, 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 residuals

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

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

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

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}, $i = 0, k$.

The Johansen's likelihood ratio (LR) statistic for testing the hypothesis $\text{rank}(\Pi) \leq r_1$ (at most r_1 cointegrating relationships) versus the alternative $\text{rank}(\Pi) \in (r_1, r_2]$ (between r_1 and r_2 cointegrating relationships) with $r_2 > r_1$ has the form

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

the Pillai-Barlett statistic is

$$\sum_{i=r_1+1}^{r_2} \lambda_i,$$

and Hotelling-Lawley statistic is

$$\sum_{i=r_1+1}^{r_2} \frac{\lambda_i}{1 - \lambda_i},$$

see Gonzalo and Pitarakis (1995) for 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 results of Onatski and Wang (2018) from VAR(1) to VAR(k) setting[2].

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

$$(4) \qquad \mu_{\mathfrak{p},\mathfrak{q}}(x) = \frac{\mathfrak{p} + \mathfrak{q}}{2\pi} \cdot \frac{\sqrt{(x - \lambda_-)(\lambda_+ - x)}}{x(1 - x)} \mathbf{1}_{[\lambda_-,\lambda_+]},$$

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

$$(5) \qquad \lambda_\pm = \frac{1}{(\mathfrak{p} + \mathfrak{q})^2} \left(\sqrt{\mathfrak{p}(\mathfrak{p} + \mathfrak{q} - 1)} \pm \sqrt{\mathfrak{q}} \right)^2.$$

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

$$(6) \qquad \lim_{N\to\infty} \frac{T}{N} = \tau > (k + 1) \qquad and$$

$$(7) \qquad \lim_{N\to\infty} \frac{1}{N} \big(\mathrm{rank}(\Pi) + \mathrm{rank}(\Gamma_1) + \mathrm{rank}(\Gamma_2) + \cdots + \mathrm{rank}(\Gamma_{k-1}) + d_D \big) = 0.$$

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

$$(8) \qquad \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(\lambda_i) = \int_0^1 f(x)\mu_{2,\tau-k}(x)\,\mathrm{d}x, \qquad in\ probability.$$

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 (7) 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 sciences: in machine learning, recommender systems (e.g., movie preference recognition), computational mathematics, etc. We refer to Udell and Townsend (2019) for theoretical explanations of the suitability of low rank models, as well as many references to the situations

---

[2]While Onatski and Wang (2018, Theorem 1) allows the data to be VAR(k) under the restriction (7), they construct the matrix $\mathcal{C}$ involved in the statistical testing procedures as if the data was VAR(1). That is, the formal procedure is based on the misspecified VAR(1) setting, while we use the true VAR(k) procedure. As illustrated in Section 4.3, using underspecified VAR can lead to severe size distortions.

where 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 (6), the model (1) can overfit the data. We view the rank restriction (7) as a natural way to avoid overfitting[3] (cf. discussion in Wang et al. (2022) and Wang and Tsay (2022)). Section 5 illustrates that the results obtained under this assumption are consistent with behavior of large-dimensional financial datasets. Another setting when we can expect to satisfy (7) is when there are few special coordinates in X_t, e.g., some macroeconomic indicators, which mostly drive the behavior of the entire vector X_t. That would correspond to the case, where the columns of Γ_i corresponding to those indicators are non-zero while the other columns are zero.

Figure 1 illustrates Theorem 3 for independent standard normal errors and $k = 2$, $N = 150$, $T = 1500$. Parameters are $\Phi D_t = 1_N$, $\Gamma_1 = 0.95 E_{12}$, $\Pi = -0.1 E_{\cdot 1}$, where 1_N is $N \times 1$ 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 the first column of ones and zeros everywhere else. Thus, all 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 $\text{rank}(\Pi) = 1$, i.e. one cointegrating relationship.

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 (7) is crucial for the validity of the theorem and we expect that the asymptotic behavior changes in situations when (7) 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 pursue this direction.

An important corollary of Theorem 3 is that it provides 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*

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

---

[3]An alternative way to introduce a low-rank assumption into the VAR model is to, instead of using the error correction form (1), 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 if close to N. While non of our theorems directly cover the factor setting, numeric simulations in Section 4.5 indicate that the test that we develop remain useful.
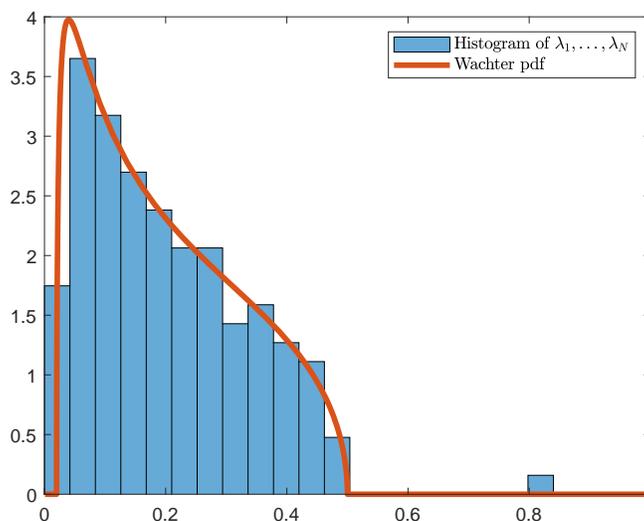
FIGURE 1. Illustration of Theorem 3: Eigenvalues and Wachter distribution. Data generating process: $\Delta X_t = 1_N + 0.95 E_{12} \Delta X_{t-1} - 0.1 E_{\cdot 1} X_{t-2} + \varepsilon_t$, $T = 1500$, $N = 150$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$.

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

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

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

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

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

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

Note that when $\rho_1 = 0$, for the Johansen LR and Hotelling-Lawley statistics we get inequalities, rather than equalities. This is because of the singularity of $\ln(1 - \lambda)$ and $\frac{\lambda}{1-\lambda}$ at $\lambda = 1$. 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., $H_1 : \mathrm{rank}(\Pi) \leq N$), 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 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, $1 \leq i \leq N$. In Section 5 we perform such
  validation on S&P100 and cryptocurrencies data sets for VAR(k) with $1 \leq k \leq 4$ 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.

In order to draw further economical and statistical conclusions and to develop precise
statistical tests and their critical values, one needs to know the corrections to the first order
asymptotic results of Theorem 3 and Corollary 4. In the next section we develop such
corrections and introduce relevant modifications.

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 is 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 asymtotic results.

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

$$(9) \qquad \Delta X_t = \mu + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-k} + \varepsilon_t, \qquad t = 1, \dots, T,$$

The null hypothesis of no cointegration is $H_0 : \text{rank}(\Pi) = 0$ or $\Pi \equiv 0$. Yet, this is not a
point hypothesis, as it does not specify Γ_i, $i = 1, \dots, k-1$. We are going to work with a
"proxy" point null approximation 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).[4] Let us introduce our "proxy" null:

$$\text{(10)} \qquad \widehat{H}_0 : \Pi = \Gamma_1 = \Gamma_2 = \cdots = \Gamma_{k-1} = 0.$$

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

$$\text{(11)} \qquad \Delta X_t = \mu + \varepsilon_t, \qquad t = 1, \ldots, T,$$

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

The complement to H_0 is $\text{rank}(\Pi) > 0$. Yet, in order to design our test we use an alternative hypothesis

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

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

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

where $\lambda_1, \lambda_2, \ldots$ are defined in Procedure 1.

Let us describe how our modified test proceeds.

