Improved Inference on the Rank of a Matrix with Applications to IV and Cointegration Models

Qihui Chen
Department of Economics
UC San Diego
qic005@ucsd.edu

Zheng Fang*
Department of Economics
Kansas State University
zfang@ksu.edu

December 25, 2016
(Click here for the latest version)

Abstract

This paper develops new methods for examining a "no greater than" inequality of the rank of a matrix and for rank determination in a general setup, which improve upon existing methods. Existing rank tests assume *a priori* that the rank is no less than the hypothesized value, which is often unrealistic. These tests when directly applied may fail to control the asymptotic null rejection rate, and the multiple testing method based on them can be conservative with the asymptotic null rejection rate strictly below the nominal level whenever the rank is less than the hypothesized value. We prove that our proposed tests have the asymptotic null rejection rate that is exactly equal to the nominal level under minimal assumptions regardless of whether the rank is less than or equal to the hypothesized value. As our simulation results show, these characteristics lead to an improved power property in general. In application to a context with stationary and nonstationary data, respectively, our tests yield improved tests for identification in linear IV models and for the existence of stochastic trend and/or cointegration with or without VAR specification. In addition, our simulation results show that the improved power property of our tests leads to an improved accuracy of the sequential testing procedure for rank determination.

Keywords: Rank inequality, Rank determination, Size control, Conservativeness, Identification test, Cointegration test.

*Qihui Chen is deeply grateful to his advisors, Andres Santos and Yixiao Sun, for their constant support and guidance. The authors would like to thank Brendan Beare, Ivana Komunjer and Qi Li for helpful discussions and comments, and Frank Kleibergen for making the data available to us.

1 Introduction

The rank of a matrix plays a fundamental role in numerous economic and statistical settings, including identification of structural parameters (Fisher, 1966), existence of common features (Engle and Kozicki, 1993) with the canonical example being that of cointegration (Engle and Granger, 1987), the rank of a (consumer) demand system (Gorman, 1981; Lewbel, 1991), specification of factor models (Ross, 1976), dimension reduction in regression analysis (Li, 1991; Bura and Yang, 2011), and model specification in time series (Aoki, 1990; Gill and Lewbel, 1992). These problems reduce to examining the following hypotheses: for an unknown matrix Π_0 of size $m \times k$ with $m \geq k$,

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

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 simply a testing problem of whether Π_0 has full rank.

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

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

In effect, this is a different testing problem and assumes *a priori* that $\mathrm{rank}(\Pi_0)$ is no less than r. Unfortunately, in the aforementioned problems, it is unrealistic to make such an assumption. As shown in Section 2.2, when in fact $\mathrm{rank}(\Pi_0) < r$, directly applying existing rank tests to (1) may fail to control the asymptotic null rejection rate, since the asymptotic distributions of test statistics can be very different from those when $\mathrm{rank}(\Pi_0) = r$. As we shall prove (see Lemma A.4), when $\mathrm{rank}(\Pi_0) < r$, the problem (1) becomes irregular in the sense that a functional characterizing the problem admits a degenerate first order derivative and is second order nondifferentiable. A general inferential framework for such functionals was not available until very recently (Fang and Santos, 2015; Chen and Fang, 2015). To the best of our knowledge, no direct tests for (1) exist in the literature.

Our method builds on the insight that (1) can be equivalently reformulated as

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

where $\phi(\Pi_0) \equiv \sum_{j=r+1}^{k} \sigma_j^2(\Pi_0)$ is the sum of the $k-r$ smallest squared singular values $\sigma_j^2(\Pi_0)$ of Π_0 (i.e., the sum of the $k-r$ smallest eigenvalues of $\Pi_0^{\intercal}\Pi_0$). For a given estimator $\hat{\Pi}_n$ of Π_0, we then employ the plug-in estimator $\tau_n^2 \phi(\hat{\Pi}_n)$ as our test statistic, where τ_n is the rate at which $\hat{\Pi}_n$ admits an asymptotic distribution. Towards invoking the Delta method, we prove, however, that the first order derivative of the map $\Pi \mapsto$

