Improved Inference on the Rank of a Matrix

Qihui Chen
School of Management and Economics
The Chinese University of Hong Kong, Shenzhen
qihuichen@cuhk.edu.cn

Zheng Fang*
Department of Economics
Texas A&M University
zfang@tamu.edu

December 4, 2018

Abstract

This paper develops a general framework for conducting inference on the rank of an unknown matrix Π_0. A defining feature of our setup is the null hypothesis of the form $\mathrm{H}_0 : \mathrm{rank}(\Pi_0) \leq r$. The problem is of first order importance because the previous literature focuses on $\mathrm{H}_0' : \mathrm{rank}(\Pi_0) = r$ by implicitly assuming away $\mathrm{rank}(\Pi_0) < r$, which may lead to invalid rank tests due to over-rejections. In particular, we show that limiting distributions of test statistics under H_0' may not stochastically dominate those under $\mathrm{rank}(\Pi_0) < r$. A multiple test on the nulls $\mathrm{rank}(\Pi_0) = 0, \ldots, r$, though valid, may be substantially conservative. We employ a testing statistic whose limiting distributions under H_0 are highly nonstandard due to the inherent irregular natures of the problem, and then construct bootstrap critical values that deliver size control and improved power. Since our procedure relies on a tuning parameter, a two-step procedure is designed to mitigate concerns on this nuisance. We additionally argue that our setup is also important for estimation. We illustrate the empirical relevance of our results through testing identification in linear IV models that allows for clustered data and inference on sorting dimensions in a two-sided matching model with transferrable utility.

KEYWORDS: Matrix rank, Bootstrap, Two-step test, Rank estimation, Identification, Matching dimension

*We are indebted to Frank Schorfheide and three anonymous referees for valuable and constructive comments that helped greatly improve this paper. We would also like to thank Yonghong An, Brendan Beare, Jiaying Gu, Yingyao Hu, Ivana Komunjer, Qi Li, Aureo de Paula, Peter Phillips, Andres Santos, Yixiao Sun, Xun Tang and seminar participants for helpful discussions and comments, and Frank Kleibergen for generously sharing the data with us.

1 Introduction

The rank of a matrix plays a number of fundamental roles in economics, not just as crucial technical identification conditions (Fisher, 1966), but also of central empirical relevance in numerous settings such as inference on cointegration rank (Engle and Granger, 1987; Johansen, 1991), specification of finite mixture models (McLachlan and Peel, 2004; Kasahara and Shimotsu, 2009) and estimation of matching dimensions (Dupuy and Galichon, 2014) – more can be found in Supplemental Appendix E. These problems reduce to examining the hypotheses: for an unknown matrix Π_0 of size $m \times k$ with $m \geq k$,

$$\mathrm{H}_0 : \mathrm{rank}(\Pi_0) \leq r \qquad \text{v.s.} \qquad \mathrm{H}_1 : \mathrm{rank}(\Pi_0) > r \ , \tag{1}$$

where $r \in \{0, \ldots, k-1\}$ is some prespecified value and $\mathrm{rank}(\Pi_0)$ denotes the rank of Π_0. If $r = k-1$, then (1) is concerned with whether Π_0 has full rank.

Despite a rich set of results in the literature, previous studies instead focus on

$$\mathrm{H}_0' : \mathrm{rank}(\Pi_0) = r \qquad \text{v.s.} \qquad \mathrm{H}_1 : \mathrm{rank}(\Pi_0) > r \ . \tag{2}$$

In effect, the testing problem (2) assumes away the possibility $\mathrm{rank}(\Pi_0) < r$, which is often unrealistic to be excluded. This, unfortunately, has drastic consequences. As elaborated through an analytic example in Section 2, a number of popular tests, including Robin and Smith (2000) and Kleibergen and Paap (2006), may over-reject for some data generating processes and under-reject for others, both having $\mathrm{rank}(\Pi_0) < r$. In particular, contrary to what appears to have been conjectured in the literature (Cragg and Donald, 1993, p.225; Johansen, 1995, p.168), our analysis suggests that *limiting distributions of tests obtained under* H_0' *may not first order stochastically dominate those under* $\mathrm{rank}(\Pi_0) < r$. Hence, ignoring the possibility $\mathrm{rank}(\Pi_0) < r$ may lead to tests that are not even first order valid.

One may nonetheless justify the setup (2) for two reasons. First, the problem (1) may be studied by a multiple test on the nulls $\mathrm{rank}(\Pi_0) = 0, 1, \ldots, r$. Our simulations show, however, that such a procedure, though valid, may be substantially conservative and have trivial power against local alternatives that are close to matrices whose rank is strictly less than r. Second, the setup (2) suits well for estimation by sequentially testing $\mathrm{rank}(\Pi_0) = j$ for $j = 0, 1, \ldots, k-1$. Crucially, however, all steps except for $j = 0$ ignore type I errors (false rejection) potentially made in previous steps, and may have limited capability of controlling type II errors (false acceptance) – see Supplemental Appendix C for more details. Hence, the setup (1) is desirable for estimation as well.

We thus conclude that developing a valid and powerful test for (1) is of first order importance. To the best of our knowledge, no direct tests to date exist in this regard.

Our objective in this paper is therefore to develop an inferential framework under the setup (1). A key insight we exploit to this end is that (1) is equivalent to

$$\mathrm{H}_0 : \phi_r(\Pi_0) = 0 \qquad \text{v.s.} \qquad \mathrm{H}_1 : \phi_r(\Pi_0) > 0 \ , \tag{3}$$

where $\phi_r(\Pi_0) \equiv \sum_{j=r+1}^{k} \sigma_j^2(\Pi_0)$ is the sum of the $k - r$ smallest squared singular values $\sigma_j^2(\Pi_0)$ of Π_0 – see Supplemental Appendix for a review on singular values. Such a reformulation is attractive because it converts an unwieldy inference problem on an integer-valued parameter (i.e., rank) into a more tractable one on a real-valued functional (i.e., a sum of singular values). Given an estimator $\hat{\Pi}_n$ of Π_0, it is thus natural to base the testing statistic on the plug-in estimator $\phi_r(\hat{\Pi}_n)$ and then invoke the Delta method. As it turns out, the formulation (3) reveals two crucial irregular natures involved, namely, ϕ_r admits a zero first order derivative under H_0 and is second order nondifferentiable precisely when $\mathrm{rank}(\Pi_0) < r$ – see Proposition 3.1 and Lemma D.5. While the null limiting distributions of $\phi_r(\hat{\Pi}_n)$ can nonetheless be derived by existing generalizations of the Delta method (Shapiro, 2000), constructions of critical values are nontrivial because the limits are non-pivotal and highly nonstandard. In particular, they depend on the true rank (among other things), upholding the importance of taking into account the possibility $\mathrm{rank}(\Pi_0) < r$. For this, we appeal to modified bootstrap schemes recently developed by Fang and Santos (2018) and Chen and Fang (2018), which yield tests for (1) that have asymptotically pointwise exact size control and are consistent. We further characterize analytically classes of local perturbations of the data generating processes under which our tests enjoy size control and nontrivial power.

A common feature of our tests is their dependence on tuning parameters, although we stress that this is only in line with the irregular natures of nonstandard problems (Chernozhukov et al., 2007; Andrews and Soares, 2010; Linton et al., 2010). While we are unable to offer a general theory guiding their choices, a two-step procedure similar to Romano et al. (2014) is proposed to mitigate potential concerns. The intuition is as follows. First, the appearance of $r_0 \equiv \mathrm{rank}(\Pi_0)$ in the limits suggests the need of a consistent rank estimator $\hat{r}_n$, which may be achieved by a sequential testing procedure coupled with a significance level α_n (serving as the tuning parameter) that tends to zero suitably. Although the estimation error of $\hat{r}_n$, i.e., the probability of false selection, is asymptotically negligible (as $\alpha_n \to 0$), that probability is positive in any finite samples. Thus, we account for false selection by fixing $\alpha_n = \beta$ rather than letting it tend to zero. Given an estimator $\hat{r}_n$ with $\liminf_{n\to\infty} P(\hat{r}_n = r_0) \geq 1 - \beta$, the two-step procedure at a significance level α is: reject H_0 if $\hat{r}_n > r$ in the first step; otherwise in the second step incorporate $\hat{r}_n$ into our bootstrap and conduct the test at the adjusted significance level $\alpha - \beta > 0$. We show in a number of simulation designs that the procedure is quite insensitive to our choices of β, even for small sample sizes. The simulation results are particularly encouraging because we fix crude choices of tuning parameters throughout the paper, and yet the performance of our methods is consistently better.

The marked size and power properties rest with several attractive features. First, since we rely on the Delta method, the theory is conceptually simple and requires mild assumptions. Essentially, all we need are a matrix estimator $\hat{\Pi}_n$ that converges weakly and a consistent bootstrap analog. In particular, the data may be non-i.i.d. and non-stationary, the convergence rate may be non-$\sqrt{n}$ and even heterogeneous across entries of $\hat{\Pi}_n$ – see Supplemental Appendix E.1, the limit $\mathcal{M}$ of $\hat{\Pi}_n$ may be non-Gaussian, the bootstrap for $\mathcal{M}$ (a crucial ingredient of our method) may be virtually any consistent resampling scheme, and no side rank conditions are directly imposed beyond those entailed by the restrictions on the population quantiles. Second, computation of our testing statistic and the critical values are quite simple as both involve only calculations of singular value decompositions – we reiterate that the need of resamping only reflects the irregular natures of the problem rather than because of an exclusive attribute of our treatment. Finally, the superior testing properties of our procedure translate to more accurate rank estimators through the aforementioned two channels, namely, reducing type I and type II errors. Simulations confirm that our methods work better when $\text{rank}(\Pi_0) < r$ or when Π_0 is close to a matrix whose rank is strictly less than r.

We illustrate the application of our framework by testing identification in linear IV models that accommodates clustered data. To draw further attention to the empirical relevance of our results, we study a two-sided bipartite matching model with transferrable utility, building upon the work of Dupuy and Galichon (2014). A central question here is: how many attributes are relevant for the matching? Under a parametric specification of the surplus function, this number is equal to the rank of the so-called affinity matrix. We show that our procedure and Kleibergen and Paap (2006) can produce quite different results with regards to several model specifications, in terms of both p-values of the tests and actual estimates of the matching dimension.

As mentioned previously, the literature has been mostly concerned with the hypotheses (2). In the context of multivariate regression, Anderson (1951) develops a likelihood ratio test based on canonical correlations. This test is restrictive in that it crucially depends on the asymptotic variance Ω_0 of $\text{vec}(\hat{\Pi}_n)$ having a Kronecker product structure. Building upon Gill and Lewbel (1992), Cragg and Donald (1996) propose a test that requires nonsingularity of Ω_0 and may be sensitive to the transformations involved. Cragg and Donald (1997) provide a test based on a constrained minimum distance criterion, which, in addition to the nonsingularity requirement of Ω_0, is in general computationally intensive. Motivated by the need to relax the nonsingularity condition, Robin and Smith (2000) employ a class of testing statistics which are asymptotically equivalent to ours, but their results only apply to the setup (2). Kleibergen and Paap (2006) study a Wald-standardized version of our statistic in order to obtain pivotal asymptotic distributions (under H_0'), but at the expense of a side rank condition. We refer the reader to Camba-Mendez and Kapetanios (2009), Portier and Delyon (2014) and Al-Sadoon (2017) for further discussions of the literature.

There are a few exceptions that study (1). Johansen (1988, 1991) obtains his likelihood ratio statistics under H_0 but only establishes their asymptotic distributions under H_0'. Shortly after, Johansen (1995, p.157-8,168) presents the limits under H_0, and essentially argues based on simulations that the asymptotic distributions under $\mathrm{rank}(\Pi_0) < r$ are first order stochastically dominated by those under H_0' and "hence not relevant for calculating the p-value". However, in light of our results in Section 2, it is unclear to us that this remains true in general, and even if this is the case, one may be still concerned with under-rejections and as a result (potentially) poor power of the tests, as documented in the literature – see Maddala and Kim (1998), Johansen (2002) and references therein. Cragg and Donald (1993, p.225) recognize the importance of studying (1), but do not derive the asymptotic distributions under H_0. Instead, they show that their statistic has first order stochastically dominant limiting laws under H_0' with somewhat restrictive conditions. Our results suggest that may not be true in general.

We now introduce some notation. The space of $m \times k$ matrices is denoted by $\mathbf{M}^{m\times k}$. For a matrix A, we write its transpose by $A^\intercal$, its trace by $\mathrm{tr}(A)$ if it is square, its vectorization by $\mathrm{vec}(A)$, and its Frobenius norm by $\|A\| \equiv \sqrt{\mathrm{tr}(A^\intercal A)}$. The identity matrix of size k is denoted I_k, the $k \times 1$ vectors of zeros and ones are respectively denoted by $\mathbf{0}_k$ and $\mathbf{1}_k$, and the $m \times k$ matrix of zeros is denoted $\mathbf{0}_{m\times k}$. We let $\mathrm{diag}(a)$ denote the diagonal matrix whose diagonal entries compose a. The jth largest singular value of a matrix $A \in \mathbf{M}^{m\times k}$ is denoted $\sigma_j(A)$. We define the set $\mathbb{S}^{m\times k} = \{A \in \mathbf{M}^{m\times k} : A^\intercal A = I_k\}$ and let $\overset{d}{=}$ signify "equal in distribution." Finally, $\lfloor a \rfloor$ is the integer part of $a \in \mathbf{R}$.

The remainder of the paper is organized as follows. Section 2 illustrates the consequences of ignoring $\mathrm{rank}(\Pi_0) < r$, and provides an overview of our tests, together with a step-by-step implementation guide. Section 3 develops our inferential framework. Section 4 presents Monte Carlo studies. Section 5 further illustrates the empirical relevance of our results by studying a matching model. Section 6 briefly concludes. Proofs are collected in a Supplemental Appendix. We also study the estimation problem, but, due to space limitation, relegate the results to Supplemental Appendix C. Finally, we have developed a Stata command `bootranktest` to test whether a matrix of the form $E[VZ^\intercal]$ has full rank – see the Supplemental Appendix for a brief description.

2 Motivations, Overview and Implementation

In this section, we first motivate the development of our theory by illustrating how serious the issue can be if one ignores the possibility $\mathrm{rank}(\Pi_0) < r$ in conducting rank tests. This is accomplished by examining the influential test proposed by Kleibergen and Paap (2006), referred to as the KP test hereafter, and its multiple testing version. Then we provide an overview of our tests, together with a step-by-step implementation guide that applies to general settings – please see Section 3.3.2 for a guide designed for the special case of testing identification in linear IV models.

To elucidate the consequences of ignoring $\text{rank}(\Pi_0) < r$, consider an example where $\Pi_0 = \mathbf{0}_{2\times 2}$ and $r = 1$ so that $\text{rank}(\Pi_0) < r$. Suppose Π_0 admits an estimator $\hat{\Pi}_n$ such that $\sqrt{n}\hat{\Pi}_n \overset{d}{=} \mathcal{M}$ for all n (rather than just asymptotically), where $\mathcal{M} \in \mathbf{M}^{2\times 2}$ satisfies $\text{vec}(\mathcal{M}) \sim N(0, \Omega_0)$ with Ω_0 nonsingular and *known*. In this case, the KP test for (2) employs critical values from $\chi^2(1)$, while the actual distribution of the KP statistic is

$$T_{n,\text{kp}} \overset{d}{=} \frac{\sigma_2^2(\mathcal{M})}{(\mathcal{Q}_2 \otimes \mathcal{P}_2)^\intercal \Omega_0 (\mathcal{Q}_2 \otimes \mathcal{P}_2)} , \tag{4}$$

where $\mathcal{P}_2$ and $\mathcal{Q}_2$ are the left and right singular vectors associated with $\sigma_2(\mathcal{M})$, both having unit length. Note the distribution of $T_{n,\text{kp}}$ depends only on Ω_0. Figure 2 plots (based on simulations) two cdfs F_1 and F_2 of $T_{n,\text{kp}}$ in (4) respectively determined by

$$\Omega_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \Omega_2 = \begin{bmatrix} 1 & 0 & 0 & -0.9\sqrt{5} \\ 0 & 1 & 0.9\sqrt{5} & 0 \\ 0 & 0.9\sqrt{5} & 5 & 0 \\ -0.9\sqrt{5} & 0 & 0 & 5 \end{bmatrix} ,$$

together with the cdf F_0 of $\chi^2(1)$. Note that F_0 is stochastically dominated by F_2 but stochastically dominates F_1, both in the first order sense. Hence, the KP test is invalid due to over-rejection when $\Omega_0 = \Omega_2$. We have thus *disproved* that the limits under $\text{rank}(\Pi_0) = r$ are first order stochastically dominant in general, a conjecture by Cragg and Donald (1993) for their statistic which they show to hold under somewhat restrictive conditions. These erratic behaviors can also be expected for the test of Robin and Smith (2000) in view of its relation to the KP test – see Supplemental Appendix B.

Alternatively, one might aim to construct a valid test for (1) by a multiple test on $\text{rank}(\Pi_0) = 0, 1, \ldots, r$. However, the validity is achieved at the expense of conservativeness – see Supplemental Appendix B, which may generate substantial power loss. To illustrate, consider the following data generating process:

$$Z = \Pi_0^\intercal V + u , \tag{5}$$

where $V, u \in N(0, I_6)$ are independent and, for $\delta \geq 0$ and $d \in \{1, \ldots, 6\}$,

$$\Pi_0 = \text{diag}(\mathbf{1}_{6-d}, \mathbf{0}_d) + \delta I_6 . \tag{6}$$

We test the hypotheses in (1) with $r = 5$ at the level $\alpha = 5\%$, and note that H_0 holds if and only if $\delta = 0$. For an i.i.d. sample $\{V_i, Z_i\}_{i=1}^{1000}$ generated according to (5), we

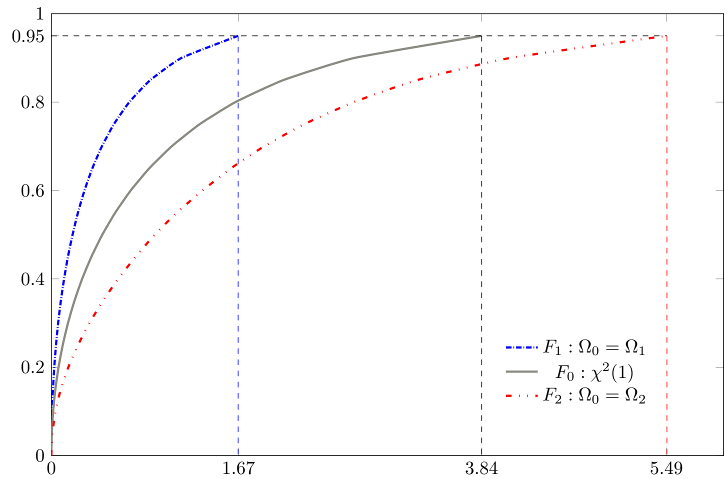

Figure 1. The cdfs of the KP statistic when $\Pi_0 = \mathbf{0}_{2\times 2}$ and $r = 1$

conduct tests based on the matrix estimator $\hat{\Pi}_n = \frac{1}{1000}\sum_{i=1}^{1000} V_i Z_i^\intercal$ for Π_0.

Figure 2 plots the power curves of the multiple KP test, labelled KP-M, based on $10,000$ simulation replications. For $d = 1$ (and so $\text{rank}(\Pi_0) = r$), the null rejection rate is 5%, while the power increases to unity as δ increases. As soon as $d > 1$ (so that $\text{rank}(\Pi_0) < r$), the power curves shift downward dramatically: the null rejection rates are close to zero and the power is well below 5% when δ is close to zero. Moreover, the power deteriorates as Π_0 becomes more degenerate in the sense that Π_0 is close to a matrix whose rank becomes smaller as d increases. This reinforces the critical importance to develop a valid and powerful test that accommodates $\text{rank}(\Pi_0) < r$.

For comparison, Figure 2 also depicts the power curves for a version of our test, labelled CF-A.[1] For the conventional setup ($d = 1$), the two tests have virtually the same reject rates across δ. Whenever $d > 1$, in stark contrast, our test effectively raises the power curves of the multiple KP test so that the null rejection rates are equal to 5%, and meanwhile the power becomes nontrivial. But it is more than that. The power improvement increases when d gets larger (and Π_0 more degenerate).

To describe our test, let $\hat{\Pi}_n$ be an estimator of $\Pi_0 \in \mathbf{M}^{m\times k}$ with $\tau_n\{\hat{\Pi}_n - \Pi_0\} \xrightarrow{L} \mathcal{M}$. The exact characterization of $\mathcal{M}$ (e.g., the covariance structure) is not required. Here, τ_n is typically $\sqrt{n}$ in cross-sectional and stationary time series settings, and may be non-$\sqrt{n}$ with non-stationary time series. Then our test statistic for (1) is $\tau_n^2\phi_r(\hat{\Pi}_n) \equiv \tau_n^2\sum_{j=r+1}^{k}\sigma_j^2(\hat{\Pi}_n)$. It turns out that, under H_0, we have: for $r_0 \equiv \text{rank}(\Pi_0)$,

$$\tau_n^2\phi_r(\hat{\Pi}_n) \xrightarrow{L} \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal \mathcal{M} Q_{0,2}) \ , \tag{7}$$

[1] In its implementation, we conduct $10,000$ simulation replications with 500 empirical bootstrap repetitions for each replication, and set $\kappa_n = n^{-1/4}$ – see the formal introduction below.

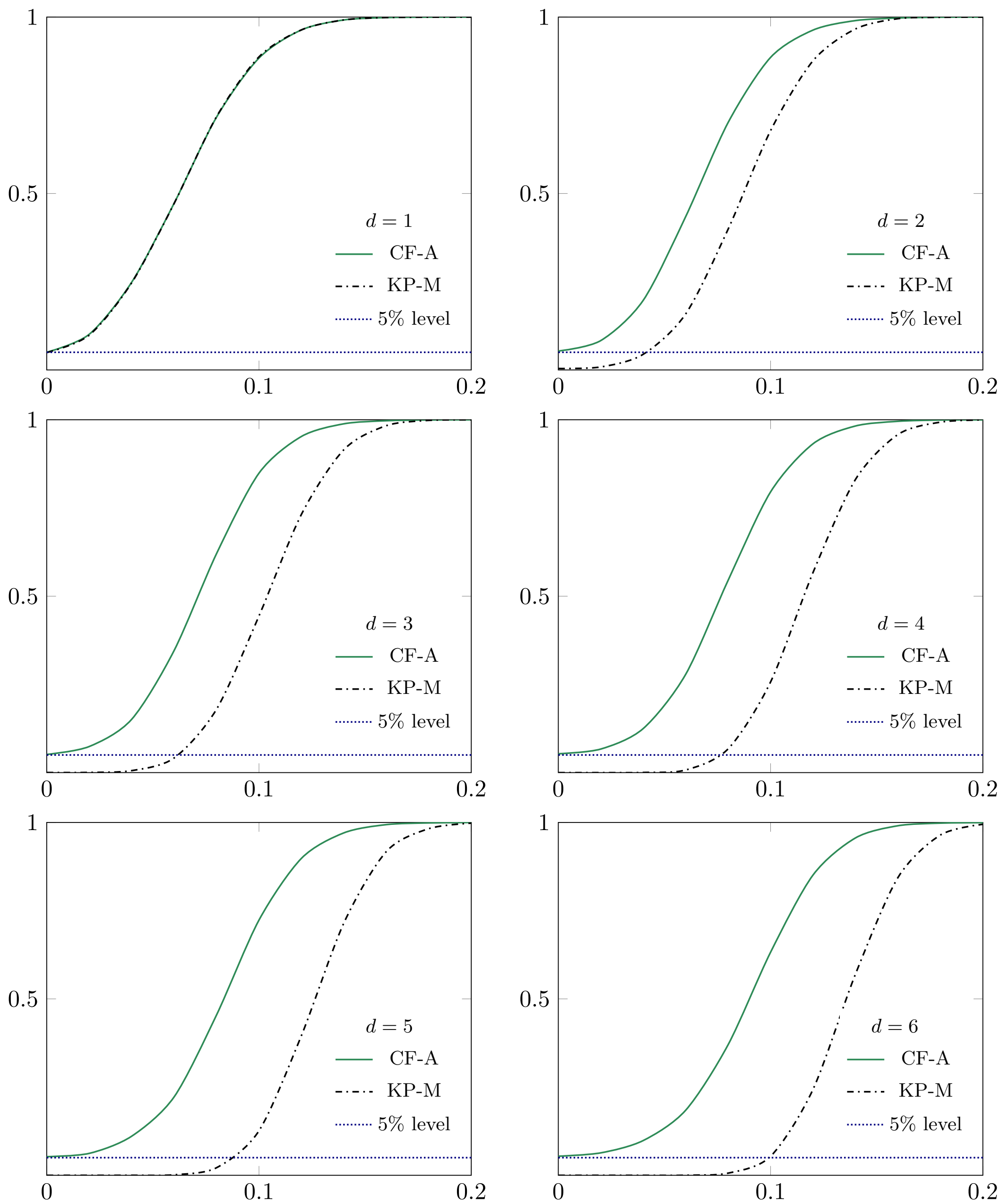

Figure 2. The rejection rates: the multiple KP test v.s. CF-A

where $P_{0,2} \in \mathbb{S}^{m\times(m-r_0)}$ and $Q_{0,2} \in \mathbb{S}^{k\times(k-r_0)}$ whose columns are respectively the left and the right singular vectors of Π_0 associated with its zero singular values. Since the limit in (7) depends on the true rank r_0 (crucially), $P_{0,2}$, $Q_{0,2}$ and $\mathcal{M}$, we estimate its law by first estimating these unknown objects, towards constructing critical values.

The rank r_0 may be consistently (under H_0) estimated by: for $\kappa_n \to 0$ and $\tau_n\kappa_n \to \infty$,

$$\hat{r}_n = \max\{j = 1, \ldots, r : \sigma_j(\hat{\Pi}_n) \geq \kappa_n\} \tag{8}$$

if the set is nonempty and $\hat{r}_n = 0$ otherwise. Heuristically, κ_n may be thought of as testing which population singular values are zero. Note that by estimating r_0 we take into account the possibility $r_0 < r$. Next, for a singular value decomposition $\hat{\Pi}_n = \hat{P}_n \hat{\Sigma}_n \hat{Q}_n^\intercal$, we may respectively estimate $P_{0,2}$ and $Q_{0,2}$ by $\hat{P}_{2,n}$ and $\hat{Q}_{2,n}$, which are respectively formed by the last $(m - \hat{r}_n)$ and $(k - \hat{r}_n)$ columns of $\hat{P}_n$ and $\hat{Q}_n$. The law of $\mathcal{M}$ may be consistently estimated by a bootstrap, say, $\hat{\mathcal{M}}_n^*$. Often, $\hat{\mathcal{M}}_n^* = \sqrt{n}\{\hat{\Pi}_n^* - \hat{\Pi}_n\}$ with $\hat{\Pi}_n^*$ computed in the same way as $\hat{\Pi}_n$ but based on a bootstrap sample. Finally, the law of the limit in (7) is estimated by the conditional distribution (given the data) of

$$\sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^\intercal \hat{\mathcal{M}}_n^* \hat{Q}_{2,n}) \ . \tag{9}$$

Given a significance level α, we then reject H_0 whenever $\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}$, where $\hat{c}_{n,1-\alpha}$ is the $1 - \alpha$ conditional quantile of (9) given the data.

While we are unable to provide an optimal choice of κ_n, a two-step test is proposed to mitigate potential concerns. In the first step, we obtain an estimator $\hat{r}_n$ satisfying $\liminf_{n\to\infty} P(\hat{r}_n = r_0) \geq 1 - \beta$ for some $\beta < \alpha$, and then reject H_0 if $\hat{r}_n > r$ and move on to the next step if $\hat{r}_n \leq r$. In the second step, we plug $\hat{r}_n$ into (9) and reject H_0 if $\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}$, where the significance level is adjusted to be $\alpha - \beta$. The estimator $\hat{r}_n$ in (8) now may not be appropriate as it appears challenging to control $P(\hat{r}_n = r_0)$. Instead, a desired estimator $\hat{r}_n$ may be obtained by a sequential testing procedure as actually employed in the literature and formalized in Supplemental Appendix C. In this regard, we stress that the KP test may be utilized and is recommended as it is tuning parameter free and does not require additional simulations.

Below we provide an implementation guide for testing (1) at significance level α.

<u>Step 1:</u> Compute a singular value decomposition $\hat{\Pi}_n = \hat{P}_n \hat{\Sigma}_n \hat{Q}_n^\intercal$.
<u>Step 2:</u> Obtain $\hat{r}_n$ as in (8) by specifying a tuning parameter κ_n (e.g. $\kappa_n = n^{-1/4}$).
<u>Step 3:</u> Draw B bootstrap samples to compute copies of $\hat{\mathcal{M}}_n^*$, denoted $\{\hat{\mathcal{M}}_{n,b}^*\}_{b=1}^B$.
<u>Step 4:</u> For $\hat{P}_{2,n}$ and $\hat{Q}_{2,n}$ formed by the last $(m - \hat{r}_n)$ and $(k - \hat{r}_n)$ columns of $\hat{P}_n$ and $\hat{Q}_n$ respectively, obtain $\hat{c}_{n,1-\alpha}$ as the $\lfloor B(1-\alpha) \rfloor$-th largest number in

$$\sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^\intercal \hat{\mathcal{M}}_{n,1}^* \hat{Q}_{2,n}) \ , \ \ldots \ , \ \sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^\intercal \hat{\mathcal{M}}_{n,B}^* \hat{Q}_{2,n}) \ . \tag{10}$$

<u>Step 5:</u> Reject H_0 if $\tau_n^2 \sum_{j=r+1}^k \sigma_j^2(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}$.

The above algorithm implements a version of our test, referred to as CF-A in Section 4. Simulation evidence suggests that it is somewhat insensitive to the choice of κ_n even in small samples. The two-step test, on the other hand, is overall the least sensitive, but may be over-sized in small samples ($n \leq 100$). We thus recommend the latter when the sample size is reasonably large. To implement it, one replaces STEPS 2 and 5 with

---

STEP 2': Obtain $\hat{r}_n$ by sequentially testing $\text{rank}(\Pi_0) = 0, 1, \ldots, k-1$ at level β (e.g., $\beta = \alpha/10$) using the KP test (based on $\hat{\Pi}_n$), i.e., $\hat{r}_n = j^*$ if accepting $\text{rank}(\Pi_0) = j^*$ is the first acceptance in the procedure, and $\hat{r}_n = k$ if all nulls are rejected. Reject H_0 if $\hat{r}_n > r$ and move on to Step 3 otherwise.
STEP 5': Reject H_0 if $\tau_n^2 \sum_{j=r+1}^k \sigma_j^2(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}$.

---

3 The Inferential Framework

In this section, we develop our inferential framework in three steps. First, we derive the differential properties of the map ϕ_r given in (3), which is nontrivial and the key to our theory. Second, given an estimator $\hat{\Pi}_n$ of Π_0, we derive the asymptotic distributions for the plug-in estimator $\phi_r(\hat{\Pi}_n)$ by invoking the Delta method. These limits turn out to be highly nonstandard whenever $\text{rank}(\Pi_0) < r$. Thus, in the third step, we construct valid and powerful rank tests by appealing to recent advances on bootstrap in irregular problems (Fang and Santos, 2018; Chen and Fang, 2018; Hong and Li, 2018). A two-step test is proposed to mitigate potential concerns on sensitivity of our tests to the choices of tuning parameters. Local properties of our tests will also be discussed.

3.1 Differential Properties

Let $\Pi_0 \in \mathbf{M}^{m\times k}$ be an unknown matrix with $m \geq k$ and $\sigma_1(\Pi_0) \geq \cdots \geq \sigma_k(\Pi_0) \geq 0$ be singular values of Π_0. Then the rank of Π_0 is equal to the number of nonzero singular values of Π_0 – see, for example, Bhatia (1997, p.5) and also Supplemental Appendix for a brief review. Hence, the hypotheses in (1) are equivalent to

$$\text{H}_0 : \phi_r(\Pi_0) = 0 \qquad \text{v.s.} \qquad \text{H}_1 : \phi_r(\Pi_0) > 0 \ , \tag{11}$$

where $\phi_r : \mathbf{M}^{m\times k} \to \mathbf{R}$ is given by

$$\phi_r(\Pi) \equiv \sum_{j=r+1}^{k} \sigma_j^2(\Pi) \ . \tag{12}$$

Heuristically, $\phi_r(\Pi)$ simply gives us the sum of the $k - r$ smallest squared singular values of Π. One may also consider other L_p-type functionals such as $\sum_{j=r+1}^k \sigma_j(\Pi)$. Our current focus, however, allows us to uncover χ^2-type limiting distributions when $\text{rank}(\Pi_0) = r$ and in this way facilitates comparisons with existing rank tests.

Towards deriving the asymptotic distributions of the plug-in estimator $\phi_r(\hat{\Pi}_n)$ for a given estimator $\hat{\Pi}_n$ of Π_0, we need to first establish suitable differentiability for the map ϕ_r. The following lemma shall prove useful in this regard.

Lemma 3.1. *For the map ϕ_r in* (12)*, we have:*

$$\phi_r(\Pi) = \min_{U \in \mathbb{S}^{k\times(k-r)}} \|\Pi U\|^2 \ . \tag{13}$$

Lemma 3.1 shows that $\phi_r(\Pi)$ can be represented as the minimum of a quadratic form over the space of orthonormal matrices in $\mathbf{M}^{m\times(k-r)}$. The special case when $r = k-1$ (corresponding to the test of Π having full rank) is a well known implication of the classical Courant-Fischer theorem, i.e., $\sigma_k^2(\Pi) = \min_{\|U\|=1} \|\Pi U\|^2$. Note that the minimum in (13) is attained and hence well defined. It turns out that ϕ_r is not fully differentiable in general but belongs to a class of directionally differentiable maps. For completeness, we next introduce the relevant notions of directional differentiability.

Definition 3.1. Let $\phi : \mathbf{M}^{m\times k} \to \mathbf{R}$ be a generic function.

(i) The map ϕ is said to be *Hadamard directionally differentiable* at $\Pi \in \mathbf{M}^{m\times k}$ if there is a map $\phi'_\Pi : \mathbf{M}^{m\times k} \to \mathbf{R}$ such that:

$$\lim_{n\to\infty} \frac{\phi(\Pi + t_n M_n) - \phi(\Pi)}{t_n} = \phi'_\Pi(M) \ , \tag{14}$$

whenever $M_n \to M$ in $\mathbf{M}^{m\times k}$ and $t_n \downarrow 0$ for $\{t_n\}$ all strictly positive.

(ii) If $\phi : \mathbf{M}^{m\times k} \to \mathbf{R}$ is Hadamard directionally differentiable at $\Pi \in \mathbf{M}^{m\times k}$, then we say that ϕ is *second order Hadamard directionally differentiable* at $\Pi \in \mathbf{M}^{m\times k}$ if there is a map $\phi''_\Pi : \mathbf{M}^{m\times k} \to \mathbf{R}$ such that:

$$\lim_{n\to\infty} \frac{\phi(\Pi + t_n M_n) - \phi(\Pi) - t_n\phi'_\Pi(M_n)}{t_n^2} = \phi''_\Pi(M) \ , \tag{15}$$

whenever $M_n \to M$ in $\mathbf{M}^{m\times k}$ and $t_n \downarrow 0$ for $\{t_n\}$ all strictly positive.

Definition 3.1(i) generalizes Hadamard (full) differentiability which additionally requires the derivative ϕ'_Π to be linear. By Proposition 2.1 in Fang and Santos (2018), linearity is precisely the gap between these two notions of differentiability – see also Shapiro (1990) for more discussions. Despite the relaxation, the Delta method remains valid even when ϕ is only Hadamard directionally differentiable (Shapiro, 1991; Dümbgen, 1993). Unfortunately, as shall be proved, the asymptotic distributions of our statistic $\phi(\hat{\Pi}_n)$ implied by this generalized Delta method are degenerate under the null. In turn, Definition 3.1(ii) formulates a suitable second order analog of the directional differentiability, which permits us to obtain nondegenerate asymptotic distributions by a (generalized) second order Delta method (Shapiro, 2000; Chen and Fang, 2018). The second order Hadamard directional differentiability becomes second order Hadamard full differentiability precisely when ϕ''_{Π} corresponds to a bilinear form.