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

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

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

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

Define

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

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

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

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

---

[4]In Section 4 we present Monte Carlo simulations which indicate that (11) can be replaced by less restrictive (and, thus, closer to H_0) hypothesis, for which asymptotic results of Section 3.2 remain valid.

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

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

$$(16) \qquad \tilde{\mathcal{C}} = \tilde{S}_{10} \tilde{S}_{00}^{-1} \tilde{S}_{01} \tilde{S}_{11}^{-1}.$$

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

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

Step 5. Form the test statistic

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

The subscript N, T in (18) indicates that we modify Johansen LR test to develop the large N, T asymptotics. This statistic after centering and rescaling will be compared with appropriate critical values to decide whether one can reject $\widehat{H}_0$ or not (see Theorem 9). Visually rejections correspond to the case when 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. I.e., 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

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

3.2. Second order asymptotics. In this section we show that 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 formally stating 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

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

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 get 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

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

which can be defined through the following proposition.

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

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

The 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.*

Theorem 8. *Fix $C > 0$ and suppose that $T, N \to \infty$ in such a way that $\frac{T}{N} \in [k+1+C^{-1}, C]$. Under the hypothesis $\widehat{H}_0$ given by (10) for (1), 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*

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

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 8. Second, in Section 9 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, 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 $\frac{T}{N} \in [k+1+C^{-1}, C]$. Under the proxy $\widehat{H}_0$ given by (10) for (1), for each finite $r = 1, 2, \ldots$ we have convergence in distribution for the largest eigenvalues defined in Eq. (17):*

$$(22) \qquad \frac{\sum_{i=1}^{r} \ln(1 - \tilde{\lambda}_i) - r \cdot c_1(N, T)}{N^{-2/3} c_2(N, T)} \xrightarrow[T, N \to \infty]{d} \sum_{i=1}^{r} \mathfrak{a}_i,$$

where

$$(23) \quad c_1(N, T) = \ln(1 - \lambda_+), \qquad c_2(N, T) = -\frac{2^{2/3} \lambda_+^{2/3}}{(1 - \lambda_+)^{1/3}(\lambda_+ - \lambda_-)^{1/3}} (\mathfrak{p} + \mathfrak{q})^{-2/3} < 0,$$

$$(24) \qquad \mathfrak{p} = 2, \qquad \mathfrak{q} = \frac{T}{N} - k, \qquad \lambda_{\pm} = \frac{1}{(\mathfrak{p} + \mathfrak{q})^2} \left[\sqrt{\mathfrak{p}(\mathfrak{p} + \mathfrak{q} - 1)} \pm \sqrt{\mathfrak{q}} \right]^2.$$

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 (6).*

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 the presence of cointegration. Note that although the asymptotic result (22) is shown under the proxy $\widehat{H}_0$, the Monte-Carlo simulations in Section 4 suggest that it extends well beyond $\widehat{H}_0$. We believe that the same asymptotic results and testing procedures continue to hold in many situations with non-zero Γ_i in Eq. (1). To be more precise, we expect (see Section 4.2 for relevant simulations) that the conclusion of Theorem 9 is still true in the setting (1) with $\Pi = 0$ subject to certain technical conditions in the spirit of small rank restriction (7). The intuition is that addition of generic small rank matrices is negligible when compared to the rest of the process and, thus, does not change the asymptotics. In order 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 relevant ΔX_{t-i} in the procedure), the presence of Γ_i can have a similar effect as the presence of non-zero 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 simulations in Section 4.3.

Formally, to perform testing one first needs to calculate the statistic $LR_{N,T}(r)$ following the Procedure 2. We recommend using small[5] 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 $(1 - \alpha)$ level. We report the quantiles for $r = 1$ in Table 1.

---

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

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)).

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. (4). 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 other eigenvalues. The separation is due to small rank of Π. However, even if the rank of Π is large, we expect the largest eigenvalue to be significantly larger than λ_+, thus, the test remains relevant, see Section 4.5. Rigorous results on the consistency of the test form an important task for future research.

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 $k = 1$. Results for VAR(k) are similar and we do not show them in much detail here. For illustration purpose and to see 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.[6] We can see that the numbers are close to the desired 5%, and for the same N and T 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 procedure for $k > 1$ is the introduction of the proxy null $\widehat{H}_0$. The aim of our next simulation is to check whether Theorem 9 can hold

---

[6]Depending on assumed order of autoregression, we have different number of regressors in Procedure 2.

(A) $\Gamma_1 = 0.95 E_{11}$.   (B) $\Gamma_1 = 0.1 E_{\cdot 1} + 0.95 E_{12}$.
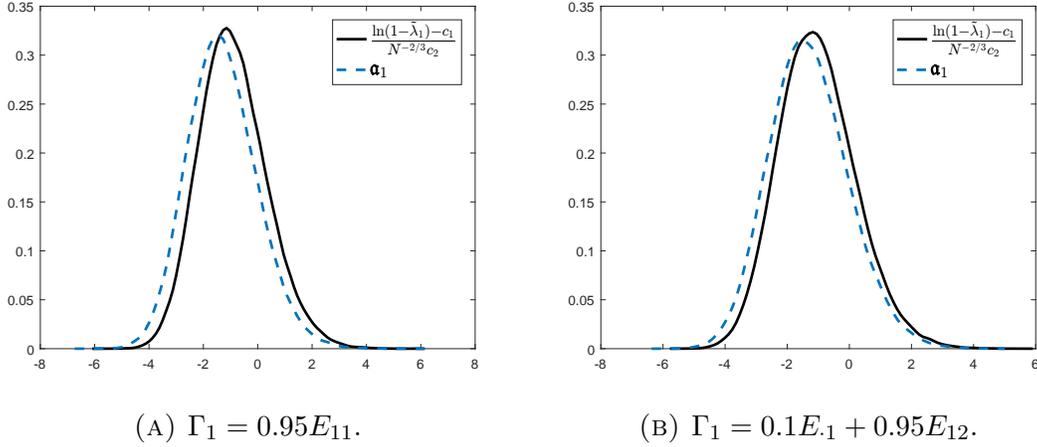
FIGURE 2. Airy$_1$ and asymptotic distribution of the rescaled $\ln(1 - \tilde\lambda_1)$ under H_0. Data generating process: $\Delta X_t = \Gamma_1 \Delta X_{t-1} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 100,000$ replications.
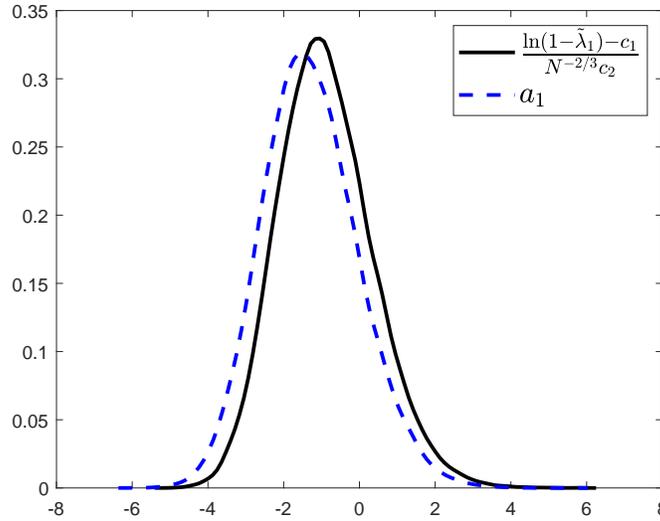

