Quantitative Economics 10 (2019), 1787–1824 1759-7331/20191787

Improved inference on the rank of a matrix

Qihui Chen
School of Management and Economics and Shenzhen Finance Institute,
The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)

Zheng Fang
Department of Economics, Texas A&M University

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 $H_0 : \text{rank}(\Pi_0) \leq r$. The problem is of first-order importance because the previous literature focuses on $H_0' : \text{rank}(\Pi_0) = r$ by implicitly assuming away $\text{rank}(\Pi_0) < r$, which may lead to invalid rank tests due to overrejections. In particular, we show that limiting distributions of test statistics under H_0' may not stochastically dominate those under $\text{rank}(\Pi_0) < r$. A multiple test on the nulls $\text{rank}(\Pi_0) = 0, \ldots, r$, 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.

JEL classification. C12, C15.

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

Qihui Chen: qihuichen@cuhk.edu.cn
Zheng Fang: zfang@tamu.edu

We are indebted to 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. Qihui Chen gratefully acknowledges financial support by the National Natural Science Foundation of China (Grant No. 71803169).

 https://doi.org/10.3982/QE1139

(2004), Kasahara and Shimotsu (2009)), and estimation of matching dimensions (Dupuy and Galichon (2014))—more can be found in Online Supplemental Appendix E which is located within the replication file (Chen and Fang (2019)). These problems reduce to examining the hypotheses: for an unknown matrix Π_0 of size $m \times k$ with $m \geq k$,

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

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

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

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 overreject for some data generating processes and underreject 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 the Online Supplemental Appendix C (Chen and Fang (2019)) 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

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

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 the Online 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 $\text{rank}(\Pi_0) < r$; 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 nonpivotal and highly nonstandard. In particular, they depend on the true rank (among other things), upholding the importance of taking into account the possibility $\text{rank}(\Pi_0) < r$. For this, we appeal to modified bootstrap schemes recently developed by Fang and Santos (2018) and Chen and Fang (2019), 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, Hong, and Tamer (2007), Andrews and Soares (2010), Linton, Song, and Whang (2010)). While we are unable to offer a general theory guiding their choices, a two-step procedure similar to Romano, Shaikh, and Wolf (2014) is proposed to mitigate potential concerns. The intuition is as follows. First, the appearance of $r_0 \equiv \text{rank}(\Pi_0)$ 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 $\hat{r}_n$, that is, 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 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 nonstationary, the convergence rate may be non-$\sqrt{n}$ and even heterogeneous across entries of $\hat{\Pi}_n$ (see the Online 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 resampling 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) developed 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) proposed a test that requires nonsingularity of Ω_0 and may be sensitive to the transformations involved. Cragg and Donald (1997) provided a test based on a constrained minimum distance criterion, which, in addition to the nonsingularity requirement of Ω_0, is in general computationally intensive. To relax the nonsingularity condition, Robin and Smith (2000) employed a class of testing statistics which are asymptotically equivalent to ours, but their results only apply to the setup (2). Kleibergen and Paap (2006) studied 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.

There are a few exceptions that study (1). Johansen (1988, 1991) obtained his likelihood ratio statistics under H_0 but only establishes their asymptotic distributions under H_0'. Shortly after, Johansen (1995, pp. 157–8, 168) presented the limits under H_0, and essentially argues based on simulations that the asymptotic distributions under $\text{rank}(\Pi_0) < r$ are first-order stochastically dominated by those under H_0' and "hence not relevant for calculating the p-value." However, the counterexample given in Section 2 disproves this conjecture. Cragg and Donald (1993, p. 225) recognized 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 $\text{tr}(A)$ if it is square, its vectorization by $\text{vec}(A)$, and its Frobenius norm by $\|A\| \equiv \sqrt{\text{tr}(A^\intercal A)}$. 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 $\text{diag}(a)$ 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 $\text{rank}(\Pi_0) < r$, 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 an Online Supplemental Appendix. We also study the estimation problem, but, due to space limitation, relegate the results to Online 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 Online 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 $\text{rank}(\Pi_0) < r$ 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.

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 1 plots (based on simulations) two cdfs F_1 and F_2 of $T_{n,\text{kp}}$ in (4) respectively determined by

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

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 overrejection 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 the Online Supplemental Appendix B.

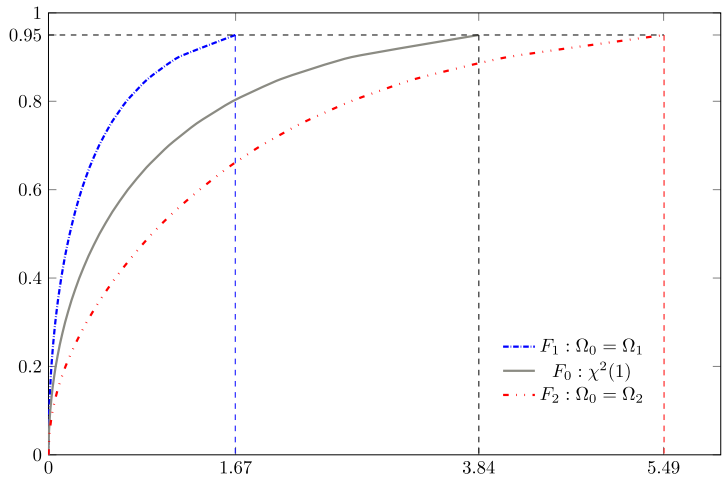

FIGURE 1. The cdfs of the KP statistic when $\Pi_0 = \mathbf{0}_{2\times 2}$ and $r = 1$.

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 the Online Supplemental Appendix B), which may generate substantial power loss. To illustrate, consider the following data generating process:

$$Z = \Pi_0^\intercal V + u, \tag{6}$$

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

$$\Pi_0 = \text{diag}(\mathbf{1}_{6-d}, \mathbf{0}_d) + \delta I_6. \tag{7}$$

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 (6), we 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 functions (against δ) of the multiple KP test, labeled KP-M. 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 accommodate $\text{rank}(\Pi_0) < r$.