The following proposition formally establishes the differentiability of ϕ_r.

Proposition 3.1. *Let $\phi_r : \mathbf{M}^{m\times k} \to \mathbf{R}$ be defined as in* (12)*.*

(i) ϕ_r is first order Hadamard directionally differentiable at any $\Pi \in \mathbf{M}^{m\times k}$ with the derivative $\phi'_{r,\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ given by

$$\phi'_{r,\Pi}(M) = \min_{U\in\Psi(\Pi)} 2\mathrm{tr}\big((\Pi U)^\intercal M U\big) \ , \tag{16}$$

where $\Psi(\Pi) \equiv \arg\min_{U\in\mathbb{S}^{k\times(k-r)}} \|\Pi U\|^2$.

(ii) ϕ_r is second order Hadamard directionally differentiable at any $\Pi \in \mathbf{M}^{m\times k}$ satisfying $\phi_r(\Pi) = 0$ with the derivative $\phi''_{r,\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ given by: for $r_0 \equiv \mathrm{rank}(\Pi)$,

$$\phi''_{r,\Pi}(M) = \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_2^\intercal M Q_2) \ , \tag{17}$$

where the columns of $P_2 \in \mathbb{S}^{m\times(m-r_0)}$ and $Q_2 \in \mathbb{S}^{k\times(k-r_0)}$ are left and right singular vectors associated with the zero singular values of Π.

Proposition 3.1(i) shows that ϕ_r is not fully differentiable in general but only directionally differentiable in the Hadamard sense. Moreover, the first order derivative is degenerate at zero whenever $\phi_r(\Pi) = 0$ as in this case $\Pi U = 0$ for any $U \in \Psi(\Pi)$. Proposition 3.1(ii) indicates that ϕ_r is second order Hadamard directionally differentiable whenever the degeneracy occurs, and, interestingly, the derivative evaluated at M is simply the sum of the $k-r$ smallest squared singular values of the $(m-r_0)\times(k-r_0)$ matrix $P_2^\intercal M Q_2$. In general, ϕ_r is not second order Hadamard fully differentiable precisely when $\mathrm{rank}(\Pi) < r$, reflecting a critical irregular nature of our setup – see Lemma D.5 for more details. To gain further intuition, suppose that $\Pi_0 = \mathrm{diag}(\pi_{0,1}, \pi_{0,2})$ and we want to test if $\mathrm{rank}(\Pi_0) \leq 1$. Then by definition

$$\phi_r(\Pi_0) = \min\{\pi_{0,1}^2, \pi_{0,2}^2\} \ . \tag{18}$$

Note that if $\mathrm{rank}(\Pi_0) \leq 1$, then $\pi_{0,1}^2 = \pi_{0,2}^2$ if and only if $\mathrm{rank}(\Pi_0) < 1$ in which case $\pi_{0,1} = \pi_{0,2} = 0$. Hence, ϕ_r is not second order differentiable at Π_0 if and only if $\mathrm{rank}(\Pi_0) < 1$ as the map $(\pi_1, \pi_2) \mapsto \min\{\pi_1, \pi_2\}$ is not differentiable precisely when $\pi_1 = \pi_2$. In any case, fortunately, ϕ_r is second order Hadamard directionally differentiable, which is sufficient to invoke the second order Delta method as we elaborate next.

3.2 The Asymptotic Distributions

With the differentiability established in Proposition 3.1, we now derive the asymptotic distributions for the plug-in statistic $\phi_r(\hat{\Pi}_n)$ where $\hat{\Pi}_n$ is a generic estimator of Π_0. This is achieved by appealing to a generalized Delta method for second order Hadamard directionally differentiable maps (Shapiro, 2000; Chen and Fang, 2018). Towards this end, we impose the following assumption.

Assumption 3.1. *There is an estimator $\hat{\Pi}_n : \{X_i\}_{i=1}^n \to \mathbf{M}^{m\times k}$ of $\Pi_0 \in \mathbf{M}^{m\times k}$ (with $m \geq k$) satisfying $\tau_n\{\hat{\Pi}_n - \Pi_0\} \xrightarrow{L} \mathcal{M}$ for some $\tau_n \uparrow \infty$ and random matrix $\mathcal{M} \in \mathbf{M}^{m\times k}$.*

Assumption 3.1 simply requires an estimator $\hat{\Pi}_n$ of Π_0 that admits an asymptotic distribution. Note that the data need not be i.i.d., τ_n may be non-$\sqrt{n}$ and $\mathcal{M}$ can be non-Gaussian, which is important in, for example, nonstationary time series settings. Moreover, as in Robin and Smith (2000) but in contrast to Cragg and Donald (1997), the covariance matrix of $\mathrm{vec}(\mathcal{M})$ is not required to be nonsingular. Assumption 3.1 can be relaxed to accommodate settings where convergence rates across entries of $\hat{\Pi}_n$ are not homogeneous, as in cointegratoin settings – see Supplemental Appendix E.1. For ease of exposition, however, we stick to Assumption 3.1 in the main text.

Given Proposition 3.1 and Assumption 3.1, the following theorem delivers the asymptotic distributions of $\phi_r(\hat{\Pi}_n)$ by the Delta method.

Theorem 3.1. *If Assumption 3.1 holds, then we have, for any $\Pi_0 \in \mathbf{M}^{m\times k}$,*

$$\tau_n\{\phi_r(\hat{\Pi}_n) - \phi_r(\Pi_0)\} \xrightarrow{L} \min_{U\in\Psi(\Pi_0)} 2\mathrm{tr}(U^\intercal \Pi_0^\intercal \mathcal{M} U) \ . \tag{19}$$

If in addition $r_0 \equiv \mathrm{rank}(\Pi_0) \leq r$, then

$$\tau_n^2 \phi_r(\hat{\Pi}_n) \xrightarrow{L} \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal \mathcal{M} Q_{0,2}) \ , \tag{20}$$

where the columns of $P_{0,2} \in \mathbb{S}^{m\times(m-r_0)}$ and $Q_{0,2} \in \mathbb{S}^{k\times(k-r_0)}$ are respectively the left and the right singular vectors of Π_0 associated with its zero singular values.

Theorem 3.1 implies that, under H_0 (and so $\tau_n\phi_r(\hat{\Pi}_n)$ is degenerate), the statistic $\tau_n^2\phi_r(\hat{\Pi}_n)$ converges in law to a nondegenerate second order limit. Towards constructing critical values, we would then like to estimate the law of the limit. Unfortunately, as shown by Chen and Fang (2018), bootstrapping a nondegenerate second order limit is nontrivial; in particular, standard bootstrap schemes such as the nonparametric bootstrap of Efron (1979) are necessarily inconsistent even if they are consistent for $\mathcal{M}$. This predicament is further intensified by the nondifferentiability nature of the map

ϕ_r (Dümbgen, 1993; Fang and Santos, 2018), which renders the limits in (20) highly nonstandard in general. We shall thus present a consistent bootstrap shortly.

We emphasize that the limit of $\tau_n^2\phi_r(\hat{\Pi}_n)$ in Theorem 3.1 is obtained pointwise in each Π_0 under the *entire* null, regardless of whether the truth $\operatorname{rank}(\Pi_0)$ is strictly less than r or not. To the best of our knowledge, this is the first distributional result for a rank test statistic that accommodates the possibility $\operatorname{rank}(\Pi_0) < r$, at the generality of our setup. In turn, such a result permits us to develop a test that has asymptotic null rejection rates exactly equal to the significance level, and hence is more powerful.

In relating our work to the literature, we note that, if $\tau_n = \sqrt{n}$, then the plug-in statistic $\tau_n^2\phi_r(\hat{\Pi}_n)$ is precisely a Robin-Smith statistic (see (B.3)), while the KP statistic is simply a Wald-type standardization of it. Though standardization can help obtain pivotal asymptotic distributions under $r_0 = r$, this is generally not hopeful whenever $r_0 < r$. Since we shall reply on bootstrap for inference, non-pivotalness creates no problems for us. Perhaps more importantly, one may be better off without standardization because it entails invertibility of the weighting matrix in the limit, which may be hard to justify. One might nonetheless interpret the inverse in the KP statistic as a generalized inverse, but consistency of the inverse does not automatically follow from consistency of the covariance matrix estimator without further conditions (Andrews, 1987).

Finally, the limit of $\tau_n^2\phi_r(\hat{\Pi}_n)$ obtained under H_0 is in fact a weighted sum of independent $\chi^2(1)$ variables if $r_0 = r$ and $\mathcal{M}$ is centered Gaussian, showing consistency of our work with Robin and Smith (2000). To see this, note that

$$\sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal \mathcal{M} Q_{0,2}) = \sum_{j=1}^{k-r} \sigma_j^2(P_{0,2}^\intercal \mathcal{M} Q_{0,2}) \ , \tag{21}$$

which is simply the sum of all squared singular values of the $(m-r)\times(k-r)$ matrix $P_{0,2}^\intercal \mathcal{M} Q_{0,2}$, or equivalently the squared Frobenius norm of $P_{0,2}^\intercal \mathcal{M} Q_{0,2}$ (Bhatia, 1997, p.7). Consequently, the limit in (20) can be rewritten as

$$\operatorname{vec}(P_{0,2}^\intercal \mathcal{M} Q_{0,2})^\intercal \operatorname{vec}(P_{0,2}^\intercal \mathcal{M} Q_{0,2}) = \operatorname{vec}(\mathcal{M})^\intercal (Q_{0,2}\otimes P_{0,2})(Q_{0,2}\otimes P_{0,2})^\intercal \operatorname{vec}(\mathcal{M}) \ , \tag{22}$$

as claimed, where we exploited a property of the vec operator (Hamilton, 1994, Proposition 10.4). Our general limit in (20) characterizes the channels through which the true rank plays its role, and thus highlights the importance of studying the problem (1).

3.3 The Bootstrap Inference

Since asymptotic distributions of our statistic $\tau_n^2\phi_r(\hat{\Pi}_n)$ are not pivotal and highly nonstandard in general, in this section we thus aim to develop a consistent bootstrap. This turns out to be quite challenging due to two complications involved.

First, since under H_0 the first order derivative of ϕ_r is degenerate while second order derivative is not (by Proposition 3.1), $\phi_r(\hat{\Pi}_n^*)$ is necessarily inconsistent even if $\hat{\Pi}_n^*$ is a consistent bootstrap (in a sense defined below) in estimating the law of $\mathcal{M}$ (Chen and Fang, 2018), and this remains true in the conventional setup where $\mathrm{rank}(\Pi_0) = r$. Second, the possibility $\mathrm{rank}(\Pi_0) < r$ makes the map ϕ_r nondifferentiable – see Lemma D.5, and hence further complicates the inference (Dümbgen, 1993; Fang and Santos, 2018). One may resort to the m out of n resampling (Shao, 1994) or subsampling (Politis and Romano, 1994). However, both methods can be viewed as special cases of our general bootstrap procedure, and that, more importantly, such a perspective enables us to improve upon these existing resampling schemes and to analyze the local properties in a unified and transparent way – see Remark 3.1 and Section 3.3.1.

The insight our bootstrap builds on is that the limit $\phi''_{r,\Pi_0}(\mathcal{M})$ in Theorem 3.1 is a composition of two unknown components, namely, the limit $\mathcal{M}$ and the derivative ϕ''_{r,Π_0}. Heuristically, one may therefore obtain a consistent estimator for the law of $\phi''_{r,\Pi_0}(\mathcal{M})$ by composing a consistent bootstrap $\hat{\mathcal{M}}_n^*$ for $\mathcal{M}$ with an estimator $\hat{\phi}''_{r,n}$ of ϕ''_{r,Π_0} that is suitably "consistent." This is precisely the bootstrap initially proposed in Fang and Santos (2018) and further developed in Chen and Fang (2018) and Hong and Li (2018). In what follows, we thus commence by estimating the two components separately.

Starting with $\mathcal{M}$, we note that the law of $\mathcal{M}$ may be estimated by standard bootstrap or variants of it that suit particular settings. To formalize the notion of bootstrap consistency, we employ the bounded Lipschitz metric (van der Vaart and Wellner, 1996) and consider estimating the law of a general random element $\mathbb{G}$ in a normed space $\mathbb{D}$ with norm $\|\cdot\|_{\mathbb{D}}$ – the space $\mathbb{D}$ is either $\mathbf{M}^{m\times k}$ or $\mathbf{R}$ in this paper. Let $\mathbb{G}_n^* : \{X_i, W_{ni}\}_{i=1}^n \to \mathbb{D}$ be a generic bootstrap estimator where $\{W_{ni}\}_{i=1}^n$ are bootstrap weights independent of the data $\{X_i\}_{i=1}^n$. Then we say that the conditional law of $\mathbb{G}_n^*$ given the data is consistent for the law of $\mathbb{G}$, or simply $\mathbb{G}_n^*$ is a consistent bootstrap for $\mathbb{G}$, if

$$\sup_{f\in \mathrm{BL}_1(\mathbb{D})} \big|E_W[f(\mathbb{G}_n^*)] - E[f(\mathbb{G})]\big| = o_p(1) , \tag{23}$$

where E_W denotes expectation with respect to $\{W_{ni}\}_{i=1}^n$ holding $\{X_i\}_{i=1}^n$ fixed, and

$$\mathrm{BL}_1(\mathbb{D}) \equiv \{f : \mathbb{D} \to \mathbf{R} : \sup_{x\in\mathbb{D}} |f(x)| < \infty, |f(x) - f(y)| \le \|x - y\|_{\mathbb{D}}\, \forall\, x, y \in \mathbb{D}\} . \tag{24}$$

Given the metric, we now proceed by imposing

Assumption 3.2. *(i) $\hat{\mathcal{M}}_n^* : \{X_i, W_{ni}\}_{i=1}^n \to \mathbf{M}^{m\times k}$ is a bootstrap estimator with $\{W_{ni}\}_{i=1}^n$ independent of $\{X_i\}_{i=1}^n$; (ii) $\hat{\mathcal{M}}_n^*$ is a consistent bootstrap for $\mathcal{M}$.*

Assumption 3.2(i) introduces the bootstrap estimator $\hat{\mathcal{M}}_n^*$, which may be constructed from nonparametric bootstrap, multiplier bootstrap, general exchangeable bootstrap, block bootstrap, score bootstrap, the m out of n resampling or subsampling. The presence of $\{W_{ni}\}_{i=1}^n$ simply characterizes the bootstrap randomness given the data – see Praestgaard and Wellner (1993). For $\hat{\Pi}_n^*$ a bootstrap analog of $\hat{\Pi}_n$, it is common to have $\hat{\mathcal{M}}_n^* = \tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\}$; if $\hat{\Pi}_{m_n}^*$ is an analog of $\hat{\Pi}_n$ constructed based on a subsample of size m_n, then one may instead have $\hat{\mathcal{M}}_n^* = \tau_{m_n}\{\hat{\Pi}_{m_n}^* - \hat{\Pi}_n\}$. Assumption 3.2(ii) requires that $\hat{\mathcal{M}}_n^*$ be consistent in estimating the law of the target limit $\mathcal{M}$.

Turning to the estimation of ϕ''_{r,Π_0}, we recall by Chen and Fang (2018) that, given Assumption 3.2, the composition $\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*)$ is a consistent bootstrap for $\phi''_{r,\Pi_0}(\mathcal{M})$ provided $\hat{\phi}''_{r,n}$ is consistent for ϕ''_{r,Π_0} in the sense that, whenever $M_n \to M$ as $n \to \infty$,

$$\hat{\phi}''_{r,n}(M_n) \xrightarrow{p} \phi''_{r,\Pi_0}(M) \ . \tag{25}$$

In this regard, there are two general constructions, namely, the numerical estimator and the analytic estimator, as we elaborate next.

The numerical estimator is simply a finite sample analog of (15) in the definition of second order derivative, i.e., we estimate ϕ''_{r,Π_0} by: for any $M \in \mathbf{M}^{m\times k}$,

$$\hat{\phi}''_{r,n}(M) = \frac{\phi_r(\hat{\Pi}_n + \kappa_n M) - \phi_r(\hat{\Pi}_n)}{\kappa_n^2} \ , \tag{26}$$

for a suitable $\kappa_n \downarrow 0$, where we have exploited $\phi'_{r,\Pi_0} = 0$ under the null. By Chen and Fang (2018), (26) meets the requirement (25) if $\kappa_n \downarrow 0$ and $\tau_n\kappa_n \to \infty$. Numerical differentiation in the general context of the Delta method dates back to Dümbgen (1993), and is recently extended by Hong and Li (2018). The numerical estimator enjoys marked simplicity and wide applicability, because it merely requires a sequence $\{\kappa_n\}$ of step sizes satisfying certain rate conditions. There is, however, no general theory to date guiding the choice of κ_n, a problem that appears challenging (Hong and Li, 2018). In this regard, it may be sensible to employ the analytic estimator instead.

The analytic estimator heavily exploits the analytic structure of the derivative ϕ''_{r,Π_0}, which, by Proposition 3.1(ii), involves three unknown objects, namely, the true rank r_0, $P_{0,2}$ and $Q_{0,2}$ – note that the columns of $P_{0,2}$ and $Q_{0,2}$ are the left and the right singular vectors associated with the zero singular values of Π_0. We may thus estimate ϕ''_{r,Π_0} by replacing these unknowns with their estimated counterparts. First, consistent estimation of ϕ''_{r,Π_0} entails consistent (under H_0) estimation of r_0. One possible construction is

$$\hat{r}_n = \max\{j = 1, \ldots, r : \sigma_j(\hat{\Pi}_n) \geq \kappa_n\} \tag{27}$$

if the set is nonempty and $\hat{r}_n = 0$ otherwise, where $\kappa_n \downarrow 0$ and $\tau_n\kappa_n \to \infty$ as $n \to \infty$. Heuristically, κ_n may be thought as testing which population singular values are zeros. In turn, for $\hat{\Pi}_n = \hat{P}_n\hat{\Sigma}_n\hat{Q}_n^\intercal$ a singular value decomposition of $\hat{\Pi}_n$, where $\hat{P}_n \in \mathbb{S}^{m\times m}$ and $\hat{Q}_n \in \mathbb{S}^{k\times k}$, and $\hat{\Sigma}_n \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order, we may then partition $\hat{P}_n = [\hat{P}_{1,n}, \hat{P}_{2,n}]$ and $\hat{Q}_n = [\hat{Q}_{1,n}, \hat{Q}_{2,n}]$ with $\hat{P}_{1,n} \in \mathbf{M}^{m\times\hat{r}_n}$ and $\hat{Q}_{1,n} \in \mathbf{M}^{k\times\hat{r}_n}$, and estimate $P_{0,2}$ and $Q_{0,2}$ by $\hat{P}_{2,n}$ and $\hat{Q}_{2,n}$ respectively. By Lemma D.6, we thus obtain a consistent estimator for ϕ''_{r,Π_0}: for any $M \in \mathbf{M}^{m\times k}$,

$$\hat{\phi}''_{r,n}(M) = \sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n}) \ . \tag{28}$$

Similar to the numerical estimator, the analytic estimator (28) also depends on a tuning parameter, but now through consistent estimation of the rank. An advantage of the latter over the former is that the choice of the tuning parameter is easier to motivate. For example, if $\hat{r}_n$ is given by (27), then κ_n has a meaningful interpretation, namely, it measures the parsimoniousness in selecting the rank. Alternatively, a consistent estimator $\hat{r}_n$ may also be obtained by sequential testing, and the tuning parameter then becomes an adjusted significance level – see Supplemental Appendix C.

Given our bootstrap construction $\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*)$, we are now in a position to describe our test for (1). Specifically, at a significance level α, we reject H_0 whenever $\tau_n^2\phi_r(\hat{\Pi}_n)$ is larger than the $(1-\alpha)$th conditional quantile $\hat{c}_{n,1-\alpha}$ of $\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*)$, i.e.,

$$\hat{c}_{n,1-\alpha} \equiv \inf\{c \in \mathbf{R} : P_W(\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*) \leq c) \geq 1-\alpha\} \ , \tag{29}$$

where P_W denotes the probability evaluated with respect to $\{W_{ni}\}_{i=1}^n$ holding the data fixed. In practice, we often approximate $\hat{c}_{n,1-\alpha}$ using the following algorithm:

STEP 1: Compute the derivative estimator $\hat{\phi}''_{r,n}$ according to either (26) or (27)-(28).

STEP 2: Generate B realizations $\{\hat{\mathcal{M}}_{n,b}^*\}_{b=1}^B$ of $\hat{\mathcal{M}}_n^*$ based on B bootstrap samples.

STEP 3: Approximate $\hat{c}_{n,1-\alpha}$ by the $\lfloor B(1-\alpha)\rfloor$ largest number in $\{\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_{n,b}^*)\}_{b=1}^B$.

Our simulations suggest that the analytic method tends to enjoy better size control.

The following theorem establishes that our test has pointwise *exact* asymptotic size control under the entire null H_0, and is consistent against any fixed alternatives.

Theorem 3.2. *Let Assumptions 3.1 and 3.2 hold, and $\hat{c}_{n,1-\alpha}$ be as in* (29) *where $\hat{\phi}''_{r,n}$ is given by either* (26) *with $\{\kappa_n\}$ satisfying $\kappa_n \downarrow 0$ and $\tau_n\kappa_n \to \infty$, or* (28) *with $\hat{r}_n \xrightarrow{p} r_0$ under H_0. If the cdf of the limiting distribution in* (20) *is continuous and strictly increasing at its $(1-\alpha)$-quantile for $\alpha \in (0,1)$, then under H_0,*

$$\lim_{n\to\infty} P(\tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}) = \alpha \ .$$

Furthermore, under H_1,

$$\lim_{n\to\infty} P(\tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}) = 1 \ .$$

Theorem 3.2 shows that our test is not conservative in the pointwise sense while accommodating the possibility $\text{rank}(\Pi_0) < r$. This roots in the simple fact that our critical values are constructed for the pointwise distributions obtained under H_0. By the same token, the power is nontrivial and tends to one against any fixed alternative. We shall further examine the local power properties in Section 3.3.1 and provide numerical evidences in Section 4. Overall, the theoretical and numerical results manifest superiority of our test in terms of size control and power performance.

These attractive testing properties enjoy the elegance of the Delta method. The assumptions are mild, mainly requiring only a matrix estimator that converges in distribution and a consistent bootstrap analog. In addition to the features that the data may be non-i.i.d. and non-stationary, the convergence rate may be non-$\sqrt{n}$, and the limit $\mathcal{M}$ may be non-Gaussian, we stress that the bootstrap for $\mathcal{M}$ may be virtually any consistent resampling scheme, and that no side rank conditions whatsoever are directly imposed beyond those entailed by the restriction that the limiting cdf is continuous and strictly increasing at $c_{1-\alpha}$. Such a quantile restriction is standard as consistent estimation of the limiting laws does not guarantee consistency of critical values – see, for example, Lemma 11.2.1 in Lehmann and Romano (2005). To appreciate how weak this condition is, consider the conventional setup (2) when $\mathcal{M}$ is Gaussian. Then each limit under H_0' is a weighted sum of independent $\chi^2(1)$ random variables – see our discussions towards the end of Section 3.2. Consequently, the quantile condition is automatically satisfied provided the covariance matrix of $\text{vec}(P_{0,2}^{\intercal}\mathcal{M}Q_{0,2})$ is nonzero (i.e., nonzero rank), which is precisely Assumption 2.4 in Robin and Smith (2000). In contrast, Kleibergen and Paap (2006) require nonsingularity of the same matrix (i.e., full rank).

Despite the irregular natures of the problem, computation of our testing statistic and the critical values are quite simple as both involve only calculations of singular value decompositions, for which there are commands in common computation softwares. In particular, $\hat{c}_{n,1-\alpha}$ in practice is set to be the $(1-\alpha)$-quantile of

$$\hat{\phi}''_{r,n}(\hat{\mathcal{M}}^*_{n,1}), \hat{\phi}''_{r,n}(\hat{\mathcal{M}}^*_{n,2}), \ldots, \hat{\phi}''_{r,n}(\hat{\mathcal{M}}^*_{n,B}) \ . \tag{30}$$

Therefore, in each repetition, the numerical and the analytic approaches simply entail singular value decompositions of $\hat{\Pi}_n + \kappa_n\hat{\mathcal{M}}^*_{n,b}$ and $\hat{P}_{2,n}^{\intercal}\hat{\mathcal{M}}^*_{n,b}\hat{Q}_{2,n}$ respectively.

A common feature of our previous two tests is their dependence on a tuning parameter – see (26) and (28). To mitigate concerns on sensitivity to the choice of tuning parameters, we next develop a two-step test by exploiting the structure in (28). The intuition is as follows. The estimator (27), though consistent, may differ from the truth in finite samples. We would thus like to control $P(\hat{r}_n = r_0)$, for which (27) may not be appropriate as it appears challenging to bound $P(\hat{r}_n = r_0)$. Instead, we may obtain a suitable estimator $\hat{r}_n$ by a sequential testing procedure – see Theorem C.1. Specifically, we sequentially test $\text{rank}(\Pi_0) = 0, 1, \ldots, k-1$ at level $\beta < \alpha$, and set $\hat{r}_n = j^*$ if accepting $\text{rank}(\Pi_0) = j^*$ is the first acceptance, and $\hat{r}_n = k$ if no acceptance occurs. In this regard, we recommend the KP test as it is tuning parameter free and does not require additional simulations.[2] Table 1 compares the empirical probabilities of $\{\hat{r}_n = r_0\}$ for $\hat{r}_n$ obtained by (27) and the sequential KP test respectively, based on the same simulation data from Section 2 when $d > 1$. The empirical probabilities for (27) are close to one when $\kappa_n = n^{-1/4}$ (as chosen in Section 2) or $\in \{n^{-1/4}, 1.5n^{-1/4}, n^{-1/5}, 1.5n^{-1/5}\}$ (omitted due to space limitation), but may be far away from one or even close to zero for other choices. On the other hand, the sequential approach leads to rank estimators with empirical probabilities approximately $1-\beta$ across our choices of β.

Table 1. Estimation of $\text{rank}(\Pi_0)$ Defined by the Model (5)-(6)

d	Choices of κ_n in (27)					Choices of β for the Sequential Method					
	$n^{-1/4}$	$n^{-1/3}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	$\alpha/5$	$\alpha/10$	$\alpha/15$	$\alpha/20$	$\alpha/25$	$\alpha/30$
2	1.0000	0.9975	1.0000	0.6679	0.9618	0.9902	0.9947	0.9965	0.9974	0.9975	0.9979
3	1.0000	0.8516	0.9988	0.2246	0.7862	0.9908	0.9951	0.9958	0.9963	0.9976	0.9980
4	0.9995	0.5550	0.9922	0.0249	0.4474	0.9877	0.9949	0.9963	0.9972	0.9976	0.9981
5	0.9977	0.2176	0.9581	0.0003	0.1420	0.9861	0.9933	0.9958	0.9968	0.9976	0.9979
6	0.9899	0.0422	0.8557	0.0000	0.0203	0.9840	0.9916	0.9946	0.9960	0.9967	0.9967

Given an estimator $\hat{r}_n$ with $P(\hat{r}_n = r_0) \geq 1-\beta$ (approximately) for some $\beta < \alpha$, the two-step test now goes as follows. In the first step, we reject H_0 if $\hat{r}_n > r$; otherwise we plug $\hat{r}_n$ into (28) in the second step and reject H_0 if $\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}$. Note that the significance level in the second step is adjusted to be $\alpha - \beta$ in order to take into account the event of false selection (which has probability β). Formally, letting

$$\psi_n = 1\{\hat{r}_n > r \text{ or } \tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}\} , \tag{31}$$

we then reject the null H_0 if $\psi_n = 1$ and fail to reject otherwise. Our next theorem shows that the two-step procedure controls size and is consistent.

Theorem 3.3. *Suppose that Assumptions 3.1 and 3.2 hold, and that the cdf of the limit distribution in* (20) *is continuous and strictly increasing at its* $(1-\alpha+\beta)$*-quantile for* $\alpha \in (0,1)$ *and* $\beta \in (0, \alpha)$*. Let* ψ_n *be the test given by* (31)*. Then, under* H_0*,*

$$\limsup_{n\to\infty} E[\psi_n] \leq \alpha$$

provided $\liminf_{n\to\infty} P(\hat{r}_n = r_0) \geq 1-\beta$*, and, under* H_1*,*

$$\lim_{n\to\infty} E[\psi_n] = 1 .$$

[2] If estimation of r_0 is one's *ultimate* goal (rather than an intermediate step for test), then it may be desirable to instead employ our tests in the sequential procedure, as existing tests may lead to estimators that are not as accurate when Π_0 is "local to degeneracy" – see Section 4 for simulation evidences.

The idea of the two-step test may be found in Loh (1985), Berger and Boos (1994), and Silvapulle (1996), and has recently been employed in the context of moment inequality models (Andrews and Barwick, 2012; Romano et al., 2014). A common feature that our test shares here is that the size control is not exact, i.e., we cannot show the size is equal to α. This raises the concern that the test may be potentially conservative. Nonetheless, it is possible to derive a lower bound of the asymptotic size which is close to α by choosing a small β – see Romano et al. (2014) for a similar feature. Summarizing, there are two (types of) test procedures: one rejects H_0 if $\tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}$ with $\hat{c}_{n,1-\alpha}$ computed according to (29), and the other one applies when one has control over $P(\hat{r}_n = r_0)$: if $\liminf_{n\to\infty} P(\hat{r}_n = r_0) \geq 1-\beta$, we rejects if $\hat{r}_n > r$ or $\tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}$. Our simulation results in Section 4 show that the two-step procedure produces results that are quite insensitive to our choice of β.

Remark 3.1. The m out of n bootstrap and the subsampling are special cases of our bootstrap procedure. For example, the former amounts to $\hat{\mathcal{M}}_n^* = \tau_{m_n}\{\hat{\Pi}_{m_n}^* - \hat{\Pi}_n\}$ with $\hat{\Pi}_{m_n}^*$ constructed based on subsamples of size m_n (obtained through resampling with replacement), and the derivative estimator $\hat{\phi}''_{r,n}$ given by (26) with $\kappa_n = m_n^{-1}$. Subsampling is precisely the same procedure except that the subsamples are obtained without replacement. In other words, these procedures estimate the derivative through (26) implicitly and automatically when the subsample size is properly chosen, combining the two-steps into one single step. By disentangling estimation of the two ingredients, however, we may better estimate both the derivative ϕ''_{r,Π_0} (through exploiting the structure of the derivative and a choice of the tuning parameter) and the law of the limit $\mathcal{M}$ (using full samples), which may in turn lead to efficiency improvement. Moreover, such a perspective enables us to establish conditions under which tests based on these resampling schemes have local size control and nontrivial power, properties not guaranteed in general and nontrivial to analyze otherwise (Andrews and Guggenberger, 2010). ■

3.3.1 Local Power Properties

Having established size control and consistency, we next aim to obtain a more precise characterization of the quality of our tests by studying the local power properties (Neyman, 1937). Following Cragg and Donald (1997), we thus proceed by imposing

Assumption 3.1'. *(i)* $\operatorname{rank}(\Pi_{0,n}) > r$ *for all* n*; (ii)* $\tau_n\{\Pi_{0,n} - \Pi_0\} \to \Delta$ *for some* Π_0 *with* $\operatorname{rank}(\Pi_0) \leq r$ *and nonrandom* Δ*; (iii)* $\tau_n\{\hat{\Pi}_n - \Pi_{0,n}\} \xrightarrow{L_n} \mathcal{M}$ *for some* $\tau_n \uparrow \infty$*, where* $\xrightarrow{L_n}$ *denotes convergence in law along distributions of the data associated with* $\{\Pi_{0,n}\}$*.*

Assumption 3.1'(i)(ii) formally defines $\{\Pi_{0,n}\}$ as a sequence of local alternatives that approaches some Π_0 in the null at the convergence rate τ_n, while Assumption 3.1'(iii) formalizes the notion that the *asymptotic* distributions of $\hat{\Pi}_n$ should remain unchanged in response to *small* (finite sample) perturbations of the data generating processes, a property that may be verified through, for example, the framework of limits of statistical experiments (van der Vaart, 1998; Hallin et al., 2016).

Our next result characterizes the asymptotic behaviors of the testing statistic $\tau_n^2\phi_r(\hat{\Pi}_n)$ under local alternatives that satisfy Assumption 3.1'.

Proposition 3.2. *If Assumption 3.1' holds, then it follows that*

$$\tau_n^2\phi_r(\hat{\Pi}_n) \overset{L_n}{\to} \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal(\mathcal{M}+\Delta)Q_{0,2}) \ . \tag{32}$$

Proposition 3.2 includes Theorem 3.1 as a special case with $\Pi_{0,n} = \Pi_0$ for all n so that $\Delta = 0$. The main utility of this result is to analyze the asymptotic local power function. In what follows, we focus on the one-step tests for conciseness and transparency, though analogous results hold for the two-step test ψ_n. Thus, if the local alternatives $\{\Pi_{0,n}\}$ in Assumption 3.1' approach Π_0 in the sense of contiguity (Roussas, 1972; Rothenberg, 1984),[3] then we may obtain a lower bound as follows:

$$\liminf_{n\to\infty} P_n(\tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}) \geq P(\sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal(\mathcal{M}+\Delta)Q_{0,2}) > c_{1-\alpha}) \ , \tag{33}$$

where P_n denotes probability evaluated under $\Pi_{0,n}$. While it appears challenging to prove that the asymptotic local power is nontrivial under arbitrary local alternatives, there are, nonetheless, two interesting classes under which the asymptotic local power can be proven to be nontrivial. The first case is the conventional setup where $\mathrm{rank}(\Pi_0)$ is exactly equal to the hypothesized value r and $\mathcal{M}$ is centered Gaussian. Since the derivative ϕ''_{r,Π_0} then coincides with the squared Frobenius norm – see Proposition 3.1(ii), we have along contiguous local alternatives that

$$\liminf_{n\to\infty} P_n(\tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}) \geq P(\|P_{0,2}^\intercal(\mathcal{M}+\Delta)Q_{0,2}\|^2 > c_{1-\alpha}) \ . \tag{34}$$

An application of Anderson's lemma – see, for example, Lemma 3.11.4 in van der Vaart and Wellner (1996) – then yields the claimed (weak) nontriviality, i.e.,

$$P(\|P_{0,2}^\intercal(\mathcal{M}+\Delta)Q_{0,2}\|^2 > c_{1-\alpha}) \geq P(\|P_{0,2}^\intercal\mathcal{M}Q_{0,2}\|^2 > c_{1-\alpha}) = \alpha \ . \tag{35}$$

If the localization parameter Δ is nontrivial (i.e., $\Delta \neq 0$) and belongs to the support of $\mathcal{M}$ – which is the case, for example, if the covariance matrix of $\mathrm{vec}(\mathcal{M})$ is nonsingular, then by Lemma B.4 in Chen and Santos (2018) (a strengthening of Anderson's lemma),

[3] This means that if (any) T_n is negligible (i.e., of order $o_p(1)$) under Π_0 then it remains so under $\Pi_{0,n}$. Thus, contiguity simply formalizes the notion that the effect of "small" perturbations is negligible.

the asymptotic local lower is in fact strictly nontrivial, i.e.,

$$P(\|P_{0,2}^{\intercal}(\mathcal{M}+\Delta)Q_{0,2}\|^2 > c_{1-\alpha}) > \alpha \ . \tag{36}$$

The other instance in which the asymptotic local power is nontrivial is the irregular setup where potentially $\operatorname{rank}(\Pi_0) < r$ but the local alternatives behave in a special way. Concretely, if the localization parameter Δ is more degenerate than Π_0 in the sense that either the null space of Π_0 is contained in that of Δ or the null space of $\Pi_0^{\intercal}$ is contained in that of $\Delta^{\intercal}$ – recall that the null space of a generic matrix $A \in \mathbf{M}^{m\times k}$ is the set $\{x \in \mathbf{R}^k : Ax = 0\}$, then we must have $P_{0,2}^{\intercal}\Delta Q_{0,2} = 0$ and (35) follows immediately.