FIGURE 3. Airy$_1$ and asymptotic distribution of the rescaled $\ln(1 - \tilde\lambda_1)$ under H_0. Data generating process: $\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + \varepsilon_t$, $\Gamma_1 = E_{11}$, $\Gamma_2 = -\frac{2}{9} E_{11}$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 100,000$ replications.

under less restrictive H_0 instead of $\widehat{H}_0$. For N=100, T=500 we simulate the data based on i.i.d. $\mathcal{N}(0,1)$ errors ε_{it}, zero Π, and non-zero 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 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, while in the second one we consider an asymmetric matrix $\Gamma_1 = 0.1E_{\cdot 1} + 0.95E_{12}$. Thus, the first case corresponds to Γ_1 of rank 1 and the second to rank 2. 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 result is shown in Figure 3. We interpret the outcomes of these three experiments as a strong argument towards the validity of an analogue of Theorem 9 well-beyond the $\widehat{H}_0$ setting.[7]

4.3. Order of VAR. It is essential for the experiments in the last paragraph that the data generating process is VAR(2) and 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 larger k would lead to similar results, while using incorrect VAR(1) procedure, when the data generating process is VAR(2), implies wrong centering and scaling. Moreover, one can spot in Figure 5 that underestimating order of VAR can be misinterpreted as a presence of cointegration (largest eigenvalue separated from the rest leading to the large value of the test statistic). Yet as we increase the order, the largest eigenvalue becomes inseparable from the rest, and no sign of false cointegration remains present. Thus, practitioners are encouraged to experiment with the order of VAR to make sure they are detecting cointegration and not a wrong model.[8]

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

4.4. Small ranks. To illustrate the importance of small ranks (e.g., (7) 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 6. While the shape and the scale (corresponding to $N^{2/3}$ rescaling in Theorem 9) of the distribution remains similar, the location changes. Thus, the small rank restriction of Eq. (7) is important not only in the context of Theorem 3, but also for the correct centering in possible generalizations of Theorem 9.

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

---

[7]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. (22)) predicts errors of at least const $\cdot N^{-1/3}$ in the approximations under $\widehat{H}_0$.

[8]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).
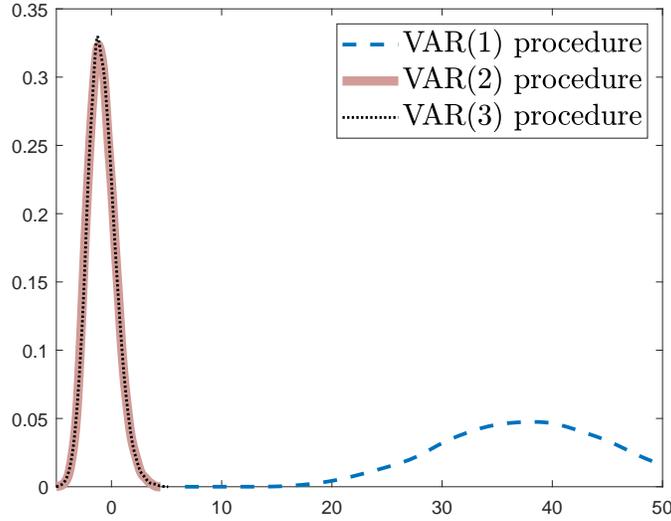
FIGURE 4. Density of rescaled $\ln(1 - \tilde{\lambda}_1)$ obtained from various procedures. The correct procedure corresponds to VAR(2), $k = 2$. Data generating process: $\Delta X_t = 0.95 E_{11} \Delta X_{t-1} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 10,000$ replications.

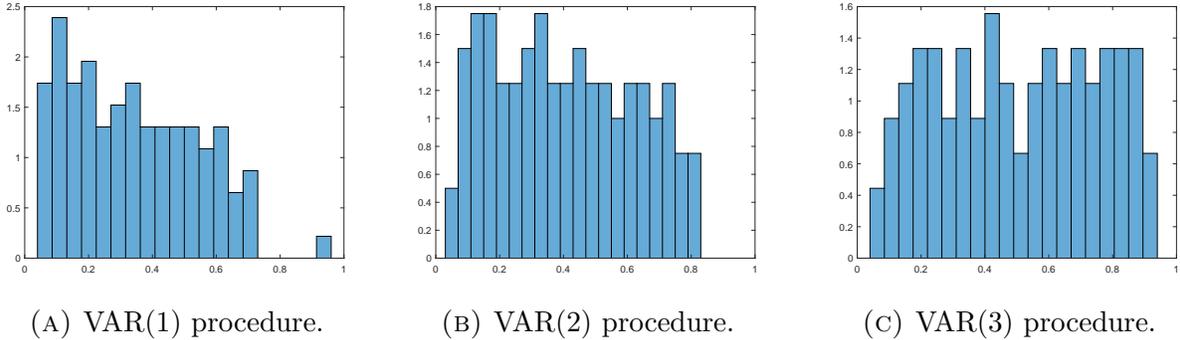(A) VAR(1) procedure.   (B) VAR(2) procedure.   (C) VAR(3) procedure.

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. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$.

For the first experiment, we use $k = 2$, $\Gamma_1 = 0$, and $\Pi = -0.95 E_{11}$. Figure 7 shows the results of the simulation. Two curves are separated, moreover, the black straight line (test 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, distinct variances are due to the fact that (as we expect from 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 $N^{1/2}$.
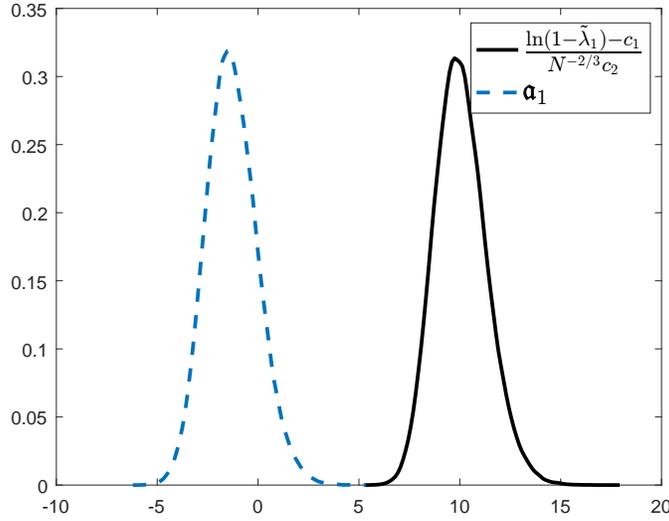
FIGURE 6. Airy$_1$ and asymptotic distribution of the rescaled $\ln(1 - \tilde{\lambda}_1)$ under H_0 with Γ_1 of full rank. Data generating process: $\Delta X_{it} = 0.95 \Delta X_{it-1} + \varepsilon_{it}$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 100,000$ replications.