To compare, we first show that three versions of our test—CF-A, CF-N, and CF-T (see below)—control size even when the KP test does not. Let $\{Z_i\}_{i=1}^{1000}$ be an i.i.d. sample in $\mathbf{M}^{2\times 2}$ such that $\text{vec}(Z_1) \sim N(\text{vec}(\Pi_0), \Omega_0)$, where $\text{vec}(\Pi_0) = \delta\Omega_0^{1/2}\text{vec}(I_2)$ with $\delta \geq 0$ and $\Omega_0 \in \{\Omega_1, \Omega_2\}$ as in (5). We test (1) with $r = 1$ based on $\hat{\Pi}_n = \frac{1}{1000}\sum_{i=1}^{1000} Z_i$, at $\alpha = 5\%$. Figure 3 shows our tests indeed control size for both choices of Ω_0, while the KP test underrejects when $\Omega_0 = \Omega_1$ and overrejects when $\Omega_0 = \Omega_2$. Note also that the KP-M test is conservative. Next, for the designs in (6) and (7), Figure 2 depicts the power

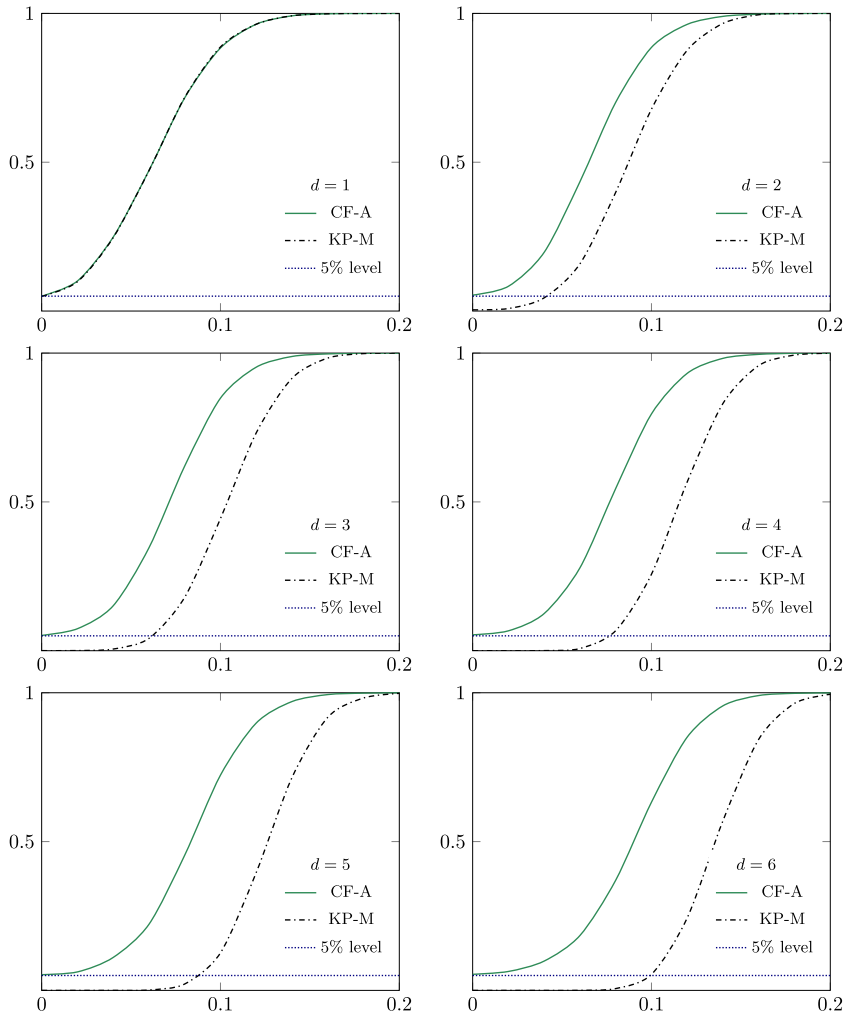

FIGURE 2. Conservativeness of the KP-M test. The number of Monte Carlo simulations is 10,000, the number of bootstrap repetitions is 500, and $\kappa_n = n^{-1/4}$ (for CF-A).

curves of CF-A. For $d = 1$, CF-A and KP-M have virtually the same rejection rates across δ. Whenever $d > 1$, our test effectively raises the power curves of the KP-M test so that the null rejection rates equal 5%, and the power becomes nontrivial, but it is more than that. The power improvement increases when d gets larger.

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 nonstationary time series. Then our test statistic for (1) is $\tau_n^2 \phi_r(\hat{\Pi}_n) \equiv$

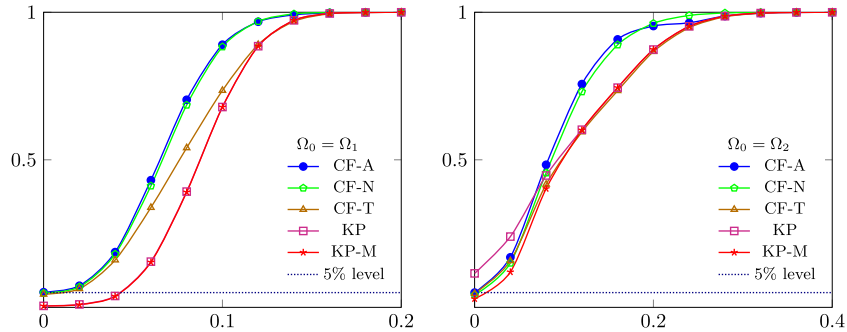

FIGURE 3. Comparisons with the KP and the KP-M tests. The number of Monte Carlo simulations is 10,000, the number of bootstrap repetitions is 1000, $\kappa_n = n^{-1/4}$ (for both CF-A and CF-N), and $\beta = \alpha/10$ (for CF-T).

$\tau_n^2 \sum_{j=r+1}^{k} \sigma_j^2(\hat{\Pi}_n)$. It turns out that, under H_0, we have: for $r_0 \equiv \operatorname{rank}(\Pi_0)$,

$$\tau_n^2 \phi_r(\hat{\Pi}_n) \xrightarrow{L} \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2\big(P_{0,2}^{\intercal} \mathcal{M} Q_{0,2}\big), \tag{8}$$

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 (8) 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, toward constructing critical values.

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

$$\hat{r}_n = \max\big\{j = 1, \ldots, r : \sigma_j(\hat{\Pi}_n) \geq \kappa_n\big\} \tag{9}$$

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 (8) is estimated by the conditional distribution (given the data) of

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

Given a significance level α, the CF-A test rejects 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 (10) given the data.

While we are unable to provide an optimal choice of κ_n, a two-step test, CF-T, 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 $\hat{r}_n \le r$. In the second step, we plug $\hat{r}_n$ into (10) 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 (9) 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 the Online 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.