In view of the irregularities of the problem (1), one may also be interested in the size control of our test. Under Assumption 3.1' but with (i) replaced by $\operatorname{rank}(\Pi_{0,n}) \leq r$ for all $n \in \mathbf{N}$ so that the contiguous perturbations occur under the null, we may obtain

$$\limsup_{n\to\infty} P_n(\tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}) \leq P(\sum_{j=r-r_0+1}^{k-r_0}\sigma_j^2(P_{0,2}^{\intercal}(\mathcal{M}+\Delta)Q_{0,2}) \geq c_{1-\alpha}) \ . \tag{37}$$

Now suppose $\operatorname{rank}(\Pi_0) = r$ but without requiring $\mathcal{M}$ to be centered nor Gaussian. Since $\phi_r(\Pi_{0,n}) = \phi_r(\Pi_0) = 0$, it follows by Assumption 3.1'(ii) and Proposition 3.1 that

$$0 = \lim_{n\to\infty} \tau_n^2\{\phi_r(\Pi_{0,n}) - \phi_r(\Pi_0)\} = \phi''_{r,\Pi_0}(\Delta) = \|P_{0,2}^{\intercal}\Delta Q_{0,2}\|^2 \ . \tag{38}$$

Hence, we have $P_{0,2}^{\intercal}\Delta Q_{0,2} = 0$, a previously analyzed special case. Consequently,

$$P(\sum_{j=r-r_0+1}^{k-r_0}\sigma_j^2(P_{0,2}^{\intercal}(\mathcal{M}+\Delta)Q_{0,2}) \geq c_{1-\alpha}) = \alpha \ , \tag{39}$$

if the quantile restrictions on $c_{1-\alpha}$ as in Theorem 3.2 hold. Size control under arbitrary local perturbations in H_0, unfortunately, appears (to us) as challenging as establishing nontrivial local power under arbitrary local alternatives. We pose these as open questions, and leave them for future study.

3.3.2 Illustration: Identification in Linear IV Models

We now illustrate how to apply our framework by testing identification in linear IV models due to their simplicity and popularity. Let $(Y, Z^{\intercal})^{\intercal} \in \mathbf{R}^{1+k}$ satisfy:

$$Y = Z^{\intercal}\beta_0 + u \ , \tag{40}$$

where $\beta_0 \in \mathbf{R}^k$ and u is an error term. Let $V \in \mathbf{R}^m$ be an instrument variable with $E[Vu] = 0$ and $m \geq k$. Then global identification of β_0 requires $E[VZ^{\intercal}]$ to be of full rank. Thus, identification of β_0 may be tested by examining (1) with

$$\Pi_0 = E[VZ^\intercal] \text{ and } r = k - 1 \ . \tag{41}$$

The hypotheses in (2) may be restrictive since it is generally unknown if $\text{rank}(\Pi_0) \geq k-1$. Analogous rank conditions also arise for global identification in simultaneous linear equation models (Koopmans and Hood, 1953; Fisher, 1961) and in models with misclassification errors (Hu, 2008), and for local identification in nonlinear/nonparametric models (Rothenberg, 1971; Roehrig, 1988; Chesher, 2003; Matzkin, 2008; Chen et al., 2014) and in DSGE models (Canova and Sala, 2009; Komunjer and Ng, 2011).

To apply our framework, let $\{V_i, Z_i\}_{i=1}^n$ be an i.i.d. sample. Then the estimator

$$\hat{\Pi}_n = \frac{1}{n}\sum_{i=1}^n V_i Z_i^\intercal \tag{42}$$

satisfies Assumption 3.1 for $\tau_n = \sqrt{n}$ and some centered Gaussian matrix $\mathcal{M}$ under suitable moment restrictions. In turn, let $\{Z_i^*, V_i^*\}_{i=1}^n$ be an i.i.d. sample drawn with replacement from $\{Z_i, V_i\}_{i=1}^n$. Then $\hat{\mathcal{M}}_n^* \equiv \sqrt{n}\{\hat{\Pi}_n^* - \hat{\Pi}_n\}$ with $\hat{\Pi}_n^*$ given by

$$\hat{\Pi}_n^* \equiv \frac{1}{n}\sum_{i=1}^n V_i^* Z_i^{*\intercal} = \frac{1}{n}\sum_{i=1}^n W_{ni} V_i Z_i^\intercal \ , \tag{43}$$

where $(W_{n1}, \ldots, W_{nn})$ is multinomial over n categories with probabilities $(n^{-1}, \ldots, n^{-1})$, satisfies Assumption 3.2 – see, for example, Theorem 23.4 in van der Vaart (1998). We have thus verified the main assumptions.

Empirical research, however, is often faced with clustered data. For example, micro-level data often cluster on geographical regions such as cities or states. To illustrate, suppose that there are (large) G clusters, and the gth cluster has observations $\{V_{gi}, Z_{gi}\}_{i=1}^{n_g}$. The data are independent across clusters but may otherwise be correlated within each cluster. Let $n \equiv \sum_{g=1}^G n_g$. In these settings, Π_0 is identified as the probability limit of

$$\hat{\Pi}_n \equiv \frac{1}{n}\sum_{g=1}^G V_g^\intercal Z_g \tag{44}$$

as $G \to \infty$, where $V_g \equiv [V_{g1}, \ldots, V_{gn_g}]^\intercal$ and $Z_g \equiv [Z_{g1}, \ldots, Z_{gn_g}]^\intercal$. Assumption 3.1 holds for $\tau_n = \sqrt{n}$ and some centered Gaussian matrix $\mathcal{M}$, by the Lindeberg-Feller type central limit theorem. Following Cameron et al. (2008), we may construct

$$\hat{\mathcal{M}}_n^* \equiv \frac{1}{n}\sum_{g=1}^G W_g\{V_g^\intercal Z_g - \hat{\Pi}_n\} \ , \tag{45}$$

where $(W_1, \ldots, W_G)$ may be a multinomial vector over G categories with probabilities

$(1/G, \ldots, /1G)$ (corresponding to the pairs cluster bootstrap) or other weights (such as those leading to th cluster wild bootstrap); see also Djogbenou et al. (2018).

For the convenience of practitioners, we next provide an implementation guide of our two-step test at significance level α.

STEP 1: (a) Sequentially test $\text{rank}(\Pi_0) = 0, 1, \ldots, k-1$ at level β (e.g., $\beta = \alpha/10$) based on $\hat{\Pi}_n$ using the KP test and obtain the rank estimator $\hat{r}_n$; (b) Reject H_0 if $\hat{r}_n = k$ and move on to the next step otherwise.

STEP 2: (a) Draw B bootstrap samples by either the empirical bootstrap or the cluster bootstrap depending on if clustering is present, construct $\{\hat{\mathcal{M}}_{n,b}^*\}_{b=1}^B$ accordingly (i.e., as in (43) or (45)), and set $\hat{c}_{1-\alpha+\beta}$ to be the $\lfloor B(1-\alpha+\beta) \rfloor$ largest number in

$$\sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^\intercal \hat{\mathcal{M}}_{n,1}^* \hat{Q}_{2,n}) , \ldots , \sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^\intercal \hat{\mathcal{M}}_{n,B}^* \hat{Q}_{2,n}) , \tag{46}$$

where $\hat{P}_{2,n}$ and $\hat{Q}_{2,n}$ are from the singular value decomposition of $\hat{\Pi}_n$ as before; (b) Reject H_0 if $n\sigma_{\min}^2(\hat{\Pi}_n) > \hat{c}_{1-\alpha+\beta}$ with $\sigma_{\min}(\hat{\Pi}_n)$ the smallest singular value of $\hat{\Pi}_n$.

For our one-step test based on (27)-(28), one may directly proceed with Step 2, but with $\hat{r}_n$ constructed from (27) and reject if $n\sigma_{\min}^2(\hat{\Pi}_n) > \hat{c}_{1-\alpha}$.

4 Simulation Studies

In this section, we examine the finite sample performance of our inferential framework by Monte Carlo simulations. First, we compare our tests with the multiple KP test in more complicated data environments with heteroskedasticity, serial correlation and different sample sizes. We shall pay special attention to the choices of tuning parameters. We refer the reader to Supplemental Appendix B where we provide additional comparisons with Kleibergen and Paap (2006) based on their simulation designs and a real dataset that they use. Second, we also conduct simulations to assess the performance of our rank estimators, obtained by a sequential testing procedure employed in the literature and formalized in Supplemental Appendix C.

We commence by considering the following linear model

$$Z_t = \Pi_0^\intercal V_t + V_{1,t} u_t , \tag{47}$$

where $Z_t \in \mathbf{R}^4$ for all t, $\{V_t\} \stackrel{\text{i.i.d.}}{\sim} N(0, I_4)$ and $\{u_t\}$ are generated according to

$$u_t = \epsilon_t - \frac{1}{4} \mathbf{1}_4 \mathbf{1}_4^\intercal \epsilon_{t-1} \tag{48}$$

with $\{\epsilon_t\} \overset{\text{i.i.d.}}{\sim} N(0, I_4)$ independent of $\{V_t\}$, and $V_{1,t}$ the first entry of V_t. Moreover, we configure Π_0 as: for $\delta \in \{0, 0.1, 0.3, 0.5\}$,

$$\Pi_0 = \text{diag}(\mathbf{1}_2, \mathbf{0}_2) + \delta I_4 \ . \tag{49}$$

We test the hypotheses in (1) for $r \in \{2, 3\}$ at level $\alpha = 5\%$. Thus, for both cases, H_0 is true if and only if $\delta = 0$, and they respectively correspond to $\text{rank}(\Pi_0) = r$ and $\text{rank}(\Pi_0) < r$ under H_0. We estimate Π_0 by $\hat{\Pi}_n = \frac{1}{n}\sum_{t=1}^{n} V_t Z_t^\intercal$ for sample sizes $n \in \{50, 100, 300, 1000\}$, and for each n, the number of simulation replications is set to be 5,000 with 500 bootstrap repetitions for each replication. As the data exhibit first order autocorrelation, we adopt the circular block bootstrap (Politis and Romano, 1992) with block size $b = 2$. To implement the multiple KP test, labelled KP-M, we estimate the variance of $\text{vec}(\hat{\Pi}_n)$ by the HACC estimator with one lag (West, 1997). To carry out our tests, we choose $\kappa_n \in \{n^{-2/5}, 1.5n^{-2/5}, n^{-1/5}, n^{-1/4}, n^{-1/3}, 1.5n^{-1/5}, 1.5n^{-1/4}, 1.5n^{-1/3}\}$ for both the numerical estimator (26) and the analytic estimator (27)-(28), and $\beta \in \{\alpha/5, \alpha/10, \alpha/15, \alpha/20, \alpha/25, \alpha/30\}$ for the two-step test. For ease of reference, we respectively label these three tests as CF-N, CF-A, and CF-T.

Table 2 summarizes the simulation results for tuning parameters in the middle range of the choices, while Tables 3 and 4 collect results for the remaining choices. For the case of $r = 2$ (so that $\text{rank}(\Pi_0) = r$ under H_0), the performance of CF-A and CF-T is comparable with that of KP-M especially when the sample size is large, though CF-T exhibits more size distortion than KP-M for $n = 50$ and CF-N appears to be somewhat sensitive to the choice of κ_n. For the case of $r = 3$ (so that $\text{rank}(\Pi_0) < r$ under H_0), KP-M is markedly under-sized even in large samples, while its local power is uniformly dominated by our three tests, across all the choices of the tuning parameters, sample sizes, and the local parameter δ. With regards to comparisons among our three tests, there are also some persistent patterns. First, CF-N overall tends to be the most over-sized especially in small samples, and the most sensitive to the choice of the tuning parameters. Second, between CF-A and CF-T, one does not seem to dominate the other. The former appears to perform better overall in terms of size control and local power in small samples, though the differences become smaller as the sample size increases. The latter, on the other hand, seems to be the least sensitive to the choice of the tuning parameters especially in the irregular case when $r = 3$, as desired. Thus, it seems sensible to employ CF-A in small samples and CF-T instead in large samples.

We now compare with Kleibergen and Paap (2006) in terms of estimation by making use of the same data generating process as specified by (5) and (6) with $\delta = 0.1$ and 0.12 so that $\text{rank}(\Pi_0) = 6$ (i.e., full rank) in both cases for all $d = 1, \ldots, 6$. Our estimation is based on the analytic derivative estimator (28) with $\hat{r}_n$ given by (27) and $\kappa_n = n^{-1/4}$ – the results for $\kappa_n = n^{-1/3}$ are similar and available upon request. In each configuration, we depict the empirical distributions of the estimators based on 5,000 simulations, 500

Table 2. Rejection rates of rank tests for the model (47) at $\alpha = 5\%^\dagger$

Sample size	CF-T			CF-A			CF-N			KP-M
	$\alpha/10$	$\alpha/15$	$\alpha/20$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	
Rejection rates for $r = 2$										
$\delta = 0$										
50	0.17	0.17	0.17	0.04	0.04	0.04	0.29	0.28	0.21	0.08
100	0.08	0.08	0.08	0.04	0.04	0.04	0.23	0.20	0.12	0.08
300	0.04	0.04	0.04	0.05	0.05	0.05	0.16	0.12	0.05	0.06
1000	0.04	0.04	0.04	0.05	0.05	0.05	0.11	0.08	0.04	0.05
$\delta = 0.1$										
50	0.23	0.23	0.23	0.08	0.08	0.08	0.37	0.35	0.27	0.13
100	0.18	0.17	0.17	0.12	0.12	0.12	0.38	0.34	0.23	0.19
300	0.34	0.34	0.34	0.35	0.35	0.35	0.57	0.51	0.36	0.44
1000	0.89	0.90	0.90	0.90	0.90	0.90	0.95	0.92	0.88	0.92
$\delta = 0.3$										
50	0.67	0.67	0.66	0.48	0.48	0.48	0.80	0.79	0.72	0.40
100	0.85	0.85	0.85	0.80	0.80	0.80	0.95	0.93	0.89	0.77
300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\delta = 0.5$										
50	0.95	0.95	0.94	0.89	0.89	0.89	0.98	0.98	0.97	0.55
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.87
300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Rejection rates for $r = 3$										
$\delta = 0$										
50	0.09	0.09	0.10	0.07	0.06	0.04	0.14	0.14	0.12	0.01
100	0.06	0.07	0.07	0.06	0.06	0.03	0.12	0.12	0.09	0.01
300	0.04	0.05	0.05	0.05	0.05	0.03	0.09	0.08	0.06	0.01
1000	0.05	0.05	0.05	0.06	0.06	0.05	0.08	0.07	0.05	0.00
$\delta = 0.1$										
50	0.12	0.12	0.12	0.10	0.09	0.05	0.18	0.18	0.16	0.01
100	0.12	0.13	0.13	0.13	0.11	0.06	0.21	0.19	0.16	0.02
300	0.25	0.26	0.27	0.32	0.29	0.16	0.38	0.36	0.31	0.09
1000	0.63	0.65	0.67	0.82	0.81	0.59	0.84	0.82	0.77	0.54
$\delta = 0.3$										
50	0.43	0.44	0.45	0.39	0.33	0.25	0.57	0.56	0.52	0.12
100	0.61	0.63	0.64	0.66	0.57	0.50	0.80	0.79	0.74	0.43
300	0.96	0.96	0.96	0.98	0.96	0.96	1.00	0.99	0.99	0.96
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\delta = 0.5$										
50	0.76	0.77	0.78	0.68	0.64	0.63	0.88	0.88	0.84	0.37
100	0.92	0.93	0.93	0.92	0.91	0.91	0.99	0.99	0.98	0.79
300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

† The three values under CF-T are the choices of β, and those under CF-A and CF-N are the choices of κ_n as in 27 and (26) respectively.

Table 3. Additional results on rejection rates of rank tests for the model (47) with $r = 2$, at $\alpha = 5\%^{\dagger}$

Sample	CF-T			CF-A					CF-N					KP-M
size	$\alpha/5$	$\alpha/25$	$\alpha/30$	$1.5n^{-1/5}$	$1.5n^{-1/4}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	$1.5n^{-1/5}$	$1.5n^{-1/4}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	
								$\delta = 0$						
50	0.16	0.17	0.17	0.08	0.05	0.04	0.04	0.04	0.31	0.30	0.29	0.13	0.25	0.08
100	0.08	0.08	0.08	0.04	0.04	0.04	0.04	0.04	0.26	0.24	0.20	0.06	0.15	0.08
300	0.03	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.20	0.18	0.11	0.02	0.06	0.06
1000	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.15	0.12	0.06	0.02	0.03	0.05
								$\delta = 0.1$						
50	0.23	0.23	0.24	0.11	0.09	0.08	0.08	0.08	0.39	0.39	0.36	0.17	0.31	0.13
100	0.17	0.17	0.17	0.12	0.12	0.12	0.12	0.12	0.41	0.40	0.34	0.12	0.26	0.19
300	0.34	0.34	0.34	0.35	0.35	0.35	0.35	0.35	0.63	0.60	0.50	0.21	0.37	0.44
1000	0.88	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.97	0.96	0.91	0.80	0.87	0.92
								$\delta = 0.3$						
50	0.67	0.66	0.66	0.49	0.48	0.48	0.48	0.48	0.81	0.81	0.79	0.61	0.76	0.40
100	0.86	0.84	0.84	0.80	0.80	0.80	0.80	0.80	0.95	0.95	0.94	0.79	0.90	0.77
300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
								$\delta = 0.5$						
50	0.95	0.94	0.94	0.89	0.89	0.89	0.89	0.89	0.98	0.99	0.98	0.93	0.97	0.55
100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.87
300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

† The three values under CF-T are the choices of β, and those under CF-A and CF-N are the choices of κ_n as in 27 and (26) respectively.

Table 4. Additional results on rejection rates of rank tests for the model (47) with $r = 3$, at $\alpha = 5\%^{\dagger}$

Sample	CF-T			CF-A					CF-N					KP-M
size	$\alpha/5$	$\alpha/25$	$\alpha/30$	$1.5n^{-1/5}$	$1.5n^{-1/4}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	$1.5n^{-1/5}$	$1.5n^{-1/4}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	
								$\delta = 0$						
50	0.08	0.10	0.10	0.08	0.08	0.06	0.02	0.05	0.14	0.14	0.14	0.10	0.14	0.01
100	0.05	0.07	0.07	0.06	0.06	0.06	0.01	0.04	0.13	0.13	0.12	0.07	0.10	0.01
300	0.04	0.05	0.05	0.05	0.05	0.05	0.01	0.03	0.10	0.09	0.07	0.04	0.06	0.01
1000	0.04	0.05	0.05	0.06	0.06	0.05	0.01	0.05	0.10	0.08	0.06	0.04	0.05	0.00
								$\delta = 0.1$						
50	0.10	0.13	0.13	0.11	0.17	0.09	0.03	0.07	0.18	0.18	0.18	0.13	0.17	0.01
100	0.10	0.13	0.14	0.13	0.13	0.11	0.04	0.07	0.21	0.21	0.20	0.11	0.17	0.02
300	0.21	0.28	0.28	0.32	0.32	0.28	0.10	0.16	0.41	0.39	0.35	0.23	0.31	0.09
1000	0.58	0.68	0.68	0.88	0.82	0.76	0.54	0.58	0.86	0.84	0.81	0.68	0.77	0.54
								$\delta = 0.3$						
50	0.40	0.45	0.45	0.47	0.44	0.35	0.23	0.28	0.57	0.57	0.56	0.44	0.54	0.12
100	0.56	0.65	0.65	0.75	0.72	0.58	0.49	0.51	0.81	0.80	0.79	0.65	0.76	0.43
300	0.95	0.96	0.96	0.99	0.99	0.96	0.96	0.96	1.00	1.00	0.99	0.97	0.99	0.96
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
								$\delta = 0.5$						
50	0.74	0.78	0.79	0.80	0.74	0.65	0.63	0.63	0.89	0.89	0.88	0.77	0.87	0.37
100	0.91	0.93	0.93	0.96	0.94	0.92	0.91	0.91	0.99	0.99	0.99	0.95	0.98	0.79
300	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

† The three values under CF-T are the choices of β, and those under CF-A and CF-N are the choices of κ_n as in 27 and (26) respectively.

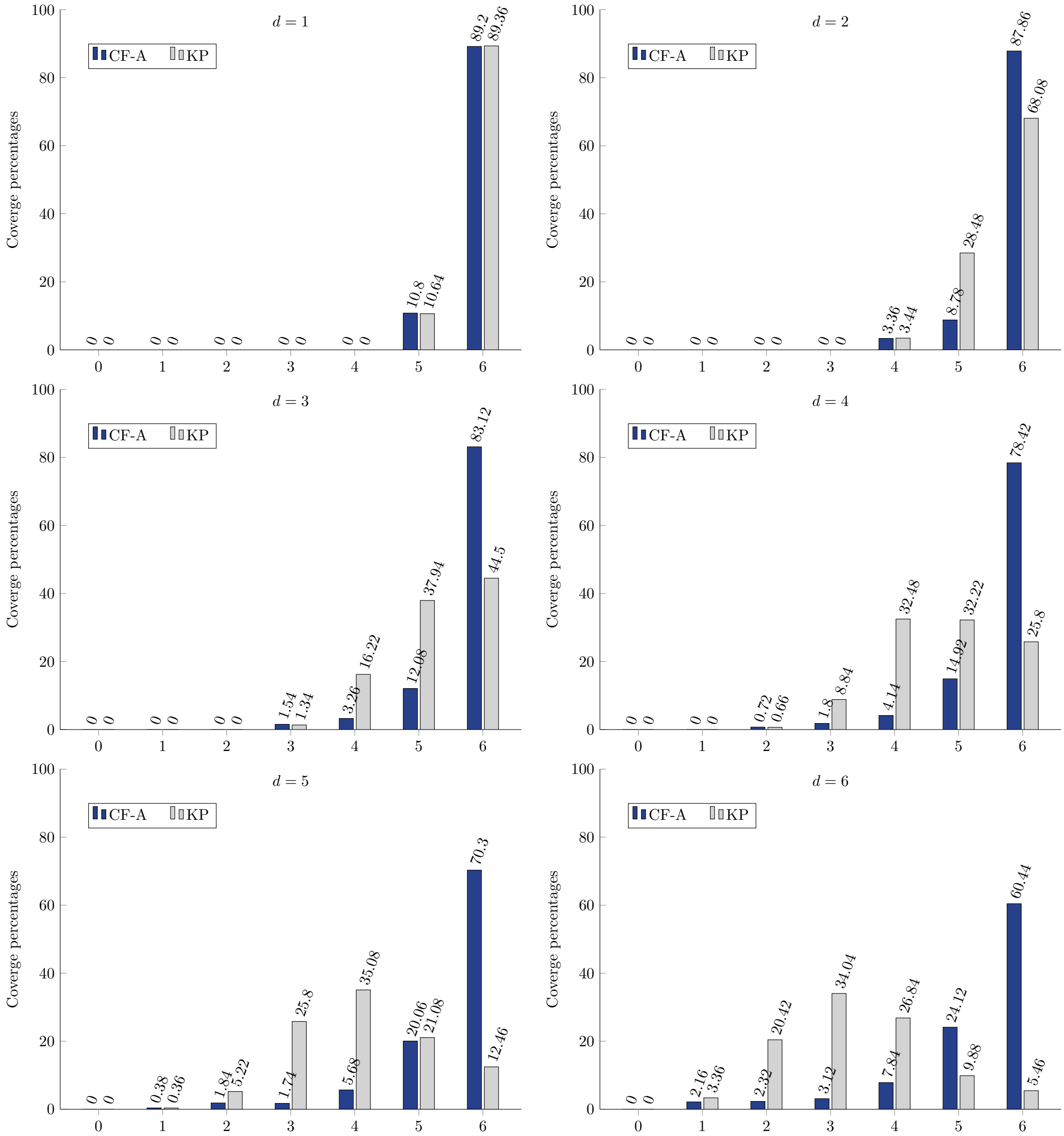

Figure 3. The rank estimation: $\text{rank}(\Pi_0) = 6$, $\alpha = 5\%$ and $\delta = 0.1$

bootstrap repetitions for each simulation, and $\alpha = 5\%$. As shown by Figures 3 and 4, our rank estimators, labelled CF-A, pick up the truth with probabilities higher than the KP estimators, uniformly over $d \in \{2, \ldots, 6\}$ and $\delta \in \{0.1, 0.12\}$; when $d = 1$, the two sets of estimators are very similar. Note that the empirical probabilities of $\hat{r}_n = r_0$ are lower in Figure 3 than in Figure 4 because Π_0 is closer to a lower rank matrix (due to a smaller value of δ), and in each figure, the probabilities for both sets of estimators decrease as Π_0 becomes more degenerate (as d increases). There are two additional interesting persistent patterns. First, the distributions of the KP estimators are more spread out and tend to underestimate the true rank, especially when d is large, i.e., when Π_0 is local to a matrix whose rank is small. This is in accord with the trivial power of the KP test in this scenario – see Figure 2. Second, the probability of our rank

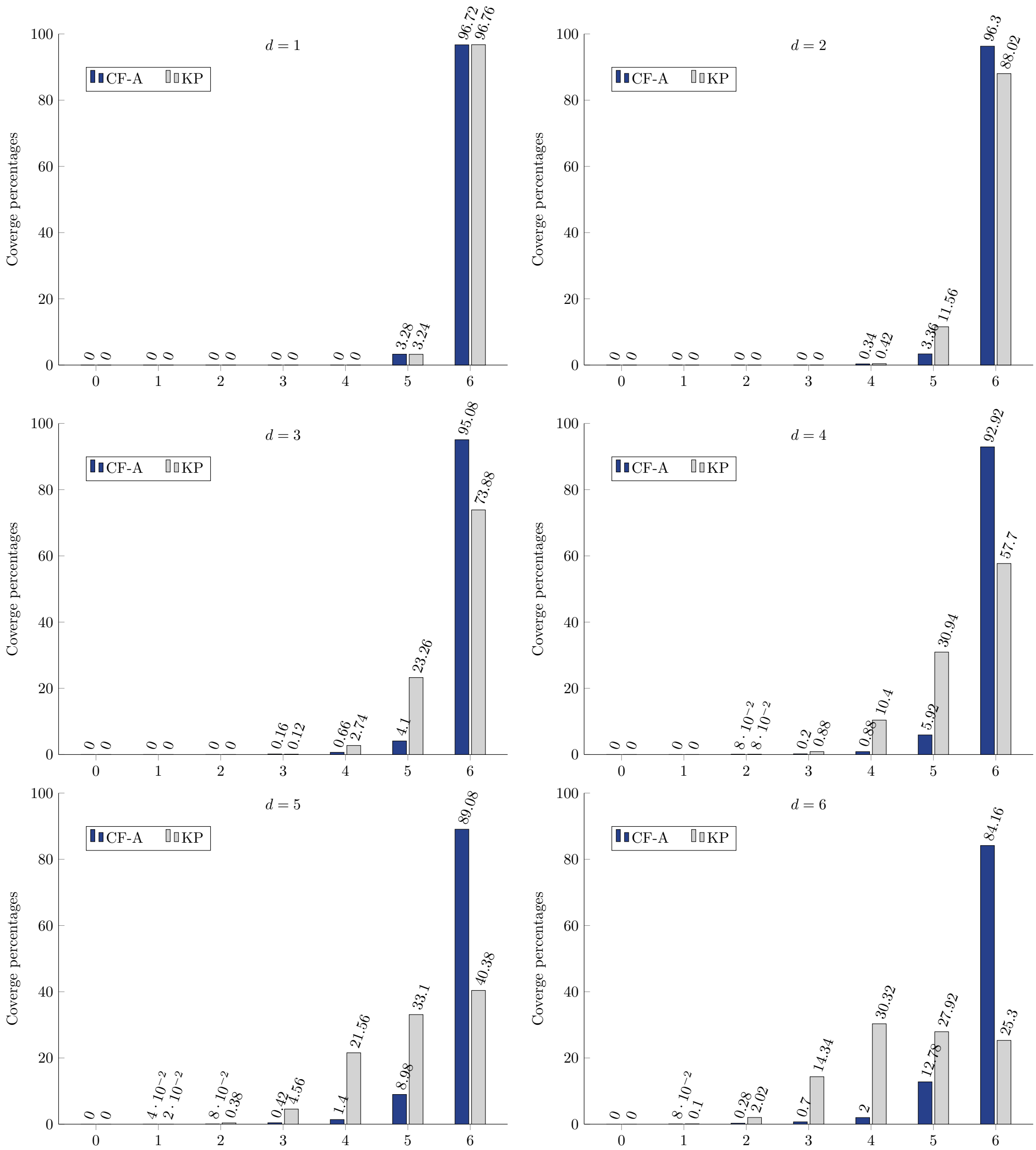

Figure 4. The rank estimation: $\text{rank}(\Pi_0) = 6$, $\alpha = 5\%$ and $\delta = 0.12$

estimators equal to the truth can exceed that of the KP rank estimator by as high as 57.84%, and in 5 out of the 12 data generating processes considered, the probabilities of our rank estimator covering the truth are at least 48.70% higher. Once again, this happens especially when Π_0 is local to a matrix whose rank is small. These observations suggest that our estimators are more robust to local-to-degeneracy.

5 Saliency Analysis in Matching Models

In this section, we study a one-to-one, bipartite matching model with transferable utility, where a central question is how many attributes are statistically relevant for the sorting of agents (Dupuy and Galichon, 2014; Ciscato et al., 2018). As shall be seen shortly, this question can be answered by appealing to our framework developed previously. Following the literature, we shall call the two sets of agents men and women, though the theory obviously extends under the general setup.

5.1 The Model Setup and Saliency Analysis

Let $X \in \mathcal{X} \subset \mathbf{R}^m$ and $Y \in \mathcal{Y} \subset \mathbf{R}^k$ be vectors of attributes of men and women respectively, with P_0 and Q_0 the probability distributions of X and Y respectively. A matching is then characterized by a probability distribution π on $\mathcal{X} \times \mathcal{Y}$ such that its density $f_\pi(x, y)$ describes the probability of occurrence of a couple with attributes (x, y). Since we only consider matched couples and matching is one-to-one, π must have marginals P_0 and Q_0. A defining feature of the transferrable utility framework is that matched couples behave unitarily, i.e., there is a single surplus function $s : \mathcal{X} \times \mathcal{Y} \to \mathbf{R}$ generated by the matching, and how the surplus is shared between the spouses is endogenous. A final ingredient crucial to the matching game is the equilibrium concept. As standard in the literature, we employ the notion of stability (Gale and Shapley, 1962), and call a matching stable if (i) no matched individual would rather be single and (ii) no pair of individuals would *both* like being matched together better than their current situation. It is well known that stability (a game theoretical concept) and surplus maximization (a social planner's problem) are equivalent (Shapley and Shubik, 1971; Chiappori et al., 2010). Consequently, the matching π_0 in equilibrium can be characterized by the centralized problem:

$$\max_{\pi \in \mathbf{\Pi}(P_0, Q_0)} E_\pi[s(X, Y)] \ , \tag{50}$$

where $\mathbf{\Pi}(P_0, Q_0)$ is the family of distributions on $\mathcal{X} \times \mathcal{Y}$ with marginals P_0 and Q_0.

Without further appropriate modelling, the optimal transport problem (50) implies pure matching under regularity conditions (Becker, 1973; Chiappori et al., 2010), i.e., a certain type of men is for sure going to be matched with a certain type of women. One empirical strategy to reconcile such unrealistic predictions with data is to incorporate unobserved heterogeneity into the surplus function. Following Choo and Siow (2006) and Chiappori et al. (2017), we assume that

$$s(x, y) = \Phi(x, y) + \epsilon_m(y) + \epsilon_w(x) \ , \tag{51}$$

where $\Phi(x, y)$ is the systematic part of the surplus, and $\epsilon_m(y)$ and $\epsilon_w(x)$ are unobserved random shocks. Note that $\epsilon_m(y)$ and $\epsilon_w(x)$ enter the surplus function additively and separably, which is by no means a haphazard restriction: it makes an otherwise extremely difficult problem more tractable (Chiappori and Salanié, 2016; Chiappori, 2017). Nonparametric identification of both Φ and the error distributions, however, remains a challenging task. Following Dagsvik (2000) and Choo and Siow (2006), we thus further assume that the errors follow the type-I extreme value distribution, though we note that such distributional assumption can be completely dispensed with (Galichon and Salanié, 2015). The matching distribution π_0 can in turn be characterized by

$$\max_{\pi \in \mathbf{\Pi}(P_0, Q_0)} E_\pi[\Phi(X,Y)] - E_\pi[\log f_\pi(X,Y)] \;, \tag{52}$$

and Φ is nonparametrically identified (Galichon and Salanié, 2015). For the purpose of estimation, we further assume that, for some $A_0 \in \mathbf{M}^{m\times k}$ and any $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$\Phi(x,y) \equiv \Phi_{A_0}(x,y) = x^\intercal A_0 y \;, \tag{53}$$

where A_0 is called the affinity matrix. Such a parametric specification has also been employed by Galichon and Salanié (2010, 2015) and Dupuy and Galichon (2014).

Heuristically, the (i,j)th entry a_{ij} of A_0 measures the strength of mutual attractiveness between attributes x_i and y_j. The rank of A_0 provides valuable information on the number of dimensions on which sorting occurs, and helps construct indices of mutual attractiveness (Dupuy and Galichon, 2014, 2015). Following Dupuy and Galichon (2014) and Galichon and Salanié (2015), we estimate A_0 by matching moments:

$$E_{\pi(A_0,P_0,Q_0)}[XY^\intercal] = E[XY^\intercal] \;, \tag{54}$$

where $\pi_0 \equiv \pi(A_0,P_0,Q_0)$ is the matching distribution in equilibrium. By Lemma D.11, if X and Y are finitely discrete-valued with probability mass functions p_0 and q_0, then equation (54) defines not only a unique A_0, but also an implicit map $(p_0, q_0, E[XY^\intercal]) \mapsto A(p_0, q_0, E[XY^\intercal]) \equiv A_0$ which is Hadamard differentiable. This has two immediate implications. First, the estimator $\hat{A}_n$ defined by the sample analog of (54), i.e.,

$$E_{\pi(\hat{A}_n,\hat{p}_n,\hat{q}_n)}[XY^\intercal] = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i^\intercal \;, \tag{55}$$

where $\hat{p}_n$ and $\hat{q}_n$ are sample analogs of p_0 and q_0 respectively, is asymptotically normal. Second, the bootstrap estimator $\hat{A}_n^*$ defined by the bootstrap analog of (55), i.e.,

$$E_{\pi(\hat{A}_n^*,\hat{p}_n^*,\hat{q}_n^*)}[XY^\intercal] = \frac{1}{n}\sum_{i=1}^{n} X_i^* Y_i^{*\intercal} \;, \tag{56}$$

where $\hat{p}_n^*$ and $\hat{q}_n^*$ are bootstrap analogs of $\hat{p}_n$ and $\hat{q}_n$ respectively, is consistent in estimating the asymptotic distribution of $\hat{A}_n$. We have thus verified the main assumptions in order to apply our framework. We note in passing that it appears challenging to verify Assumption 3.2 when X and Y are continuous, and we believe it should be based on arguments different from those above.

Alternatively, Dupuy and Galichon (2014) estimate the rank of A_0 by employing the test of Kleibergen and Paap (2006), which they call the saliency analysis. There are two motivations of using our inferential procedure. First, as argued previously, the KP test is designed for the more restrictive setup (2) and can be invalid and/or conservative for the hypotheses in (1). Consequently, estimation of rank(A_0) by sequentially conducting the KP tests may be less accurate. Second, the KP test relies on an estimator of the asymptotic variance of $\hat{A}_n$ which appears to be somewhat complicated – see the formula (B18) in Dupuy and Galichon (2014), while one generic merit of bootstrap inference is to avoid analytic complications by repetitive resampling (Horowitz, 2001).