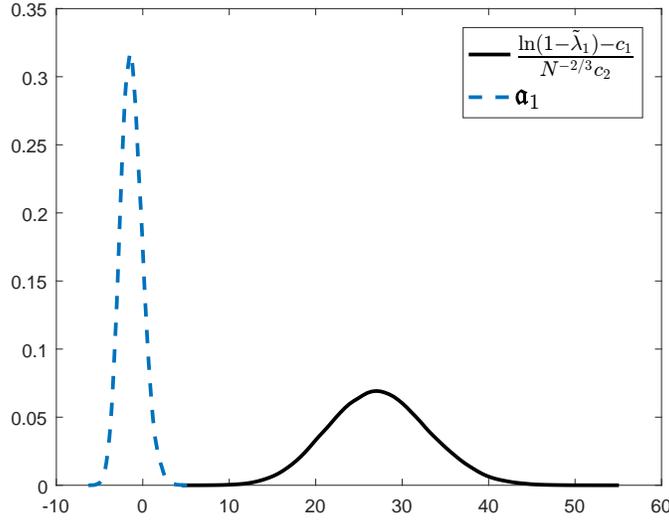

FIGURE 7. Airy$_1$ and asymptotic distribution of the rescaled $\ln(1 - \tilde{\lambda}_1)$ under H_1. Data generating process: $\Delta X_t = -0.95 E_{11} X_{t-2} + \varepsilon_t$, $\varepsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 500$, $N = 100$, $MC = 100,000$ replications.

For the second experiment, we use $k = 2$, $N = 150$, $T = 1500$, $\Gamma_1 = 0.95 E_{12}$, and $\Pi = -0.8 I_\rho X_{t-2}$, where I_ρ is a matrix with ones on the first ρ diagonal elements and zeros elsewhere. The histograms of eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N$ for $\rho = 5, 50, 150$ are shown in Figure 8. As ρ becomes large, we are no longer in the framework of Theorem 3 and the result
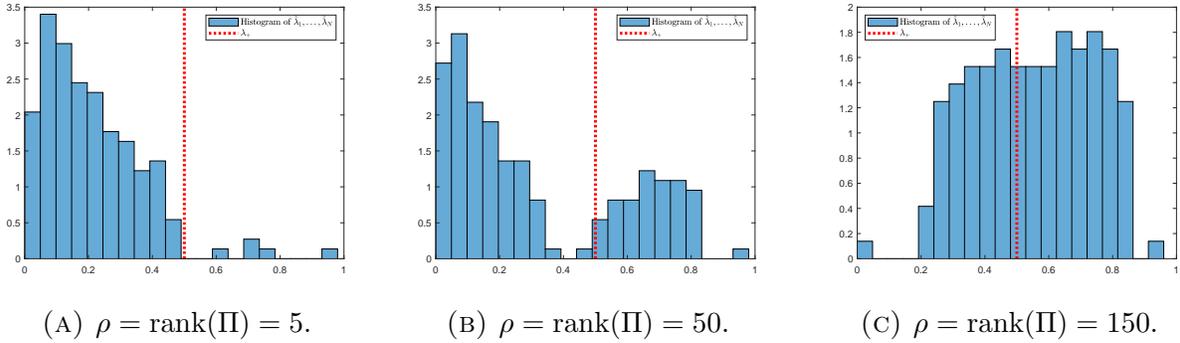
(A) $\rho = \operatorname{rank}(\Pi) = 5$.   (B) $\rho = \operatorname{rank}(\Pi) = 50$.   (C) $\rho = \operatorname{rank}(\Pi) = 150$.

FIGURE 8. Separation of eigenvalues from λ_+ under various ranks of Π. Data generating process: $\Delta X_t = 1_N + 0.95 E_{12}\Delta X_{t-1} - 0.8 I_\rho X_{t-2} + \varepsilon_t$, $\varepsilon_{it} \sim$ 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.

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[9] 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., $\Pi = 0$ hypothesis) on 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 S&P100 over ten years: $01.01.2010 - 01.01.2020$, which gives us 522 observations across time. The more detailed description of the variables can be found in Bykhovskaya and Gorin (2022, Section 6).

For the S&P100 data set we use the Procedure 2 to calculate $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N$. 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 9.

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 modelling for the S&P data. We do not see any outliers in the largest eigenvalues, which would appear if there was 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 data of Keilbar and Zhang (2021) (25 series from github

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

(A) VAR(1).

(B) VAR(2).
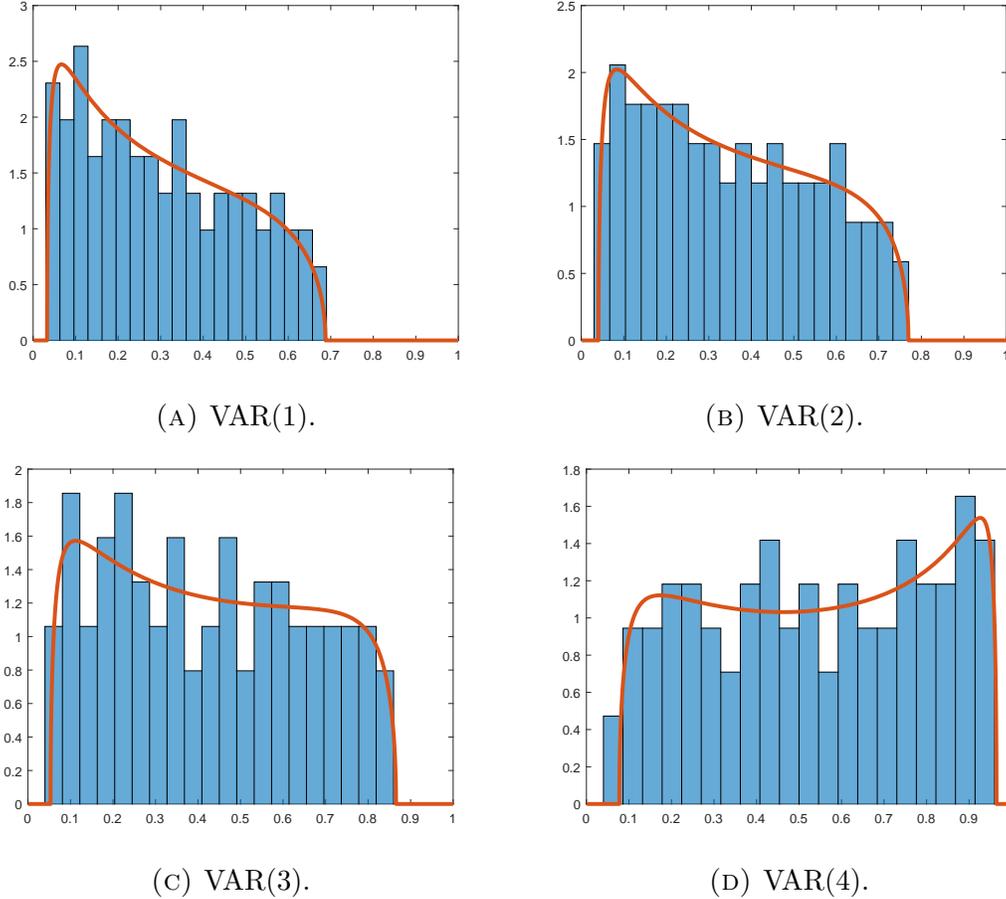
(C) VAR(3).

(D) VAR(4).

FIGURE 9. Eigenvalues from S&P data (blue histogram) and Wachter distribution (orange line) based on various VAR(k) settings.

repository). Logarithms of daily prices for two years (from Oct 5, 2017 to Oct 4, 2019) are shown in Figure 10. The results of our testing Procedure 2 to calculate $\tilde{\lambda}_1, \dots, \tilde{\lambda}_N$ are shown in Figure 11.

Similarly to the S&P example in the previous subsection, we see a match between the eigenvalues and the Wachter distribution.[10] Yet, there is a major difference between Figures 9 and 11: 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-value below 0.01 for all four choices of 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 cryptocurrency market is still very inefficient and, thus, has numerous trading possibilities. The presence of cointegration can be one such inefficiency.