$\phi(\Pi)$ is null at $\Pi = \Pi_0$ under H_0, necessitating a second order analysis. Since the asymptotic distributions (under the composite null) implied by the second order Delta method (Shapiro, 2000) are highly nonstandard, we appeal to the bootstrap procedure recently developed by Fang and Santos (2015) and Chen and Fang (2015) in order to obtain valid critical values and conduct inference. We also extend the results to accommodate the case when the convergence rates of $\hat{\Pi}_n$ are not homogenous across its columns as in VAR models with stochastic trend and cointegration (see Appendix B).

There are several attractive features of our tests. First, since we rely on the Delta method, the theory is conceptually simple and requires minimal assumptions. Essentially, all we need are a matrix estimator $\hat{\Pi}_n$ that converges weakly and a consistent bootstrap analog $\hat{\Pi}_n^*$. As a matter of fact, our tests apply to various data generating processes. Second, implementation of the procedure is computationally easy, only involving calculation of singular value decompositions. Finally, since construction of the critical values is based on bootstrapping the asymptotic distributions pointwise in Π_0, the resulting tests have the asymptotic null rejection rate that is exactly equal to the nominal level regardless of whether $\mathrm{rank}(\Pi_0) = r$ or $\mathrm{rank}(\Pi_0) < r$. As our simulation results show, these characteristics lead to good power properties of our tests in general. In application to a context with stationary and nonstationary data, respectively, our tests yield new and powerful tests for identification in linear IV models (Fisher, 1966) and for the existence of stochastic trend and/or cointegration with or without VAR specification (Engle and Granger, 1987).

As an alternative to the direct application, one may instead adapt existing rank tests into multiple testing procedures, since H_0 holds if and only if $\mathrm{H}_0^{(q)}$ holds for some $0 \leq q \leq r$. Specifically, the multiple testing method rejects H_0 if and only if all $\mathrm{H}_0^{(q)}$ are rejected and otherwise fails to reject. However, as demonstrated in Sections 2.2 and 4.1, the method can be severely conservative when $\mathrm{rank}(\Pi_0) > r$ and Π_0 is close to a matrix with rank strictly less than r, with the asymptotic null rejection rate strictly below the nominal level when $\mathrm{rank}(\Pi_0) < r$. This is in sharp contrast to our tests, which by design achieve asymptotic null rejection rates exactly equal to the nominal level and hence improve the power properties. In an application to testing for identification in stochastic discount factor models, compared to the multiple testing method based the Kleibergen and Paap (2006) test, our tests suggest much weaker evidence of non-identification of the risk premia parameters.

In some settings such as the rank of a demand system, specification of factor models and model specification in time series, the main concern boils down to determining the true rank of a matrix. To determine $\mathrm{rank}(\Pi_0)$, one may implement the sequential testing procedure, following Johansen (1995), based on rank tests for (1) or (2). Interestingly, efficient rank determination does not require the ability of detecting whether $\mathrm{rank}(\Pi_0)$ is strictly less than a hypothesized value. This explains the prevalence of existing rank tests in rank determination. Nevertheless, the power of detecting whether $\text{rank}(\Pi_0)$ is strictly greater than hypothesized values plays an important role in the procedure. Our simulation results show that the improved power property of our tests leads to an improved accuracy of the sequential testing procedure for rank determination.