5.2 Data and Empirical Results

We use the same data source as Dupuy and Galichon (2014), i.e., the 1993-2002 waves of the DNB Household Survey, to estimate preferences in the marriage market in Dutch. The panel contains rich information about individual attributes such as demographic variables (e.g., education), anthropometry parameters (e.g., height and weight), personality traits (e.g., emotional stability, extraversion, conscientiousness, agreeableness, autonomy) and risk attitude – see Nyhus (1996) for more detailed descriptions of the data. In order to apply our framework, we have discretized the variables in the following way: (i) BMI[4] is converted into a trinary variable according to the international BMI classification, i.e., BMI is set to be 1 if BMI <18.50, 2 if 18.50 $\leq$ BMI $<$ 24.99, and 3 if BMI $\geq$ 24.99; (ii) Five personal traits variables and risk aversion are also converted into trinary data by taking the value 1 if they are below the corresponding 25% quantiles, 2 if they are between the 25% and the 75% quantiles, and 3 if they are strictly larger than the 75% quantiles; (iii) Education remains unchanged since it is discrete as it is. We make use of the same sample as Dupuy and Galichon (2014) which has 1158 couples, but only include subsets of the 10 attribute variables that they considered to reduce the computational burden – see Table 5. Following Dupuy and Galichon (2014) still, we demean and standardize the data beforehand, and then compute the optimal matching distribution by the iterative projection fitting procedure (Rüschendorf, 1995).

For each model specification, we study two problems: testing singularity of the corresponding affinity matrix and estimating its true rank. In carrying out our inferential procedures, we estimate the derivative through either (28) or (26), for which we choose the tuning parameter $\kappa_n \in \{n^{-1/5}, n^{-1/4}, n^{-1/3}\}$. The corresponding results are labelled as CF-A and CF-N respectively. We also implement the two-step procedure with $\beta \in \{\alpha/10, \alpha/15, \alpha/20\}$, labelled as CF-T. The significance level is $\alpha = 5\%$. As shown

[4] The body mass index (BMI) is defined as the body mass divided by the square of the body height, which provides a simple numeric measure of a person's thinness.

Table 5. Model specifications

Model	Attributes included
(1)	Education, BMI, Risk aversion
(2)	Education, BMI, Risk aversion, Conscientiousness
(3)	Education, BMI, Risk aversion, Extraversion
(4)	Education, BMI, Risk aversion, Agreeableness
(5)	Education, BMI, Risk aversion, Emotional stability
(6)	Education, BMI, Risk aversion, Autonomy
(7)	Education, BMI, Risk aversion, Conscientiousness, Extraversion
(8)	Education, BMI, Risk aversion, Conscientiousness, Autonomy
(9)	Education, BMI, Risk aversion, Extraversion, Autonomy

Table 6. Empirical results

		The p-values for full rank tests[†]									
Model	Maximum rank	CF-T			CF-A			CF-N			KP-M[‡]
		$\alpha/10$	$\alpha/15$	$\alpha/20$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	
(1)	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(2)	4	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.03
(3)	4	0.04	0.04	0.04	0.01	0.04	0.18	0.01	0.02	0.04	0.25
(4)	4	0.88	0.88	0.88	0.86	0.88	0.92	0.85	0.86	0.87	0.94
(5)	4	0.23	0.08	0.08	0.03	0.08	0.23	0.02	0.03	0.06	0.35
(6)	4	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.03
(7)	5	0.02	0.02	0.02	0.00	0.02	0.14	0.00	0.00	0.01	0.19
(8)	5	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.03
(9)	5	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.22
		Estimates of the true rank ($\alpha = 5\%$)									
(1)	3	3	3	3	3	3	3	3	3	3	3
(2)	4	4	4	4	4	4	4	4	4	4	4
(3)	4	4	4	4	4	4	3	4	4	4	3
(4)	4	3	3	3	3	3	3	3	3	3	3
(5)	4	3	3	3	4	3	3	4	4	3	3
(6)	4	4	4	4	4	4	4	4	4	4	4
(7)	5	5	5	5	5	5	4	5	5	5	3
(8)	5	5	5	5	5	5	5	5	5	5	5
(9)	5	5	5	5	5	5	5	5	5	5	3

[†] The three values under CF-T are the choices of β, and those under CF-A and CF-N are the choices of κ_n as in 27 and (26) respectively.

[‡] The p-value for KP-M is given by the smallest significance level such that the null hypothesis is rejected, which is equal to the maximum p-value of all Kleibergen and Paap (2006)'s tests implemented by the multiple testing method.

by Table 6, our three inferential procedures yield overall consistent results, with the exception of models (3), (5) and (7). For example, for model (3), all our procedures estimate the rank to be 4, except CF-A with $\kappa_n = n^{-1/3}$ which estimates the rank to be 3. Such discrepancies may be due to the choices of tuning parameters or finite sample variations. Nonetheless, what is comforting to us is that, in the three models, the majority of the 9 estimates point to the same rank. We also note that the p-values and estimates of the rank based on CF-T are the same across all three choices of β, for all model specifications except for model (5).

There are, however, noticeable differences between our results and those obtained by the KP test. First, there are sizable discrepancies between the p-values of our tests and those for the KP-M tests, especially for model specifications (3), (5), (7) and (9). Second, in terms of estimation, there are also marked differences. For example, for model (9), our tests unanimously estimate the rank to be 5, while the KP test estimates the rank to be 3. Similar patterns occur for models (3) and (7) for which the KP test provides a smaller rank estimator. Inspecting these differences, it seems that Extraversion is not important for matching in the Dutch marriage market according to the KP results, while our results show that it is important. Overall, we obtain estimates different from those based on Kleibergen and Paap (2006) in 3 out of the 9 model specifications.

6 Conclusion

In this paper, we have developed a general framework for conducting inference on the rank of a matrix Π_0. The problem is of first order importance because we have shown, through an analytic example and simulation evidences, that existing tests may be invalid due to over-rejections when in truth $\text{rank}(\Pi_0)$ is strictly less than the hypothesized value r, while their multiple testing versions, though valid, can be substantially conservative. We have then developed a testing procedure that has asymptotic exact size control, is consistent, and meanwhile accommodates the possibility $\text{rank}(\Pi_0) < r$. A two-step test is proposed to mitigate the concern on tuning parameters. We also characterized classes of local perturbations under which our tests have local size control and nontrivial local power. These attractive testing properties in turn lead to more accurate rank estimators. We illustrated the empirical relevance of our results by conducting inference on the rank of an affinity matrix in a two-sided matching model.

We stress that our framework is limited to matrices of fixed dimensions and inapplicable to examples where the dimensions diverge as sample size increases. This is because Assumption 3.1 is being violated in these settings, as Π_0 typically does not admit weakly convergent estimators. While we find extensions allowing varying dimensions important in, for example, many IV problems and high dimensional factor models, a thorough treatment is beyond the scope of this paper and hence left for future study.

References

Al-Sadoon, M. M. (2017). A unifying theory of tests of rank. *Journal of Econometrics*, **199** 49–62.

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

Andrews, D. W. (1987). Asymptotic results for generalized wald tests. *Econometric Theory*, **3** 348–358.

Andrews, D. W. K. and Barwick, P. J. (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica*, **80** 2805–2826.

Andrews, D. W. K. and Guggenberger, P. (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory*, **26** 426–468.

Andrews, D. W. K. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** 119–157.

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political economy*, **81** 813–846.

Berger, R. L. and Boos, D. D. (1994). p values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*, **89** 1012–1016.

Bhatia, R. (1997). *Matrix Analysis*. Springer.

Camba-Mendez, G. and Kapetanios, G. (2009). Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling. *Econometric Reviews*, **28** 581–611.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, **90** 414–427.

Canova, F. and Sala, L. (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, **56** 431 – 449.

Chen, Q. and Fang, Z. (2018). Inference on functionals under first order degeneracy. *Available at SSRN: https://ssrn.com/abstract=2662758.*

Chen, X., Chernozhukov, V., Lee, S. and Newey, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica*, **82** 785–809.

Chen, X. and Santos, A. (2018). Overidentification in regular models. *Econometrica*, **86** 1771–1817.

Chernozhukov, V., Hong, H. and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, **75** 1243–1284.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, **71** 1405–1441.

Chiappori, P.-A. (2017). *Matching with Transfers: The Economics of Love and Marriage*. Princeton University Press.

Chiappori, P.-A., McCann, R. J. and Nesheim, L. P. (2010). Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory*, **42** 317–354.

Chiappori, P.-A. and Salanié, B. (2016). The econometrics of matching models. *Journal of Economic Literature*, **54** 832–861.

Chiappori, P.-A., Salanié, B. and Weiss, Y. (2017). Partner choice, investment in children, and the marital college premium. *American Economic Review*, **107** 2109–67.

Choo, E. and Siow, A. (2006). Who marries whom and why. *Journal of Political Economy*, **114** 175–201.

Ciscato, E., Galichon, A. and Goussé, M. (2018). Like attract like? a structural comparison of homogamy across same-sex and different-sex households. *Available at SSRN: https://ssrn.com/abstract=2530724*.

Cragg, J. G. and Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, **9** 222–240.

Cragg, J. G. and Donald, S. G. (1996). On the asymptotic properties of ldu-based tests of the rank of a matrix. *Journal of the American Statistical Association*, **91** 1301–1309.

Cragg, J. G. and Donald, S. G. (1997). Inferring the rank of a matrix. *Journal of Econometrics*, **76** 223 – 250.

Dagsvik, J. K. (2000). Aggregation in matching markets. *International Economic Review*, **41** 27–58.

Djogbenou, A. A., MacKinnon, J. G. and Nielsen, M. O. (2018). Asymptotic theory and wild bootstrap inference with clustered errors. Working Papers 1406, Queen's University, Department of Economics. URL `https://ideas.repec.org/p/qed/wpaper/1406.html`.

DÜMBGEN, L. (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, **95** 125–140.

DUPUY, A. and GALICHON, A. (2014). Personality traits and the marriage market. *Journal of Political Economy*, **122** 1271–1319.

DUPUY, A. and GALICHON, A. (2015). Canonical correlation and assortative matching: A remark. *Annals of Economics and Statistics* 375–383.

EFRON, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, **7** 1–26.

ENGLE, R. F. and GRANGER, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, **55** 251–276.

FANG, Z. and SANTOS, A. (2018). Inference on directionally differentiable functions. *The Review of Economic Studies*. URL `http://dx.doi.org/10.1093/restud/rdy049`.

FISHER, F. M. (1961). Identifiability criteria in nonlinear systems. *Econometrica*, **29** 574–590.

FISHER, F. M. (1966). *The Identification Problem in Econometrics*. McGraw-Hill.

GALE, D. and SHAPLEY, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, **69** 9–15.

GALICHON, A. and SALANIÉ, B. (2010). Matching with trade-offs: Revealed preferences over competing characteristics. *Available at SSRN: https://ssrn.com/abstract=1640380*.

GALICHON, A. and SALANIÉ, B. (2015). Cupid's invisible hand: Social surplus and identification in matching models. *Available at SSRN: https://ssrn.com/abstract=1804623*.

GILL, L. and LEWBEL, A. (1992). Testing the rank and definiteness of estimated matrices with applications to factor, state-space and arma models. *Journal of the American Statistical Association*, **87** 766–776.

HALLIN, M., VAN DEN AKKER, R. and WERKER, B. J. (2016). Semiparametric error-correction models for cointegration with trends: Pseudo-Gaussian and optimal rank-based tests of the cointegration rank. *Journal of Econometrics*, **190** 46 – 61.

HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton University Press, New Jersey.

HONG, H. and LI, J. (2018). The numerical Delta method and bootstrap. *Journal of Econometrics*, **206** 379 – 394.

Horowitz, J. L. (2001). The bootstrap. In *Handbook of Econometrics V* (J. J. Heckman and E. Leamer, eds.). Elsevier, 3159–3228.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, **144** 27–61.

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

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

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

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

Kasahara, H. and Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, **77** 135–175.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, **133** 97 – 126.

Komunjer, I. and Ng, S. (2011). Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica*, **79** 1995–2032.

Koopmans, T. C. and Hood, W. C. (1953). The estimation of simultaneous linear economic relationships. In *Studies in Econometric Method, Cowles Commission Monograph*, vol. 14. 112–199.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3rd ed. Springer Verlag.

Linton, O., Song, K. E. and Whang, Y.-J. (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, **154** 186 – 202.

Loh, W.-Y. (1985). A new method for testing separate families of hypotheses. *Journal of the American Statistical Association*, **80** 362–368.

Maddala, G. S. and Kim, I.-M. (1998). *Unit Roots, Cointegration, and Structural Change*. 4, Cambridge University Press.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica*, **76** 945–978.

McLachlan, G. and Peel, D. (2004). *Finite Mixture Models*. John Wiley & Sons.

Neyman, J. (1937). "Smooth test" for goodness of fit. *Scandinavian Actuarial Journal*, **1937** 149–199.

Nyhus, E. K. (1996). The VSB-CentER savings project: Data collection methods, questionnaires and sampling procedures. In *VSB-CentER Savings project progress report*, vol. 42.

Politis, D. N. and Romano, J. P. (1992). A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap*. John Wiley New York, 263–270.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, **22** 2031–2050.

Portier, F. and Delyon, B. (2014). Bootstrap testing of the rank of a matrix via least-squared constrained estimation. *Journal of the American Statistical Association*, **109** 160–172.

Praestgaard, J. and Wellner, J. (1993). Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, **21** 2053–2086.

Robin, J.-M. and Smith, R. J. (2000). Tests of rank. *Econometric Theory*, **16** 151–175.

Roehrig, C. S. (1988). Conditions for identification in nonparametric and parametric models. *Econometrica*, **56** 433–447.

Romano, J. P., Shaikh, A. and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica*, **82** 1979–2002.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica* 577–591.

Rothenberg, T. J. (1984). Approximating the distributions of econometric estimators and test statistics. In *Handbook of Econometrics*, vol. 2. 881–935.

Roussas, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics*, vol. 63. Cambridge University Press.

Rüschendorf, L. (1995). Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, **23** 1160–1174.

Shao, J. (1994). Bootstrap sample size in nonregular cases. In *Proceedings of the American Mathematical Society*, vol. 122. 1251–1262.

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, **66** 477–487.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research*, **30** 169–186.

Shapiro, A. (2000). Statistical inference of stochastic optimization problems. In *Probabilistic Constrained Optimization* (S. P. Uryasev, ed.), vol. 49 of *Nonconvex Optimization and Its Applications*. Springer, 282–307.

Shapley, L. S. and Shubik, M. (1971). The assignment game I: The core. *International Journal of Game Theory*, **1** 111–130.

Silvapulle, M. J. (1996). A test in the presence of nuisance parameters. *Journal of the American Statistical Association*, **91** 1690–1693.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag.

West, K. D. (1997). Another heteroskedasticity-and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics*, **76** 171–191.

Supplemental Appendix to "Improved Inference on the Rank of a Matrix"

Qihui Chen
School of Management and Economics
The Chinese University of Hong Kong, Shenzhen
qihuichen@cuhk.edu.cn

Zheng Fang
Department of Economics
Texas A&M University
zfang@tamu.edu

December 4, 2018

For convenience of the reader, we commence by gathering some notation that appear in the paper, most of which are standard in the literature.

$\mathbf{M}^{m\times k}$	The space of $m \times k$ real matrices for $m, k \in \mathbf{N}$.
I_k	The identity matrix of size k.
$\mathbf{0}_k, \mathbf{1}_k$	The $k \times 1$ vectors of zeros and ones.
$A^\intercal$	The transpose of a matrix $A \in \mathbf{M}^{m\times k}$.
$\mathrm{tr}(A)$	The trace of a square matrix $A \in \mathbf{M}^{k\times k}$.
$\mathrm{vec}(A)$	The column vectorization of $A \in \mathbf{M}^{m\times k}$.
$\|A\|$	The Frobenius norm of a matrix $A \in \mathbf{M}^{m\times k}$.
$\sigma_j(A)$	The jth largest singular value of a matrix $A \in \mathbf{M}^{m\times k}$.
$\mathbb{S}^{m\times k}$	A subset of $\mathbf{M}^{m\times k}$: $\mathbb{S}^{m\times k} \equiv \{U \in \mathbf{M}^{m\times k} : U^\intercal U = I_k\}$.
$C(T)$	The space of continuous functions on a (topological) space T.
$\varphi : \mathbb{D} \twoheadrightarrow \mathbb{E}$	A correspondence from a set $\mathbb{D}$ to another set $\mathbb{E}$.

Due to the fundamental role played by the singular value decomposition in the paper, we next provide a brief review and emphasize facts that are relevant to our development. Conceptually, the singular value decomposition generalizes the spectral decomposition to arbitrary (possibly rectangular) matrices. Let $\Pi \in \mathbf{M}^{m\times k}$ with $m \geq k$. Then the singular value decomposition of Π is

$$\Pi = P\Sigma Q^\intercal , \tag{A.1}$$

where $P \in \mathbb{S}^{m\times m}$ and $Q \in \mathbb{S}^{k\times k}$ are orthornormal, and $\Sigma \in \mathbf{M}^{m\times k}$ is a diagonal matrix with its diagonal entries in descending order – throughout the paper such a decomposition format is silently understood. The columns of P, called the left singular vectors of Π, are eigenvectors of $\Pi\Pi^\intercal$ (which is symmetric), the columns of Q, called the right singular vectors of Π, are eigenvectors of $\Pi^\intercal\Pi$ (which is also symmetric), and the diagonal entries of Σ, called the singular values, are the corresponding square roots of the eigenvalues of $\Pi\Pi^\intercal$ and also of $\Pi^\intercal\Pi$. Such a decomposition allows us to conclude that $\text{rank}(\Pi)$ is precisely equal to the number of nonzero singular values.

The matrix Σ is uniquely determined, though not the matrices P and Q. If $\text{rank}(\Pi) = r_0$, then we may partition P as $P = [P_1, P_2]$ such that P_1 consists of precisely the first r_0 columns of P that are associated with the nonzero singular values of Π; similarly we may partition Q as $Q = [Q_1, Q_2]$. Then the null space of Π is precisely the column space of Q_2, and the null space of $\Pi^\intercal$ is precisely the column space of P_2. Moreover, P_2 and Q_2 are uniquely determined respectively up to postmultiplication by $(m-r_0)\times(m-r_0)$ and $(k-r_0)\times(k-r_0)$ orthonormal matrices. Fortunately, the singular values $\sigma_j(P_2^\intercal M Q_2)$ (as in (20)) for any $M \in \mathbf{M}^{m\times k}$ are invariant to such transformations.

For convenience of applied researchers who work with Stata, we have developed a command `bootranktest` that may be used to test whether a matrix of the form $E[VZ^\intercal]$ has full rank based on our two-step test. In the first step, we use the KP test to obtain the rank estimator by choosing $\beta = 0.05/15$. Its syntax is as follows:

```
bootranktest (varlist1) (varlist2) [if] [in]
```

where `varlist1` should have more variables than `varlist2`. As of now, this command is designed for i.i.d. data and employs Efron (1979)'s empirical bootstrap with 500 bootstrap repetitions. We plan to refine it by adding more features in future.

The remainder of the supplement is organized as follows. Appendix A presents the proofs of our main results. Appendix B provides additional details and discussions regarding comparisons with Kleibergen and Paap (2006), while Appendix C derives some estimation results based on a sequential testing procedure. Appendix D contains some supporting lemmas. Additional examples are presented in Appendix E where special attention is paid to inference on cointegration rank.

Appendix A Proofs of Main Results

Proof of Lemma 3.1: The proof is based on a simple application of the representation of extremal partial trace. Recall that $\sigma_1^2(\Pi), \ldots, \sigma_k^2(\Pi)$ are eigenvalues of $\Pi^\intercal\Pi$ in descending order. Let $d \equiv k - r$. It follows by Proposition 1.3.4 in Tao (2012) that

$$\phi_r(\Pi) = \sum_{j=r+1}^{k} \sigma_j^2(\Pi) = \inf_{u_1,\ldots,u_d} \sum_{j=1}^{d} u_j^\intercal \Pi^\intercal \Pi u_j \ , \tag{A.2}$$

where the infimum is taken over all $u_1, \ldots, u_d \in \mathbf{R}^k$ that are orthonormal. Noting $U \equiv [u_1, \ldots, u_d] \in \mathbb{S}^{k\times d}$, we obtain by (A.2) and the definition of Frobenius norm that

$$\phi_r(\Pi) = \inf_{U\in\mathbb{S}^{k\times d}} \operatorname{tr}(U^\intercal\Pi^\intercal\Pi U) = \inf_{U\in\mathbb{S}^{k\times d}} \|\Pi U\|^2 . \tag{A.3}$$

The infimum in (A.3) is achieved on $\mathbb{S}^{k\times d}$ because $U \mapsto \|\Pi U\|^2$ is continuous, and $\mathbb{S}^{k\times d}$ is compact since it is closed and bounded. This completes the proof of the lemma. ■

Proof of Proposition 3.1: Let $d \equiv k - r$, and define $\phi_1 : \mathbf{M}^{m\times k} \to C(\mathbb{S}^{k\times d})$ by $\phi_1(\Pi)(U) = \|\Pi U\|^2$, and $\phi_2 : C(\mathbb{S}^{k\times d}) \to \mathbf{R}$ by $\phi_2(f) = \min\{f(U) : U \in \mathbb{S}^{k\times d}\}$, so that $\phi_r = \phi_2 \circ \phi_1$ by Lemma 3.1. For part (i), we proceed by verifying first order Hadamard directional differentiability of ϕ_1 and ϕ_2, and then conclude by the chain rule.

Let $\{M_n\} \subset \mathbf{M}^{m\times k}$ be a sequence satisfying $M_n \to M \in \mathbf{M}^{m\times k}$, and $t_n \downarrow 0$ as $n \to \infty$. For each $n \in \mathbf{N}$, define $g_n : \mathbb{S}^{k\times d} \to \mathbf{R}$ by

$$g_n(U) = \frac{\|(\Pi + t_nM_n)U\|^2 - \|\Pi U\|^2}{t_n} = \frac{\|\Pi U + t_nM_nU\|^2 - \|\Pi U\|^2}{t_n} ,$$

and $g : \mathbb{S}^{k\times d} \to \mathbf{R}$ by $g(U) = 2\operatorname{tr}((\Pi U)^\intercal MU)$. Then by simple algebra we have

$$\begin{aligned}\sup_{U\in\mathbb{S}^{k\times d}} |g_n(U) - g(U)| &= \sup_{U\in\mathbb{S}^{k\times d}} |2\operatorname{tr}((\Pi U)^\intercal(M_n - M)U) + t_n\|M_nU\|^2| \\ &\le \sup_{U\in\mathbb{S}^{k\times d}} \{2\|\Pi U\|\|(M_n - M)U\| + t_n\|M_nU\|^2\} ,\end{aligned} \tag{A.4}$$

where the inequality follows by the triangle inequality and the Cauchy-Schwarz inequality for the trace operator. For the right hand side of (A.4), we further have

$$\begin{aligned}&\sup_{U\in\mathbb{S}^{k\times d}} \{2\|\Pi U\|\|(M_n - M)U\| + t_n\|M_nU\|^2\} \\ &\quad\le \sup_{U\in\mathbb{S}^{k\times d}} \{2\|\Pi\|\|U\|\|M_n - M\|\|U\| + t_n\|M_n\|^2\|U\|^2\} = o(1) ,\end{aligned} \tag{A.5}$$

where we exploited the sub-multiplicativity of Frobenius norm and the facts that $\|U\| = \sqrt{d}$, $M_n \to M$ and $t_n \downarrow 0$ as $n \to \infty$. We thus conclude from (A.4) and (A.5) that $g_n \to g$ uniformly in $C(\mathbb{S}^{k\times d})$, or equivalently ϕ_1 is first order Hadamard directionally differentiable at Π with derivative $\phi'_{1,\Pi} : \mathbf{M}^{m\times k} \to C(\mathbb{S}^{k\times d})$ given by

$$\phi'_{1,\Pi}(M)(U) = 2\operatorname{tr}((\Pi U)^\intercal MU) . \tag{A.6}$$

On the other hand, Theorem 3.1 in Shapiro (1991) implies that $\phi_2 : C(\mathbb{S}^{k\times d}) \to \mathbf{R}$ is first order Hadamard directionally differentiable at any $f \in C(\mathbb{S}^{k\times d})$ with derivative

$\phi'_{2,f} : C(\mathbb{S}^{k\times d}) \to \mathbf{R}$ given by: for $\Psi(f) \equiv \arg\min_{U\in\mathbb{S}^{k\times d}} f(U)$,

$$\phi'_{2,f}(h) = \min_{U\in\Psi(f)} h(U) \ . \tag{A.7}$$

Combining (A.6), (A.7) and the chain rule (Shapiro, 1990, Proposition 3.6), we may now conclude that $\phi_r : \mathbf{M}^{m\times k} \to \mathbf{R}$ is first order Hadamard directionally differentiable at any $\Pi \in \mathbf{M}^{m\times k}$ with the derivative $\phi'_{r,\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ given by

$$\phi'_{r,\Pi}(M) = \phi'_{2,\phi_1(\Pi)} \circ \phi'_{1,\Pi}(M) = \min_{U\in\Psi(\Pi)} 2\mathrm{tr}((\Pi U)^\intercal MU) \ .$$

This completes the proof of part (i) of the proposition.

For part (ii), note that $\phi_r(\Pi) = 0$ implies that $\Pi U = 0$ for all $U \in \Psi(\Pi)$ and hence $\phi'_{r,\Pi}(M) = 0$ for all $M \in \mathbf{M}^{m\times k}$. Recall that $\{M_n\} \subset \mathbf{M}^{m\times k}$ with $M_n \to M \in \mathbf{M}^{m\times k}$ and $t_n \downarrow 0$ as $n \to \infty$. By Lemma 3.1 we have

$$\begin{aligned}|\phi_r(\Pi + t_nM_n) - \phi_r(\Pi + t_nM)| = \big|\min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM_n)U\| - \min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM)U\|\big| \\ \times \big(\min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM_n)U\| + \min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM)U\|\big) \ , \end{aligned} \tag{A.8}$$

where the equality also exploited the elementary formula $a^2 - b^2 = (a + b)(a - b)$. For the first term on the right hand side of (A.8), we have

$$\big|\min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM_n)U\| - \min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM)U\|\big| \le t_n\sqrt{d}\|M_n - M\| = o(t_n) \ , \tag{A.9}$$

where the inequality follows by the Lipschitz continuity of the infimum operator, the triangle inequality, $\|\cdot\|$ being sub-multiplicative, and $\|U\| = \sqrt{d}$ for $U \in \mathbb{S}^{k\times d}$. For the second term on the right hand side of (A.8), we have: for any fixed $U^* \in \Psi(\Pi)$,

$$\begin{aligned}\min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM_n)U\| + \min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM)U\| \le \|(\Pi + t_nM_n)U^*\| \\ + \|(\Pi + t_nM)U^*\| \le t_n\|M_n\|\|U^*\| + t_n\|M\|\|U^*\| = O(t_n) \ , \end{aligned} \tag{A.10}$$

where we exploited $\Pi U^* = 0$, the sub-multiplicativity of Frobenius norm, $\|U^*\| = \sqrt{d}$ and $M_n \to M$ as $n \to \infty$. Combining (A.8)-(A.10), we thus obtain

$$|\phi_r(\Pi + t_nM_n) - \phi_r(\Pi + t_nM)| = o(t_n^2) \ . \tag{A.11}$$

Next, for $\epsilon > 0$, let $\Psi(\Pi)^\epsilon \equiv \{U \in \mathbb{S}^{k\times d} : \min_{U'\in\Psi(\Pi)} \|U' - U\| \le \epsilon\}$ and $\Psi(\Pi)_1^\epsilon \equiv \{U \in \mathbb{S}^{k\times d} : \min_{U'\in\Psi(\Pi)} \|U' - U\| \ge \epsilon\}$. In what follows we consider the nontrivial case when $\Pi \neq 0$ and $M \neq 0$. Then we must have $\Psi(\Pi) \subsetneqq \mathbb{S}^{k\times d}$ and hence $\Psi(\Pi)_1^\epsilon \neq \emptyset$ for ϵ sufficiently small. Let $\sigma^+_{\min}(\Pi)$ denote the smallest positive singular value of Π which exists since $\Pi \neq 0$, and $\Delta \equiv 3\sqrt{2}[\sigma^{+}_{\min}(\Pi)]^{-1}\max_{U\in\mathbb{S}^{k\times d}}\|MU\| > 0$ since $M \neq 0$. Then it follows that for all n sufficiently large

$$\begin{aligned}\min_{U\in\Psi(\Pi)_1^{t_n\Delta}}\|(\Pi+t_nM)U\| &\geq \min_{U\in\Psi(\Pi)_1^{t_n\Delta}}\|\Pi U\| - t_n\max_{U\in\mathbb{S}^{k\times d}}\|MU\| \\ &\geq \frac{\sqrt{2}}{2}t_n\sigma^{+}_{\min}(\Pi)\Delta - t_n\max_{U\in\mathbb{S}^{k\times d}}\|MU\| > t_n\max_{U\in\mathbb{S}^{k\times d}}\|MU\| \\ &\geq \min_{U\in\Psi(\Pi)}\|(\Pi+t_nM)U\| \geq \sqrt{\phi_r(\Pi+t_nM)}\ , \end{aligned} \tag{A.12}$$

where the first inequality follows by the triangle inequality and the fact that $\Psi(\Pi)_1^{t_n\Delta} \subset S^{k\times d}$, the second inequality follows by Lemma D.1, the third inequality is due to the definition of Δ, and the fourth inequality holds by the fact that $\Pi U = 0$ for $U \in \Psi(\Pi)$. By (A.12), we thus obtain that, for all n sufficiently large

$$\phi_r(\Pi+t_nM) = \min_{U\in\Psi(\Pi)^{t_n\Delta}}\|(\Pi+t_nM)U\|^2\ . \tag{A.13}$$

Now, for fixed $U \in \Psi(\Pi)$, $\Delta > 0$ and $t \in \mathbf{R}$, let $\Gamma^{\Delta} \equiv \{V \in \mathbf{M}^{k\times d} : \|V\| \leq \Delta\}$ and $\Gamma^{\Delta}_U(t) \equiv \{V \in \Gamma^{\Delta} : U + tV \in \mathbb{S}^{k\times d}\} = \{V \in \Gamma^{\Delta} : V^{\intercal}U + U^{\intercal}V = -tV^{\intercal}V\}$. Define a correspondence $\varphi : \mathbf{R} \twoheadrightarrow \mathbb{S}^{k\times d}\times\Gamma^{\Delta}$ by $\varphi(t) = \{(U,V) : U \in \Psi(\Pi), V \in \Gamma^{\Delta}_U(t)\}$. Then the right hand side of (A.13) can be written as

$$\begin{aligned}\min_{U\in\Psi(\Pi)^{t_n\Delta}}\|(\Pi+t_nM)U\|^2 &= \min_{(U,V)\in\varphi(t_n)}\|(\Pi+t_nM)(U+t_nV)\|^2 \\ &= t_n^2\min_{(U,V)\in\varphi(t_n)}\|\Pi V + MU\|^2 + o(t_n^2)\ , \end{aligned} \tag{A.14}$$

where we exploited $\Pi U = 0$ for all $U \in \Psi(\Pi)$ and $\|MV\| \leq \|M\|\Delta$ for all $V \in \Gamma^{\Delta}$. By Lemma D.2, $\varphi(t)$ is continuous at $t = 0$. Since φ is obviously compact-valued, we may then obtain by Theorem 17.31 in Aliprantis and Border (2006) that

$$\begin{aligned}\min_{(U,V)\in\varphi(t_n)}\|\Pi V + MU\|^2 &= \min_{(U,V)\in\varphi(0)}\|\Pi V + MU\|^2 + o(1) \\ &= \min_{U\in\Psi(\Pi)}\min_{V\in\mathbf{M}^{k\times d}}\|\Pi V + MU\|^2 + o(1)\ , \end{aligned} \tag{A.15}$$

where the second equality holds by letting Δ sufficiently large in view of Lemma D.3. Combining (A.13), (A.14) and (A.15) then yields

$$\phi_r(\Pi+t_nM) = t_n^2\min_{U\in\Psi(\Pi)}\min_{V\in\mathbf{M}^{k\times d}}\|\Pi V + MU\|^2 + o(t_n^2)\ . \tag{A.16}$$

The proposition now follows from result (A.16) and Lemma D.4. ■

Proof of Theorem 3.1: The first and second results are respectively straightforward implications of Theorems 2.1 in Fang and Santos (2018) and Chen and Fang (2018) by noting that $\phi'_{r,\Pi_0} = 0$ under H_0. In particular, their Assumptions 2.1 are satisfied in view of Proposition 3.1 and their Assumptions 2.2 are satisfied by Assumption 3.1. ■

Proof of Theorem 3.2: By the rate conditions on $\{\kappa_n\}$ and Assumption 3.1, the numerical estimator (26) satisfies the condition (25) by Proposition 3.1 in Chen and Fang (2018), while the analytic estimator (28)-(27) does so by Lemma D.6. In turn, following exactly the same proof of Corollary 3.2 in Fang and Santos (2018), we obtain that $\hat{c}_{n,1-\alpha} \xrightarrow{p} c_{1-\alpha}$ by Assumption 3.2 and the quantile restrictions on $c_{1-\alpha}$. Thus, under H_0, the first claim follows from combining Theorem 3.1, Slutsky's lemma, $c_{1-\alpha}$ being a continuity point of the limiting law and the portmanteau theorem.

For the second claim, Consider first the numerical estimator (26). Note that by Assumption 3.2, $\hat{\mathcal{M}}_n^* = O_{P_W}(1)$ in P_X-probability. Together with Assumption 3.1, $\kappa_n = o(1)$ as $n \to \infty$ and continuity of ϕ_r, we in turn see that, in P_X-probability,

$$\phi_r(\hat{\Pi}_n + \kappa_n \hat{\mathcal{M}}_n^*) = O_{P_W}(1) \ . \tag{A.17}$$

By the definition of $\hat{c}_{n,1-\alpha}$, it follows from (A.17) and $\phi_r(\hat{\Pi}_n) \geq 0$ that

$$\kappa_n^2 \hat{c}_{n,1-\alpha} \leq O_{P_W}(1) \tag{A.18}$$

in P_X-probability. By Assumption 3.1 and continuity of ϕ_r at Π_0, we have: under H_1,

$$\phi_r(\hat{\Pi}_n) \xrightarrow{p} \phi_r(\Pi_0) > 0 \ . \tag{A.19}$$

Combining results (A.18) and (A.19), together with $\tau_n\kappa_n \to \infty$, we thus conclude that

$$P(\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}) = P((\tau_n\kappa_n)^2 \phi_r(\hat{\Pi}_n) > \kappa_n^2 \hat{c}_{n,1-\alpha}) = 1 \ . \tag{A.20}$$

For the analytic estimator, let $\hat{d}_n \equiv k - \hat{r}_n$ and $d \equiv k - r$. By Lemma 3.1, we have

$$\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*) = \min_{U \in \mathbb{S}^{\hat{d}_n \times d}} \|\hat{P}_{2,n}^\intercal \hat{\mathcal{M}}_n^* \hat{Q}_{2,n} U\|^2 \leq \|\hat{\mathcal{M}}_n^*\|^2 mkd \ , \tag{A.21}$$

where the second inequality exploited $\|\hat{P}_{2,n}^\intercal\|^2 \|\hat{Q}_{2,n}\|^2 \leq mk$ and $\|U\|^2 = d$. Since $\hat{\mathcal{M}}_n^* = O_{P_W}(1)$ in P_X-probability by Assumption 3.2, it follows from (A.21) that

$$\hat{c}_{n,1-\alpha} \leq O_{P_W}(1) \tag{A.22}$$

in P_X-probability. Combining (A.19) and (A.22), together with $\tau_n \to \infty$, we thus obtain

$$P(\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}) = 1 \ . \tag{A.23}$$

This completes the proof of the second claim. ■

Proof of Theorem 3.3: For notational simplicity, define

$$A_n = \{\hat{r}_n > r\} \ , B_n = \{\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}\} \ , C_n = \{\hat{r}_n = r_0\} \ . \tag{A.24}$$

It follows that, under H_0,

$$\begin{aligned}\limsup_{n\to\infty} E[\psi_n] &\le \limsup_{n\to\infty} P((A_n \cup B_n) \cap C_n) + \limsup_{n\to\infty} P((A_n \cup B_n) \cap C_n^c) \\ &\le \limsup_{n\to\infty} P(A_n \cap C_n) + \limsup_{n\to\infty} P(B_n \cap C_n) + \limsup_{n\to\infty} P(C_n^c) \\ &\le 0 + \alpha - \beta + \beta = \alpha \ ,\end{aligned} \tag{A.25}$$

where we exploited $A_n \cap C_n = \emptyset$ under H_0, $\limsup_{n\to\infty} P(B_n \cap C_n) \le \alpha - \beta$ by Theorem 3.2, and $\limsup_{n\to\infty} P(C_n^c) \le \beta$. This completes the proof of the first claim. For the second claim of the theorem, note that

$$\liminf_{n\to\infty} E[\psi_n] \ge \liminf_{n\to\infty} P(\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}) = 1 \ , \tag{A.26}$$

where the equality follows by the proof of Theorem 3.2. ■

Proof of Proposition 3.2: By Assumption 3.1'(ii)(iii), we have

$$\tau\{\hat{\Pi}_n - \Pi_0\} = \tau_n\{\hat{\Pi}_n - \Pi_{0,n}\} + \tau_n\{\Pi_{0,n} - \Pi_0\} \xrightarrow{L} \mathcal{M} + \Delta \ . \tag{A.27}$$

This in turn allows us to conclude by Proposition 3.1 and $\phi_r(\Pi_0) = 0$. ■

Appendix B Comparisons with the KP Test

In this section, we first review the KP test for the reader's convenience, and then provide additional results regarding comparisons with Kleibergen and Paap (2006).