―――――――――

[10]The Wachter distribution depends on the order of VAR, k, and on the ratio T/N. Thus, orange curves in Figures 9 and 11 have different shapes and supports.

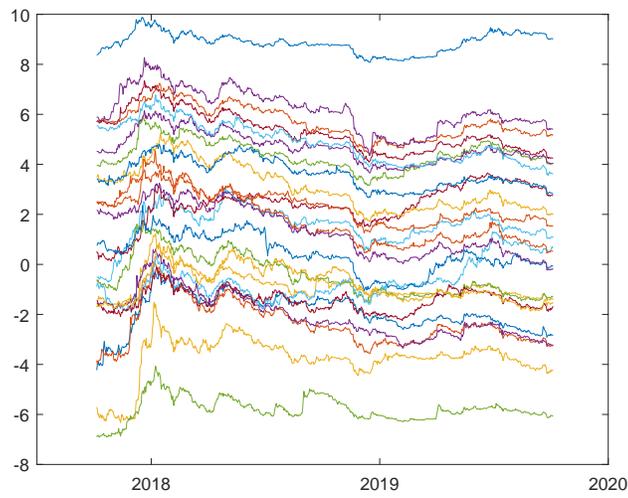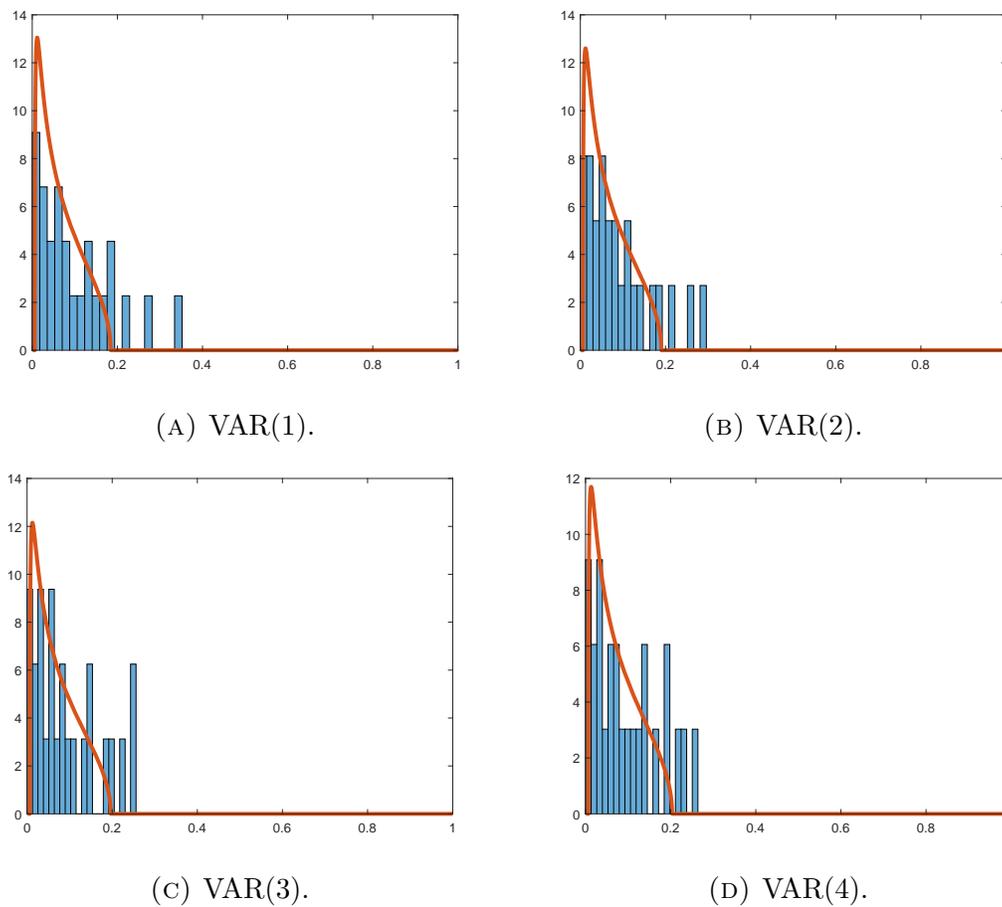FIGURE 10. Time series of daily log prices for 25 cryptocurrencies.

(A) VAR(1).

(B) VAR(2).

(C) VAR(3).

(D) VAR(4).

FIGURE 11. Eigenvalues from cryptocurrencies data (blue histogram) and Wachter distribution (orange line) based on various VAR(k) settings.

6. Conclusion

High-dimensional data becomes more and more widespread in economics and other sciences. Thus, the appropriate machinery is needed. We believe that the use of the random matrix theory is inevitable for the development of the area: as soon as dimensions are high, random matrices kick in. Along these lines, in our paper the central role is played by random matrix objects: Wachter distribution, Airy$_1$ point process, Jacobi ensemble.

The present paper focused on non-stationary high-dimensional VARs and presented the asymptotic limit of the Johansen LR test for cointegration as well as its modifications. Because the limit is non-random, the appropriate second order statistic was derived and the 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 would be to be able to test whether the cointegration rank is r for $r > 0$, i.e. find the true rank of cointegration. Heuristics for finding the correct value of r can be already seen in our simulations and data sets, cf. Figures 1, 9, and 11: when there are no cointegrations, all eigenvalues (squared canonical correlations) are to the left from 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. Figuring out the exact conditions under which this heuristics is correct forms an important problem for the future research.

7. Appendix. Proofs 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

$$(25) \qquad \frac{p}{N} = \frac{1}{2}(\mathfrak{p} - 1), \quad \mathfrak{p} \geq 1, \qquad \frac{q}{N} = \frac{1}{2}(\mathfrak{q} - 1), \quad \mathfrak{q} \geq 1,$$

where $\mathfrak{p}$ and $\mathfrak{q}$ are two parameters, which stay bounded away from 1 and from ∞ as $N \to \infty$.[11] We further define the *equilibrium measure* $\mu_{\mathfrak{p},\mathfrak{q}}$ of the Jacobi ensemble through:

$$(26) \qquad \mu_{\mathfrak{p},\mathfrak{q}}(x)\, dx = \frac{\mathfrak{p} + \mathfrak{q}}{2\pi} \cdot \frac{\sqrt{(x - \lambda_-)(\lambda_+ - x)}}{x(1 - x)} \mathbf{1}_{[\lambda_-, \lambda_+]}\, dx,$$

---

[11]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.