As mentioned previously, the literature has been mostly concerned with the hypotheses (2). In the context of multivariate regression, Anderson (1951) proposed a likelihood ratio test based on canonical correlations. This test is restrictive in the sense that it crucially depends on a Kronecker product structure of the covariance matrix of a matrix estimator. Building on the LDU decomposition approach in Gill and Lewbel (1992), Cragg and Donald (1996) proposed a test with the test statistic being a quadratic form of the vectorization of a submatrix in the LDU decomposition that is sensitive to variable ordering. In Cragg and Donald (1997), the authors provided a test based on a constrained minimum χ^2 distance criterion, which is computationally intensive because it involves minimization over the set of all matrices with rank r. Moreover, both tests rely on the condition that the asymptotic covariance matrix of the vectorization of the matrix estimator is nonsingular, which we do not require in our analysis. Motivated by the need to relax this nonsingularity condition, Robin and Smith (2000) developed a test based on functionals of the characteristics of a suitably transformed matrix. However, their test depends on a rank condition that is "empirically nonverifiable". All these rank tests may fail to control the asymptotic null rejection rate when directly applied to the hypotheses (1).

Moreover, Kleibergen and Paap (2006) proposed a test based on singular value decomposition of a transformed matrix with the test statistic having the $\chi^2((m-r)(k-r))$ asymptotic distribution under $\mathrm{H}_0^{(r)}$. Despite overcoming many of the deficiencies of previous tests, this test still requires some covariance matrix nonsingular because it is based on a Wald statistic, which we do not require in our analysis. More importantly, this rank test also has the aforementioned drawback when directly applied to the hypotheses (1). There are, nonetheless, a few exceptions that study (1), notably Cragg and Donald (1993) who considered a special case of Cragg and Donald (1997). However, the asymptotic distribution of the test statistic when $\text{rank}(\Pi_0) < r$ is not available, though Cragg and Donald (1993) established that the asymptotic null distribution when $\text{rank}(\Pi_0) = r$ is least favorable under somewhat restrictive conditions. Thus, when $\text{rank}(\Pi_0) > r$ and Π_0 is close to a matrix with rank strictly less than r, their test can be conservative. We refer the reader to Camba-Mendez and Kapetanios (2009), Portier and Delyon (2014) and Al-Sadoon (2015) for further discussions of the literature.

We now introduce some notation. We denote by $\mathbf{M}^{m\times k}$ the space of $m \times k$ real matrices for $m, k \in \mathbf{N}$. For a matrix $A \in \mathbf{M}^{m\times k}$, we write the transpose of A by $A^\intercal$, the trace of A by $\text{tr}(A)$ if $m = k$, the column vectorization of A by $\text{vec}(A)$, and the Frobenius norm of A by $\|A\|$, i.e., $\|A\| \equiv \sqrt{\text{tr}(A^\intercal A)}$. We let I_k denote the $k \times k$ identity matrix for $k \in \mathbf{N}$.

The remainder of the paper is organized as follows. Section 2 presents related examples to illustrate the importance of the problem, and demonstrates the drawback of existing rank tests and the conservativeness of the multiple testing method. Section 3 develops the test statistic, establishes its asymptotic distribution, and proposes a bootstrap procedure for inference. Section 4 presents Monte Carlo studies, applies our method to study identification in stochastic discount factor models, and demonstrates the accuracy improvement of the sequential testing procedure for rank determination based on our tests. Section 5 briefly concludes. All the proofs are collected in the appendices.

2 Examples and Motivation

In this section, we first present related examples in which the testing problem (1) is of importance. In order to motivate the development of our tests, we then demonstrate that existing rank tests when directly applied to (1) can fail to control the asymptotic null rejection rate, and that the multiple testing method can be conservative.

2.1 Examples

The first example is what motivated this paper in the first place.