Step 1: Compute a singular value decomposition $\hat{\Pi}_n = \hat{P}_n \hat{\Sigma}_n \hat{Q}_n^{\intercal}$.

Step 2: Obtain $\hat{r}_n$ as in (9) for a chosen κ_n (e.g., $\kappa_n = n^{-1/4}$).

Step 3: Bootstrap B times and compute copies of $\hat{\mathcal{M}}_n^*$, denoted $\{\hat{\mathcal{M}}_{n,b}^*\}_{b=1}^B$.

Step 4: 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, set $\hat{c}_{n,1-\alpha}$ to be the $\lfloor B(1-\alpha) \rfloor$-th largest value in

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

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

Compared to CF-N which is based on the numerical differentiation (Hong and Li (2018)) (see Sections 3 and 4 for more details), CF-A is somewhat insensitive to the choice of κ_n even in small samples. The two-step test CF-T, on the other hand, is overall the least sensitive, but may be oversized in small samples ($n \le 100$). Thus, for practical purposes, we 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 $\operatorname{rank}(\Pi_0) = 0, 1, \ldots, k-1$ at level β (e.g., $\beta = \alpha/10$) using the KP test (based on $\hat{\Pi}_n$), that is, $\hat{r}_n = j^*$ if accepting $\operatorname{rank}(\Pi_0) = j^*$ 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 $\operatorname{rank}(\Pi_0) < r$. 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 (2019), 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 the Online Supplemental Appendix for a brief review. Hence, the hypotheses in (1) are equivalent to

$$\mathrm{H}_0 : \phi_r(\Pi_0) = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \phi_r(\Pi_0) > 0,$$

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

$$\phi_r(\Pi) \equiv \sum_{j=r+1}^{k} \sigma_j^2(\Pi). \tag{11}$$

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 $\operatorname{rank}(\Pi_0) = r$ and in this way facilitates comparisons with existing rank tests.

Toward 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* (11)*, we have*

$$\phi_r(\Pi) = \min_{U \in \mathbb{S}^{k\times(k-r)}} \|\Pi U\|^2. \tag{12}$$

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, that is, $\sigma_k^2(\Pi) = \min_{\|U\|=1} \|\Pi U\|^2$. Note that the minimum in (12) 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

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

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

$$\lim_{n\to\infty} \frac{\phi(\Pi + t_n M_n) - \phi(\Pi) - t_n \phi'_\Pi(M_n)}{t_n^2} = \phi''_\Pi(M), \tag{13}$$

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

For simplicity, we shall drop the qualifier "Hadamard" in what follows, with the understanding that both full differentiability and directional differentiability (both first and second order) are meant in the Hadamard sense. Definition 3.1(i) generalizes (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 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 (2019)). The second-order directional differentiability becomes second-order 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* (11).

(i) *ϕ_r is first-order 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*

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

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

(ii) *ϕ_r is second-order 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 $r_0 \equiv \operatorname{rank}(\Pi)$,*

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

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. 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 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 fully differentiable precisely when $\operatorname{rank}(\Pi)<r$, reflecting a critical irregular nature of our setup; see Lemma D.5 for more details. To gain further intuition, suppose that $\Pi_0=\operatorname{diag}(\pi_{0,1},\pi_{0,2})$ and we want to test if $\operatorname{rank}(\Pi_0)\le 1$. Then by definition

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

Note that if $\operatorname{rank}(\Pi_0)\le 1$, then $\pi_{0,1}^2=\pi_{0,2}^2$ if and only if $\operatorname{rank}(\Pi_0)<1$ 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 $\operatorname{rank}(\Pi_0)<1$ 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 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 directionally differentiable maps (Shapiro (2000), Chen and Fang (2019)). Toward 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 $m\ge k$) 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 $\operatorname{vec}(\mathcal{M})$ 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 the Online 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 $\Pi_0\in\mathbf{M}^{m\times k}$,*

$$\tau_n\{\phi_r(\hat\Pi_n)-\phi_r(\Pi_0)\}\xrightarrow{L}\min_{U\in\Psi(\Pi_0)}2\operatorname{tr}\big(U^\intercal\Pi_0^\intercal\mathcal{M}U\big).$$

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

$$\tau_n^2\phi_r(\hat\Pi_n)\xrightarrow{L}\sum_{j=r-r_0+1}^{k-r_0}\sigma_j^2\big(P_{0,2}^\intercal\mathcal{M}Q_{0,2}\big),\tag{14}$$

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. Toward constructing critical values, we would then like to estimate the law of the limit. Unfortunately, as shown by Chen and Fang (2019), 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 (14) 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 rank of Π_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, nonpivotalness 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

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

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 (14) can be rewritten as

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

as claimed, where we exploited a property of the vec operator (Hamilton (1994, Proposition 10.4)). Our general limit in (14) 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 the 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 (2019)), and this remains true in the conventional setup where $\text{rank}(\Pi_0) = r$. Second, the possibility $\text{rank}(\Pi_0) < r$ 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 (2019) 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

$$\sup_{f\in \text{BL}_1(\mathbb{D})} \left|E_W\left[f\left(\mathbb{G}_n^*\right)\right] - E\left[f(\mathbb{G})\right]\right| = o_p(1),$$

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

$$\text{BL}_1(\mathbb{D}) \equiv \left\{f : \mathbb{D} \to \mathbf{R} : \sup_{x\in\mathbb{D}}\left|f(x)\right| \le 1, \left|f(x) - f(y)\right| \le \|x - y\|_{\mathbb{D}}\ \forall x, y \in \mathbb{D}\right\}.$$

Given the metric, we now proceed by imposing the following.