To describe the KP test, let $\hat{\Pi}_n$ be an estimator for $\Pi_0 \in \mathbf{M}^{m\times k}$ such that

$$\sqrt{n}\{\mathrm{vec}(\hat{\Pi}_n) - \mathrm{vec}(\Pi_0)\} \xrightarrow{L} N(0, \Omega_0) \ , \tag{B.1}$$

where the covariance matrix Ω_0 admits a consistent estimator $\hat{\Omega}_n$. Let $\hat{\Pi}_n = \hat{P}_n \hat{\Sigma}_n \hat{Q}_n^\intercal$ be a singular value decomposition of $\hat{\Pi}_n$, where $\hat{P}_n \in \mathbb{S}^{m\times m}$, $\hat{Q}_n \in \mathbb{S}^{k\times k}$, and $\hat{\Sigma}_n \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. For r the hypothesized value in (2), rewrite $\hat{P}_n = [\hat{P}_{1,n}, \hat{P}_{2,n}]$ and $\hat{Q}_n = [\hat{Q}_{1,n}, \hat{Q}_{2,n}]$ with $\hat{P}_{1,n} \in \mathbf{M}^{m\times r}$ and $\hat{Q}_{1,n} \in \mathbf{M}^{k\times r}$, and let $\hat{\Sigma}_{2,n}$ be the right bottom $(m-r)\times(k-r)$ submatrix of $\hat{\Sigma}_n$. Then the testing statistic proposed by Kleibergen and Paap (2006) for the hypotheses (2) is

$$T_{n,\mathrm{kp}} = n \cdot \mathrm{vec}(\hat{\Sigma}_{2,n})^\intercal [(\hat{Q}_{2,n} \otimes \hat{P}_{2,n})^\intercal \hat{\Omega}_n (\hat{Q}_{2,n} \otimes \hat{P}_{2,n})]^{-1} \mathrm{vec}(\hat{\Sigma}_{2,n}) \ , \tag{B.2}$$

where $\otimes$ signifies the Kronecker product, and the inverse is assumed to exist asymptotically. A special case of the testing statistic designed by Robin and Smith (2000) shares exactly the same form but without the weighting matrix,[1] i.e.,

$$T_{n,\mathrm{rs}} = n \cdot \mathrm{vec}(\hat{\Sigma}_{2,n})^\intercal \mathrm{vec}(\hat{\Sigma}_{2,n}) \ . \tag{B.3}$$

Kleibergen and Paap (2006) show that if $\mathrm{rank}(\Pi_0) = r$, then

$$T_{n,\mathrm{kp}} \xrightarrow{L} \chi^2((m-r)(k-r)) \ . \tag{B.4}$$

Thus, the KP test rejects the null H_0' in (2) at the significance level α if $T_{n,\mathrm{kp}}$ is larger than the $(1-\alpha)$-quantile of $\chi^2((m-r)(k-r))$.

In Section 2, we have shown that the KP test may be invalid since the χ^2-limit of the KP statistic is derived under H_0', ignoring the possibility $\mathrm{rank}(\Pi_0) < r$. As an alternative, one may construct a valid test for (1) by a multiple test on $\mathrm{rank}(\Pi_0) = 0, 1, \ldots, r$. Indeed, to show the validity of a multiple test, let $\psi_{n,r}$ be a nonrandomized test for hypotheses of the form (2) that rejects the null if $\psi_{n,r} = 1$ and fails to reject if $\psi_{n,r} = 0$. Moreover, suppose that $\psi_{n,r}$ is a consistent test that has asymptotic null rejection rates exactly equal to α. Then one may design a valid multiple test ψ_n for (1) by setting $\psi_n = \prod_{j=0}^{r} \psi_{n,j}$, i.e., ψ_n rejects H_0 if and only if all $\psi_{n,j}$'s reject. It follows that ψ_n has size control because, under H_0 and for $r_0 \equiv \mathrm{rank}(\Pi_0)$,

$$\limsup_{n\to\infty} E[\psi_n] = \limsup_{n\to\infty} P(\psi_{n,0} = 1, \ldots, \psi_{n,r} = 1) \leq \limsup_{n\to\infty} P(\psi_{n,r_0} = 1) = \alpha \ , \tag{B.5}$$

and that ψ_n is also consistent because, under H_1,

$$\begin{aligned} \liminf_{n\to\infty} E[\psi_n] &= \liminf_{n\to\infty} P(\psi_{n,0} = 1, \ldots, \psi_{n,r} = 1) \\ &\geq 1 - \sum_{j=1}^{r} [1 - \liminf_{n\to\infty} P(\psi_{n,j} = 1)] = 1 \ , \end{aligned} \tag{B.6}$$

where the inequality holds by the Boole's inequality and consistency of each $\psi_{n,j}$. This shows that ψ_n is valid, in fact consistent but may be conservative. The source of conservativeness of ψ_n is inherent in the inequality of (B.5) which is generically strict. Moreover, ψ_n is conservative whenever $\psi_{n,r}$ is, because

$$\limsup_{n\to\infty} E[\psi_n] = \limsup_{n\to\infty} P(\psi_{n,0} = 1, \ldots, \psi_{n,r} = 1) \leq \limsup_{n\to\infty} P(\psi_{n,r} = 1) < \alpha \ . \tag{B.7}$$

The remainder of this section is devoted to additional comparisons of our tests with the KP test based on the simulation designs and empirical application in Kleibergen

[1] Robin and Smith (2000) propose a class of testing statistics (indexed by functions h in their paper) which are asymptotically equivalent.

and Paap (2006). First, following those authors, we consider, for $R_t \in \mathbf{R}^{10}, F_t \in \mathbf{R}^4$,

$$R_t = \Pi_0 F_t + \varepsilon_t \ , \tag{B.8}$$

where $\{F_t\} \overset{i.i.d.}{\sim} N(0, \Sigma_F)$ and $\{\epsilon_t\}$ are independently generated according to

$$\varepsilon_t = v_t + \Gamma v_{t-1} \tag{B.9}$$

with $\{v_t\} \overset{i.i.d.}{\sim} N(0, \Sigma_v)$. We are interested in Π_0 which is specified as

$$\Pi_0 = \beta\alpha^\intercal + \delta\Pi_1 \ , \tag{B.10}$$

where $\delta \in \mathbf{R}$, $\alpha \in \mathbf{R}^4$, $\beta \in \mathbf{R}^{10}$ and $\Pi_1 \in \mathbf{M}^{10\times 4}$. Kleibergen and Paap (2006) try a wide range of values for δ; we shall focus on $\delta = 0, 0.01, \ldots, 0.1$ since we are concerned with local power. Other unknown parameters involved are configured to be exactly the same as those in Kleibergen and Paap (2006):

- Σ_F is specified as the sample correlation matrix of $\{F_t\}_{t=1}^n$, where $\{F_t\}_{t=1}^n$ is the real data to be studied for the empirical application;
- $\alpha = (0.0813, -0.0271, -0.6203, -0.0460)^\intercal$;
- $\beta = (-0.3411, -0.1277, -0.3838, -0.5312, -0.2728, -0.3527, -0.2188, -0.293, -0.2035, -0.3427)^\intercal$;
- $\Pi_1 = \bar{\Pi}_n - \beta\alpha^\intercal$, where $\bar{\Pi}_n = \sum_{t=1}^n R_t F_t^\intercal (\sum_{t=1}^n F_t F_t^\intercal)^{-1}$ with $\{R_t, F_t\}_{t=1}^n$ being the real data in the empirical application;
- Γ is specified as

$$\Gamma = \begin{bmatrix} 0.0312 & 0.0255 & -0.0185 & 0.0591 & 0.0389 & 0.0953 & -0.1515 & 0.2286 & -0.0806 & -0.1659 \\ 0.0346 & -0.0166 & -0.0608 & 0.0743 & 0.0794 & -0.0043 & -0.2194 & 0.2959 & -0.0043 & 0.0016 \\ -0.0304 & 0.0624 & -0.1347 & 0.1054 & -0.0369 & -0.0187 & -0.0989 & 0.3571 & 0.0133 & -0.1731 \\ -0.0414 & 0.0951 & 0.0029 & -0.0497 & -0.0586 & 0.0910 & -0.0903 & 0.1850 & 0.0616 & -0.0865 \\ -0.0570 & -0.0845 & 0.0606 & -0.0143 & -0.1971 & 0.0528 & 0.0403 & 0.1935 & -0.0114 & 0.1141 \\ -0.0649 & -0.0738 & 0.0030 & 0.0335 & 0.0346 & -0.0432 & -0.0787 & 0.2199 & -0.0266 & -0.0013 \\ -0.0334 & -0.1163 & -0.0139 & -0.0218 & -0.0390 & 0.0128 & -0.0645 & 0.1299 & 0.1105 & 0.0097 \\ -0.1029 & 0.0368 & 0.0737 & -0.0005 & -0.1686 & 0.0254 & 0.0184 & 0.0966 & -0.0176 & 0.0596 \\ -0.1153 & 0.0008 & 0.0373 & 0.0185 & -0.0927 & 0.1029 & 0.0546 & 0.0529 & -0.1792 & 0.0798 \\ -0.0737 & -0.0669 & 0.0500 & 0.1466 & -0.1359 & 0.0617 & 0.1090 & 0.0402 & -0.0659 & -0.0440 \end{bmatrix} ;$$

- Σ_v is specified as

$$\Sigma_v = \frac{1}{100}\begin{bmatrix} 0.19 & 0.09 & 0.07 & 0.05 & 0.04 & 0.03 & 0.02 & -0.01 & 0.00 & -0.01 \\ 0.09 & 0.11 & 0.06 & 0.05 & 0.04 & 0.04 & 0.03 & 0.01 & 0.02 & 0.01 \\ 0.07 & 0.06 & 0.10 & 0.05 & 0.04 & 0.04 & 0.03 & 0.03 & 0.02 & 0.01 \\ 0.05 & 0.05 & 0.05 & 0.08 & 0.04 & 0.04 & 0.04 & 0.03 & 0.02 & 0.01 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.08 & 0.05 & 0.05 & 0.05 & 0.04 & 0.03 \\ 0.03 & 0.04 & 0.04 & 0.04 & 0.05 & 0.08 & 0.06 & 0.05 & 0.05 & 0.03 \\ 0.02 & 0.03 & 0.03 & 0.04 & 0.05 & 0.06 & 0.08 & 0.06 & 0.05 & 0.03 \\ -0.01 & 0.01 & 0.03 & 0.03 & 0.05 & 0.05 & 0.06 & 0.10 & 0.07 & 0.05 \\ 0.00 & 0.02 & 0.02 & 0.02 & 0.04 & 0.05 & 0.05 & 0.07 & 0.09 & 0.04 \\ -0.01 & 0.01 & 0.01 & 0.01 & 0.03 & 0.03 & 0.03 & 0.05 & 0.04 & 0.07 \end{bmatrix} .$$

Given the above configurations, we test the hypotheses $\mathrm{H}_0 : \mathrm{rank}(\Pi_0) \leq r$ v.s. $\mathrm{H}_1 : \mathrm{rank}(\Pi_0) > r$ for $r = 3$ at $\alpha = 5\%$. Thus, H_0 holds if and only if $\delta = 0$, in which case $\mathrm{rank}(\Pi_0) < r$. Note that Kleibergen and Paap (2006) instead consider $\mathrm{H}_0' : \mathrm{rank}(\Pi_0) = 1$ v.s. $\mathrm{H}_1' : \mathrm{rank}(\Pi_0) > 1$ so that the possibility $\mathrm{rank}(\Pi_0) < 1$ is excluded. We estimate Π_0 based a sample $\{R_t, F_t\}_{t=1}^n$ of size $n = 330$ (as in Kleibergen and Paap (2006)) that is generated according to the process (B.8). The number of simulation replications is set to be 5,000, while the number of block bootstrap repetitions (with block size 2) is 500 for each simulation replication. We implement the three of our tests in same manner as we did in Section 4, and compare with the multiple KP test (based on the HACC estimator for the long run variance), although the results for the direct application of the KP test are similar and available upon request.

Table B.1 summarizes the simulation results. We find patterns similar to those exhibited in Table 2. In particular, the multiple KP test is severely undersized, and its local power is overall dominated by our tests, though again the test based on numerical derivative estimators (CF-N) is somewhat sensitive to the choices of the step size. The two-step test (CF-T) and the test based on numerical derivative estimators (CF-A), on the other hand, show strong insensitivity to the choices of the tuning parameters.

Finally, following Kleibergen and Paap (2006), we study a stochastic discount factor model based on the conditional capital asset pricing model proposed in the influential work of Jagannathan and Wang (1996). Suppose that $R_t \in \mathbf{R}^m$ is a vector of returns on m assets at time t and $F_t \in \mathbf{R}^k$ is a vector of k common factors at time t. According to the stochastic discount factor model, R_t and F_t are related through

$$E[R_{t+1}F_{t+1}^{\intercal}\gamma_0|\mathcal{I}_t] = \mathbf{1}_m \ , \qquad \text{(B.11)}$$

where $\mathcal{I}_t$ represents information at time t, and $\gamma_0 \in \mathbf{R}^k$ is a vector of risk premia. If

Table B.1. Rejection rates of rank tests for the model (B.8) with $r = 3$, at $\alpha = 5\%$[†]

δ	CF-T						CF-A							
	$\alpha/5$	$\alpha/10$	$\alpha/15$	$\alpha/20$	$\alpha/25$	$\alpha/30$	$n^{-1/5}$	$1.5n^{-1/5}$	$n^{-1/4}$	$1.5n^{-1/4}$	$n^{-1/3}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$
0.00	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.01	0.06	0.07	0.07	0.07	0.08	0.08	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
0.02	0.10	0.10	0.10	0.10	0.10	0.10	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
0.03	0.20	0.20	0.20	0.20	0.20	0.20	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59
0.04	0.35	0.35	0.35	0.35	0.35	0.35	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
0.05	0.54	0.54	0.53	0.53	0.53	0.53	0.87	0.87	0.87	0.87	0.87	0.87	0.79	0.87
0.06	0.71	0.71	0.71	0.71	0.71	0.71	0.95	0.95	0.95	0.95	0.95	0.95	0.89	0.95
0.07	0.85	0.85	0.85	0.84	0.84	0.84	0.98	0.98	0.98	0.98	0.96	0.98	0.96	0.96
0.08	0.93	0.93	0.93	0.93	0.92	0.92	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99
0.09	0.97	0.97	0.97	0.97	0.97	0.97	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99
0.10	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

δ	KP-M	CF-N							
		$n^{-1/5}$	$1.5n^{-1/5}$	$n^{-1/4}$	$1.5n^{-1/4}$	$n^{-1/3}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$
0.00	0.00	0.05	0.05	0.05	0.05	0.04	0.05	0.03	0.04
0.01	0.05	0.21	0.25	0.18	0.22	0.11	0.17	0.06	0.11
0.02	0.10	0.22	0.29	0.16	0.24	0.09	0.15	0.03	0.09
0.03	0.20	0.28	0.37	0.22	0.31	0.12	0.20	0.07	0.13
0.04	0.36	0.39	0.49	0.33	0.42	0.24	0.32	0.14	0.24
0.05	0.53	0.54	0.62	0.49	0.56	0.40	0.48	0.28	0.41
0.06	0.69	0.69	0.75	0.66	0.71	0.58	0.65	0.45	0.58
0.07	0.80	0.83	0.86	0.81	0.84	0.74	0.80	0.63	0.74
0.08	0.87	0.92	0.93	0.90	0.92	0.86	0.90	0.77	0.87
0.09	0.91	0.96	0.97	0.96	0.97	0.94	0.96	0.88	0.94
0.10	0.93	0.99	0.99	0.98	0.99	0.98	0.98	0.95	0.98

[†] The six values under CF-T are the choices of β, and those under CF-A and CF-N are the choices of κ_n as in 27 and (26) respectively.

Table B.2. The p-values for different tests

Panel A: our tests[†]									
Block	CF-T			CF-A			CF-N		
size	$\alpha/10$	$\alpha/15$	$\alpha/20$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$
$b = 1$	0.15	0.15	0.15	0.08	0.08	0.08	0.11	0.12	0.12
$b = 2$	0.15	0.15	0.15	0.10	0.09	0.09	0.11	0.11	0.12
$b = 3$	0.18	0.18	0.18	0.10	0.10	0.10	0.13	0.13	0.14
$b = 4$	0.16	0.16	0.16	0.08	0.08	0.08	0.13	0.13	0.14
Panel B: the KP-M test[‡]									
0.91									

[†] The three values under CF-T are the choices of β, and those under CF-A and CF-N are the choices of κ_n as in 27 and (26) respectively.

[‡] The p-value for KP-M is given by the smallest significance level such that the null hypothesis is rejected, which is equal to the maximum p-value of all Kleibergen and Paap (2006)'s tests implemented by the multiple testing method.

$\{R_t, F_t\}$ is governed by a stationary linear process:

$$R_t = \Pi_0 F_t + \varepsilon_t \tag{B.12}$$

where $E[\epsilon_{t+1} F_{t+1} | \mathcal{I}_t] = 0$ and $E[F_{t+1} F_{t+1}^\intercal]$ is nonsingular, then γ_0 is identified if and only if the coefficient matrix Π_0 is of full rank. For this, we may test $\mathrm{H}_0 : \mathrm{rank}(\Pi_0) \leq r$ v.s. $\mathrm{H}_1 : \mathrm{rank}(\Pi_0) > r$ with $r = k - 1$.

We use the same data set as in Kleibergen and Paap (2006). There are returns R_t on 10 portfolios and 4 factors in F_t with monthly observations from July 1963 to December 1990, so $m = 10$, $k = 4$ and $n = 330$. The factors in F_t consist of constant, the return on a value-weighted portfolio, a corporate bond yield spread and a measure of per capita labor income growth. We estimate Π_0 by

$$\hat{\Pi}_n = \sum_{t=1}^{n} R_t F_t^\intercal \Big(\sum_{t=1}^{n} F_t F_t^\intercal\Big)^{-1} . \tag{B.13}$$

Since the return sequence $\{R_t\}$ exhibits first order autocorrelation, we thus follow Kleibergen and Paap (2006) and compute the KP statistic by employing the HACC estimator with one lag (West, 1997) for the long run covariance matrix. We implement our CF-T, CF-A and CF-N tests by adopting the block bootstrap (Lahiri, 2003) with block size $b = 1, 2, 3, 4$, employing the same choices of tuning parameters as before, and setting the number of bootstrap repetitions to be 1,000.

Table B.2 reports the p-values of CF-T, CF-A and CF-N, as well as that of the KP-M test. The differences between our p-values and those of the KP-M tests are substantial: ours are uniformly less than 20% while the latter are over 90%. Thus, while the KP-M test strongly support the null, our tests are inconclusive depending on the significance levels and of course also the choices of the tuning parameters. It is worth noting that our three tests are quite insensitive across all choices of tuning parameters and the block size; in particular, the p-values of CF-T and CF-A are invariant to these choices.

Appendix C Estimation of the Rank

There are settings as evident in Examples E.1 and E.3-E.5 in Appendix E.2 where one would like to construct an estimate of the rank. The need of rank estimation is further reinforced should one deem our test based on (28) desirable. Following Cragg and Donald (1997) and Robin and Smith (2000), we adopt a sequential testing procedure that has been previously employed in the literature of model selection (Pötscher, 1983; Bauer et al., 1988; Hosoya, 1989).[2]

Specifically, one may progressively test if the true rank is equal to 0, 1,. . . , $k-1$ and set the estimator $\hat{r}_n$ to be the smallest $r \in \{0, 1, \ldots, k-1\}$ that cannot be rejected if such a r exists and to be k if it does not. The conventional setup (2) then suits well to this end because the possibility of $\text{rank}(\Pi_0)$ strictly smaller than the hypothesized value is "ruled out" in each step by previous test(s). However, we argue that accommodating the possibility $\text{rank}(\Pi_0) < r$, as we do in what follows, may once again lead to more reliable results. Heuristically, there are two possible errors involved in the procedure, namely, falsely rejecting a true null (i.e., type I error) and not rejecting a false null (i.e., type II error). Sequentially testing nulls of the form (2) ignores type I errors potentially made in previous steps, and may have trivial or poor power when Π_0 is local to a matrix whose rank is "small", i.e., the capability of controlling type II error is limited. These are the two channels through which our rank estimator improves upon existing ones. Given a confidence level $1-\alpha$, we formally define the rank estimator $\hat{r}_n$ as

$$\hat{r}_n = \min\{r = 0, \ldots, k-1 : \tau_n^2 \phi_r(\hat{\Pi}_n) \leq \hat{c}_{n,1-\alpha}(r)\} \tag{C.1}$$

if the set is nonempty, and $\hat{r}_n = k$ if the set is empty, where $\hat{c}_{n,1-\alpha}(r)$ is defined by (29) for which we also make its dependence on r explicit.

The following theorem shows that the estimator $\hat{r}_n$ in (C.1) picks up the true rank with probability at least $1-\alpha$ (asymptotically).

Theorem C.1. *Let Assumptions 3.1 and 3.2 hold, and the cdf of the limiting law in* (20) *when $r = r_0$ be continuous and strictly increasing at its $(1-\alpha)$-quantile for $\alpha \in (0, 1)$.*

[2] One may alternatively employ information criteria as in Cragg and Donald (1997). We do not pursue this possibility here in order to coherently present what is essential to our paper.

Then the rank estimator $\hat{r}_n$ *defined by* (C.1) *satisfies*

$$\lim_{n\to\infty} P(\hat{r}_n = r_0) = \begin{cases} 1-\alpha & \text{if } r_0 < k \\ 1 & \text{if } r_0 = k \end{cases}, \tag{C.2}$$

$\lim_{n\to\infty} P(\hat{r}_n < r_0) = 0$*, and* $\lim_{n\to\infty} P(\hat{r}_n > r_0) = \alpha$ *(for* $r_0 < k$*).*

Theorem C.1 implies that the procedure will select an estimator that is no smaller than the truth (asymptotically), and the probability of choosing a larger value (i.e., false selection) is controlled by the significance level α – see Johansen (1995) for related results in cointegration settings. These properties are intrinsically connected to the size control and consistency of our test. Moreover, by Theorem C.1, the sequential procedure can be utilized in our two-step test to provide a preliminary rank estimator, although we stress that existing tests can also be employed in this regard – the downside of these tests is that they may yield less accurate rank estimators as argued previously.

While the construction of a "confidence set/singleton" is of interest in its own right, one may also be interested in obtaining a consistent estimator, for which the probability of false selection should be negligible. One such an estimator is given by (27) or Lemma D.7, where a tuning parameter is involved. This estimator is somewhat crude in that the probability of false selection is unclear and appears challenging to control. Employing the sequential procedure, we may achieve consistency while controlling the estimation error. As suggested by (C.2) and noted in the literature (Pötscher, 1983), we must adjust the significance level $\alpha = \alpha_n$ according to the sample size so that $\alpha_n \to 0$ at a suitable rate, in order to obtain a consistent estimator. This turns out to be nontrivial in the current setup (where $\text{rank}(\Pi_0) \le r$ is tested in each step) as we elaborate next.

If one sequentially tests $\text{rank}(\Pi_0) = r$ for $r = 0, \ldots, k-1$ based on, for example, Cragg and Donald (1997) or Kleibergen and Paap (2006), the critical values are then obtained from chi-squared distributions. The rate at which α_n should tend to zero in order to deliver consistency has been well understood in this case by exploiting the analytic expansions of the cdfs of chi-squares – see Theorem 5.8 in Pötscher (1983) for this result, Cragg and Donald (1997) for an application of it in rank estimation, and Andrews (1999) in moment selection. There are, unfortunately, two challenges for us. First, the limiting distributions whose critical values we aim to approximate is highly nonstandard in general, and as a result, deriving rate conditions on α_n through analytic expansions appears challenging to us. Second, our critical values are obtained through bootstrap, and we believe that it is nontrivial to control the sample uncertainty embodied in these critical values as $\alpha_n \downarrow 0$. Nonetheless, we show that the our rank estimator is consistent under the same rate conditions on α_n as in Cragg and Donald (1997) and Robin and Smith (2000). To formalize our discussions below, we thus impose:

Assumption C.1. $\{\alpha_n\}_{n=1}^{\infty}$ *satisfy (i)* $\alpha_n \downarrow 0$*, and (ii)* $\tau_n^{-2} \log \alpha_n \to 0$.

Assumption C.1 is quite mild in that it merely requires that, loosely speaking, α_n approach zero slower than exponentially decaying rates (not too fast). In this way, it encompasses a wide range of choices for α_n. Given the adjusted significance level α_n, we may now formally define the rank estimator to be

$$\tilde{r}_n = \min\{r = 0, \ldots, k-1 : \tau_n^2 \phi_r(\hat{\Pi}_n) \leq \hat{c}_{n,1-\alpha_n}(r)\} \tag{C.3}$$

if the set is nonempty, and $\tilde{r}_n = k$ if the set is empty.

The next theorem establishes if Assumption C.1 holds and $\mathcal{M}$ is Gaussian (in addition to previous assumptions), then the estimator $\tilde{r}_n$ is indeed consistent.

Theorem C.2. *Suppose that Assumptions 3.1, 3.2 and C.1 hold. Let $\tilde{r}_n$ be given by* (C.3)*. If $\mathcal{M}$ is Gaussian but not constant (in $\mathbf{M}^{m\times k}$), then $\lim_{n\to\infty} P(\tilde{r}_n = r_0) = 1$.*

We reiterate that Theorem C.2 may be of use not only in estimation problems but also in conducting our rank test based on the analytic derivative estimator – see (28) and Lemma D.6. On a technical note, the Gaussianity condition plays an instrumental but not essential role. Concretely, it allows us to relate the significance levels to the corresponding critical values through a concentration inequality for Gaussian random vectors/matrices – see Lemmas D.8-D.10. Thus, this condition can be relaxed whenever a suitable concentration inequality for $\mathcal{M}$ is available (Ledoux, 2001).

Proof of Theorem C.1: For notational simplicity, define: for $r = 0, \ldots, k-1$,

$$A_{n,r} = \{\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}(r)\} \ , \tag{C.4}$$

i.e., $A_{n,r}$ are the events of rejecting the nulls. Consider the first the case when $r_0 = k$. Then we must have $\{\hat{r}_n = r_0\} = A_{n,0} \cap A_{n,1} \cap \cdots \cap A_{n,k-1}$ and hence

$$\begin{aligned}\liminf_{n\to\infty} P(\hat{r}_n = r_0) &= \liminf_{n\to\infty} P(A_{n,0} \cap A_{n,1} \cap \cdots \cap A_{n,k-1}) \\ &\geq 1 - \sum_{r=0}^{k-1} [1 - \liminf_{n\to\infty} P(A_{n,r})] = 1 \ ,\end{aligned} \tag{C.5}$$

where the inequality follows from the Boole's inequality, and the last step is because of the consistency result of Theorem 3.2.

Next, suppose $r_0 < k$. Then $\{\hat{r}_n = r_0\} = A_{n,0} \cap \cdots \cap A_{n,r_0-1} \cap A_{n,r_0}^c$ and hence

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n = r_0) &= \limsup_{n\to\infty} P(A_{n,0} \cap \cdots \cap A_{n,r_0-1} \cap A_{n,r_0}^c) \\ &\leq \limsup_{n\to\infty} P(A_{n,r_0}^c) = 1 - \liminf_{n\to\infty} P(A_{n,r_0}) = 1 - \alpha \ ,\end{aligned} \tag{C.6}$$

where the last step follows from the first claim of Theorem 3.2. Moreover,

$$\begin{aligned}\liminf_{n\to\infty} P(\hat{r}_n = r_0) &= \liminf_{n\to\infty} P(A_{n,0}\cap\cdots\cap A_{n,r_0-1}\cap A_{n,r_0}^c)\\ &\geq 1-\sum_{r=0}^{r_0-1}[1-\liminf_{n\to\infty} P(A_{n,r})] - \limsup_{n\to\infty} P(A_{n,r_0}) = 1-\alpha\ , \end{aligned}\tag{C.7}$$

where we exploited the size control and the consistency results in Theorem 3.2.

Turning to the second claim, note that if $\hat{r}_n < r_0$, then $r_0 > 0$ and $\{\hat{r}_n < r_0\} \subset A_{n,0}^c \cup\cdots\cup A_{n,r_0-1}^c$. It follows that

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n < r_0) &\leq \limsup_{n\to\infty} P(A_{n,0}^c\cup\cdots\cup A_{n,r_0-1}^c)\\ &\leq \sum_{r=0}^{r_0-1}\limsup_{n\to\infty} P(A_{n,r}^c) = \sum_{r=0}^{r_0-1}[1-\liminf_{n\to\infty} P(A_{n,r})] = 0\ , \end{aligned}\tag{C.8}$$

where the last step is because of the consistency result of Theorem 3.2. The last claim is a simple implication of the first two claims. We are thus done. ■

Proof of Theorem C.2: For notational simplicity, define: for $r = 0,\ldots,k-1$,

$$A_{n,r} = \{\tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha_n}(r)\}\ . \tag{C.9}$$

First, note that $\tilde{r}_n < r_0$ if and only if $r_0 \geq 1$ and

$$\{\tilde{r}_n = r\} = A_{n,0}\cap\cdots\cap A_{n,r-1}\cap A_{n,r}^c \tag{C.10}$$

for some $r = 0,\ldots,r_0-1$. Fix $r \in \{0,1,\ldots,r_0-1\}$. It follows from (C.10) that

$$P(\tilde{r}_n = r) \leq P(A_{n,r}^c) = 1 - P(\phi_r(\hat{\Pi}_n) > \frac{\hat{c}_{n,1-\alpha_n}}{\tau_n^2}) \to 0\ , \tag{C.11}$$

where we exploited $\hat{c}_{n,1-\alpha_n}/\tau_n^2 = o_p(1)$ by Assumption C.1(ii), Lemma D.10 and $\phi_r(\hat{\Pi}_n) \xrightarrow{p} \phi_r(\Pi_0) > 0$ by the continuous mapping theorem and $\text{rank}(\Pi_0) \equiv r_0 > r$. Since the result (C.11) is true for any $r = 0,\ldots,r_0-1$, we thus obtain

$$\limsup_{n\to\infty} P(\tilde{r}_n < r_0) = 0\ . \tag{C.12}$$

Next, note that $\tilde{r}_n > r_0$ if and only if $r_0 \leq k-1$ and either the relation (C.10) holds for some $r = r_0+1,\ldots,k-1$ or the following event occurs

$$\{\tilde{r}_n = k\} = A_{n,0}\cap\cdots\cap A_{n,k-1}\cap A_{n,k}\ . \tag{C.13}$$

Hence, $\{\tilde{r}_n = r\} \subset A_{n,r_0}$ for all $r = r_0 + 1, \ldots, k$. Fix $r = \{r_0 + 1, \ldots, k\}$. We thus have

$$P(\tilde{r}_n = r) \le P(A_{n,r_0}) = P(\tau_n^2 \phi_{r_0}(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha_n}) \ . \tag{C.14}$$

Fix $\epsilon \in (0,1)$ so that $c_{1-\epsilon}$ is a continuity point of the cdf F of $\phi''_{r,\Pi_0}(\mathcal{M})$. This can be done without loss of generality because the set of discontinuity points is countable. By Assumption C.1(i), it holds that: for all n sufficiently large,

$$F(c_{1-\epsilon}) = 1 - \epsilon < 1 - \alpha_n \ , \tag{C.15}$$

and hence $c_{1-\alpha_n} > c_{1-\epsilon}$. In turn, we obtain from (C.15) and Assumption C.1(i) that there exists some $\delta > 0$ satisfying: for all n sufficiently large,

$$F(c_{1-\epsilon}) + \delta < 1 - \alpha_n \ . \tag{C.16}$$

Note that if $\hat{c}_{n,1-\alpha_n} \le c_{1-\epsilon}$, then we obtain from (C.16) that

$$F(c_{1-\epsilon}) + \delta < 1 - \alpha_n \le \hat{F}_n(\hat{c}_{n,1-\alpha_n}) \le \hat{F}_n(c_{1-\epsilon}) \ . \tag{C.17}$$

By Lemma 10.11 in Kosorok (2008), we may thus conclude that

$$\limsup_{n\to\infty} P(\hat{c}_{n,1-\alpha_n} \le c_{1-\epsilon}) \le \limsup_{n\to\infty} P(\hat{F}_n(c_{1-\epsilon}) - F(c_{1-\epsilon}) > \delta) = 0 \ . \tag{C.18}$$

Combination of results (C.14) and (C.18), together with Assumption 3.1, now yields

$$\limsup_{n\to\infty} P(\tilde{r}_n = r) \le \limsup_{n\to\infty} P(\tau_n^2 \phi_{r_0}(\hat{\Pi}_n) > c_{1-\epsilon}) = 1 - F(c_{1-\epsilon}) \le \epsilon \ . \tag{C.19}$$

Since $\epsilon > 0$ and $r \in \{r_0 + 1, \ldots, k\}$ are both arbitrary, it follows from (C.19) that

$$\limsup_{n\to\infty} P(\tilde{r}_n > r_0) = 0 \ . \tag{C.20}$$

The theorem now follows from results (C.12) and (C.20) since

$$\liminf_{n\to\infty} P(\tilde{r}_n = r_0) \ge 1 - \limsup_{n\to\infty} P(\tilde{r}_n < r_0) - \limsup_{n\to\infty} P(\tilde{r}_n > r_0) = 1 \ . \qquad \blacksquare$$

Appendix D Auxiliary Lemmas

Lemma D.1. *Suppose* $\Pi \in \mathbf{M}^{m\times k}$ *with* $\Pi \neq 0$ *and rank*$(\Pi) \le r$. *For* $\epsilon > 0$, *let* $\Psi(\Pi)_1^\epsilon$ *be given as in the proof of Proposition 3.1. Let* $\sigma^+_{\min}(\Pi)$ *be the smallest positive singular value of* Π. *Then for all sufficiently small* $\epsilon > 0$, *we have*

$$\min_{U \in \Psi(\Pi)_1^\epsilon} \|\Pi U\| \ge \frac{\sqrt{2}}{2}\sigma^+_{\min}(\Pi)\epsilon \ .$$