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

$$(27) \qquad \lambda_\pm = \frac{1}{(\mathfrak{p} + \mathfrak{q})^2} \left(\sqrt{\mathfrak{p}(\mathfrak{p} + \mathfrak{q} - 1)} \pm \sqrt{\mathfrak{q}} \right)^2.$$

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

$$(28) \qquad c_\pm = \frac{(\mathfrak{p} + \mathfrak{q})}{2} \frac{\sqrt{\lambda_+ - \lambda_-}}{\lambda_\pm(1 - \lambda_\pm)},$$

and note that

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

where the normalization $\frac{1}{\pi}\sqrt{|x - \lambda_\pm|}$ was chosen to match the behavior of the Wigner semi-circle 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 (25) 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) $\displaystyle \lim_{N \to \infty} \left| \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} - \mu_{\mathfrak{p},\mathfrak{q}} \right| = 0$, *weakly in probability.*

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

$$(29) \qquad \lim_{N \to \infty} \left| \frac{1}{N} \sum_{i=1}^{N} f(x_i) - \int_0^1 f(x) \mu_{\mathfrak{p},\mathfrak{q}}(x) \mathrm{d}x \right| = 0.$$

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

$$(30) \qquad \lim_{N \to \infty} \left\{ N^{2/3} c_+^{2/3} (x_i - \lambda_+) \right\}_{i=1}^{\infty} \to \{\mathfrak{a}_i\}_{i=1}^{\infty}.$$

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

(3) *We also have convergence in distribution for the smallest eigenvalues[12]*

$$(31) \qquad \lim_{N \to \infty} \left\{ N^{2/3} c_-^{2/3} (\lambda_- - x_{N+1-i}) \right\}_{i=1}^{\infty} \to \{\mathfrak{a}_i\}_{i=1}^{\infty}.$$

8. Appendix. Proofs 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 ($\beta = 2$).

---

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

Theorem 12. *Fix $k = 1, 2, \ldots$ 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 $O\mathcal{V}$, $O^2\mathcal{V}, \ldots, O^{k-1}\mathcal{V}$. 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 $P_1 P_2 P_1$ 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*

$$(32) \qquad \det(\mathcal{M})^{\frac{\beta}{2}N + \beta - 2} \det(I_N - \mathcal{M})^{\frac{\beta}{2}(\mathcal{T} - (k+1)N + 1) - 1} \, d\mathcal{M}, \quad 0 \leq \mathcal{M} \leq I_N, \qquad \beta = 1, 2.$$

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 $O(I_\mathcal{T} + O)^{-1} = (I_T + O^{-1})^{-1}$ 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:*

$$(33) \qquad \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{Q} & -\mathbf{Q}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{Q} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{Q}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix}, \qquad \mathbf{Q} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1},$$

Proof. Direct computation. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

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*

$$(34) \qquad \mathcal{R} = (I_N - O)(I_N + O)^{-1} = \frac{I_N - O}{I_N + O}, \quad so\ that \quad O = \frac{I_N - \mathcal{R}}{I_N + \mathcal{R}}$$

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

Proof. The formulas (34) 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, $\det(O) = \det\left(\frac{I_N - \mathcal{R}}{I_N + \mathcal{R}}\right) = 1$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

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 (41) below implies that the density of $\widehat{W}$ is proportional to

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

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 (34) with (33) we have

$$(35) \quad \mathcal{R} = \frac{I_\mathcal{T} - O}{I_\mathcal{T} + O} = \frac{2}{I_\mathcal{T} + O} - I_\mathcal{T}$$

$$= \begin{pmatrix} 2Q - I_M & -2QB(I_N + D)^{-1} \\ -2(I_N + D)^{-1}CQ & 2(I_N + D)^{-1} + 2(I_N + D)^{-1}CQB(I_N + D)^{-1} - I_N \end{pmatrix},$$

$$Q = (I_M + A - B(I_N + D)^{-1}C)^{-1}.$$

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,

$$(36) \quad \frac{I_M - \tilde{\mathcal{R}}}{I_M + \tilde{\mathcal{R}}} = \frac{2I_M - 2Q}{2Q} = Q^{-1} - I_M = A - B(I_N + D)^{-1}C = \tilde{O}.$$

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

$$(37) \quad \det(I_\mathcal{T} - \mathcal{R}^2)^{-\frac{1}{2}\mathcal{T}+\frac{1}{2}}d\mathcal{R} = \det(I_\mathcal{T} - \mathcal{R})^{1-\mathcal{T}}d\mathcal{R} = \det(I_\mathcal{T} + \mathcal{R})^{1-\mathcal{T}}d\mathcal{R},$$

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 (35) 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

$$(38) \qquad \widehat{W} = 2(I_M + \tilde{\mathcal{R}})^{-1}QB(I_N + D)^{-1} = B(I_N + D)^{-1}.$$

We also define

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

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

$$(39) \qquad \det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det \mathbf{A} \cdot \det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}),$$

we rewrite (37) as

$$(40) \quad \det(I_{\mathcal{T}} - \mathcal{R})^{1-\mathcal{T}}\, d\tilde{\mathcal{R}}\, dW\, d\mathcal{R}_2 = \det \begin{pmatrix} I_M - \tilde{\mathcal{R}} & W \\ -W^* & I_N - \mathcal{R}_2 \end{pmatrix}^{1-\mathcal{T}} d\tilde{\mathcal{R}}\, dW\, d\mathcal{R}_2$$

$$= \det(I_M - \tilde{\mathcal{R}})^{1-\mathcal{T}} \det(I_N - \mathcal{R}_2 + W^*(I_M - \tilde{\mathcal{R}})^{-1}W)^{1-\mathcal{T}}\, d\tilde{\mathcal{R}}\, dW\, d\mathcal{R}_2$$

$$= \det\left(I_M - \tilde{\mathcal{R}}\right)^{1-\mathcal{T}} \det\left(I_N - \mathcal{R}_2 + \widehat{W}^*(I_M + \tilde{\mathcal{R}})\widehat{W}\right)^{1-\mathcal{T}}\, d\tilde{\mathcal{R}}\, dW\, d\mathcal{R}_2$$

Further, notice that

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

Hence, the last line of (40) is transformed into

$$(41) \quad \det\left(I_M - \tilde{\mathcal{R}}\right)^{1-\mathcal{T}} \det\left(I_N + \widehat{W}^*\widehat{W}\right)^{1-\mathcal{T}} \det\left(I_N + \widehat{\mathcal{R}}_2\right)^{1-\mathcal{T}}\, d\tilde{\mathcal{R}}\, dW\, d\mathcal{R}_2$$

$$= \det\left(I_M + \tilde{\mathcal{R}}\right)^{1-M} \det\left(I_N + \widehat{W}^*\widehat{W}\right)^{1/2-M-N/2} \det\left(I_N + \widehat{\mathcal{R}}_2\right)^{1-M-N}\, d\tilde{\mathcal{R}}\, d\widehat{W}\, d\widehat{\mathcal{R}}_2,$$

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

$$(42) \qquad \left|\frac{\partial(QZ)}{\partial Z}\right| = |\det Q|^m,$$

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

$$(43) \qquad \left| \frac{\partial(QZQ^*)}{\partial Z} \right| = |\det Q|^{n-1}.$$

The first identity (42) 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 (41) 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 $\det\left(1 + \tilde{\mathcal{R}}\right)^{1-M} d\tilde{\mathcal{R}}$. Comparing with (37) 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 (36), we conclude that $\tilde{O}$ is independent with $\widehat{W}$, which is precisely $B(I_N + D)^{-1}$ by (38). $\qquad \square$

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 $P_1 P_2 P_1$ 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 $P_1 P_2 P_1$. 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 $P_1 P_2 P_1$ is transformed into $O^r P_1 P_2 P_1 O^{-r}$. 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 $P_1 P_2 P_1$ as if we replaced $O \mapsto A^* O A$. In both Steps 1 and 2 O had the same distribution as $A^* O A$, hence,

the eigenvalues were unchanged. But in general, if A is correlated with O in a non-trivial way, then the distribution might change.[13]

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 $P_1 P_2 P_1$ 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 $P_1 P_2 P_1$ are distributed as (32).

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

$$O \binom{w_2}{w_1} = \binom{u_2}{u_1}.$$

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.

---

[13]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 $\mathcal{V}$.

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 (32) unchanged, hence, we can invoke the induction assumption, thus, finishing the proof. $\square$

9. Appendix. Proofs 3: A perturbation of the Jacobi ensemble.

In this section we use Theorem 12 to prove Theorem 8.