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 (2019) 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 $n \to \infty$,

$$\hat{\phi}''_{r,n}(M_n) \xrightarrow{p} \phi''_{r,\Pi_0}(M). \tag{15}$$

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 (13) in the definition of second-order derivative, that is, we estimate ϕ''_{r,Π_0} by: for any $M \in \mathbf{M}^{m\times k}$,

$$\hat{\phi}''_{r,n}(M) = \frac{\phi_r(\hat{\Pi}_n + \kappa_n M) - \phi_r(\hat{\Pi}_n)}{\kappa_n^2}, \tag{16}$$

for a suitable $\kappa_n \downarrow 0$, where we have exploited $\phi'_{r,\Pi_0} = 0$ under the null. By Chen and Fang (2019), (16) meets the requirement (15) 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. The key is consistent estimation of r_0: given a consistent estimator $\hat{r}_n$ of r_0, we may then obtain estimators $\hat{P}_{2,n}$ and $\hat{Q}_{2,n}$ of $P_{0,2}$ and $Q_{0,2}$, respectively, in a straightforward manner as described in Section 2. One possible construction of $\hat{r}_n$ is given by (9). Alternatively, $\hat{r}_n$ may also be constructed by sequential testing, and the tuning parameter then becomes an adjusted significance level; see the Online Supplemental Appendix C. In any case, by Lemma D.6, we may then obtain a consistent estimator for ϕ''_{r,Π_0}: for any $M \in \mathbf{M}^{m\times k}$,

$$\hat{\phi}''_{r,n}(M) = \sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2\big(\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n}\big). \tag{17}$$

Similar to the numerical estimator, the analytic estimator (17) 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 (9), then κ_n has a meaningful interpretation, namely, it measures the parsimoniousness in selecting the rank.

Given a significance level α, we now formally define our critical value $\hat{c}_{n,1-\alpha}$ as

$$\hat{c}_{n,1-\alpha} \equiv \inf\{c \in \mathbf{R} : P_W(\hat{\phi}''_{r,n}(\hat{\mathcal{M}}^*_n) \leq c) \geq 1-\alpha\}, \tag{18}$$

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 $\hat{\phi}''_{r,n}$ by either (16), or (9) and (17).

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 (18) where $\hat{\phi}''_{r,n}$ is given by either (16) with $\{\kappa_n\}$ satisfying $\kappa_n \downarrow 0$ and $\tau_n\kappa_n \to \infty$, or (17) with $\hat{r}_n \xrightarrow{p} r_0$ under H_0. If the cdf of the limiting distribution in (14) is continuous and strictly increasing at its $(1-\alpha)$-quantile for $\alpha \in (0,1)$, then under H_0,*

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

Furthermore, under H_1,

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

Theorem 3.2 shows that our test is not conservative in the pointwise sense while accommodating the possibility $\mathrm{rank}(\Pi_0) < r$. 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.

In addition to the attractive features mentioned after Assumption 3.1, 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 toward the end of Section 3.2. Consequently, the quantile condition is automatically satisfied provided the covariance matrix

TABLE 1. Estimation of rank(Π_0) defined by the models (6)–(7).

	Choices of κ_n in (9)					Choices of β for the Sequential Method					
d	$n^{-1/4}$	$n^{-1/3}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	$\alpha/5$	$\alpha/10$	$\alpha/15$	$\alpha/20$	$\alpha/25$	$\alpha/30$
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

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) required 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

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

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 $\hat{P}^{\intercal}_{2,n}\hat{\mathcal{M}}^*_{n,b}\hat{Q}_{2,n}$, respectively.

A common feature of our previous two tests is their dependence on a tuning parameter; see (16) and (17). To mitigate concerns on sensitivity to the choice of tuning parameters, we next develop a two-step test by exploiting the structure in (17). The intuition is as follows. The estimator (9), though consistent, may differ from the truth in finite samples. We would thus like to control $P(\hat{r}_n = r_0)$, for which (9) 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.[1] Table 1 compares the empirical probabilities of $\{\hat{r}_n = r_0\}$ for $\hat{r}_n$ obtained by (9) and the sequential KP test respectively, based on the same simulation data from Section 2 when $d > 1$. The empirical probabilities for (9) 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.

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

[1] 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.

plug $\hat{r}_n$ into (17) 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

$$\psi_n = 1\{\hat{r}_n > r \text{ or } \tau_n^2\phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}\}, \tag{19}$$

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 (14) 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* (19). *Then, under* H_0,

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

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

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

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, Shaikh, and Wolf (2014)). A common feature that our test shares here is that the size control is not exact, that is, 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, Shaikh, and Wolf (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 (18), and the other one applies when one has control over $P(\hat{r}_n = r_0)$: if $\liminf_{n\to\infty} P(\hat{r}_n = r_0) \ge 1-\beta$, we reject 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 (16) 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 (16) 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 the following.

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, van den Akker, and Werker (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*

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

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)),[2] then we may obtain a lower bound as follows:

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

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 is, nonetheless, an interesting case under which the asymptotic local power can be proven to be nontrivial. This is the conventional setup where $\operatorname{rank}(\Pi_0)$ 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

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

[2]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.

An application of Anderson's lemma (see, e.g., Lemma 3.11.4 in van der Vaart and Wellner (1996)) then yields

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

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 $\operatorname{vec}(\mathcal{M})$ is nonsingular, then by Lemma B.4 in Chen and Santos (2018) (a strengthening of Anderson's lemma), the asymptotic local lower is in fact nontrivial, that is,

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

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

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

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

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

Hence, we have $P_{0,2}^{\intercal}\Delta Q_{0,2} = 0$, and consequently,

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

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

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

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

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

The hypotheses in (2) may be restrictive since it is generally unknown if $\operatorname{rank}(\Pi_0) \geq k-1$. 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, Chernozhukov, Lee, and Newey (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

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

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

$$\hat{\Pi}_n^* \equiv \frac{1}{n}\sum_{i=1}^n V_i^* Z_i^{*\intercal} = \frac{1}{n}\sum_{i=1}^n W_{ni} V_i Z_i^\intercal, \tag{20}$$

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, microlevel data often cluster on geographical regions such as cities or states. To illustrate, suppose that there are G clusters where G is large, 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

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

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, Gelbach, and Miller (2008), we may construct

$$\hat{\mathcal{M}}_n^* \equiv \frac{1}{n}\sum_{g=1}^G W_g\{V_g^\intercal Z_g - \hat{\Pi}_n\}, \tag{21}$$

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 the cluster wild bootstrap); see also Djogbenou, MacKinnon, and Nielsen (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 (20) or (21)), and set $\hat{c}_{1-\alpha+\beta}$ to be the $\lfloor B(1-\alpha+\beta)\rfloor$ largest number in

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

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 $n\sigma^2_{\min}(\hat{\Pi}_n) > \hat{c}_{1-\alpha+\beta}$ with $\sigma_{\min}(\hat{\Pi}_n)$ the smallest singular value of $\hat{\Pi}_n$.