Proof: Let $\Pi = P\Sigma Q^\intercal$ be a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$, $Q \in \mathbb{S}^{k\times k}$, and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Let $d \equiv k - r$ and $d_0 \equiv k - r_0$ with $r_0 \equiv \mathrm{rank}(\Pi)$. For $U \in \mathbb{S}^{k\times d}$, let $U_Q \equiv Q^\intercal U$ and write $U_Q^\intercal = [U_Q^{(1)\intercal}, U_Q^{(2)\intercal}]$ such that $U_Q^{(1)} \in \mathbf{M}^{r_0\times d}$. Then we have that for $U \in \mathbb{S}^{k\times d}$,

$$\|\Pi U\| = \|P\Sigma Q^\intercal U\| = \|\Sigma U_Q\| \geq \sigma^+_{\min}(\Pi)\|U_Q^{(1)}\| \ , \tag{D.1}$$

where the second equality follows by $P^\intercal P = I_m$, and the inequality follows by the fact that Σ is diagonal with diagonal entries in descending order with $\sigma^+_{\min}(\Pi) = \sigma_{r_0}(\Pi)$ the smallest positive entry. Let $U_Q^{(2)} = P_U^{(2)}\Sigma_U^{(2)}Q_U^{(2)\intercal}$ be a singular value decomposition of $U_Q^{(2)}$ where $Q_U^{(2)} \in \mathbb{S}^{d\times d}$, $P_U^{(2)} \in \mathbb{S}^{d_0\times d_0}$ and $\Sigma_U^{(2)} \in \mathbf{M}^{d_0\times d}$. Since $r_0 \leq r$ and hence $d_0 \geq d$, it follows that, for $U \in \mathbb{S}^{k\times d}$,

$$\|U_Q^{(2)}\|^2 = \sum_{j=1}^{d}\sigma_j^2(U_Q^{(2)}) \leq \sum_{j=1}^{d}\sigma_j(U_Q^{(2)}) = \mathrm{tr}([I_d, \mathbf{0}_{r-r_0}]\Sigma_U^{(2)}) \ , \tag{D.2}$$

where the inequality follows by the fact that $\sigma_j(U_Q^{(2)}) \in [0,1]$ as singular values of $U_Q^{(2)}$ due to $U_Q^{(2)\intercal}U_Q^{(2)} + U_Q^{(1)\intercal}U_Q^{(1)} = I_d$, and the second equality follows by noting that the diagonal entries of $\Sigma_U^{(2)}$ are singular values of $U_Q^{(2)}$. Since $\|U_Q^{(1)}\|^2 + \|U_Q^{(2)}\|^2 = \|U_Q\|^2 = d$, thus combining (D.1) and (D.2) yields that for $U \in \mathbb{S}^{k\times d}$,

$$\|\Pi U\| \geq \sigma^+_{\min}(\Pi)\sqrt{d - \mathrm{tr}([I_d, \mathbf{0}_{r-r_0}]\Sigma_U^{(2)})} \ . \tag{D.3}$$

Since $\|U_Q^{(1)}\|^2 + \|\Sigma_U^{(2)}\|^2 = \|U_Q^{(1)}\|^2 + \|U_Q^{(2)}\|^2 = d$ and $\|[I_d, \mathbf{0}_{r-r_0}]^\intercal\|^2 = d$, then simple algebra yields that for $U \in \mathbb{S}^{k\times d}$,

$$2(d - \mathrm{tr}([I_d, \mathbf{0}_{d-r_0}]\Sigma_U^{(2)})) = \|U_Q^{(1)}\|^2 + \|\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r_0}]^\intercal\|^2 \ . \tag{D.4}$$

Write $Q = [Q_1, Q_2]$ such that $Q_1 \in \mathbf{M}^{k\times r_0}$. Since $Q_1^\intercal Q_1 = I_{r_0}$, $Q_2^\intercal Q_2 = I_{d_0}$ and $Q_1^\intercal Q_2 = 0$ as well as $P_U^{(2)}$ and $Q_U^{(2)}$ are orthonormal, we then have that for $U \in \mathbb{S}^{k\times d}$,

$$\|U_Q^{(1)}\|^2 + \|\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r_0}]^\intercal\|^2 = \|Q_1U_Q^{(1)} + Q_2P_U^{(2)}(\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r_0}]^\intercal)Q_U^{(2)\intercal}\|^2 \ . \tag{D.5}$$

Since $U_Q^{(1)} = Q_1^\intercal U$ and $U_Q^{(2)} = Q_2^\intercal U$ by construction and $Q_1Q_1^\intercal U + Q_2Q_2^\intercal U = U$ by $QQ^\intercal = I_k$, we then have that, for $U \in \mathbb{S}^{k\times d}$,

$$|Q_1U_Q^{(1)} + Q_2P_U^{(2)}(\Sigma_2 - [I_d, \mathbf{0}_{r-r_0}]^\intercal)Q_U^{(2)\intercal}\|^2 = \|U - Q_2P_U^{(2)}[I_d, \mathbf{0}_{r-r_0}]^\intercal Q_U^{(2)\intercal}\|^2 \ . \tag{D.6}$$

Noting that $Q_2P_U^{(2)}[I_d, \mathbf{0}_{r-r_0}]^\intercal Q_U^{(2)\intercal} \in \Psi(\Pi)$, we have by (D.4)-(D.6) that, for $U \in \mathbb{S}^{k\times d}$,

$$2(d - \mathrm{tr}([I_d, \mathbf{0}_{r-r_0}]\Sigma_U^{(2)})) \geq \min_{U' \in \Psi(\Pi)} \|U - U'\|^2 \ . \tag{D.7}$$

Since $\Pi \neq 0$, then $\Psi(\Pi)_1^\epsilon \neq \emptyset$ for all sufficiently small $\epsilon > 0$. Fix such an $\epsilon > 0$. By the definition of $\Psi(\Pi)_1^\epsilon$, combining (D.3) and (D.7) yields that for all $U \in \Psi(\Pi)_1^\epsilon$,

$$\|\Pi U\| \geq \frac{\sqrt{2}}{2}\sigma_{\min}^{+}(\Pi) \min_{U' \in \Psi(\Pi)} \|U - U'\| \geq \frac{\sqrt{2}}{2}\sigma_{\min}^{+}(\Pi)\epsilon \ . \tag{D.8}$$

Then the lemma follows by applying minimum over $\Psi(\Pi)_1^\epsilon$ to both sides of (D.8) and noting that the result continues to hold for all sufficiently small $\epsilon > 0$. ■

Lemma D.2. *The correspondence φ in the proof of Proposition 3.1 is continuous at* 0.

Proof: Fix $U_0 \in \Psi(\Pi)$, and define the correspondence $\bar{\varphi} : \mathbf{R} \twoheadrightarrow \Gamma^\Delta$ by $\bar{\varphi}(t) = \Gamma^\Delta_{U_0}(t)$, where $\Psi(\Pi)$, Γ^Δ and $\Gamma^\Delta_{U_0}(t)$ are given in the proof of Proposition 3.1. Let $d \equiv k - r$. For each t_n and each $V_0 \in \bar{\varphi}(0)$, define $f : \Gamma^\Delta \to \mathbf{M}^{k\times d}$ by

$$f(V) = V_0 - \frac{t_n}{2} U_0 V^\intercal V \ .$$

Since f is continuous and Γ^Δ is compact, f is a compact map in the sense of Granas and Dugundji (2003). By Theorem 0.2.3 in Granas and Dugundji (2003), one of the following two cases must happen: (i) f has a fixed point $V_{1n} \in \Gamma^\Delta$, and (ii) there exists some $V_{2n} \in \Gamma^\Delta$ such that $\|V_{2n}\| = \Delta$ and $V_{2n} = \lambda_n f(V_{2n})$ with $\lambda_n \equiv \frac{\Delta}{\|f(V_{2n})\|} \in (0, 1)$. In case (i), since $U_0 \in \Psi(\Pi)$, $V_0 \in \bar{\varphi}(0)$ and $f(V_{1n}) = V_{1n}$, we have by simple algebra:

$$V_{1n}^\intercal U_0 + U_0^\intercal V_{1n} = (V_0 - \frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n})^\intercal U_0 + U_0^\intercal (V_0 - \frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n}) = -t_n V_{1n}^\intercal V_{1n} \ . \tag{D.9}$$

This together with $V_{1n} \in \Gamma^\Delta$ implies that $V_{1n} \in \bar{\varphi}(t_n)$. Moreover, since $f(V_{1n}) = V_{1n}$, $\|U_0\| = \sqrt{d}$ and $V_{1n} \in \Gamma^\Delta$, then by the sub-multiplicativity of Frobenius norm we have

$$\|V_{1n} - V_0\| = \|\frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n}\| \leq \frac{t_n}{2}\sqrt{d}\Delta^2 \ . \tag{D.10}$$

In case (ii), since $U_0 \in \Psi(\Pi)$, $\lambda_n^2 V_0 \in \bar{\varphi}(0)$ and $\lambda_n V_{2n} = \lambda_n^2 f(V_{2n})$, then by analogous calculations as in (D.9), we have

$$(\lambda_n V_{2n})^\intercal U_0 + U_0^\intercal (\lambda_n V_{2n}) = -t_n (\lambda_n V_{2n})^\intercal (\lambda_n V_{2n}) \ .$$

This together with $\lambda_n V_{2n} \in \Gamma^\Delta$ due to $\lambda_n \in (0, 1)$ and $V_{2n} \in \Gamma^\Delta$ implies that $\lambda_n V_{2n} \in \bar{\varphi}(t_n)$. Moreover, since $\lambda_n V_{2n} = \lambda_n^2 f(V_{2n})$, similar to (D.10) we have:

$$\|\lambda_n V_{2n} - V_0\| \leq \|\lambda_n^2 f(V_{2n}) - \lambda_n^2 V_0\| + |\lambda_n^2 - 1|\|V_0\| \leq \frac{t_n}{2}\sqrt{d}\Delta^2 + |\lambda_n^2 - 1|\Delta \ , \tag{D.11}$$

where the first inequality follows the triangle inequality and the second inequality follows since $\lambda_n \in (0, 1)$. Now, for each $n \in \mathbf{N}$, define V_n^* to be V_{1n} if case (i) happens and $\lambda_n V_{2n}$ otherwise. Let $\delta_n \equiv 1$ if case (i) happens and $\delta_n \equiv \lambda_n$ otherwise. Then $V_n^* \in \Gamma^\Delta_{U_0}(t_n)$

for all $n \in \mathbf{N}$, and combination of (D.10) and (D.11) yields

$$\|V_n^* - V_0\| \leq \frac{t_n}{2}\sqrt{d}\Delta^2 + |\delta_n^2 - 1|\Delta \to 0 \ ,$$

where we exploited the fact that if V_{2n} exists infinitely often, $\delta_n = \lambda_n = \frac{\Delta}{\|f(V_{2n})\|} \to 1$ due to $f(V_{2n}) \to V_0$ as $n \to \infty$ and $\|V_0\| \leq \Delta$, and $t_n \to 0$ as $n \to \infty$. It follows that $\bar{\varphi}(t)$ is lower hemicontinuous at $t = 0$ by Theorem 17.21 in Aliprantis and Border (2006).

The lower hemicontinuity of $\varphi(t)$ at $t = 0$ follows easily from that of $\bar{\varphi}(t)$ again by Theorem 17.21 in Aliprantis and Border (2006). To see this, let $t_n \to 0$ and $(U_0, V_0) \in \varphi(0)$. Define (U_n^*, V_n^*) to be $U_n^* = U_0$ and V_n^* be as in previous construction for all $n \in \mathbf{N}$. Clearly, $(U_n^*, V_n^*) \to (U_0, V_0)$, implying that $\varphi(t)$ is lower hemicontinuous at $t = 0$. Since $\varphi(t)$ is contained in the compact set $\mathbb{S}^{k\times d} \times \Gamma^{\Delta}$ for all t, $\varphi(t)$ is upper hemicontinuous at $t = 0$ by Theorem 17.20 in Aliprantis and Border (2006). We have therefore showed that $\varphi(t)$ is continuous at $t = 0$. ■

Lemma D.3. *Suppose* $\Pi \in \mathbf{M}^{m\times k}$ *with rank*$(\Pi) \leq r$, *and* $M \in \mathbf{M}^{m\times k}$ *with* $M \neq 0$. *Let* $\Psi(\Pi) = \arg\min_{U \in \mathbb{S}^{k\times(k-r)}} \|\Pi U\|^2$, *and for* $U \in \Psi(\Pi)$ *and* $\Delta > 0$ *let* $\Gamma_U^{\Delta}(0)$ *be as in the proof of Proposition 3.1. For* Δ *sufficiently large, it follows that for all* $U \in \Psi(\Pi)$,

$$\min_{V \in \Gamma_U^{\Delta}(0)} \|\Pi V + MU\|^2 = \min_{V \in \mathbf{M}^{k\times(k-r)}} \|\Pi V + MU\|^2 \ .$$

Proof: The conclusion is trivial if $\Pi = 0$. Suppose that $\Pi \neq 0$ and let $d \equiv k - r$. Let $r_0 = \text{rank}(\Pi)$ and $\Pi = P\Sigma Q^{\intercal}$ be a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$, $Q \in \mathbb{S}^{k\times k}$, and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Since $\Pi \neq 0$ and $r_0 \leq r$, we may write $\Sigma = [\Sigma_1, 0]$ with $\Sigma_1 \in \mathbf{M}^{m\times r_0}$ of full rank so that

$$\min_{V \in \mathbf{M}^{k\times d}} \|\Pi V + MU\|^2 = \min_{V \in \mathbf{M}^{r_0\times d}} \|[P\Sigma_1 V + MU\|^2 \ . \tag{D.12}$$

By the projection theorem, the minimum on the right hand side of (D.12) is attained at some point, say $V_1^* \in \mathbf{M}^{r_0\times d}$. Moreover, V_1^* is uniformly bounded over $U \in \Psi(\Pi)$. Let $V^* \equiv Q[V_1^{*\intercal}, 0]^{\intercal} \in \mathbf{M}^{k\times d}$, then the minimum on the left hand side of (D.12) is attained at V^*. Decompose Q as $Q = [Q_1, Q_2]$, where $Q_1 \in \mathbf{M}^{k\times r_0}$. Then $V^* = Q_1 V_1^* \in \Gamma_U^{\Delta}(0)$ for all $U \in \Psi(\Pi)$, when Δ is sufficiently large. It implies that the minimum on the right hand side of (D.12) is attained within $\Gamma_U^{\Delta}(0)$ as well for all $U \in \Psi(\Pi)$, when Δ is sufficiently large. This implies that when Δ is sufficiently large,

$$\min_{V \in \Gamma_U^{\Delta}(0)} \|\Pi V + MU\|^2 \leq \min_{V \in \mathbf{M}^{k\times d}} \|\Pi V + MU\|^2$$

for all $U \in \Psi(\Pi)$. The reverse inequality is simply true since $\Gamma_U^{\Delta}(0) \subset \mathbf{M}^{k\times d}$ all $U \in \Psi(\Pi)$ and all $\Delta > 0$. This completes the proof of the lemma. ■

Lemma D.4. *If $r_0 \equiv \operatorname{rank}(\Pi) \leq r$, then for any $M \in \mathbf{M}^{m\times k}$,*

$$\min_{U\in\Psi(\Pi)} \min_{V\in\mathbf{M}^{k\times d}} \|\Pi V + MU\|^2 = \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_2^{\intercal} M Q_2) \ , \tag{D.13}$$

where $\Psi(\Pi) = \arg\min_{U\in\mathbb{S}^{k\times(k-r)}} \|\Pi U\|^2$.

Proof: Let $d \equiv k - r$ and $d_0 \equiv k - r_0$. Noting that the column vectors in Q_2 form a orthonormal basis for the null space of Π_0, we may rewrite $\Psi(\Pi)$ as $\Psi(\Pi) = \{Q_2 V : V \in \mathbb{S}^{d_0\times d}\}$. This, together with the projection theorem, implies

$$\phi''_{r,\Pi}(M) = \min_{V\in\mathbb{S}^{d_0\times d}} \|(I - \Pi(\Pi^{\intercal}\Pi)^{-}\Pi^{\intercal})MQ_2V\|^2 \ , \tag{D.14}$$

where A^{-} denotes the Moore-Penrose inverse of a generic matrix A. By the singular value decomposition of Π, we have

$$\begin{aligned}(I - \Pi(\Pi^{\intercal}\Pi)^{-}\Pi^{\intercal})P &= P - P\Sigma Q^{\intercal}(Q\Sigma^{\intercal}P^{\intercal}P\Sigma Q^{\intercal})^{-}Q\Sigma^{\intercal}P^{\intercal}P \\ &= P - P\Sigma Q^{\intercal}Q(\Sigma^{\intercal}P^{\intercal}P\Sigma)^{-}Q^{\intercal}Q\Sigma^{\intercal}P^{\intercal}P = P - P\Sigma(\Sigma^{\intercal}\Sigma)^{-}\Sigma^{\intercal} = [0, P_2] \ , \end{aligned} \tag{D.15}$$

where the second equality exploited Theorem 20.5.6 in Harville (2008), the third equality follows from P and Q being orthonormal, and the fourth equality is obtained by carrying out the Moore-Penrose inverse by Exercise 2.7.4 in Magnus and Neudecker (2007) and noting that Σ is diagonal. In view of (D.15), we have

$$\begin{aligned}&\min_{V\in\mathbb{S}^{d_0\times d}} \|(I - \Pi(\Pi^{\intercal}\Pi)^{-}\Pi^{\intercal})MQ_2V\|^2 = \min_{V\in\mathbb{S}^{d_0\times d}} \|[0, P_2]P^{\intercal}MQ_2V\|^2 \\ &= \min_{V\in\mathbb{S}^{d_0\times d}} \|P_2P_2^{\intercal}MQ_2V\|^2 = \min_{V\in\mathbb{S}^{d_0\times d}} \|P_2^{\intercal}MQ_2V\|^2 = \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_2^{\intercal}MQ_2) \ , \end{aligned} \tag{D.16}$$

where the third equality follows from $P_2^{\intercal}P_2 = I_{m-r_0}$ and the final equality follows from Lemma 3.1. Combining (D.14) and (D.16) concludes the proof of the lemma. ■

Lemma D.5. *Suppose rank$(\Pi) \leq r$ and let $\phi''_{r,\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ be given as in Proposition 3.1. If rank$(\Pi) = r$, there exists a bilinear map $\Phi''_{r,\Pi} : \mathbf{M}^{m\times k} \times \mathbf{M}^{m\times k} \to \mathbf{R}$ such that $\phi''_{r,\Pi}(M) = \Phi''_{r,\Pi}(M, M)$ for all $M \in \mathbf{M}^{m\times k}$; if rank$(\Pi) < r$, such a $\Phi''_{r,\Pi}$ does not exist.*

Proof: Let $\Pi = P\Sigma Q^{\intercal}$ is a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$ whose last $m - r$ columns constitutes P_2, $Q \in \mathbb{S}^{k\times k}$ whose last $k - r$ columns constitutes Q_2,, and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Let $d \equiv k - r$. If $\operatorname{rank}(\Pi) = r$, then Lemma D.4 and Lemma 3.1 imply

$$\phi''_{r,\Pi}(M) = \min_{V\in\mathbb{S}^{d\times d}} \|P_2^{\intercal}MQ_2V\|^2 = \|P_2^{\intercal}MQ_2\|^2 \ ,$$

for all $M \in \mathbf{M}^{m\times k}$, which is a quadratic form corresponding to the bilinear form $\Phi''_{r,\Pi}(M_1, M_2) \equiv \mathrm{tr}(Q_2^\intercal M_1^\intercal P_2 P_2^\intercal M_2 Q_2)$ for $M_1 \in \mathbf{M}^{m\times k}$ and $M_2 \in \mathbf{M}^{m\times k}$.

Next, assume that $\mathrm{rank}(\Pi) < r$. Suppose for the sake of a contradiction that there exists a bilinear map $\Phi''_{r,\Pi}$ corresponding to $\phi''_{r,\Pi}$. Bilinearity of $\Phi''_{r,\Pi}$ then implies that

$$\phi''_{r,\Pi}(M_1) + \phi''_{r,\Pi}(M_2) = \frac{\phi''_{r,\Pi}(M_1 + M_2) + \phi''_{r,\Pi}(M_1 - M_2)}{2} \tag{D.17}$$

for all $M_1 \in \mathbf{M}^{m\times k}$ and $M_2 \in \mathbf{M}^{m\times k}$. Let $r_0 \equiv \mathrm{rank}(\Pi)$ and $d_0 \equiv k - r_0$. If $M = P_2 H Q_2^\intercal$ for some $H \in \mathbf{M}^{(m-r_0)\times d_0}$, then Lemma D.4 and Lemma 3.1 imply

$$\phi''_{r,\Pi}(M) = \sigma^2_{r-r_0+1}(H) + \cdots + \sigma^2_{d_0}(H) \ . \tag{D.18}$$

Now, let $H_1 \in \mathbf{M}^{(m-r_0)\times d_0}$ be diagonal with the (j,j)th entry equal to 1 for $j = 1, \ldots, d_0$ and $H_2 \in \mathbf{M}^{(m-r_0)\times d_0}$ be diagonal with the (j,j)th entry equal to -1 for $j = 1$ and 1 for $j = 2, \ldots, d_0$. Set $M_i = P_2 H_i Q_2^\intercal$ for $i = 1, 2$, the result in (D.18) implies $\phi''_{r,\Pi}(M_1) = \phi''_{r,\Pi}(M_2) = k - r$, $\phi''_{r,\Pi}(M_1 + M_2) = 4(k-r) - 4$ and $\phi''_{r,\Pi}(M_1 - M_2) = 0$. It follows that

$$2(k-r) = \phi''_{r,\Pi}(M_1) + \phi''_{r,\Pi}(M_2) \neq \frac{\phi''_{r,\Pi}(M_1 + M_2) + \phi''_{r,\Pi}(M_1 - M_2)}{2} = 2(k-r) - 2 \ ,$$

which contradicts the result (D.17). Thus, the second result of the lemma follows. ■

Lemma D.6. *Suppose Assumption 3.1 holds. Let $\hat{\phi}''_{r,n}$ be the analytic estimator given by* (28)*. If $\hat{r}_n \xrightarrow{p} r_0 \equiv \mathrm{rank}(\Pi_0)$ and $r_0 \le r < k$, then condition* (25) *holds.*

PROOF: For notational simplicity, let $d \equiv k - r$ and $\hat{d}_n \equiv k - \hat{r}_n$. Fix a sequence $\{M_n\}$ such that $M_n \to M$ as $n \to \infty$. By Lemma 3.1, we have:

$$\begin{aligned}|\hat{\phi}''_{r,n}(M_n) - \hat{\phi}''_{r,n}(M)| = \big| &\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^\intercal M_n \hat{Q}_{2,n} U\| - \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^\intercal M \hat{Q}_{2,n} U\|\big| \\ &\times \big(\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^\intercal M_n \hat{Q}_{2,n} U\| + \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^\intercal M \hat{Q}_{2,n} U\|\big) \ ,\end{aligned} \tag{D.19}$$

where the inequality follows by the formula $(a^2 - b^2) = (a+b)(a-b)$. For the first term on the right hand side of (D.19), we have

$$\begin{aligned}\big| \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} &\|\hat{P}_{2,n}^\intercal M_n \hat{Q}_{2,n} U\| - \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^\intercal M \hat{Q}_{2,n} U\|\big| \\ &\le \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^\intercal (M_n - M) \hat{Q}_{2,n} U\| \le \sqrt{kmd}\|M_n - M\| = o_p(1) \ ,\end{aligned} \tag{D.20}$$

where the first inequality follows by the Lipschitz continuity of the min operator and the triangle inequality, the second inequality holds by the submultiplicativity of Frobenius norm, $\|\hat{P}_{2,n}\| \le \sqrt{m}$, $\|\hat{Q}_{2,n}\| \le \sqrt{k}$ and $\|U\| = \sqrt{r}$ for all $U \in \mathbb{S}^{\hat{d}_n\times d}$, and the equality is because $M_n \to M$. For the second term on the right hand side of (D.19), once again exploiting the sub-multiplicability of the Frobenius norm, $\|\hat{P}_{2,n}\| \le \sqrt{m}$, $\|\hat{Q}_{2,n}\| \le \sqrt{k}$, $\|U\| = \sqrt{r}$ for all $U \in \mathbb{S}^{\hat{d}_n \times d}$ and $M_n \to M$, we have that

$$\min_{U \in \mathbb{S}^{\hat{d}_n \times d}} \|\hat{P}_{2,n}^{\intercal} M_n \hat{Q}_{2,n} U\| + \min_{U \in \mathbb{S}^{\hat{d}_n \times d}} \|\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n} U\| \le \sqrt{kmd}\|M_n\| + \sqrt{kmd}\|M\| = O(1) \ . \quad \text{(D.21)}$$

Combining results (D.19)-(D.21), then we obtain

$$|\hat{\phi}''_{r,n}(M_n) - \hat{\phi}''_{r,n}(M)| = o_p(1) \ . \quad \text{(D.22)}$$

In view of (D.22), it thus suffices to show that

$$|\hat{\phi}''_{r,n}(M) - \phi''_{r,\Pi_0}(M)| \equiv |\sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n}) - \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^{\intercal} M Q_{0,2})| = o_p(1) \ . \quad \text{(D.23)}$$

Let $\hat{q}_j$ be the jth column of $\hat{Q}_{2,n}$. Since $Q_0 \in \mathbb{S}^{k\times k}$, we may write $\hat{q}_j = Q_0 \hat{u}_j$ for some (random) $\hat{u}_j \in \mathbb{S}^{k\times 1}$. Noting that $\hat{q}_j$ is an eigenvector of $\hat{\Pi}_n^{\intercal}\hat{\Pi}_n$ associated with the eigenvalue $\sigma_{r_0+j}^2(\hat{\Pi}_n)$ when $\hat{r}_n = r_0$ and that $P(\hat{r}_n = r_0) \to 1$ as given, we have

$$[\hat{\Pi}_n^{\intercal}\hat{\Pi}_n - \Pi_0^{\intercal}\Pi_0 - (\sigma_{r_0+j}^2(\hat{\Pi}_n) - \sigma_{r_0+j}^2(\Pi_0))I_k + \Pi_0^{\intercal}\Pi_0 - \sigma_{r_0+j}^2(\Pi_0)I_k]Q_0\hat{u}_j = [\hat{\Pi}_n^{\intercal}\hat{\Pi}_n - \sigma_{r_0+j}^2(\hat{\Pi}_n)I_k]\hat{q}_j = o_p(1) \ . \quad \text{(D.24)}$$

Observe that $\|\hat{\Pi}_n^{\intercal}\hat{\Pi}_n - \Pi_0^{\intercal}\Pi_0\| = o_p(1)$ and $|\sigma_{r_0+j}^2(\hat{\Pi}_n) - \sigma_{r_0+j}^2(\Pi_0)| = o_p(1)$ by the continuous mapping theorem, the Weyl inequality (Tao, 2012, Exercise 1.3.22(iv)) and Assumption 3.1, we then conclude from (D.24) that

$$o_p(1) = [\Pi_0^{\intercal}\Pi_0 - \sigma_{r_0+j}^2(\Pi_0)I_k]Q_0\hat{u}_j = Q_0\Sigma_0^{\intercal}\Sigma_0\hat{u}_j \ , \quad \text{(D.25)}$$

where we exploited the singular value decomposition $\Pi_0 = P_0\Sigma_0 Q_0^{\intercal}$, and the fact that $\sigma_{r_0+j}^2(\Pi_0) = 0$. Since the first r_0 diagonal elements of $\Sigma_0^{\intercal}\Sigma_0$ are positive and Q_0 is nonsingular, we may conclude from result (D.25) that the first r_0 elements of $\hat{u}_j$ are $o_p(1)$ and moreover by the definition of $\hat{q}_j$ that for some random $U_2 \in \mathbb{S}^{(k-r_0)\times(k-r_0)}$,

$$\hat{Q}_{2,n} = Q_{0,2}U_2 + o_p(1) \ , \quad \text{(D.26)}$$

By an analogous argument, we have that for some random $V_2 \in \mathbb{S}^{(m-r_0)\times(m-r_0)}$,

$$\hat{P}_{2,n} = P_{0,2}V_2 + o_p(1) \ . \quad \text{(D.27)}$$

Combining results (D.26) and (D.27) and the continuous mapping theorem yields

$$\|\hat{P}_{2,n}^{\intercal} M\hat{Q}_{2,n} - V_2^{\intercal} P_{0,2}^{\intercal} MQ_{0,2}U_2\| = o_p(1) \ . \tag{D.28}$$

Thus, (D.23) follows from (D.28), the continuous mapping theorem and the fact that the singular values of $V_2^{\intercal} P_{0,2}^{\intercal} MQ_{0,2}U_2$ are equal to those of $P_{0,2}^{\intercal} MQ_{0,2}$. ■

Lemma D.7. *Suppose Assumption 3.1 holds. Let $\hat{r}_n$ be the maximal $j \in \{1, \ldots, k\}$ such that $\sigma_j(\hat{\Pi}_n) \geq \kappa_n$ if such a j exists and $\hat{r}_n = 0$ otherwise. If $\kappa_n \downarrow 0$ and $\tau_n\kappa_n \to \infty$, then it follows that*

$$\lim_{n\to\infty} P(\hat{r}_n = r_0) = 1 \ .$$

Proof: On the one hand, note that if $\hat{r}_n > r_0$, then we must have $r_0 \leq k - 1$, $\sigma_{r_0+1}(\hat{\Pi}_n) \geq \kappa_n$ and $\sigma_{r_0+1}(\Pi_0) = 0$. In turn, it follows that

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n > r_0) &\leq \limsup_{n\to\infty} P(|\sigma_{r_0+1}(\hat{\Pi}_n) - \sigma_{r_0+1}(\Pi_0)| \geq \kappa_n) \\ &\leq \limsup_{n\to\infty} P(\|\tau_n\{\hat{\Pi}_n - \Pi_0\}\| \geq \tau_n\kappa_n) = 0 \ ,\end{aligned} \tag{D.29}$$

where the second inequality is by the Weyl inequality (Tao, 2012, Exercise 1.3.22(iv)), and the equality follows from $\|\tau_n\{\hat{\Pi}_n - \Pi_0\}\| = O_p(1)$ by Assumption 3.1 and $\tau_n\kappa_n \to \infty$ as given. On the other hand, if $\hat{r}_n < r_0$, then $r_0 > 0$ and $\sigma_{r_0}(\hat{\Pi}_n) < \kappa_n$. Hence,

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n < r_0) &\leq \limsup_{n\to\infty} P(|\sigma_{r_0}(\hat{\Pi}_n) - \sigma_{r_0}(\Pi_0)| > -\kappa_n + \sigma_{r_0}(\Pi_0)) \\ &\leq \limsup_{n\to\infty} P(\|\tau_n\{\hat{\Pi}_n - \Pi_0\}\| \geq \tau_n\sigma_{r_0}(\Pi_0)(1 - \kappa_n/\sigma_{r_0}(\Pi_0)) = 0 \ ,\end{aligned} \tag{D.30}$$

where the first inequality exploited $\kappa_n < \sigma_{r_0}(\Pi_0)$ for all n sufficiently large by $\kappa_n \downarrow 0$, the second inequality again follows by the Weyl inequality (Tao, 2012, Exercise 1.3.22(iv)) and also $\sigma_{r_0}(\Pi_0) > 0$, and the equality is because $\|\tau_n\{\hat{\Pi}_n - \Pi_0\}\| = O_p(1)$ by Assumption 3.1, $\tau_n \to \infty$ and $\kappa_n \downarrow 0$. Combining (D.29) and (D.30) yields

$$\limsup_{n\to\infty} P(\hat{r}_n \neq r_0) \leq \limsup_{n\to\infty} P(\hat{r}_n < r_0) + \limsup_{n\to\infty} P(\hat{r}_n > r_0) = 0 \ .$$

This completes the proof of the lemma. ■

Lemma D.8. *Let $\mathbb{G} \in \mathbf{R}^k$ follow $N(\mu, \Omega_0)$ and $g : \mathbf{R}^k \to \mathbf{R}$ be a Lipschitz map with Lipschitz constant L. Then, for M the median of $g(\mathbb{G})$ and any $x > 0$*

$$P(g(\mathbb{G}) - M > x) \leq \frac{1}{2}\exp\{-\frac{1}{2}\frac{x^2}{C^2}\} \tag{D.31}$$

for some $C > 0$ depending on L and $\|\Omega_0\|$.

Proof: This is a mild extension of Lemma A.2.2 in van der Vaart and Wellner (1996), and we include a proof here only for completeness. Since $\mathbb{G} \sim N(\mu, \Omega_0)$, we may write $\mathbb{G} \stackrel{d}{=} \Omega_0^{1/2} Z + \mu$ for some $Z \sim N(0, I_k)$. Define a map $h : \mathbf{R}^k \to \mathbf{R}$ by $h(z) = g(\Omega_0^{1/2} z + \mu)$ for any $z \in \mathbf{R}^k$. Then by Lipschitz continuity of g we have: for any $z_1, z_2 \in \mathbf{R}^k$,

$$\begin{aligned} |h(z_1) - h(z_2)| = |g(\Omega_0^{1/2} z_1 + \mu) - g(\Omega_0^{1/2} z_2 + \mu)| &\le L \|\Omega_0^{1/2} z_1 - \Omega_0^{1/2} z_2\| \\ &\le L \|\Omega_0^{1/2}\| \|z_1 - z_2\| \le L \|\Omega_0\|^{1/2} \|z_1 - z_2\| \ , \end{aligned} \tag{D.32}$$

where the fact $\|\Omega_0^{1/2}\| \le \|\Omega_0\|^{1/2}$ follows from Theorem X.1.1 in Bhatia (1997). By replacing $L\|\Omega_0\|^{1/2}$ with $(L\|\Omega_0\|^{1/2}) \vee 1$ if necessary, we may assume $C \equiv L\|\Omega_0\|^{1/2} > 0$ without loss of generality. Since M is the median of $g(\mathbb{G})$ and hence also of $h(Z)$, we conclude that M/C is the median of $h(Z)/C$. It follows from Lemma A.2.2 in van der Vaart and Wellner (1996) that: for any $x > 0$,

$$P(g(\mathbb{G}) - M > x) = P(\frac{h(Z)}{C} - \frac{M}{C} > \frac{x}{C}) \le \frac{1}{2} \exp\{-\frac{1}{2} \frac{x^2}{C^2}\} \ . \tag{D.33}$$

This completes the proof of the lemma. ■

For the next two lemmas, we let $\text{BL}_1(\mathbf{R})$ be the set of real-valued Lipschitz functions on $\mathbf{R}$ with levels and Lipschitz constants both bounded by one.

Lemma D.9. *Let $T_n^* : \{X_i, W_{ni}\}_{i=1}^n \to \mathbf{R}$ be a bootstrap estimator for the distribution of $g(\mathbb{G})$ such that $\mathbb{G} \in \mathbf{R}^k$ is Gaussian, $g : \mathbf{R}^k \to \mathbf{R}$ is a Lipschitz map, and,*

$$\sup_{f \in \text{BL}_1(\mathbf{R})} |E_W[f(T_n^*)] - E[f(g(\mathbb{G}))]| = o_p(1) \ . \tag{D.34}$$

Suppose Assumption C.1 holds. Let $\hat{c}_{n,1-\alpha_n}$ be $(1-\alpha_n)$ conditional quantiles of T_n^* given the data. If the cdf of $g(\mathbb{G})$ is continuous and strictly increasing on $[r_0, \infty)$ for some $r_0 \in \mathbf{R}$, then $\hat{c}_{n,1-\alpha_n}/\tau_n \xrightarrow{p} 0$.