Recall the cyclic shift[14] operator L_c acting in T–dimensional space. Let V be the $(T-1)$–dimensional space orthogonal to the vector $(1, 1, \ldots, 1)$, i.e., $V = \{(x_1, \ldots, x_T) \mid x_1 + \cdots +$

---

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

$x_T = 0\}$. 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 $\tilde{L}\mathcal{U}, \tilde{L}^2\mathcal{U}, \ldots, \tilde{L}^{k-1}\mathcal{U}$. 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 $P_1 P_2 P_1$ 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, $1 \leq t \leq T$. 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 $t = 1, 2, \ldots, T$, 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 $(1, \ldots, 1)$, 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.

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

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

The last matrix is a product of two projectors:[15] 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:

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

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 $(1, 1, \ldots, 1)$). 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.

   We can write

$$(44) \quad \tilde{X}_t = X_{t-1} - \frac{t-1}{T}(X_T - X_0) = X_0 + (t-1)\mu + \sum_{\tau=1}^{t-1}\varepsilon_\tau - \frac{t-1}{T}\left(T\mu + \sum_{\tau=1}^{T}\varepsilon_\tau\right)$$

$$= X_0 + \sum_{\tau=1}^{t-1}\varepsilon_\tau - \frac{t-1}{T}\sum_{\tau=1}^{T}\varepsilon_\tau.$$

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

---

[15]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.

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. $\square$

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*

$$(45) \qquad \lim_{N\to\infty} \mathrm{Prob}\left(\|O_1 - O_2\|_2 < \frac{1}{N^{1-\varepsilon}}\right) = 1.$$

Then for each $\varepsilon > 0$ we have

$$(46) \qquad \lim_{N\to\infty} \mathrm{Prob}\left(\|f^{k,N,\mathcal{T};\mathcal{V}}(O_1) - f^{k,N,\mathcal{T};\mathcal{V}}(O_2)\|_\infty < \frac{1}{N^{1-\varepsilon}}\right) = 1.$$

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 $\mathcal{T} = (\mathcal{T} - N) + N$:

$$(47) \qquad O_1 = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \qquad O_2 = \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix}.$$

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*

$$(48) \quad \lim_{N\to\infty} \mathrm{Prob}\left(\|(A_1 - B_1(I_N + D_1)^{-1}C_1) - (A_2 - B_2(I_N + D_2)^{-1}C_2)\|_2 < \frac{1}{N^{1-\varepsilon}}\right) = 1.$$

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*

$$(49) \qquad \lim_{N \to \infty} \text{Prob}\left(\|U_1 - U_2\|_2 < \frac{1}{N^{1-\varepsilon}} \right) = 1.$$

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

$$(50) \qquad f^{k,N,\mathcal{T};U\mathcal{V}}(O) = f^{k,N,\mathcal{T};\mathcal{V}}(U^*OU).$$

Note that conjugations (by the same orthogonal matrix for O_1 and O_2) leave the three conditions of Proposition 20 unchanged, hence, the (50) 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 $P_1 P_2 P_1$, 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 $\mathcal{V}, O\mathcal{V}, O^2\mathcal{V}, \ldots, O^{k-2}\mathcal{V}$. 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 $P_1 P_2 P_1$.

Using the block notations (47), steps 4-7 in the proof of Theorem 12 imply the following almost sure identities:

$$(51) \qquad \tilde{f}^{k,N,\mathcal{T};\mathcal{V}_0}(O_1) = \tilde{f}^{k-1,N,\mathcal{T}-N;\langle B_1 \rangle}(A_1 - B_1(I_N + D_1)^{-1}C_1),$$

$$(52) \qquad \tilde{f}^{k,N,\mathcal{T};\mathcal{V}_0}(O_2) = \tilde{f}^{k-1,N,\mathcal{T}-N;\langle B_2 \rangle}(A_2 - B_2(I_N + D_2)^{-1}C_2).$$

We would like to check that the right-hand sides of (51) and (52) are close by using the induction assumption.

Using (50) and Lemma 24, we rewrite the right-hand sides of (51) and (52) as:
$$(53)$$
$$\tilde{f}^{k-1,N,\mathcal{T}-N;\tilde{\mathcal{V}}_0}\left(U_1^*(A_1 - B_1(I_N+D_1)^{-1}C_1)U_1 \right); \qquad \tilde{f}^{k-1,N,\mathcal{T}-N;\tilde{\mathcal{V}}_0}\left(U_2^*(A_2 - B_2(I_N+D_2)^{-1}C_2)U_2 \right).$$

Let us check that we can apply the induction assumption to deduce that the expressions of (53) 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

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

Hence, using the $(k-1)$ statement, the expressions in (53) are close to each other as $N \to \infty$ and, therefore, (51) is close to (52). $\qquad \square$

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 (36):

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

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. $\qquad \square$

Proof of Lemma 23. Note that whenever X is a submatrix of Y, we have $\|X\|_2 \leq \|Y\|_2$. 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, $\|B_1\|_2 \leq 1$ and $\|C_1\|_2 \leq 1$.

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:

$$(54) \qquad \det Y^{-1/2} \det(I_N - Y)^{\frac{\mathcal{T}}{2} - N - 1/2} \, dY, \qquad 0 < Y < I_N.$$

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 Theorem 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

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

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

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

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 $i = 1, 2, \ldots, \mathcal{T} - N$.

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,

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

and similarly for X_2.

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

$$(55) \qquad \lim_{N \to \infty} \text{Prob}\left(\|X_1 - X_2\|_2 < \frac{1}{N^{1-\varepsilon}}\right) = 1.$$

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[16]. 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:

$$(56) \qquad \det \Lambda^{\frac{\mathcal{T}}{2} - N - 1/2} \det(I_N - \Lambda)^{-1/2} \, d\Lambda, \qquad 0 < \Lambda < I_N.$$

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 Theorem 11, which implies that the eigenvalues of Λ are bounded away from 0 as $N \to \infty$. The claim is proven.

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

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

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).

---

[16]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: $\sqrt{x} = \sqrt{1 + (x-1)} = 1 + \frac{x-1}{2} - \frac{1}{4}(x-1)^2 + \ldots$ to deduce the Lipschitz property.

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 $X_1^1, X_1^2, \ldots, X_1^N$. 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 instability 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

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

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

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

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

(58) $$W_2^* W_2 = \begin{pmatrix} Y_2^* Y_2 + Z_2^* Z_2 & Z_2^* \\ Z_2 & I_M \end{pmatrix}.$$

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

$$(59) \qquad W_2^* W_2 = \begin{pmatrix} I_N & 0 \\ Q_2 & I_M \end{pmatrix} \begin{pmatrix} G_2 & 0 \\ 0 & H_2 \end{pmatrix} \begin{pmatrix} I_N & Q_2^* \\ 0 & I_M \end{pmatrix} = \begin{pmatrix} G_2 & G_2 Q_2^* \\ Q_2 G_2 & Q_2 G_2 Q_2^* + H_2 \end{pmatrix}.$$

Comparing with (58) we conclude that

$$(60) \qquad G_2 = Y_2^* Y_2 + Z_2^* Z_2 = F^* F, \quad Q_2 = Z_2 (F^* F)^{-1}, \quad H_2 = I_M - Z_2 (F^* F)^{-1} Z_2^*.$$