For our one-step test based on (9) and (17), one may directly proceed with Step 2, but with $\hat{r}_n$ constructed from (9) and reject if $n\sigma^2_{\min}(\hat{\Pi}_n) > \hat{c}_{1-\alpha}$.

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 the Online 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 the Online Supplemental Appendix C.

We commence by considering the following linear model:

$$Z_t = \Pi_0^\intercal V_t + V_{1,t}u_t, \tag{22}$$

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

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

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 $\delta \in \{0, 0.1, 0.3, 0.5\}$,

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

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 5000 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, labeled 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 in (16) and the analytic estimator in (9) and (17), and $\beta \in \{\alpha/5, \alpha/10, \alpha/15, \alpha/20, \alpha/25, \alpha/30\}$ for the two-step test. As in Section 2, 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 undersized 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 oversized 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 (6) and (7) 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 (17) with $\hat{r}_n$ given by (9) 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 5000 simulations, 500 bootstrap repetitions for each simulation, and $\alpha = 5\%$. As shown by Figures 4 and 5, our rank estimators, labeled 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 4 than in Figure 5 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, that is, 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 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

TABLE 2. Rejection rates of rank tests for the model (22) at $\alpha = 5\%$.

	CF-T			CF-A			CF-N			
Sample Size	$\alpha/10$	$\alpha/15$	$\alpha/20$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	KP-M
	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

Note: 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 (9) and (16), respectively.

TABLE 3. Additional results on rejection rates of rank tests for the model (22) with $r = 2$, at $\alpha = 5\%$.

	CF-T			CF-A					CF-N					
Sample Size	$\alpha/5$	$\alpha/25$	$\alpha/30$	$1.5n^{-1/5}$	$1.5n^{-1/4}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	$1.5n^{-1/5}$	$1.5n^{-1/4}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	KP-M
							$\delta = 0$							
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

Note: 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 (9) and (16), respectively.

TABLE 4. Additional results on rejection rates of rank tests for the model (22) with $r = 3$, at $\alpha = 5\%$.

	CF-T			CF-A					CF-N					
Sample Size	$\alpha/5$	$\alpha/25$	$\alpha/30$	$1.5n^{-1/5}$	$1.5n^{-1/4}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	$1.5n^{-1/5}$	$1.5n^{-1/4}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$	KP-M
							$\delta = 0$							
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

Note: 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 (9) and (16), respectively.

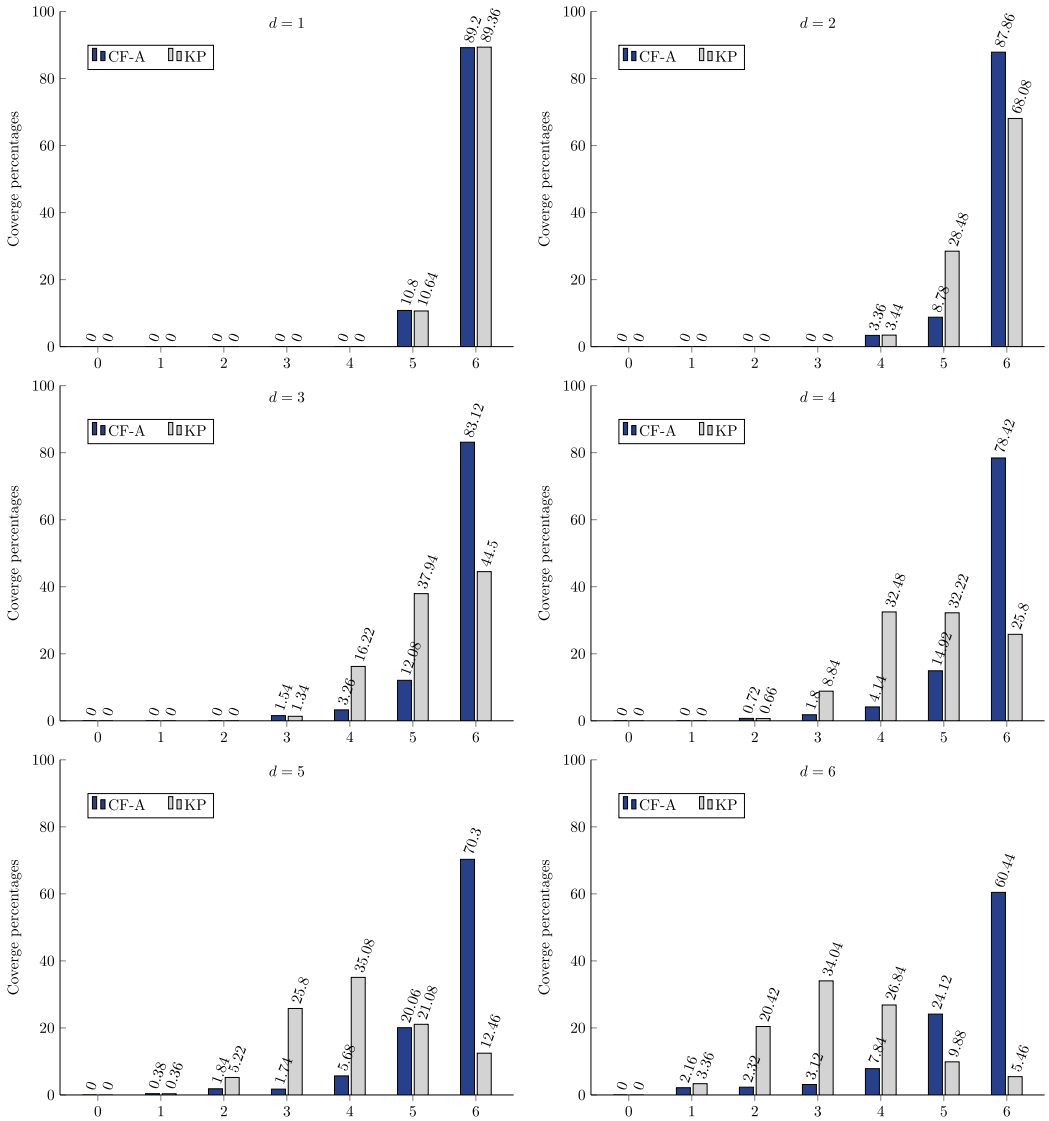

FIGURE 4. The rank estimation: $\text{rank}(\Pi_0) = 6$, $\alpha = 5\%$, and $\delta = 0.1$.