Example 2.1 (Identification). Let $Y \in \mathbf{R}$ and $Z \in \mathbf{R}^k$ be random variables satisfying

$$Y = Z^{\intercal}\beta_0 + u \ . \tag{4}$$

Let $W \in \mathbf{R}^m$ be instrument variables such that $E[Wu] = 0$ with $m \geq k$. Then identification of the coefficient β_0 reduces to whether $E[WZ^{\intercal}]$ is of full rank. Thus, testing for identification of β_0 reduces to examining the hypotheses (1) with

$$\Pi_0 = E[WZ^{\intercal}] \text{ and } r = k-1 \ . \tag{5}$$

We cannot restrict ourselves to examine the hypotheses (2), since it is unrealistic to assume $\text{rank}(\Pi_0) \geq k-1$ unless $k = 1$. More generally, (local) identification in parametric, semiparametric and nonparametric models can often be expressed in terms of some matrices being of full rank (Fisher, 1961; Rothenberg, 1971; Roehrig, 1988; Chesher, 2003; Matzkin, 2008; Chen et al., 2014). For identification in DSGE models, see, for example, Canova and Sala (2009) and Komunjer and Ng (2011). In addition, when $W = Z$, then Π_0 is a positive semidefinite matrix and the concern becomes the existence of perfect multicolinearity among Z. ■

The next example concerns the existence of stochastic trend and/or cointegration in a vector autoregression (VAR) system (Engle and Granger, 1987; Johansen, 1991).

Example 2.2 (VAR Trend/Cointegration)**.** Let $\{Y_t\}$ be a $k \times 1$ time series such that each component of Y_t is integrated of order 0 or 1, that is, each component is a stationary or unit root process. Assume the entire vector is a VAR(1) process

$$Y_t = \Phi_0 Y_{t-1} + u_t \ , \tag{6}$$

where u_t are white noise with nonsingular covariance matrix Σ. The error-correction representation of (6) is given by (Hamilton, 1994, p.580):

$$\Delta Y_t = (\Phi_0 - I_k) Y_{t-1} + u_t \ . \tag{7}$$

Then the existence of stochastic trend for Y_t means that $\Phi_0 - I_k$ is not of full rank. Thus, testing for the existence of stochastic trend reduces to examining the hypotheses (1) with

$$\Pi_0 = \Phi_0 - I_k \text{ and } r = k - 1 \ . \tag{8}$$

It is unrealistic to assume that there is at most one linearly independent stochastic trend (i.e., $\text{rank}(\Pi_0) \geq k - 1$) unless $k = 1$, so we cannot instead focus on examining the hypotheses (2). In addition, the existence of cointegrating relations for Y_t means that $\Phi_0 - I_k$ is nonzero.[1] Thus, testing for the existence of cointegration reduces to examining the hypotheses (1) with $r = 0$. We confine our attention to the class of VAR(1) models with white noise errors for simplicity, but our framework applies more broadly to VAR(p) processes with dependent and heteroskedastic errors. ■

Our results allow us to study stochastic trend and cointegration nonparametrically. The following example concerns the existence of stochastic trend and/or cointegration without a VAR specification (Engle and Granger, 1987; Bierens, 1997; Shintani, 2001).

Example 2.3 (Nonparametric Trend/Cointegration)**.** Let $\{Y_t\}$ be a $k \times 1$ time series such that each component of Y_t is integrated of order 0 or 1, that is, each component is a stationary or unit root process. Let the first difference of Y_t follow a linear process

$$\Delta Y_t = C(L) u_t \equiv \sum_{j=0}^{\infty} C_j u_{t-j} \ , \tag{9}$$

where u_t are white noise with nonsingular covariance matrix Σ, and $C_0 = I_k$. Since the long run covariance matrix of ΔY_t is equal to $C(1)\Sigma C(1)^\intercal$, then existence of cointegrating relations for Y_t means that the long run covariance matrix of ΔY_t is not of full rank.

[1] Recall that Y_t is said to be cointegrated if there exists nonzero $\lambda \in \mathbf{R}^k$ such that $\lambda^\intercal Y_t$ is stationary.