Proof: Let $\hat{F}_n$ be the conditional cdf of T_n^* given $\{X_i\}_{i=1}^n$, and F be the cdf of $g(\mathbb{G})$. By Lemma 10.11 in Kosorok (2008), we have

$$\sup_{t \in [r_0, \infty)} |\hat{F}_n(t) - F(t)| = o_p(1) \ . \tag{D.35}$$

By the definition of quantiles, we thus obtain from (D.35) that, for any $r \in [r_0, \infty)$,

$$\begin{aligned} \limsup_{n \to \infty} P(\hat{c}_{n,1-\alpha_n} \le r) &\le \limsup_{n \to \infty} P(\hat{F}_n(r) \ge 1 - \alpha_n) \\ &= \limsup_{n \to \infty} P(o_p(1) + F(r) \ge 1 - \alpha_n) = 0 \ . \end{aligned} \tag{D.36}$$

where we exploited the facts that $F(r) < 1$ by strict monotonicity of F on $[r_0, \infty)$ and that $\alpha_n \downarrow 0$. Next, fix $\epsilon > 0$. Combination of (D.36) and Lemma D.8 yields

$$\alpha_n < 1 - \hat{F}_n(\hat{c}_{n,1-\alpha_n} - \epsilon) = P(g(\mathbb{G}) > \hat{c}_{n,1-\alpha_n} - \epsilon) + o_p(1)$$
$$\leq \frac{1}{2}\exp\{-\frac{1}{2}\frac{(\hat{c}_{n,1-\alpha_n} - \epsilon - c_{0.5})^2}{C^2}\} + o_p(1) \ , \quad \text{(D.37)}$$

for some $C > 0$ and $c_{0.5}$ the 0.5-quantile of $g(\mathbb{G})$. It follows from (D.37) that

$$(\frac{\hat{c}_{n,1-\alpha_n}}{\tau_n} - \frac{\epsilon}{\tau_n} - \frac{c_{0.5}}{\tau_n})^2 \leq 2C^2(-\frac{\log \alpha_n}{\tau_n^2} + \frac{\log o_p(1)}{\tau_n^2} - \frac{\log 2}{\tau_n^2}) \ . \quad \text{(D.38)}$$

By Assumption C.1(ii), $\tau_n \uparrow \infty$ and $\log o_p(1) \xrightarrow{p} -\infty$ as $n \to \infty$, we may then conclude the proof of the lemma from result (D.38). ■

Lemma D.10. *Suppose Assumptions 3.1, 3.2 and C.1 hold. Let $\hat{c}_{n,1-\alpha}$ be defined by* (29) *for $\alpha \in (0,1)$ where $\kappa_n \to 0$ and $\tau_n\kappa_n \to \infty$ if $\hat{\phi}''_{r,n}$ is defined by* (26) *but no restrictions on $\hat{r}_n$ if $\hat{\phi}''_{r,n}$ is defined by* (28). *If $\mathcal{M}$ is Gaussian but not constant, then $\hat{c}_{n,1-\alpha_n}/\tau_n^2 \xrightarrow{p} 0$.*

PROOF: Consider first the case when $\hat{c}_{n,1-\alpha}$ is defined by the analytic derivative estimator. By Lemma 3.1 and simple manipulations, we have

$$\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*)^{1/2} \leq \|\hat{P}_{2,n}^{\intercal}\hat{\mathcal{M}}_n^*\hat{Q}_{2,n}\| \leq (mk)^{1/2}\|\hat{\mathcal{M}}_n^*\| \ . \quad \text{(D.39)}$$

Let $\tilde{c}_{n,1-\alpha}$ be the $(1-\alpha)$th conditional quantile of $\|\hat{\mathcal{M}}_n^*\|$ for each $\alpha \in (0,1)$. Since $\mathcal{M}$ is Gaussian and the variance of $\mathrm{vec}(\mathcal{M})$ is nonzero, $\|\mathcal{M}\|^2$ is equal in law to a weighted sum of independent $\chi^2(1)$ random variables. It follows that the cdf $\|\mathcal{M}\|$ is continuous and strictly increasing on $\mathbf{R}_+$. In turn, by Proposition 10.7 in Kosorok (2008), Assumptions 3.2 and C.1, we obtain from Lemma D.9 that $\tilde{c}_{n,1-\alpha}/\tau_n \xrightarrow{p} 0$. By result (D.39) and equivariance of quantiles to monotone transformations, we may then conclude that

$$\frac{\hat{c}_{n,1-\alpha_n}}{\tau_n^2} \leq \frac{\tilde{c}^2_{n,1-\alpha_n}}{\tau_n^2} = o_p(1) \ . \quad \text{(D.40)}$$

Next, turn to the case when $\hat{c}_{n,1-\alpha}$ is defined by the numerical derivative estimator. For each $\alpha \in (0,1)$, let $\bar{c}_{n,1-\alpha}$ be the conditional quantile (given the data) of

$$\kappa_n\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*) = \frac{\phi_r(\hat{\Pi}_n + \kappa_n\hat{\mathcal{M}}_n^*) - \phi_r(\hat{\Pi}_n)}{\kappa_n} \ . \quad \text{(D.41)}$$

By Assumptions 3.1, 3.2 and the rates conditions on κ_n as given, we may employ Proposition 3.1 and Theorem 3.3 in Chen and Fang (2018) to conclude that

$$\sup_{f\in \mathrm{BL}_1(\mathbf{R})} |E_W[f(\kappa_n\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*))] - E[f(\phi'_{r,\Pi_0}(\mathcal{M}))]| = o_p(1) \ . \quad \text{(D.42)}$$

By simple algebra we may obtain that: for any $M_1, M_2 \in \mathbf{M}^{m\times k}$,

$$|\phi'_{r,\Pi_0}(M_1) - \phi'_{r,\Pi_0}(M_2)| = |\min_{U\in\Psi(\Pi)} 2\mathrm{tr}(U^\intercal\Pi^\intercal M_1 U) - \min_{U\in\Psi(\Pi)} 2\mathrm{tr}(U^\intercal\Pi^\intercal M_2 U)|$$
$$\leq \max_{U\in\Psi(\Pi_0)} 2\|\Pi_0 U\|\|(M_1 - M_2)U\| \leq 2\sqrt{k}\|\Pi_0\|\|M_1 - M_2\| \ . \quad \text{(D.43)}$$

By result (D.36) and Lemma D.9, we thus have $\bar{c}_{n,1-\alpha_n}/\tau_n = o_p(1)$ and hence

$$\frac{\hat{c}_{n,1-\alpha_n}}{\tau_n^2} \leq \frac{\bar{c}_{n,1-\alpha_n}}{\tau_n}\frac{1}{\tau_n\kappa_n} = o_p(1) \ , \quad \text{(D.44)}$$

since $\tau_n\kappa_n \to \infty$. This completes the proof of the lemma. ■

We next present lemmas that are relevant to Section 5 and proceed by imposing:

Assumption D.1. *(i) The supports of X and Y are finite; (ii) the Jacobian matrix of* $\mathrm{vec}(E_{\pi(A,p,q)}[XY^\intercal])$ *with respect to* $\mathrm{vec}(A)$ *at A_0 is nonsingular.*

Assumption D.1(i) formalizes the setup that the matching attributes are finitely valued. Assumption D.1(ii) is a technical condition, as implicitly imposed in Galichon and Salanié (2010) and Dupuy and Galichon (2014) who showed that the Jacobian coincides with the Fisher information matrix for A_0.

Next, let the supports $\mathcal{X} = \{x_1, \ldots, x_I\}$ and $\mathcal{Y} = \{y_1, \ldots, y_J\}$. Then, we may identify p_0 and q_0 as vectors in $(0,1)^I$ and $(0,1)^J$ respectively.

Lemma D.11. *If Assumption D.1 holds, then the implicit map $A : (0,1)^I \times (0,1)^J \times \mathbf{M}^{m\times k} \to \mathbf{M}^{m\times k}$ defined by* (54)*, i.e., $A(p_0, q_0, E[XY^\intercal]) = A_0$, is Hadamard differentiable on some open neighborhood of the truth $(p_0, q_0, E[XY^\intercal])$.*

Proof: First, note that A is uniquely defined by Lemma 3 in Dupuy and Galichon (2014). Next, define a map $\Psi : \mathbf{M}^{m\times k} \times (0,1)^I \times (0,1)^J \times \mathbf{M}^{m\times k} \to \mathbf{R}^{mk}$ by:

$$\Psi(A, p, q, \Sigma) \equiv \mathrm{vec}(E_{\pi(A,p,q)}[X^\intercal Y] - \Sigma) \ . \quad \text{(D.45)}$$

By Assumption D.1 and Lemma D.12, Ψ is continuously differentiable on some open neighborhood of the truth $(A_0, p_0, q_0, E[XY^\intercal])$ – note in particular that X and Y are finitely supported. In turn, Assumption D.1(ii) allows us to invoke the implicit function theorem, see, for example, Theorem 9.28 in Rudin (1976), to conclude the proof. ■

Lemma D.12. *If Assumption D.1(i) holds, then the map $(A_0, p_0, q_0) \mapsto \pi(A_0, p_0, q_0)(x, y)$ defined by* (52) *where Φ is specified as in* (53) *uniquely exists and is continuously differentiable on some open neighborhood of the truth (A_0, p_0, q_0), for each $(x, y) \in \mathcal{X} \times \mathcal{Y}$.*

Proof: First, we may rewrite the maximization problem (52) as

$$\max_{\pi} \sum_{i=1}^{I}\sum_{j=1}^{J} \pi_{ij} x_i^{\intercal} A_0 y_j - \sum_{i=1}^{I}\sum_{j=1}^{J} \pi_{ij} \log \pi_{ij} \ , \tag{D.46}$$

subject to: for all $i = 1, \dots, I$ and all $j = 1, \dots, J$,

$$\sum_{j=1}^{J} \pi_{ij} = p_{0,i} \ , \quad \sum_{i=1}^{I} \pi_{ij} = q_{0,j} \ , \tag{D.47}$$

where $p_{0,i} = P(X = x_i)$ and $q_{0,j} = P(Y = y_j)$ for all $i = 1, \dots, I$ and all $j = 1, \dots, J$. By defining $x \log x = 0$ if $x = 0$, it is simple to see that the objective function in (D.46) is continuous. Since the constraints define a compact domain for π, it follows that an optimal matching distribution π_0 always exists. The uniqueness of π_0 follows from strict concavity of the objective function since $x \mapsto x \log x$ is strictly convex. Moreover, the right derivative of the objective function at 0 is infinite – see equation (D.50) below or Galichon and Salanié (2010, p.5), implying that the optimal π_0 must satisfy $0 < \pi_{0,ij} < 1$ for all i and j. Exploiting the constraints in (D.47), together with the facts that p_0, q_0 and π are pmfs, the constrained optimization can be converted into an unconstrained one in which the objective function in (D.46) is a function of $\{p_{0,i}\}_{i=1}^{I-1}$, $\{q_{0,j}\}_{j=1}^{J-1}$ and $\{\pi_{0,ij}\}_{i=1,j=1}^{I-1,J-1}$ only, with $\pi_{0,iJ} = p_{0,i} - \sum_{j=1}^{J-1} \pi_{0,ij}$, $\pi_{0,Ij} = q_{0,j} - \sum_{i=1}^{I-1} \pi_{0,ij}$ for all $i = 1, \dots, I-1$ and $j = 1, \dots, J-1$, and

$$\begin{aligned} \pi_{0,IJ} &= 1 - \sum_{i=1}^{I-1}\sum_{j=1}^{J-1} \pi_{0,ij} - \sum_{i=1}^{I-1} \pi_{0,iJ} - \sum_{j=1}^{J-1} \pi_{0,Ij} \\ &= 1 + \sum_{i=1}^{I-1}\sum_{j=1}^{J-1} \pi_{0,ij} - \sum_{i=1}^{I-1} p_{0,i} - \sum_{j=1}^{J-1} q_{0,j} \ . \end{aligned} \tag{D.48}$$

It follows that the unique maximizer π_0 must satisfy the first order condition:

$$\begin{aligned} & x_i^{\intercal} A_0 y_j - x_I^{\intercal} A_0 y_j - x_i^{\intercal} A_0 y_J + x_I^{\intercal} A_0 y_J - 1 - \log \pi_{0,ij} \\ & \qquad + 1 + \log \pi_{0,Ij} + 1 + \log \pi_{0,iJ} - 1 - \log \pi_{0,IJ} = 0 \ , \end{aligned} \tag{D.49}$$

or equivalently

$$\begin{aligned} & x_i^{\intercal} A_0 y_j - x_I^{\intercal} A_0 y_j - x_i^{\intercal} A_0 y_J + x_I^{\intercal} A_0 y_J \\ & \qquad - \log \pi_{0,ij} + \log \pi_{0,Ij} + \log \pi_{0,iJ} - \log \pi_{0,IJ} = 0 \ , \end{aligned} \tag{D.50}$$

for all $i = 1, \dots, I-1$ and $j = 1, \dots, J-1$, where $\pi_{0,iJ}$, $\pi_{0,Ij}$ and $\pi_{0,IJ}$ are functions of $\{p_{0,i}\}_{i=1}^{I-1}$, $\{q_{0,j}\}_{j=1}^{J-1}$ and $\{\pi_{0,ij}\}_{i=1,j=1}^{I-1,J-1}$ as defined previously.

Let us stack the equations in (D.50) along $i = 1, \dots, m$ sequentially for fixed

$j = 1, \ldots, k$, and let $d^* \equiv (I-1)(J-1)$. The left side of (D.50) is then a $\mathbf{R}^{d^*}$-valued function of A_0, $\{p_{0,i}\}_{i=1}^{I-1}$, $\{q_{0,j}\}_{j=1}^{J-1}$ and $\{\pi_{0,ij}\}_{i=1,j=1}^{I-1,J-1}$, which is obviously continuously differentiable. Moreover, the derivative of the left side in (D.50) with respect to $\mathrm{vec}(\{\pi_{0,ij}\}_{i=1,j=1}^{I-1,J-1})$ is then a matrix of size $d^* \times d^*$ which is given by: for $\mathbb{J}_d$ a generic $d \times d$ matrix of ones,

$$-\underline{\pi}_0 - \underline{\pi}_{0,J} \otimes \mathbb{J}_{I-1} - \mathbb{J}_{J-1} \otimes \underline{\pi}_{0,I} - \pi_{0,IJ}^{-1}\mathbb{J}_{d^{*2}} \ , \tag{D.51}$$

with $\underline{\pi}_{0,I} \equiv \mathrm{diag}(\{\pi_{0,iJ}^{-1}\}_{i=1}^{I-1})$, $\underline{\pi}_{0,J} \equiv \mathrm{diag}(\{\pi_{0,Ij}^{-1}\}_{j=1}^{J-1})$ and $\underline{\pi}_0 \equiv \mathrm{diag}(\mathrm{vec}(\{\pi_{0,ij}^{-1}\}_{i=1,j=1}^{I-1,J-1}))$. Note that $\underline{\pi}_0$ is positive definite while $\underline{\pi}_{0,J} \otimes \mathbb{J}_{I-1}$, $\mathbb{J}_{J-1} \otimes \underline{\pi}_{0,I}$ and $\pi_{0,IJ}^{-1} \otimes \mathbb{J}_{d^{*2}}$ are positive semidefinite, so the matrix in (D.51) is invertible. The conclusion now follows from the implicit function theorem – see, for example, Theorem 9.28 in Rudin (1976). ■

Appendix E Cointegration and Additional Examples

In this section, we present additional examples where knowledge on the rank of a matrix is of interest. We single out the treatment of inference on cointegration rank because (i) it is prominent in applied macroeconomics and (ii) Assumption 3.1 may take a generalized form in this case where the convergence rates are heterogenous across entries of $\hat{\Pi}_n$ (but still falls within the scope of the Delta method).

E.1 Inference on Cointegration Rank

Let $\{Y_t\}$ be a time series in $\mathbf{R}^k$ such that all its entries are unit root processes. For ease of exposition and to hight what is essential to our theory, we limit ourselves to processes without deterministic terms throughout. By the Granger representation theorem, the number h_0 of independent cointegrating vectors is precisely equal to $k - \mathrm{rank}(\Omega_0)$ with Ω_0 the long run variance of ΔY_t.[3] Within this (nonparametric) system framework, one may be interested in testing

$$\mathrm{H}_0 : h_0 \geq h \qquad \text{v.s.} \qquad \mathrm{H}_1 : h_0 < h \tag{E.1}$$

for a given integer $h = 1, \ldots, k-1$, which is equivalent to (1) with $\Pi_0 = \Omega_0$ and $r = k-h$. The special case with $h = 1$ is concerned with testing the null of cointegration. Problems of similar nature have been studied by Stock and Watson (1988), Harris (1997), Snell (1999) and Nyblom and Harvey (2000), but they test the null $h_0 = h$ instead. Note that the nonparametric tests of Bierens (1997) and Shintani (2001) are not directly applicable to (E.1) because they test the null $h_0 = h$ against $h_0 > h$ (larger cointegration rank).

[3]By definition, Y_t is said to be cointegrated if there is nonzero vector $\lambda \in \mathbf{R}^k$ such that $\lambda^\intercal Y_t$ is stationary, in which case λ is called a cointegrating vector.

The testing problem (E.1) is not only of interest in its own right (Hayashi, 2000), but also important as a complement to tests against larger cointegration rank especially in view of the potentially poor power of the latter tests, as forcefully argued by Kwiatkowski et al. (1992) and Maddala and Kim (1998). Nevertheless, through VAR or error-correction representations, our framework can accommodate the hypotheses

$$\mathrm{H}_0 : h_0 \leq h \qquad \text{v.s.} \qquad \mathrm{H}_1 : h_0 > h \ . \tag{E.2}$$

To see this, suppose that the error-correction representation of $\{Y_t\}$ is given by

$$\Delta Y_t = \Phi_0 Y_{t-1} + \sum_{j=1}^{p-1} \Phi_j \Delta Y_{t-j} + \epsilon_t \ , \tag{E.3}$$

for some white noise $\{\epsilon_t\}$. Then $h_0 = \mathrm{rank}(\Phi_0)$ by the Granger representation theorem, and hence (E.2) is equivalent to (1) with $\Pi_0 = \Phi_0$ and $r = h$. The setup (E.2) is studied in the seminal work of Johansen (1988, 1991), for which Johansen proposes the celebrated maximum eigenvalue test and the trace test, and derives their asymptotic distributions under $h_0 = h$. The general limits under $h_0 \leq h$ are presented in Johansen (1995) but no critical values are provided.

Below we study the problems (E.1) and (E.2) separately as their treatments require different arguments, and proceed with the former.

Nonparametric Cointegration Test

We start by estimating the long run variance $\Pi_0 \equiv \Omega_0$ based on the periodogram. Specifically, for a kernel/density function $K : \mathbf{R} \to \mathbf{R}_+$, let

$$\hat{\Pi}_n = \frac{2\pi}{n} \sum_{j=-\lfloor (n-1)/2 \rfloor}^{\lfloor n/2 \rfloor} K_{b_n}(\omega_j) I_{\Delta Y,n}(\omega_j) \ , \tag{E.4}$$

where $\lfloor a \rfloor$ is the integer part of $a \in \mathbf{R}$, $K_{b_n}(\cdot) = K(\cdot/b_n)/b_n$, $\omega_j = 2\pi j/n$ with $j = -\lfloor (n-1)/2 \rfloor, \ldots, \lfloor n/2 \rfloor$ are the natural frequencies, $b_n \to 0$ is a suitable bandwidth, and $\omega \mapsto I_{\Delta Y,n}(\omega) \in \mathbf{M}^{k\times k}$ is the periodogram of $\{\Delta Y_t\}_{t=1}^n$, i.e.,

$$I_{\Delta Y,n}(\omega) = \frac{1}{2\pi n} (\sum_{t=1}^n \Delta Y_t e^{-it\omega})(\sum_{t=1}^n \Delta Y_t e^{-it\omega})^{\mathrm{H}}$$

for M^{H} denoting the Hermitian transpose of a generic complex matrix M. Then $\hat{\Pi}_n$ is asymptotically normal at the rate $\sqrt{nb_n}$ under regularity conditions – see, for example, Hannan (1970, Theorem V.5.11) and Brillinger (2001, Theorem 7.4.4) for classical treatments, and Phillips et al. (2006) and Politis (2011) for recent developments.

We construct the estimator $\hat{\mathcal{M}}_n^*$ employing the multivariate linear process bootstrap recently developed by Jentsch and Politis (2015). Let $\hat{\Gamma}_{n,j} = \sum_{t=1}^{n-j} \Delta Y_t \Delta Y_{t+j}^\intercal / n$ for $j \geq 0$ and $\hat{\Gamma}_{n,j} = \hat{\Gamma}_{n,-j}^\intercal$ if $j < 0$. Define $\hat{V}_n \in \mathbf{M}^{nk \times nk}$ to be a block matrix whose (i,j)th block is given by $\varrho((i-j)/l_n)\hat{\Gamma}_{n,i-j}$ for $i,j = 1,\ldots,n$, where $\varrho : \mathbf{R} \to \mathbf{R}$ is a flat-top kernel and l_n is a banding parameter (Politis, 2001, 2011). The matrix $\hat{V}_n$ serves as an estimator of the covariance matrix of $\Delta_n = [\Delta Y_1^\intercal, \ldots, \Delta Y_n^\intercal]^\intercal \in \mathbf{R}^{nk}$. One may modify $\hat{V}_n$ if necessary to ensure that it is positive definite (Jentsch and Politis, 2015, p.1124). Let $Z_n = L_n^{-1}\Delta_n \in \mathbf{R}^{nk}$ where L_n is from the Cholesky decomposition $\hat{V}_n = L_n L_n^\intercal$, $Z_{n,i}$ the ith entry of Z_n, and $\bar{Z}_{n,i} = (Z_{n,i} - \bar{Z}_n)/\hat{\sigma}_n$ for $\bar{Z}_n = \sum_{i=1}^{nk} Z_{n,i}/(nk)$ and $\hat{\sigma}_n^2 = \sum_{i=1}^{nk}(Z_{n,i} - \bar{Z}_n)^2/(nk)$. Now, draw an i.i.d. sample $\{Z_{n,i}^*\}_{i=1}^{nk}$ from $\{\bar{Z}_{n,i}\}_{i=1}^{nk}$ with replacement. Define $Z_n^* \in \mathbf{R}^{nk}$ whose ith entry is $Z_{n,i}^*$, and let $\Delta_n^* = L_n Z_n^* \in \mathbf{R}^{nk}$.

Finally, our bootstrap sample $\{\Delta Y_t^*\}_{t=1}^n$ is such that $\Delta_n^* = (\Delta Y_1^{*\intercal}, \ldots, \Delta Y_n^{*\intercal})^\intercal$, and then the bootstrap estimator $\hat{\Pi}_n^*$ is defined analogously to $\hat{\Pi}_n$ but with $\{\Delta Y_t\}_{t=1}^n$ replaced by $\{\Delta Y_t^*\}_{t=1}^n$. In order to construct a bootstrap estimator $\hat{\mathcal{M}}_n^*$ that satisfies Assumption 3.2, we need to properly center $\hat{\Pi}_n^*$, as is well understood for bootstrap in nonparametric settings (Hall, 1992). To this end, define

$$\tilde{\Pi}_n = \sum_{j=-(n-1)}^{n-1} \varrho(\frac{j}{l_n})\hat{\Gamma}_{n,j} \ . \tag{E.5}$$

Under regularity conditions, the bootstrap consistency of $\hat{\mathcal{M}}_n^* \equiv \sqrt{nb_n}\{\hat{\Pi}_n^* - \tilde{\Pi}_n\}$ is formally established by Theorem 4.2 in Jentsch and Politis (2015). We refer the reader to Jentsch and Kreiss (2010), Politis and Romano (1993), Politis and Romano (1994), and Berkowitz and Diebold (1998) for alternative resampling schemes.

Cointegration Test in Error-Correction Models

Now consider the error-correction model, and suppose that $\{\epsilon_t\}$ is a white noise having nonsingular covariance matrix Σ_0. Since $h_0 = \text{rank}(\Phi_0)$ under (E.3), the problem (E.2) is equivalent to (1) by identifying Π_0 with Φ_0 and r with $h \in \{0, 1, \ldots, k-1\}$. The special case $h = 0$ reduces to a test of no cointegration against existence of cointegration. For ease and transparency of our exposition, suppose $p = 1$ and hence there are no lagged variables ΔY_{t-j} in (E.3); the general case can be handled in a straightforward manner by combining our arguments below with Lemma A.6 in Liao and Phillips (2015).

We proceed with some clarifications on notation. Let $\Pi_0 = P_0\Sigma_0 Q_0^\intercal$ be a singular value decomposition of Π_0; write $P_0 = [P_{0,1}, P_{0,2}]$ and $Q_0 = [Q_{0,1}, Q_{0,2}]$ where $P_{0,1} \in \mathbb{S}^{k\times r_0}$ and $Q_{0,1} \in \mathbb{S}^{k\times r_0}$ with $r_0 \equiv \text{rank}(\Pi_0)$. On the other hand, it is more common to have $\Pi_0 = \alpha_0\beta_0^\intercal$ where $\alpha_0 \in \mathbf{M}^{k\times r_0}$ (whose columns are called adjustment coefficients) and $\beta_0 \in \mathbf{M}^{k\times r_0}$ (whose columns are cointegrating vectors) both have full rank r_0 (Johansen, 1995). As pointed out by Johansen (1988, 1991), α_0 and β_0 are not identified, but their column spaces are. To flesh out the connections between these two sets of notation, let $\Sigma_{0,1} \in \mathbf{M}^{r_0\times r_0}$ be the left top block of Σ_0. By direct calculations, we obtain $\Pi_0 = P_{0,1}\Sigma_{0,1}Q_{0,1}^\intercal$. Consequently, we may take $\alpha_0 = P_{0,1}\Sigma_{0,1}$ and $\beta_0 = Q_{0,1}$. In turn, the corresponding orthogonal complement versions $\alpha_{0,\perp} \in \mathbf{M}^{k\times(k-r_0)}$ and $\beta_{0,\perp} \in \mathbf{M}^{k\times(k-r_0)}$ (both of full rank) can be taken to be $\alpha_{0,\perp} = P_{0,2}$ and $\beta_{0,\perp} = Q_{0,2}$, satisfying $\alpha_{0,\perp}^\intercal\alpha_0 = \mathbf{0}_{(k-r_0)\times r_0}$ and $\beta_{0,\perp}^\intercal\beta_0 = \mathbf{0}_{(k-r_0)\times r_0}$ as required. Finally, define

$$B_0 \equiv \begin{bmatrix} \beta_0^\intercal \\ \alpha_{0,\perp}^\intercal \end{bmatrix} = \begin{bmatrix} Q_{0,1}^\intercal \\ P_{0,2}^\intercal \end{bmatrix} .$$

In applying our inferential framework, we need to construct a matrix estimator $\hat{\Pi}_n$ that converges weakly and a consistent bootstrap analog. In what follows, let $\{Y_t\}_{t=0}^n$ be a time series sample in $\mathbf{R}^k$ that is generated according to (E.3) (with $p = 1$).

<u>Asymptotic Distributions:</u> For this, we employ the OLS estimator:

$$\hat{\Pi}_n = (\sum_{t=1}^n \Delta Y_t Y_{t-1}^\intercal)(\sum_{t=1}^n Y_{t-1}Y_{t-1}^\intercal)^{-1} . \tag{E.6}$$

Under standard regularity conditions, Lemma A.2 in Liao and Phillips (2015), together with the continuous mapping theorem, implies that

$$\{\hat{\Pi}_n - \Pi_0\}B_0^{-1}D_nB_0 \xrightarrow{L} \mathcal{M} \equiv \mathcal{M}_1 + \mathcal{M}_2 , \tag{E.7}$$

where $D_n \equiv \operatorname{diag}(\sqrt{n}\mathbf{1}_{r_0}, n\mathbf{1}_{k-r_0})$, $\operatorname{vec}(\mathcal{M}_1^\intercal) \sim N(0, \Sigma_0 \otimes (Q_{0,1}\Sigma_1^{-1}Q_{0,1}^\intercal))$ with $\Sigma_1 \equiv \operatorname{Var}(Q_{0,1}^\intercal Y_t)$, and $\mathcal{M}_2 \in \mathbf{M}^{k\times k}$ is such that

$$\mathcal{M}_2 \sim \Sigma_0^{1/2}\int_0^1 dB_k(t)B_k(t)^\intercal\Sigma_0^{1/2}P_{0,2}(P_{0,2}^\intercal\Sigma_0^{1/2}\int_0^1 B_k(t)B_k(t)^\intercal dt\, \Sigma_0^{1/2}P_{0,2})^{-1}P_{0,2}^\intercal \tag{E.8}$$

with $B_k(\cdot)$ is a k-dimensional standard Brownian motion defined on the unit interval.

Inspecting result (E.7), it seems that Assumption 3.1 is being violated because the "convergence rate" $B_0^{-1}D_nB_0$ is not a scalar. However, this creates no conceptual difficulties if we interpret τ_n there as linear maps $\tau_n : \mathbf{M}^{m\times k} \to \mathbf{M}^{m\times k}$ – such an insight has been noted in van der Vaart and Wellner (1996, p.413). In the current setup, we have $\tau_n : \mathbf{M}^{m\times k} \to \mathbf{M}^{m\times k}$ defined by: for any $M \in \mathbf{M}^{k\times k}$,

$$\tau_n M \equiv \tau_n(M) = MB_0^{-1}D_nB_0 . \tag{E.9}$$

Therefore, in order to invoke the Delta method, the only question that remains is: is our map ϕ_r as defined in (12) suitably differentiable with respect to these linear maps? The answer is affirmative, as shown by Proposition E.1 stated at the end of this subsection.

In particular, if $\mathrm{rank}(\Pi_0) \le r \equiv h$, then we have

$$\lim_{n\to\infty} \frac{\phi_r(\Pi_0 + \tau_n^{-1}M_n) - \phi_r(\Pi_0)}{n^{-2}} = \phi''_{r,\Pi_0}(M) \equiv \sum_{j=h-h_0+1}^{k-h_0} \sigma_j^2(P_{0,2}^\intercal M Q_{0,2}) , \qquad \text{(E.10)}$$

whenever $M_n \to M$ as $n \to \infty$. By a modification of the Delta method – see Proposition E.2, we thus obtain from (E.10) and (E.7) that, under H_0 in (E.2),

$$n^2\phi_r(\hat{\Pi}_n) \xrightarrow{L} \phi''_{r,\Pi_0}(\mathcal{M}) \equiv \sum_{j=h-h_0+1}^{k-h_0} \sigma_j^2(P_{0,2}^\intercal \mathcal{M} Q_{0,2}) . \qquad \text{(E.11)}$$

The limit in (E.11) shows the importance of acknowledging the generic possibility that the true cointegration rank h_0 may be strictly less than the hypothesized value h.

In positioning our work in the literature, we note that existing tests are mainly based on the following standardized version of $\hat{\Pi}_n$ (Hubrich et al., 2001; Al-Sadoon, 2017):

$$\hat{\Pi}_{s,n} = (\sum_{t=1}^{n} \Delta Y_t \Delta Y_t^\intercal)^{-1/2} \hat{\Pi}_n (\sum_{t=1}^{n} Y_{t-1} Y_{t-1}^\intercal)^{1/2} . \qquad \text{(E.12)}$$

For example, the classical trace statistic of Johansen (1991, 1988) is given by

$$\mathrm{LR}_n(\hat{\Pi}_{s,n}) = -n \sum_{j=r+1}^{k} \log(1 - \sigma_j^2(\hat{\Pi}_{s,n})) . \qquad \text{(E.13)}$$

whose asymptotic distribution under $\mathrm{H}_0' : h_0 = h$ is: for $d_0 \equiv k - r_0$,

$$\mathrm{tr}\Big(\int_0^1 dB_{d_0}(t) B_{d_0}(t)^\intercal \big(\int_0^1 B_{d_0}(t) B_{d_0}(t)^\intercal dt \big)^{-1} \int_0^1 B_{d_0}(t) dB_{d_0}(t)^\intercal \Big) . \qquad \text{(E.14)}$$

The asymptotic distribution under $\mathrm{H}_0 : h_0 \le h$, however, is different from (E.14) in general – see Johansen (1995, p.157-8,168). This may adversely affect the trace test through channels as discussed in the main text, and hence in turn provides an alternative explanation on why its finite sample performance can be poor, as documented in the literature (Maddala and Kim, 1998; Johansen, 2002).

<u>Bootstrap Inference:</u> The limiting distribution in (E.11) is highly nonstandard, and in particular depends on the true rank h_0. In order to apply our bootstrap procedure, we need to estimate both the "derivative" ϕ''_{r,Π_0} and the limit $\mathcal{M}$. Estimation of ϕ''_{r,Π_0} is no more special than what we have discussed in Section 3.3. For example, one may estimate ϕ''_{r,Π_0} by (28) with $\hat{r}_n$ given by (27). Since $\sqrt{n}\{\hat{\Pi}_n - \Pi_0\}$ converges in distribution by (E.11), we thus still have $\hat{r}_n \xrightarrow{p} r_0 = h_0$ provided $\kappa_n \to 0$ and $\sqrt{n}\kappa_n \to \infty$ by Lemma D.7. Condition (25) in turn follows from Lemma D.6. In fact, one can show along the lines in the proof of Lemma D.7 that it suffices to have $\kappa_n \to 0$ and $n\kappa_n \to \infty$.

Given an estimator $\hat{\phi}''_{r,n}$ of ϕ''_{r,Π_0}, we may thus approximate the law of $\phi''_{r,\Pi_0}(\mathcal{M})$ in (E.11) by the conditional law (given the data) of $\hat{\phi}''_{r,n}(\hat{\mathcal{M}}^*_n)$ as long as $\hat{\mathcal{M}}^*_n$ is consistent for $\mathcal{M}$. To this end, we employ a residual-based bootstrap following van Giersbergen (1996), Swensen (2006) and Cavaliere et al. (2012), who study bootstrap cointegration tests for $\mathrm{H}'_0 : h_0 = h$ based on error-correction models. Although these rank tests are potentially subject to the deficiencies illustrated in Sections 2 and C, their work show that the residual bootstrap procedure produces bootstrap samples that mimic the data well, a property we exploit directly. Moreover, in order to properly account for the possibility $h_0 \equiv \mathrm{rank}(\Pi_0) < h$, we need a (preliminary) estimator $\hat{r}_n$ for h_0 that is consistent under both H_0 and H_1. For example, in view of Lemma D.7, we may take

$$\hat{r}_n = \max\{j = 1, \ldots, k : \sigma_j(\hat{\Pi}_n) \geq \kappa_n\} , \tag{E.15}$$

if the set is nonempty and $\hat{r}_n = 0$ otherwise, where $\kappa_n \to 0$ and $n\kappa_n \to \infty$. The residual bootstrap now goes as follows.

Step 1: Given an estimator $\hat{r}_n$ that is consistent for h_0 under both H_0 and H_1, calculate the reduced rank estimator $\tilde{\Pi}_n$ following the maximum likelihood approach of Johansen (1988, 1991), and obtain the residuals $\{\hat{\epsilon}_t\}$ as well as their centered versions $\{\bar{\epsilon}_t\}$, i.e., $\bar{\epsilon}_t \equiv \hat{\epsilon}_t - n^{-1}\sum_{t=1}^n \hat{\epsilon}_t$.

Step 2: Check if $|I_k - \lambda(\tilde{\Pi}_n + I_k)| = 0$ has roots on or outside the unit circle, and if $\tilde{P}_{2,n}^{\intercal}\tilde{Q}_{2,n}$ has full rank, where the columns in $\tilde{P}_{2,n}$ and $\tilde{Q}_{2,n}$ are left and right singular vectors of $\tilde{\Pi}_n$ associated with the smallest $k - \hat{r}_n$ singular values. If so, proceed to the next step – see Remark 1 in Swensen (2006) for discussions.