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, Galichon, and Goussé (2018)). As shall be seen shortly, this question can be answered by appealing to our framework developed

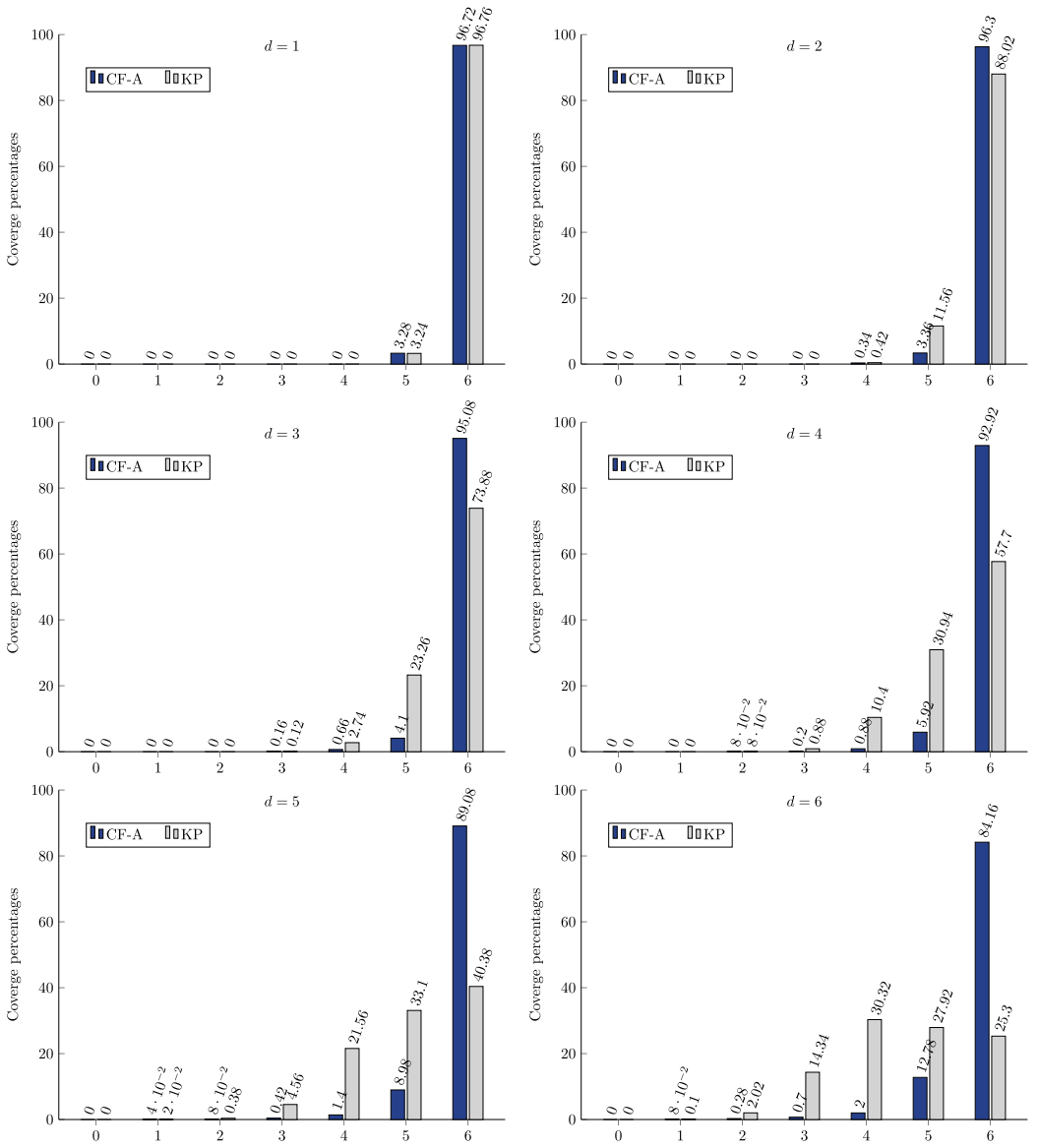

FIGURE 5. The rank estimation: $\text{rank}(\Pi_0) = 6$, $\alpha = 5\%$, and $\delta = 0.12$.

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, that is, 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, McCann, and Nesheim (2010)). Consequently, the matching π_0 in equilibrium can be characterized by the centralized problem:

$$\max_{\pi \in \boldsymbol{\Pi}(P_0, Q_0)} E_\pi\big[s(X, Y)\big], \tag{23}$$

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

Without further appropriate modeling, the optimal transport problem (23), implies pure matching under regularity conditions (Becker (1973), Chiappori, McCann, and Nesheim (2010)), that is, 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, Salanié, and Weiss (2017), we assume that

$$s(x, y) = \Phi(x, y) + \boldsymbol{\epsilon}_m(y) + \boldsymbol{\epsilon}_w(x),$$

where $\Phi(x, y)$ is the systematic part of the surplus, and $\boldsymbol{\epsilon}_m(y)$ and $\boldsymbol{\epsilon}_w(x)$ are unobserved random shocks. Note that $\boldsymbol{\epsilon}_m(y)$ and $\boldsymbol{\epsilon}_w(x)$ 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

$$\max_{\pi \in \boldsymbol{\Pi}(P_0, Q_0)} E_\pi\big[\Phi(X, Y)\big] - E_\pi\big[\log f_\pi(X, Y)\big], \tag{24}$$

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 $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$\Phi(x, y) \equiv \Phi_{A_0}(x, y) = x^\intercal A_0 y, \tag{25}$$

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:

$$E_{\pi(A_0,P_0,Q_0)}\big[XY^\intercal\big] = E\big[XY^\intercal\big], \tag{26}$$

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 (26) 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 differentiable. This has two immediate implications. First, the estimator $\hat{A}_n$ defined by the sample analog of (26), that is,

$$E_{\pi(\hat{A}_n,\hat{p}_n,\hat{q}_n)}\big[XY^\intercal\big] = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i^\intercal, \tag{27}$$

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 (27), that is,

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

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 $\operatorname{rank}(A_0)$ 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), that is, 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:

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

(i) BMI[3] is converted into a trinary variable according to the international BMI classification, that is, BMI is set to be 1 if BMI < 18.50, 2 if $18.50 \leq$ BMI < 24.99, and 3 if BMI ≥ 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 (17) or (16), for which we choose the tuning parameter $\kappa_n \in \{n^{-1/5}, n^{-1/4}, n^{-1/3}\}$. The corresponding results are labeled as CF-A and CF-N, respectively. We also implement the two-step procedure with $\beta \in \{\alpha/10, \alpha/15, \alpha/20\}$, labeled as CF-T. The significance level is $\alpha = 5\%$. As shown 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

[3]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 6. Empirical results.