Thus, testing for the existence of cointegration reduces to examining the hypotheses (1) with

$$\Pi_0 = \sum_{t=-\infty}^{\infty} E[\Delta Y_t \Delta Y_0] \text{ and } r = k - 1 \ . \tag{10}$$

We cannot restrict ourselves to examine the hypotheses (2), since it is unrealistic to assume there is at most one linearly independent cointegration vectors (i.e., $\text{rank}(\Pi_0) \geq k-1$) unless $k = 1$. In addition, the existence of stochastic trend for Y_t means that $\Phi_0 - I_k$ is nonzero. Thus, testing for the existence of stochastic trend reduces to examining the hypotheses (1) with $r = 0$. ■

Cointegration is just one particular example of the more general notion of common features (Engle and Kozicki, 1993). Our fourth example pertains to the existence of general common features.

Example 2.4 (Common Features)**.** Let $\{Y_t\}$ be a $k \times 1$ time series. According to Engle and Kozicki (1993), a feature that is present in each component of Y_t is said to be common to Y_t if there exists a nonzero linear combination of Y_t that fails to have the feature. Suppose that $\{Y_t\}$ is generated according to

$$Y_t = \Gamma_0^\intercal Z_t + \Xi_0^\intercal W_t + u_t \ , \tag{11}$$

where W_t can be thought of as control variables, and Z_t is an $m \times 1$ vector reflecting the feature under consideration with $m \geq k$. For example, testing for the existence of common serial correlation would set Z_t to be lags of Y_t, and testing for the existence of common conditionally heteroskedastic factors would set Z_t to be relevant factors. We refer to Engle and Kozicki (1993), Engle and Susmel (1993) and Dovonon and Renault (2013) for details of these and other examples. By the definition of common feature and the specification of (11), existence of common features means that Γ_0 is not of full rank. Thus, testing for the existence of common features reduces to examining the hypotheses (1) with

$$\Pi_0 = \Gamma_0 \text{ and } r = k - 1 \ . \tag{12}$$

Since the number of common features is unknown *a priori*, we cannot restrict ourself to examine the hypotheses (2) by assuming $\text{rank}(\Pi_0) \geq k - 1$ unless $k = 1$. ■

The concerns in the remaining examples reduce to determining the true rank of a matrix, which relies on examining a sequence of hypotheses (1) or (2). Our fifth example is directly related to the rank of demand systems, a notion developed by Gorman (1981) for exactly aggregable demand systems and generalized by Lewbel (1991) to all demand systems.

Example 2.5 (Consumer Demand)**.** An Engel curve is the function describing the allocation of an individual's consumption expenditures with the prices of all goods fixed, and the rank of a demand system is the dimension of the space spanned by the Engel curves of the system (Lewbel, 1991). Suppose that there are k goods in the system and the Engel curve is given by

$$Y = \Gamma_0 G(Z) + u \ , \tag{13}$$

where Y is a $k \times 1$ vector of budget shares on the k goods, Z is total expenditure, $G(\cdot)$ is a $r_0 \times 1$ vector of unknown function with $r_0 \leq k$, and u is a vector zero mean random variables independent of Z. Assume Γ_0 is of full rank, then the rank r_0 of the demand system is equal to the rank of Γ_0. Let $Q(\cdot)$ be a $m \times 1$ vector of known functions with $m \geq k$. Then the rank of Γ_0 is equal to the rank of

$$\Pi_0 = E[Q(Z)Y^\intercal] \ , \tag{14}$$

if $E[Q(Z)G(Z)^\intercal]$ is of full rank. Thus, determining the rank r_0 of the demand system reduces to determining the rank of Π_0. The rank of the demand system provides evidence on consistency of consumer behaviors with utility maximization, and has implications for welfare comparisons and aggregation across goods and across consumers (Lewbel, 1991, 2006; Barnett and Serletis, 2008). ■

Factor analysis has been widely used in modeling variations, covariance and dynamics of time series (Anderson, 2003; Lam and Yao, 2012). Our next example shows the importance of matrix rank determination in identifying the number of factors in factor analysis.