Step 3: Construct a bootstrap sample $\{Y_t^*\}_{t=1}^n$ recursively from

$$\Delta Y_t^* = \tilde{\Pi}_n Y_{t-1}^* + \epsilon_t^* ,$$

with the initial value Y_0 and ϵ_t^* being generated from $\{\bar{\epsilon}_t\}_{t=1}^n$ by the nonparametric bootstrap. Calculate the bootstrap least square estimator

$$\hat{\Pi}_n^* = \sum_{t=1}^n \Delta Y_t^* Y_{t-1}^{*\intercal} (\sum_{t=1}^n Y_{t-1}^* Y_{t-1}^{*\intercal})^{-1} . \tag{E.16}$$

Let $\hat{B}_n$ be the analog of B_0 based on $\tilde{\Pi}_n$, and $\hat{D}_n$ the analog of D_n based on $\hat{r}_n$. Following the proof of Lemma A.2 of Liao and Phillips (2015), we have: almost surely,

$$\hat{\mathcal{M}}_n^* \equiv \{\hat{\Pi}_n^* - \tilde{\Pi}_n\}\hat{B}_n^{-1}\hat{D}_n\hat{B}_n \xrightarrow{L^*} \mathcal{M} . \tag{E.17}$$

Given an estimator $\hat{\phi}''_{r,n}$ satisfying (25) and the bootstrap estimator $\hat{\mathcal{M}}^*_n$ as in (E.17), we may finally estimate the limit in (E.11) by the conditional law (given the data) of

$$\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*) \equiv \hat{\phi}''_{r,n}(\{\hat{\Pi}_n^* - \tilde{\Pi}_n\}\hat{B}_n^{-1}\hat{D}_n\hat{B}_n) . \tag{E.18}$$

Let $\hat{c}_{n,1-\alpha}$ be the conditional $(1-\alpha)$-quantile of $\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*)$ given the data. Then our test for (E.2) that rejects H_0 if $n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}$ has asymptotic size control and is consistent, along the lines in Theorem 3.2.

To conclude, we present results that establish weak convergence of our statistic.

Proposition E.1. *Let $\phi_r : \mathbf{M}^{k\times k} \to \mathbf{R}$ be defined as in* (12) *with $m = k$ and $\Pi_0 \in \mathbf{M}^{k\times k}$ satisfy $\phi_r(\Pi_0) = 0$. Then, for $r_0 \equiv \mathrm{rank}(\Pi_0)$ and $T_n \equiv \mathrm{diag}(t_n\mathbf{1}_{r_0}, t_n^2\mathbf{1}_{k-r_0})$ with $t_n > 0$,*

$$\lim_{n\to\infty} \frac{\phi_r(\Pi_0 + M_nT_nB_0)}{t_n^4} = \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal MQ_{0,2}) ,$$

whenever $t_n \downarrow 0$ and $\{M_n\} \subset \mathbf{M}^{k\times k}$ satisfies $M_nB_0 \to M \in \mathbf{M}^{m\times k}$ as $n\to\infty$.

Proof: Let $\{M_n\} \subset \mathbf{M}^{k\times k}$ be such that $M_nB_0 \to M$ and $t_n \downarrow 0$ as $n\to\infty$. Thus we may write $M_n = [M_{n,1}, M_{n,2}]$ and $M = M_1 + M_2$ such that $M_{n,1} \in \mathbf{M}^{k\times r_0}$ and

$$M_{n,1}Q_{0,1}^\intercal \to M_1 , \quad M_{n,2}P_{0,2}^\intercal \to M_2 . \tag{E.19}$$

Clearly, $M_1U = 0$ for all $U \in \Psi(\Pi_0)$. For $\epsilon > 0$, let $\Psi(\Pi_0)^\epsilon$ and $\Psi(\Pi_0)_1^\epsilon$ be given in the proof of Proposition 3.1. In what follows we consider the nontrivial case when $\Pi_0 \neq 0$ and $M_2 \neq 0$. Let $d = k - r$. Then $\Psi(\Pi_0) \subsetneqq \mathbb{S}^{k\times d}$ and hence $\Psi(\Pi_0)_1^\epsilon \neq \emptyset$ for ϵ sufficiently small. Let $\sigma_{\min}^+(\Pi_0)$ be the smallest positive singular value of Π_0, which exists since $\Pi_0 \neq 0$. Let $\Delta \equiv 5\sqrt{2}[\sigma_{\min}^+(\Pi_0)]^{-1}(\max_{U\in\mathbb{S}^{k\times d}} \|M_2U\| + \max_{U\in\mathbb{S}^{k\times d}}\|M_1U\|) > 0$, which holds since $M_2 \neq 0$. Then it follows that, for all n sufficiently large,

$$\begin{aligned}\min_{U\in\Psi(\Pi_0)_1^{t_n\Delta}} \|(\Pi_0 + M_nT_nB_0)U\| &\geq \min_{U\in\Psi(\Pi_0)_1^{t_n\Delta}} \|\Pi_0U\| - \max_{U\in\mathbb{S}^{k\times d}} \|M_nT_nB_0U\| \\ &\geq \frac{\sqrt{2}}{2}t_n\sigma_{\min}^+(\Pi_0)\Delta - t_n\max_{U\in\mathbb{S}^{k\times d}}\|M_{n,1}Q_{0,1}^\intercal U\| - t_n^2\max_{U\in\mathbb{S}^{k\times d}}\|M_{n,2}P_{0,2}^\intercal U\| \\ &> t_n^2\max_{U\in\mathbb{S}^{k\times d}}\|M_{n,2}P_{0,2}^\intercal U\| \geq \min_{U\in\Psi(\Pi_0)}\|(\Pi_0 + M_nT_nB_0)U\| \\ &\geq \sqrt{\phi_r(\Pi_0 + M_nT_nB_0)} ,\end{aligned} \tag{E.20}$$

where the first inequality follows by the Lipschitz continuity of the min operator, the triangle inequality and the fact that $\Psi(\Pi_0)_1^{t_n\Delta} \subset \mathbb{S}^{k\times d}$, the second inequality follows by Lemma D.1 and the triangle inequality, the third inequality follows by the definition of Δ, $t_n \downarrow 0$, $M_{n,1}Q_{0,1}^\intercal \to M_1$, $M_{n,2}P_{0,2}^\intercal \to M_2$ as $n\to\infty$ and the simple fact that $2a - a_n > 0$ for all n large if $a_n \to a > 0$, the fourth inequality holds by the facts that $\Pi_0U = 0$ and $Q_{0,1}^\intercal U = 0$ for $U \in \Psi(\Pi_0)$, and the last by Lemma 3.1.

Next, let Γ^{Δ} and the correspondence $\varphi : \mathbf{R} \twoheadrightarrow \mathbb{S}^{k\times d} \times \Gamma^{\Delta}$ be given as in the proof of Proposition 3.1 for $\Delta > 0$. Then it follows that

$$\max_{U\in\Psi(\Pi_0)^{t_n\Delta}} \|M_n T_n B_0 U\| \leq t_n \max_{(U,V)\in\varphi(t_n)} \|(M_{n,1}Q_{0,1}^{\intercal})(U+t_nV)\| + t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_{0,2}^{\intercal}U\|$$
$$\leq t_n^2 \max_{V\in\Gamma^{\Delta}} \|M_{n,1}Q_{0,1}^{\intercal}V\| + t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_{0,2}^{\intercal}U\| , \qquad \text{(E.21)}$$

where the first inequality follows by the triangle inequality, $M_n = [M_{n,1}, M_{n,2}]$ and $\Psi(\Pi_0)^{t_n\Delta} \subset \mathbb{S}^{k\times d}$, and the second inequality follows from $Q_{0,1}^{\intercal}U = 0$ for $U \in \Psi(\Pi_0)$ and $\varphi(t_n) \subset \Psi(\Pi_0) \times \Gamma^{\Delta}$. By analogous arguments as in (E.20), we have, for all n large,

$$\min_{U\in\Psi(\Pi_0)_1^{t_n^{3/2}\Delta}\cap\Psi(\Pi_0)^{t_n\Delta}} \|(\Pi_0 + M_nT_nB_0)U\| \geq \min_{U\in\Psi(\Pi_0)_1^{t_n^{3/2}\Delta}} \|\Pi_0U\| - \max_{U\in\Psi(\Pi_0)^{t_n\Delta}} \|M_nT_nB_0U\|$$
$$\geq \frac{\sqrt{2}}{2} t_n^{3/2}\sigma_{\min}^{+}(\Pi_0)\Delta - t_n^2 \max_{V\in\Gamma^{\Delta}} \|M_{n,1}Q_{0,1}^{\intercal}V\| - t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_{0,2}^{\intercal}U\|$$
$$> t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_{0,2}^{\intercal}U\| \geq \min_{U\in\Psi(\Pi_0)} \|(\Pi_0 + M_nT_nB_0)U\|$$
$$\geq \sqrt{\phi_r(\Pi_0 + M_nT_nB_0)} , \qquad \text{(E.22)}$$

where the first inequality follows by the Lipschitz continuity of the min operator, the triangle inequality, $\Psi(\Pi_0)_1^{t_n^{3/2}\Delta} \cap \Psi(\Pi_0)^{t_n\Delta} \subset \Psi(\Pi_0)_1^{t_n^{3/2}\Delta}$ and $\Psi(\Pi_0)_1^{t_n^{3/2}\Delta} \cap \Psi(\Pi_0)^{t_n\Delta} \subset \Psi(\Pi_0)^{t_n\Delta}$, the second inequality follows by (E.21) and Lemma D.1, the third inequality follows by the definition of Δ and Γ^{Δ}, $t_n \downarrow 0$, $M_{n,1}Q_{0,1}^{\intercal} \to M_1$ and $M_{n,2}P_{0,2}^{\intercal} \to M_2$ as $n \to \infty$, the fourth inequality holds by the facts that $\Pi_0U = 0$ and $Q_{0,1}^{\intercal}U = 0$ for $U \in \Psi(\Pi_0)$. In turn, by analogous arguments, we have, for all n sufficiently large,

$$\min_{U\in\Psi(\Pi_0)_1^{t_n^2\Delta}\cap\Psi(\Pi_0)^{t_n^{3/2}\Delta}} \|(\Pi_0 + M_nT_nB_0)U\| > \sqrt{\phi_r(\Pi_0 + M_nT_nB_0)} . \qquad \text{(E.23)}$$

Combining (E.20), (E.22), (E.23) and Lemma 3.1, we thus obtain that, for all n large,

$$\phi_r(\Pi_0 + M_nT_nB_0) = \min_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(\Pi_0 + M_nT_nB_0)U\|^2 . \qquad \text{(E.24)}$$

Now, for the right hand side of (E.24), we have

$$\Big|\min_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(\Pi_0 + M_nT_nB_0)U\|^2 - \min_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(\Pi_0 + t_nM_1 + t_n^2M_2)U\|^2\Big|$$
$$\leq (O(t_n^2) + O(t_n^2)) \max_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(t_n(M_{1,n}Q_{0,1}^{\intercal} - M_1) + t_n^2(M_{2,n}P_{0,2}^{\intercal} - M_2))U\| , \qquad \text{(E.25)}$$

where the inequality follows by the formula $a^2-b^2 = (a+b)(a-b)$, the Lipschitz inequality of the min operator, the triangle inequality, and the facts that $\min_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(\Pi_0 + M_nT_nB_0)U\| = O(t_n^2)$ and $\min_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(\Pi_0 + MT_nB_0)U\| = O(t_n^2)$. For the second term on the right hand side of (E.25), we have

$$\begin{aligned}
&\max_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(t_n(M_{1,n}Q_{0,1}^\intercal - M_1) + t_n^2(M_{2,n}P_{0,2}^\intercal - M_2))U\| \\
&\le t_n \max_{(U,V)\in\varphi(t_n^2)} \|(M_{n,1}Q_{0,1}^\intercal - M_1)(U + t_n^2 V)\| + t_n^2 \max_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(M_{n,2}P_{0,2}^\intercal - M_2)U\| \\
&\le \max_{V\in\Gamma^\Delta} t_n^3\|(M_{n,1}Q_{0,1}^\intercal - M_1)V\| + t_n^2 \max_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(M_{n,2}P_{0,2}^\intercal - M_2)U\| = o(t_n^2)\ , \quad \text{(E.26)}
\end{aligned}$$

where the first inequality follows by the triangle inequality and the definition of $\varphi(t_n^2)$, the second inequality follows by the fact that $Q_{0,1}^\intercal U = 0$ and $M_1 U = 0$ for $U \in \Psi(\Pi_0)$ and $\varphi(t_n^2) \subset \Psi(\Pi_0) \times \Gamma^\Delta$, and the equality follows by applying the sub-multiplicativity of Frobenius norm and the facts that $M_{n,1}Q_{0,1}^\intercal \to M_1$ and $M_{n,2}P_{0,2}^\intercal \to M_2$ as $n \to \infty$. Combining results (E.24), (E.25) and (E.26), we then obtain

$$\phi_r(\Pi_0 + M_n T_n B_0) = \min_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(\Pi_0 + t_n M_1 + t_n^2 M_2)U\|^2 + o(t_n^4)\ . \quad \text{(E.27)}$$

Next, the first term on the right hand side of (E.27) can be written as

$$\begin{aligned}
\min_{U\in\Psi(\Pi_0)^{t_n^2\Delta}} \|(\Pi_0 + t_n M_1 + t_n^2 M_2)U\|^2 &= \min_{(U,V)\in\varphi(t_n^2)} \|(\Pi_0 + t_n M_1 + t_n^2 M_2)(U + t_n^2 V)\|^2 \\
&= t_n^4 \min_{(U,V)\in\varphi(t_n^2)} \|\Pi_0 V + MU\|^2 + o(t_n^4)\ , \quad \text{(E.28)}
\end{aligned}$$

where the second equality follows by the facts that $\Pi_0 U = 0$ and $M_1 U = 0$ for $U \in \Psi(\Pi_0)$, and $\|V\| \le \Delta$ for all $V \in \Gamma^\Delta$. By analogous arguments in (A.15), we have

$$\min_{(U,V)\in\varphi(t_n^2)} \|\Pi_0 V + MU\|^2 = \min_{U\in\Psi(\Pi_0)} \min_{V\in\mathbf{M}^{k\times d}} \|\Pi_0 V + MU\|^2 + o(1)\ . \quad \text{(E.29)}$$

Combining (E.27), (E.28) and (E.29), we may conclude that

$$\begin{aligned}
\lim_{n\to\infty} \frac{\phi_r(\Pi_0 + M_n T_n B_0)}{t_n^4} &= \min_{U\in\Psi(\Pi_0)} \min_{V\in\mathbf{M}^{k\times d}} \|\Pi_0 V + MU\| \\
&= \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal M Q_2)\ , \quad \text{(E.30)}
\end{aligned}$$

where the second equality follows by Lemma D.4, as desired. ■

Proposition E.2. *Suppose that there is an estimator $\hat{\Pi}_n : \{X_i\}_{i=1}^n \to \mathbf{M}^{k\times k}$ for $\Pi_0 \in \mathbf{M}^{k\times k}$ such that $\{\hat{\Pi}_n - \Pi_0\}B_0^{-1}D_n B_0 \xrightarrow{L} \mathcal{M}$ for some $\tau_n \uparrow \infty$ and random matrix $\mathcal{M} \in \mathbf{M}^{k\times k}$, where $D_n \equiv diag(\tau_n \mathbf{1}_{r_0}, \tau_n^2 \mathbf{1}_{k-r_0})$. If $\mathrm{rank}(\Pi_0) \le r$, then we have*

$$\tau_n^4 \phi_r(\hat{\Pi}_n) \xrightarrow{L} \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal \mathcal{M} Q_{0,2})\ .$$

Proof: For each $n \in \mathbf{N}$, define $g_n : \mathbf{M}^{k\times k} \to \mathbf{R}$ by

$$g_n(M) \equiv \tau_n^4 \phi_r(\Pi_0 + M D_n^{-1} B_0) \; . \tag{E.31}$$

By Proposition E.1, $g_n(M_n) \to \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^{\intercal} M Q_{0,2})$ whenever $M_n B_0 \to M$ as $n \to \infty$. In turn, since $\tau_n^4 \phi_r(\hat{\Pi}_n) = g_n((\hat{\Pi}_n - \Pi_0)B_0^{-1} D_n)$, the proposition follows by Theorem 1.11.1(i) in van der Vaart and Wellner (1996). ∎

E.2 Additional Examples

Our first example in this section arises in finite mixture models of dynamic discrete choices where a problem of both theoretical and practical importance is inference on the number of types (McLachlan and Peel, 2004; Kasahara and Shimotsu, 2009). It is also related to incomplete information games with multiple equiliria studied in Xiao (2018).

Example E.1 (Finite Mixtures, Discrete Choices and Multiple Equilibria)**.** Consider an individual with characteristic $Z_t \in \mathcal{Z} \equiv \{z_1, \ldots, z_d\}$ who makes a choice $S_t \in \mathcal{S} \equiv \{0, 1\}$ depending on his/her unknown (to econometricians) type, at time $t = 1, 2$. Suppose that there are γ_0 (finite) types. Under regularity conditions, Kasahara and Shimotsu (2009) establish a lower bound for γ_0, i.e., $\gamma_0 \geq \text{rank}(\Pi_0)$ with

$$\Pi_0 = \begin{bmatrix} 1 & \tilde{p}_{X_2}(1, z_1) & \cdots & \tilde{p}_{X_2}(1, z_d) \\ \tilde{p}_{X_1}(1, z_1) & \tilde{p}_{X_1,X_2}(1, z_1; 1, z_1) & \cdots & \tilde{p}_{X_1,X_2}(1, z_1; 1, z_d) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_{X_1}(1, z_d) & \tilde{p}_{X_1,X_2}(1, z_d; 1, z_1) & \cdots & \tilde{p}_{X_1,X_2}(1, z_d; 1, z_d) \end{bmatrix} , \tag{E.32}$$

where $X_t \equiv (Z_t, S_t) \in \mathcal{X} \equiv \mathcal{Z} \times \mathcal{S}$ for $t = 1, 2$, $\tilde{p}_{X_1}(1, z) \equiv \sum_{x_2 \in \mathcal{X}} \tilde{p}_{X_1,X_2}(1, z; x_2)$, $\tilde{p}_{X_2}(1, z) \equiv \sum_{x_1 \in \mathcal{X}} \tilde{p}_{X_1,X_2}(x_1; 1, z)$, and, for any $x = (z, s)$ and $x' = (z', s')$ in $\mathcal{X}$,

$$\tilde{p}_{X_1,X_2}(z, s; z', s') \equiv \frac{p_{X_1,X_2}(z, s; z', s')}{p_{Z_2|X_1}(z'; z, s)}$$

with p_{X_1,X_2} the probability mass function (pmf) of (X_1, X_2) and $p_{Z_2|X_1}$ the conditional pmf of Z_2 given X_1. Under additional conditions, Kasahara and Shimotsu (2009) show in fact $\gamma_0 = \text{rank}(\Pi_0)$. We focus on discrete variables and two periods for ease of exposition, but the results extend to more general cases by Remark 2(iv) and Propositions 3 and 8 in Kasahara and Shimotsu (2009). The number of types is crucial for the specification of mixture distributions, and yet inference on γ_0 without parametric assumptions on the component distributions has been understood to be challenging (Kasahara and Shimotsu, 2009, 2014; Bonhomme et al., 2016). By further restricting each component distribution to have independent marginals, Kasahara and Shimotsu (2014) and Bonhomme et al. (2016) accomplish nonparametric estimation of γ_0 based on the rank test of

Kleibergen and Paap (2006). Interestingly, Xiao (2018) derives a similar nonparametric identification result for the number of equilibria in incomplete information games and obtains a consistent estimator based on the rank test of Robin and Smith (2000). ■

Our second example pertains to the existence of general common features (Engle and Kozicki, 1993), which conceptually includes cointegration as a special case.

Example E.2 (Common Features)**.** Let $\{Y_t\}$ be a $k \times 1$ time series. According to Engle and Kozicki (1993), a feature that is present in each component of Y_t is said to be common to Y_t if there exists a nonzero linear combination of Y_t that fails to have the feature. To fix ideas, suppose that $\{Y_t\}$ is generated according to

$$Y_t = \Gamma_0^\intercal Z_t + \Xi_0^\intercal W_t + u_t \ , \tag{E.33}$$

where W_t can be thought of as control variables, and Z_t is an $m \times 1$ vector reflecting the feature under consideration with $m \geq k$. For example, testing for the existence of common serial correlation would set Z_t to be lags of Y_t, and testing for the existence of common conditionally heteroskedastic factors would set Z_t to be relevant factors. We refer to Engle and Kozicki (1993), Engle and Susmel (1993) and Dovonon and Renault (2013) for details of these and other examples. By the specification of (E.33), existence of common features means that Γ_0 is not of full rank. Thus, testing for the existence of common features reduces to examining the hypotheses in (1) with

$$\Pi_0 = \Gamma_0 \text{ and } r = k - 1 \ . \tag{E.34}$$

Since the number of common features is generally unknown *a priori*, the assumption $\text{rank}(\Pi_0) \geq k - 1$ that underlies the hypotheses in (2) may again be unrealistic. ■

Our next example involves estimation of the rank of demand systems, a notion developed by Gorman (1981) for exactly aggregable demand systems and generalized by Lewbel (1991) to all demand systems.

Example E.3 (Consumer Demand)**.** An Engel curve is the function describing the allocation of an individual's consumption expenditures with the prices of all goods fixed, and the rank of a demand system is the dimension of the space spanned by the Engel curves of the system (Lewbel, 1991). Suppose that there are k goods in the system and that the Engel curve is given by

$$Y = \Gamma_0 G(Z) + u \ , \tag{E.35}$$

where Y is a $k \times 1$ vector of budget shares on the k goods, Z is the total expenditure, $G(\cdot)$ is a $r_0 \times 1$ vector of unknown function with $r_0 \leq k$, and u is an error term. The rank of the demand system ie precisely r_0, and in fact also equal to the rank of

$$\Pi_0 = E[Q(Z)Y^\intercal] \ , \tag{E.36}$$

where $Q(\cdot)$ is an $m \times 1$ vector of known functions with $m \geq k$, under suitable conditions. Estimation of the rank of the demand system is important because it provides evidence on consistency of consumer behaviors with utility maximization, and has implications for welfare comparisons and aggregation across goods and across consumers (Lewbel, 1991, 2006; Barnett and Serletis, 2008). ■

Our fourth example shows the importance of rank estimation in identifying the number of factors in factor models (Anderson, 2003; Lam and Yao, 2012).

Example E.4 (Factor Analysis)**.** Let $Y \in \mathbf{R}^d$ be generated by the following model

$$Y = \mu_0 + \Lambda_0 F + u \ , \tag{E.37}$$

where F is a $r_0 \times 1$ vector of unobserved common factors with $r_0 \leq d$, and u is an error term. Partition $Y = [Y_1^\intercal, Y_2^\intercal, Y_3^\intercal]^\intercal$ for $Y_1 \in \mathbf{R}^m$ and $Y_2 \in \mathbf{R}^k$ with some $r_0 \leq k \leq m < d$ and $m + k \leq d$, and also $\Lambda_0 = [\Lambda_{0,1}^\intercal, \Lambda_{0,2}^\intercal, \Lambda_{0,3}^\intercal]^\intercal$ with $\Lambda_{0,1}$ and $\Lambda_{0,2}$ having m and k rows. Then under appropriate restrictions, the rank of $\mathrm{Var}(F)$ is equal to the rank of

$$\Pi_0 = \mathrm{Cov}(Y_1, Y_2) \ . \tag{E.38}$$

Thus, determining the number r_0 of the common factors reduces to estimation of the rank of Π_0. Such a problem also arises in the interbattery factor analysis (Gill and Lewbel, 1992), the dynamic analysis of time series (Lam and Yao, 2012), and finance and macroeconomics (Bai and Ng, 2002, 2007). ■

Our final example is taken from Gill and Lewbel (1992), and manifests how matrix rank determination is useful in model selection in time series models.

Example E.5 (Model Selection)**.** Let $\{Y_t\}$ be a $p \times 1$ weakly stationary time series, which has the following state space representation:

$$Y_t = \Gamma_0 Z_t + u_t \ , \ Z_t = \Lambda_0 Z_{t-1} + \epsilon_t \ , \tag{E.39}$$

where Z_t is a $r_0 \times 1$ vector of state variables, and u_t and ϵ_t are error terms. It turns out that the number r_0 of state variables is equal to the rank of the Hankel matrix

$$\Pi_0 = E(\begin{bmatrix} Y_{t+1} \\ \vdots \\ Y_{t+b} \end{bmatrix} \begin{bmatrix} Y_t^\intercal & \cdots & Y_{t-b+1}^\intercal \end{bmatrix}) \ , \tag{E.40}$$

for b sufficiently large (Aoki, 1990, p.52). Consequently, determining the number of state variables r_0 to model Y_t reduces to determining the rank of Π_0. When Y_t is a scalar and follows an ARMA(p_1, p_2) model, then Y_t has a state space representation with the number r_0 of state variables equal to $\max(p_1, p_2)$ (Aoki, 1990). Thus, determining the rank of the Hankel matrix is crucial for model specification in these contexts. ■

For simplicity, we verify the main assumptions only for Example E.1. Let $\{X_{it}\}_{i=1}^n$ be a sample generated by the mixture model with $X_{it} = (Z_{it}, S_{it})$ for $t = 1, 2$. Then we estimate Π_0 by its empirical analog:

$$\hat{\Pi}_n = \begin{bmatrix} 1 & \tilde{p}_{X_2,n}(1, z_1) & \cdots & \tilde{p}_{X_2,n}(1, z_d) \\ \tilde{p}_{X_1,n}(1, z_1) & \tilde{p}_{X_1,X_2,n}(1, z_1; 1, z_1) & \cdots & \tilde{p}_{X_1,X_2,n}(1, z_1; 1, z_d) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_{X_1,n}(1, z_d) & \tilde{p}_{X_1,X_2,n}(1, z_d; 1, z_1) & \cdots & \tilde{p}_{X_1,X_2,n}(1, z_d; 1, z_d) \end{bmatrix}, \tag{E.41}$$

where $\tilde{p}_{X_1,n}(1, z) \equiv \sum_{x_2 \in \mathcal{X}} \tilde{p}_{X_1,X_2,n}(1, z; x_2)$, $\tilde{p}_{X_2,n}(1, z) \equiv \sum_{x_1 \in \mathcal{X}} \tilde{p}_{X_1,X_2,n}(x_1; 1, z)$, and, for any $x = (z, s)$ and $x' = (z', s')$ in $\mathcal{X}$,

$$\tilde{p}_{X_1,X_2,n}(z, s; z', s') \equiv \frac{\hat{p}_{X_1,X_2,n}(z, s; z', s')}{\hat{p}_{Z_2|X_1,n}(z'; z, s)}$$

with $\hat{p}_{X_1,X_2,n}$ the empirical pmf of $\{(X_{i1}, X_{i2})\}_{i=1}^n$ and $\hat{p}_{Z_2|X_1,n}(\cdot\,; z, s)$ the empirical conditional pmf of $\{Z_{i2}\}$ given $X_1 = (z, s)$. Since sums and ratios are differentiable maps (at nonzero denominators as assumed in Kasahara and Shimotsu (2009)), a simple application of the Delta method shows that $\hat{\Pi}_n$ satisfies Assumption 3.1 with $\tau_n = \sqrt{n}$ and $\mathcal{M}$ some centered Gaussian matrix, under standard regularity conditions.

Next, suppose that $\{(X_{i1}, X_{i2})\}_{i=1}^n$ are i.i.d. across i for ease of exposition. Let $\{(X_{i1}^*, X_{i2}^*)\}_{i=1}^n$ be an i.i.d. sample drawn with replacement from $\{(X_{i1}, X_{i2})\}_{i=1}^n$. Then we propose the bootstrap estimator $\hat{\Pi}_n^*$ as follows:

$$\hat{\Pi}_n^* = \begin{bmatrix} 1 & \tilde{p}^*_{X_2,n}(1, z_1) & \cdots & \tilde{p}^*_{X_2,n}(1, z_d) \\ \tilde{p}^*_{X_1,n}(1, z_1) & \tilde{p}^*_{X_1,X_2,n}(1, z_1; 1, z_1) & \cdots & \tilde{p}^*_{X_1,X_2,n}(1, z_1; 1, z_d) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}^*_{X_1,n}(1, z_d) & \tilde{p}^*_{X_1,X_2,n}(1, z_d; 1, z_1) & \cdots & \tilde{p}^*_{X_1,X_2,n}(1, z_d; 1, z_d) \end{bmatrix}, \tag{E.42}$$

where $\tilde{p}^*_{X_1,n}(1, z) \equiv \sum_{x_2 \in \mathcal{X}} \tilde{p}^*_{X_1,X_2,n}(1, z; x_2)$, $\tilde{p}^*_{X_2,n}(1, z) \equiv \sum_{x_1 \in \mathcal{X}} \tilde{p}^*_{X_1,X_2,n}(x_1; 1, z)$, and, for any $x = (z, s)$ and $x' = (z', s')$ in $\mathcal{X}$,

$$\tilde{p}^*_{X_1,X_2,n}(z, s; z', s') \equiv \frac{\hat{p}^*_{X_1,X_2,n}(z, s; z', s')}{\hat{p}^*_{Z_2|X_1,n}(z'; z, s)}$$

with $\hat{p}^*_{X_1,X_2,n}$ and $\hat{p}^*_{Z_2|X_1,n}(\cdot\,; z, s)$ the bootstrap analogs of $\hat{p}_{X_1,X_2,n}$ and $\hat{p}_{Z_2|X_1,n}(\cdot\,; z, s)$ respectively based on $\{(X_{i1}^*, X_{i2}^*)\}_{i=1}^n$. Assumption 3.2 now follows from the Delta method for bootstrap, as a result of the same differentiability mentioned previously – see, for example, Theorem 23.5 in van der Vaart (1998).

References

Al-Sadoon, M. M. (2017). A unifying theory of tests of rank. *Journal of Econometrics*, **199** 49–62.

Aliprantis, C. D. and Border, K. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 3rd ed. Springer Verlag.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley New York.

Andrews, D. W. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, **67** 543–563.

Aoki, M. (1990). *State Space Modeling of Time Series*. 2nd ed. Springer.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70** 191–221.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, **25** 52–60.

Barnett, W. A. and Serletis, A. (2008). Measuring consumer preferences and estimating demand systems. In *Quantifying Consumer Preferences*. 1–35.

Bauer, P., Pötscher, B. M. and Hackl, P. (1988). Model selection by multiple test procedures. *Statistics*, **19** 39–44.

Berkowitz, J. and Diebold, F. X. (1998). Bootstrapping multivariate spectra. *The Review of Economics and Statistics*, **80** 664–666.

Bhatia, R. (1997). *Matrix Analysis*. Springer.

Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics*, **77** 379–404.

Bonhomme, S., Jochmans, K. and Robin, J.-M. (2016). Non-parametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **78** 211–229.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*, vol. 36. Siam.

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

Chen, Q. and Fang, Z. (2018). Inference on functionals under first order degeneracy. *Available at SSRN: https://ssrn.com/abstract=2662758*.

Cragg, J. G. and Donald, S. G. (1997). Inferring the rank of a matrix. *Journal of Econometrics*, **76** 223 – 250.

Dovonon, P. and Renault, E. (2013). Testing for common conditionally heteroskedastic factors. *Econometrica*, **81** 2561–2586.

Dupuy, A. and Galichon, A. (2014). Personality traits and the marriage market. *Journal of Political Economy*, **122** 1271–1319.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, **7** 1–26.

Engle, R. F. and Kozicki, S. (1993). Testing for common features. *Journal of Business & Economic Statistics*, **11** 369–380.

Engle, R. F. and Susmel, R. (1993). Common volatility in international equity markets. *Journal of Business & Economic Statistics*, **11** 167–176.

Fang, Z. and Santos, A. (2018). Inference on directionally differentiable functions. *The Review of Economic Studies*. URL `http://dx.doi.org/10.1093/restud/rdy049`.

Galichon, A. and Salanié, B. (2010). Matching with trade-offs: Revealed preferences over competing characteristics. *Available at SSRN: https://ssrn.com/abstract=1640380*.

Gill, L. and Lewbel, A. (1992). Testing the rank and definiteness of estimated matrices with applications to factor, state-space and arma models. *Journal of the American Statistical Association*, **87** 766–776.

Gorman, W. M. (1981). Some Engel curves. In *Essays in the Theory and Measurement of Consumer Behaviour: In Honour of Sir Richard Stone*. Cambridge University Press, 7–29.

Granas, A. and Dugundji, J. (2003). *Fixed Point Theory*. Springer.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.

Hannan, E. J. (1970). *Multiple Time Series*. Wiley.

Harris, D. (1997). Principal components analysis of cointegrated time series. *Econometric Theory*, **13** 529–557.

Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*. Springer.

Hayashi, F. (2000). *Econometrics*. Princeton University Press, New Jersey.

Hosoya, Y. (1989). Hierarchical statistical models and a generalized likelihood ratio test. *Journal of the Royal Statistical Society. Series B. Methodological*, **51** 435–447.

Hubrich, K., Lütkepohl, H. and Saikkonen, P. (2001). A review of systems cointegration tests. *Econometric Reviews*, **20** 247–318.

Jagannathan, R. and Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *The Journal of Finance*, **51** 3–53.

Jentsch, C. and Kreiss, J.-P. (2010). The multiple hybrid bootstrap—resampling multivariate linear processes. *Journal of Multivariate Analysis*, **101** 2320–2345.

Jentsch, C. and Politis, D. N. (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *The Annals of Statistics*, **43** 1117–1140.

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

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

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

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

Kasahara, H. and Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, **77** 135–175.

Kasahara, H. and Shimotsu, K. (2014). Non-parametric identification and estimation of the number of components in multivariate mixtures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76** 97–111.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, **133** 97 – 126.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, **54** 159–178.

LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data.* Springer.

LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, **40** 694–726.

LEDOUX, M. (2001). *The Concentration of Measure Phenomenon.* American Mathematical Society.

LEWBEL, A. (1991). The rank of demand systems: Theory and nonparametric estimation. *Econometrica*, **59** 711–730.

LEWBEL, A. (2006). Engel curves. In *The New Palgrave Dictionary of Economics*, vol. 2.

LIAO, Z. and PHILLIPS, P. C. (2015). Automated estimation of vector error correction models. *Econometric Theory*, **31** 581–646.

MADDALA, G. S. and KIM, I.-M. (1998). *Unit Roots, Cointegration, and Structural Change.* 4, Cambridge University Press.

MAGNUS, J. R. and NEUDECKER, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* 3rd ed. John Wiley & Sons.

MCLACHLAN, G. and PEEL, D. (2004). *Finite Mixture Models.* John Wiley & Sons.

NYBLOM, J. and HARVEY, A. (2000). Tests of common stochastic trends. *Econometric Theory*, **16** 176–199.

PHILLIPS, P. C., SUN, Y. and JIN, S. (2006). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review*, **47** 837–894.

POLITIS, D. (2001). On nonparametric function estimation with infinite-order flat-top kernels. In *Probability and Statistical Models with Applications.* Chapman and Hall/CRC, 484–498.

POLITIS, D. N. (2011). Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, **27** 703–744.

POLITIS, D. N. and ROMANO, J. P. (1993). Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis*, **47** 301–328.

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, **22** 2031–2050.

PÖTSCHER, B. M. (1983). Order estimation in ARMA-models by Lagrangian multiplier tests. *The Annals of Statistics*, **11** 872–885.

Robin, J.-M. and Smith, R. J. (2000). Tests of rank. *Econometric Theory*, **16** 151–175.

Rudin, W. (1976). *Principles of Mathematical Analysis*. 3rd ed. McGraw-Hill New York.

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, **66** 477–487.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research*, **30** 169–186.

Shintani, M. (2001). A simple cointegrating rank test without vector autoregression. *Journal of Econometrics*, **105** 337–362.

Snell, A. (1999). Testing for r versus $r - 1$ cointegrating vectors. *Journal of Econometrics*, **88** 151–191.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American statistical Association*, **83** 1097–1107.

Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models1. *Econometrica*, **74** 1699–1714.

Tao, T. (2012). *Topics in Random Matrix Theory*. American Mathematical Society.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag.

van Giersbergen, N. P. (1996). Bootstrapping the trace statistic in VAR models: Monte Carlo results and applications. *Oxford Bulletin of Economics and Statistics*, **58** 391–408.

West, K. D. (1997). Another heteroskedasticity-and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics*, **76** 171–191.

Xiao, R. (2018). Identification and estimation of incomplete information games with multiple equilibria. *Journal of Econometrics*, **203** 328 – 343.