Model	Maximum Rank	CF-T			CF-A			CF-N			KP-M‡
		$\alpha/10$	$\alpha/15$	$\alpha/20$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	
		The p-values for full rank tests†									
(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

Note: † 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 (9) and (16), 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's (2006) tests implemented by the multiple testing method.

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 overrejections 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. [1790]

Anderson, T. W. (1951), "Estimating linear restrictions on regression coefficients for multivariate normal distributions." *The Annals of Mathematical Statistics*, 22, 327–351. [1790]

Andrews, D. W. (1987), "Asymptotic results for generalized Wald tests." *Econometric Theory*, 3, 348–358. [1799]

Andrews, D. W. K. and P. J. Barwick (2012), "Inference for parameters defined by moment inequalities: A recommended moment selection procedure." *Econometrica*, 80, 2805–2826. [1804]

Andrews, D. W. K. and P. Guggenberger (2010), "Asymptotic size and a problem with subsampling and with the m out of n bootstrap." *Econometric Theory*, 26, 426–468. [1804]

Andrews, D. W. K. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78, 119–157. [1789]

Becker, G. S. (1973), "A theory of marriage: Part I." *Journal of Political economy*, 81, 813–846. [1815]

Berger, R. L. and D. D. Boos (1994), "p values maximized over a confidence set for the nuisance parameter." *Journal of the American Statistical Association*, 89, 1012–1016. [1804]

Bhatia, R. (1997), *Matrix Analysis*. Springer-Verlag, New York, NY. [1796, 1799]

Camba-Mendez, G. and G. Kapetanios (2009), "Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling." *Econometric Reviews*, 28, 581–611. [1790]

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008), "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics*, 90, 414–427. [1807]

Canova, F. and L. Sala (2009), "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics*, 56, 431–449. [1807]

Chen, Q. and Z. Fang (2019), "Inference on functionals under first order degeneracy." *Journal of Econometrics*, 210, 459–481. [1789, 1795, 1797, 1798, 1799, 1800, 1801]

Chen, Q. and Z. Fang (2019), "Supplement to 'Improved inference on the rank of a matrix'." *Quantitative Economics Supplemental Material*, 10, https://doi.org/10.3982/QE1139. [1788]

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014), "Local identification of nonparametric and semiparametric models." *Econometrica*, 82, 785–809. [1807]

Chen, X. and A. Santos (2018), "Overidentification in regular models." *Econometrica*, 86, 1771–1817. [1806]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75, 1243–1284. [1789]

Chesher, A. (2003), "Identification in nonseparable models." *Econometrica*, 71, 1405–1441. [1807]

Chiappori, P.-A. (2017), *Matching With Transfers: The Economics of Love and Marriage*. Princeton University Press, New Jersey, NJ. [1815]

Chiappori, P.-A., R. J. McCann, and L. P. Nesheim (2010), "Hedonic price equilibria, stable matching, and optimal transport: Equivalence, topology, and uniqueness." *Economic Theory*, 42, 317–354. [1815]

Chiappori, P.-A. and B. Salanié (2016), "The econometrics of matching models." *Journal of Economic Literature*, 54, 832–861. [1815]

Chiappori, P.-A., B. Salanié, and Y. Weiss (2017), "Partner choice, investment in children, and the marital college premium." *American Economic Review*, 107, 2109–2167. [1815]

Choo, E. and A. Siow (2006), "Who marries whom and why." *Journal of Political Economy*, 114, 175–201. [1815]

Ciscato, E., A. Galichon, and M. Goussé (2018), "Like attract like? A structural comparison of homogamy across same-sex and different-sex households." Available at SSRN, https://ssrn.com/abstract=2530724. [1813]

Cragg, J. G. and S. G. Donald (1993), "Testing identifiability and specification in instrumental variable models." *Econometric Theory*, 9, 222–240. [1788, 1790, 1791]

Cragg, J. G. and S. G. Donald (1996), "On the asymptotic properties of ldu-based tests of the rank of a matrix." *Journal of the American Statistical Association*, 91, 1301–1309. [1790]

Cragg, J. G. and S. G. Donald (1997), "Inferring the rank of a matrix." *Journal of Econometrics*, 76, 223–250. [1790, 1798, 1805]

Dagsvik, J. K. (2000), "Aggregation in matching markets." *International Economic Review*, 41, 27–58. [1815]

Djogbenou, A. A., J. G. MacKinnon, and M. O. Nielsen (2018), "Asymptotic theory and wild bootstrap inference with clustered errors." Working Papers 1406, Queen's University, Department of Economics, https://ideas.repec.org/p/qed/wpaper/1406.html. [1807]

Dümbgen, L. (1993), "On nondifferentiable functions and the bootstrap." *Probability Theory and Related Fields*, 95, 125–140. [1797, 1799, 1800, 1801]

Dupuy, A. and A. Galichon (2014), "Personality traits and the marriage market." *Journal of Political Economy*, 122, 1271–1319. [1788, 1790, 1813, 1815, 1816, 1817]

Dupuy, A. and A. Galichon (2015), "Canonical correlation and assortative matching: A remark." *Annals of Economics and Statistics*, 375–383. [1816]

Efron, B. (1979), "Bootstrap methods: Another look at the jackknife." *The Annals of Statistics*, 7, 1–26. [1799]

Engle, R. F. and C. W. J. Granger (1987), "Co-integration and error correction: Representation, estimation, and testing." *Econometrica*, 55, 251–276. [1787]

Fang, Z. and A. Santos (2018), "Inference on directionally differentiable functions." *The Review of Economic Studies*, 86, 377–412. [1789, 1795, 1797, 1799, 1800]

Fisher, F. M. (1961), "Identifiability criteria in nonlinear systems." *Econometrica*, 29, 574–590. [1807]

Fisher, F. M. (1966), *The Identification Problem in Econometrics*. McGraw-Hill, New York, NY. [1787]

Gale, D. and L. S. Shapley (1962), "College admissions and the stability of marriage." *The American Mathematical Monthly*, 69, 9–15. [1815]

Galichon, A. and B. Salanié (2010), "Matching with trade-offs: Revealed preferences over competing characteristics." Available at SSRN, https://ssrn.com/abstract=1640380. [1815]