Since the spectral norm of a submartix is at most the spectral norm of the matrix, (57) implies that

$$(61) \quad \lim_{N \to \infty} \mathrm{Prob} \left(\| Y_2 - I_N \|_2 < \frac{1}{N^{1-\varepsilon}} \right) = 1, \qquad \lim_{N \to \infty} \mathrm{Prob} \left(\| Z_2 - 0_{M \times N} \|_2 < \frac{1}{N^{1-\varepsilon}} \right) = 1.$$

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

$$(62) \qquad \tilde{U}_2 = W_2 \cdot \begin{pmatrix} I_N & -Q_2^* \\ 0 & I_M \end{pmatrix} \cdot \begin{pmatrix} G_2^{-1/2} & 0 \\ 0 & H_2^{-1/2} \end{pmatrix}.$$

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 (59) 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 (62) and see that they are close to identical matrices by (61). 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. $\square$

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 [17]

$$(63) \qquad \mathrm{Prob}\left(\max_{1 \le i,j < \mathcal{T}} |D - D^{\mathrm{rand}}|_{ij} > \frac{1}{\mathcal{T}^{1-\delta}} \right) < c_1(\beta) \cdot \mathcal{T} \cdot \exp\left(-c_2(\beta) \frac{\mathcal{T}^{2\delta}}{\log \mathcal{T}} \right).$$

We remark that since D and D^{rand} are block-diagonal, the bound on the maximum matrix element of their difference is equivalent to a similar bound for any other norm, e.g., for the 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 $O_2D^{\mathrm{rand}}O_2^*$. The invariance of the uniform measure on the orthogonal group $SO(N)$ (or on the unitary group $U(N)$ if $\beta = 2$) with respect to right/left multiplications, implies the distributional identities:

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

---

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

The right-hand sides of the identities provide the desired coupling and (63) 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 $O_2 D^{\mathrm{rand}} O_2^*$ and the second matrix being $O_2 D O_2^*$. □

10. Appendix. Proofs 4: Small rank perturbations

In this section we prove Theorem 3 by combining Theorem 8 with Theorem 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. (11). 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:*

(64)
$$P_0 = R_0^*(R_0 R_0^*)^{-1} R_0, \quad P_k = R_k^*(R_k R_k^*)^{-1} R_k, \quad \tilde{P}_0 = \tilde{R}_0^*(\tilde{R}_0 \tilde{R}_0^*)^{-1} \tilde{R}_0, \quad \tilde{P}_k = \tilde{R}_k^*(\tilde{R}_k \tilde{R}_k^*)^{-1} \tilde{R}_k.$$

Then under the assumptions (6), (7) *of Theorem 3 we have*

(65)
$$\lim_{N \to \infty} \frac{1}{N} \mathrm{rank} \left(P_0 P_k P_0 - \tilde{P}_0 \tilde{P}_k \tilde{P}_0 \right) = 0.$$

Proof. Throughout the proof we assume that the matrices $R_i R_i^*$ 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 $\mathrm{rank}(A - B) \leq \mathfrak{r}$;
- $\mathrm{rank}(CA - CB) \leq \mathrm{rank}(A - B)$;
- $\mathrm{rank}(A + B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$;
- If matrices A and B are invertible, then $\mathrm{rank}(A^{-1} - B^{-1}) = \mathrm{rank}(A - B)$.

We refer to the time series defined by (1) as X_t and to the time series defined by (11) as $\mathcal{X}_t$. We form two $N \times T$ matrix X and $\mathcal{X}$ with columns X_t and $\mathcal{X}_t$, $t = 1, \ldots, T$, respectively. Our first task is to show that $\frac{1}{N} \mathrm{rank}(X - \mathcal{X}) \to 0$ as $N \to \infty$.

For that we subtract (11) from (1) to get:

(66)
$$\Delta X_t - \Delta \mathcal{X}_t = \Pi X_{t-k} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t - \mu, \quad t = 1, 2, \ldots, T.$$

In the matrix form, (66) 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

(66)) at most

(67) $$\text{rank}(\Pi) + \sum_{i=1}^{k-1} \text{rank}(\Gamma_i) + d_D + 1.$$

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

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

Hence, the rank of $X - \mathcal{X}$ is at most (67) and $\frac{1}{N}\text{rank}(X - \mathcal{X}) \to 0$ by assumption (7) 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 (7). 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,

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

Since $P_0 P_k P_0$ 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, (65) follows from (68). $\qquad\square$

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 Theorem 11 we conclude that the empirical measure of $\tilde{\lambda}_i$ converges:

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

We would like to show that $\tilde{\lambda}_i$ can be replaced by λ_i in (69). Note that although the spectra of matrices $\mathcal{C}$ and $\tilde{\mathcal{C}}$ are real, but these matrices are not symmetric. Similarly to (65), $\frac{1}{N}\text{rank}(\mathcal{C} - \tilde{\mathcal{C}}) \to 0$, 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 (65).

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 $S_{ij} = R_i R_j^*$ and using the notations (64), we conclude that the eigenvalues of $\mathcal{C}$ are the same as N largest eigenvalues of $P_k P_0$. Since $P_0^2 = P_0$, they are also the same as N largest eigenvalues of $P_k P_0 P_0$ and the same as those of $P_0 P_k P_0$. 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 $\mathfrak{r} = \mathrm{rank}(P_0 P_k P_0 - \tilde{P}_0 \tilde{P}_k \tilde{P}_0)$. 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

$$(70) \qquad \lambda_{m-\mathfrak{r}} \geq \tilde{\lambda}_m \geq \lambda_{m+\mathfrak{r}}, \quad 1 \leq m - \mathfrak{r} \leq m + \mathfrak{r} \leq N.$$

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

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

Hence, (69) implies

$$(71) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i} = \mu_{2,\tau-k}. \qquad \square$$

References

Anderson, T. W., "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 1951, *22* (3), 327–351.

Bai, J. and S. Ng, "Large Dimensional Factor Analysis," *Foundations and Trends in Econometrics*, 2008, *3* (2), 89–163.

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

Breitung, J. 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," Berlin, Heidelberg: Springer, 2008, pp. 279–322.

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

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

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

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

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

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

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

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

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

Han, C., 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*, 2018, *24* (4B), 3447–3468.

Han, X., G. M. Pan, and B. Zhang, "The Tracy-Widom law for the Largest Eigenvalue of F Type Matrix," *The Annals of Statistics*, 2016, *44* (4), 1564–1592.

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

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

Johansen, S., "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 1988, *12* (2–3), 231–254.

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

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

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

Johnstone, I., "Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence," *Annals of statistics*, 2008, *36* (6), 2638–2716.

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

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

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

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

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

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

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

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

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

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

Reinsel, G. C. and S. K. Ahn, "Vector autoregressive models with unit roots and reduced rank structure: Estimation. Likelihood ratio test, and forecasting," *Journal of time series analysis*, 1992, *13* (4), 353–375.

Sims, C. A., "Macroeconomics and Reality," *Econometrica*, 1980, *48* (1), 1–48.

Swensen, A. R., "Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models," *Econometrica*, 2006, *74* (6), 1699–1714.

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

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

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

Wang, D. and R. S. Tsay, "Rate-Optimal Robust Estimation of High-Dimensional Vector Autoregressive Models," *arXiv preprint arXiv:2107.11002*, 2022.

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

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

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

(Anna Bykhovskaya) DUKE UNIVERSITY
Email address: anna.bykhovskaya@duke.edu

(Vadim Gorin) UNIVERSITY OF CALIFORNIA AT BERKELEY
Email address: vadicgor@gmail.com