Example 2.6 (Factor Analysis)**.** Let $Y \in \mathbf{R}^p$ be generated by the following model

$$Y = \mu_0 + \Lambda_0 F + u \ , \tag{15}$$

where F is a $r_0 \times 1$ vector of unobserved common factors with $E[F] = 0$ and $r_0 \leq p$, and u is an idiosyncratic error term with $E[u] = 0$. Assume $\mathrm{Var}(F)$ is of full rank, then the number r_0 of common factors is equal to the rank of $\mathrm{Var}(F)$. Let us write $Y = [Y_1^\intercal, Y_2^\intercal, Y_3^\intercal]^\intercal$ for $Y_1 \in \mathbf{R}^m$, $Y_2 \in \mathbf{R}^k$ and $Y_2 \in \mathbf{R}^{p-k-m}$ for some $r_0 \leq k \leq m < p$ and $m + k \leq p$. Write $\Lambda_0 = [\Lambda_{0,1}^\intercal, \Lambda_{0,2}^\intercal, \Lambda_{0,3}^\intercal]^\intercal$ with $\Lambda_{0,1}$ and $\Lambda_{0,2}$ having m and k rows. Given the mild condition that u is independent of F and $E[uu']$ is diagonal, the rank of $\mathrm{Var}(F)$ is equal to the rank of

$$\Pi_0 = \mathrm{Cov}(Y_1, Y_2) \ , \tag{16}$$

if $\Lambda_{0,1}$ and $\Lambda_{0,1}$ are of full rank. Thus, determining the number r_0 of these common factors reduces to determining the rank of Π_0. Such a question also arises in the interbattery factor analysis (Gill and Lewbel, 1992), the dynamic analysis of time series (Lam and Yao, 2012), and finance and macroeconomics (Bai and Ng, 2002, 2007). ■

Our final example is taken from Gill and Lewbel (1992), and manifests how matrix rank determination is useful in model selection for ARMA processes and state space models.

Example 2.7 (Model Selection)**.** Let $\{Y_t\}$ be a $p \times 1$ weakly stationary time series, which has the following state space representation:

$$Y_t = \Gamma_0 Z_t + u_t \ , \ Z_t = \Lambda_0 Z_{t-1} + \epsilon_t \ , \tag{17}$$

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

$$\Pi_0 = E(\begin{bmatrix} Y_{t+1} \\ \vdots \\ Y_{t+b} \end{bmatrix} \begin{bmatrix} Y_t^\intercal & \cdots & Y_{t-b+1}^\intercal \end{bmatrix}) \ , \tag{18}$$

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

2.2 Motivation

To proceed, we let $\alpha \in (0, 1)$ be the nominal level and $\phi_n^{(r)}$ be any one of the existing rank tests designed for the hypotheses (2), which are reviewed in the introduction.[2] It has been well established in the literature that $\lim_{n\to\infty} P(\phi_n^{(r)} = 1) = \alpha$ under $\mathrm{H}_0^{(r)}$ and $\lim_{n\to\infty} P(\phi_n^{(r)} = 1) = 1$ under $\mathrm{H}_1^{(r)}$.

When $\mathrm{rank}(\Pi_0) < r$, the asymptotic distributions of test statistics have not been established and can be very different from those when $\mathrm{rank}(\Pi_0) = r$. On the one hand, $\phi_n^{(r)}$ may fail to control the asymptotic rejection rate. In Appendix C, we prove that this is true for the Kleibergen and Paap (2006) version of $\phi_n^{(r)}$. Therefore, $\phi_n^{(r)}$ cannot be directly applied to test for the hypotheses (1). On the other hand, the asymptotic rejection rate of $\phi_n^{(r)}$ can be strictly below the nominal level, i.e, $\lim_{n\to\infty} P(\phi_n^{(r)} = 1) < \alpha$. In Appendix C, we also prove that this is true for the Kleibergen and Paap (2006) version of $\phi_n^{(r)}$. By Theorem 2 of Cragg and Donald (1993), this is also true for the Cragg and Donald (1997) version of $\phi_n^{(r)}$. In view of this, $\phi_n^{(r)}$ may alternatively be conservative