Galichon, A. and B. Salanié (2015), "Cupid's invisible hand: Social surplus and identification in matching models." Available at SSRN, https://ssrn.com/abstract=1804623. [1815, 1816]

Gill, L. and A. Lewbel (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. [1790]

Hallin, M., R. van den Akker, and B. J. Werker (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. [1805]

Hamilton, J. D. (1994), *Time Series Analysis*. Princeton University Press, New Jersey, NJ. [1799]

Hong, H. and J. Li (2018), "The numerical Delta method and bootstrap." *Journal of Econometrics*, 206, 379–394. [1795, 1800, 1801]

Horowitz, J. L. (2001), "The bootstrap." In *Handbook of Econometrics V* (J. J. Heckman and E. Leamer, eds.), 3159–3228, Elsevier, Amsterdam, The Netherlands. [1816]

Hu, Y. (2008), "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution." *Journal of Econometrics*, 144, 27–61. [1807]

Johansen, S. (1988), "Statistical analysis of cointegration vectors." *Journal of Economic Dynamics and Control*, 12, 231–254. [1790]

Johansen, S. (1991), "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models." *Econometrica*, 59, 1551–1580. [1787, 1790]

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, New York, NY. [1788, 1790]

Kasahara, H. and K. Shimotsu (2009), "Nonparametric identification of finite mixture models of dynamic discrete choices." *Econometrica*, 77, 135–175. [1788]

Kleibergen, F. and R. Paap (2006), "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133, 97–126. [1788, 1790, 1791, 1803, 1808, 1809, 1816, 1818]

Komunjer, I. and S. Ng (2011), "Dynamic identification of dynamic stochastic general equilibrium models." *Econometrica*, 79, 1995–2032. [1807]

Koopmans, T. C. and W. C. Hood (1953), "The estimation of simultaneous linear economic relationships." In *Studies in Econometric Method, Cowles Commission Monograph*, Vol. 14 (W. C. Hood and T. C. Koopmans, eds.), 112–199, John Wiley & Sons, Inc and Chapman & Hall, Limited, New York, NY and London. [1807]

Lehmann, E. L. and J. P. Romano (2005), *Testing Statistical Hypotheses*, third edition. Springer, New York, NY. [1802]

Linton, O., K. E. Song, and Y.-J. Whang (2010), "An improved bootstrap test of stochastic dominance." *Journal of Econometrics*, 154, 186–202. [1789]

Loh, W.-Y. (1985), "A new method for testing separate families of hypotheses." *Journal of the American Statistical Association*, 80, 362–368. [1804]

Matzkin, R. L. (2008), "Identification in nonparametric simultaneous equations models." *Econometrica*, 76, 945–978. [1807]

McLachlan, G. and D. Peel (2004), *Finite Mixture Models*. John Wiley & Sons, New York, NY. [1787, 1788]

Neyman, J. (1937), ""Smooth test" for goodness of fit." *Scandinavian Actuarial Journal*, 1937, 149–199. [1805]

Nyhus, E. K. (1996), "The VSB-CentER savings project: Data collection methods, questionnaires and sampling procedures." In *VSB-CentER Savings Project Progress Peport, Tilburg University*, Vol. 42. [1816]

Politis, D. N. and J. P. Romano (1992), "A circular block-resampling procedure for stationary data." In *Exploring the Limits of Bootstrap*, 263–270, John Wiley, New York, NY. [1809]

Politis, D. N. and J. P. Romano (1994), "Large sample confidence regions based on subsamples under minimal assumptions." *The Annals of Statistics*, 22, 2031–2050. [1800]

Portier, F. and B. Delyon (2014), "Bootstrap testing of the rank of a matrix via least-squared constrained estimation." *Journal of the American Statistical Association*, 109, 160–172. [1790]

Praestgaard, J. and J. Wellner (1993), "Exchangeably weighted bootstraps of the general empirical process." *The Annals of Probability*, 21, 2053–2086. [1801]

Robin, J.-M. and R. J. Smith (2000), "Tests of rank." *Econometric Theory*, 16, 151–175. [1788, 1790, 1791, 1798, 1799, 1803]

Roehrig, C. S. (1988), "Conditions for identification in nonparametric and parametric models." *Econometrica*, 56, 433–447. [1807]

Romano, J. P., A. Shaikh, and M. Wolf (2014), "A practical two-step method for testing moment inequalities." *Econometrica*, 82, 1979–2002. [1789, 1804]

Rothenberg, T. J. (1971), "Identification in parametric models." *Econometrica*, 577–591. [1807]

Rothenberg, T. J. (1984), "Approximating the distributions of econometric estimators and test statistics." In *Handbook of Econometrics*, Vol. 2, 881–935. [1805]

Roussas, G. G. (1972), *Contiguity of Probability Measures: Some Applications in Statistics*, Vol. 63. Cambridge University Press, New York, NY. [1805]

Rüschendorf, L. (1995), "Convergence of the iterative proportional fitting procedure." *The Annals of Statistics*, 23, 1160–1174. [1817]

Shao, J. (1994), "Bootstrap sample size in nonregular cases." In *Proceedings of the American Mathematical Society*, Vol. 122 1251–1262. [1800]

Shapiro, A. (1990), "On concepts of directional differentiability." *Journal of Optimization Theory and Applications*, 66, 477–487. [1797]

Shapiro, A. (1991), "Asymptotic analysis of stochastic programs." *Annals of Operations Research*, 30, 169–186. [1797]

Shapiro, A. (2000), "Statistical inference of stochastic optimization problems." In *Probabilistic Constrained Optimization* (S. P. Uryasev, ed.), Nonconvex Optimization and Its Applications, Vol. 49, 282–307, Springer, New York, NY. [1789, 1797, 1798]

Shapley, L. S. and M. Shubik (1971), "The assignment game I: The core." *International Journal of Game Theory*, 1, 111–130. [1815]

Silvapulle, M. J. (1996), "A test in the presence of nuisance parameters." *Journal of the American Statistical Association*, 91, 1690–1693. [1804]

van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge University Press, Cambridge, U.K. [1805, 1807]

van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*. Springer Verlag, New York, NY. [1800, 1806]

West, K. D. (1997), "Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator." *Journal of Econometrics*, 76, 171–191. [1809]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 22 May, 2018; final version accepted 9 May, 2019; available online 24 May, 2019.