[2]Rejection means $\phi_n^{(r)} = 1$ and acceptance means $\phi_n^{(r)} = 0$.

when directly applied to the hypotheses (1). Thus, the critical value may be adjusted to improve the power of $\phi_n^{(r)}$ for detecting H_1 when Π_0 is close to a matrix with rank strictly less than r.

Given that H_0 being false is equivalent to $\mathrm{H}_0^{(q)}$ being false for all $0 \le q \le r$, one may then consider implementing multiple existing rank tests in order to obtain tests for the hypotheses (1) such that the asymptotic null rejection rate is controlled. The multiple testing method is based on the decision rule $\phi_n = \prod_{q=0}^{r} \phi_n^{(q)}$, which means that H_0 is rejected if and only if $\mathrm{H}_0^{(q)}$ is rejected for all $0 \le q \le r$. In VAR models (see, for instance, Example 2.2), Johansen (1995, Chapter 12) used this method to test for inequality of cointegration rank. In stochastic discount factor models, Kleibergen and Paap (2006) employed this method to test for identification of the risk premia parameters. Indeed, the asymptotic null rejection rate of this method is controlled, since under H_0,

$$\lim_{n\to\infty} P(\phi_n = 1) = \lim_{n\to\infty} P(\phi_n^{(0)} = 1, \ldots, \phi_n^{(r)} = 1) \le \lim_{n\to\infty} P(\phi_n^{(\mathrm{rank}(\Pi_0))} = 1) = \alpha \ , \quad (19)$$

where the first inequality holds since $P(A) \le P(B)$ for $A \subset B$. Moreover, this method is consistent, since under H_1,

$$\lim_{n\to\infty} P(\phi_n = 1) = \lim_{n\to\infty} P(\phi_n^{(0)} = 1, \ldots, \phi_n^{(r)} = 1) \ge 1 - \sum_{q=0}^{r}(1 - \lim_{n\to\infty} P(\phi_n^{(q)} = 1)) = 1 \ ,$$

where the inequality holds by the Boole's inequality.

Unfortunately, the multiple testing method can be conservative. When $\mathrm{rank}(\Pi_0) < r$, the inequality of (19) becomes strict whenever $\lim_{n\to\infty} P(\phi_n^{(r)} = 1) < \alpha$. This is because

$$\lim_{n\to\infty} P(\phi_n = 1) = \lim_{n\to\infty} P(\phi_n^{(0)} = 1, \ldots, \phi_n^{(r)} = 1) \le \lim_{n\to\infty} P(\phi_n^{(r)} = 1) < \alpha \ , \quad (20)$$

where the first inequality holds since $P(A) \le P(B)$ for $A \subset B$. As mentioned above, this is true for the Cragg and Donald (1997) and Kleibergen and Paap (2006) version of $\phi_n^{(r)}$. Thus, the critical value of each $\phi_n^{(q)}$ may be adjusted to improve the power of the multiple testing method for detecting H_1 when Π_0 is close to a matrix with rank strictly less than r. Furthermore, due to the dependence among $\{\phi_n^{(q)}\}_{q=0}^{r}$ the inequality in both (19) and (20) may become strict. In view of this, power loss may occur in a complicated way.

To show the drawback of existing rank tests and the conservativeness of the multiple testing method, we focus on the Kleibergen and Paap (2006) test and present some simulation evidence.[3] We assume that

$$Z_i^\intercal = W_i^\intercal \Pi_0 + u_i^\intercal \ , i = 1, \ldots, n \ , \tag{21}$$

with $W_i \overset{i.i.d.}{\sim} N(0, I_6)$, $u_i \overset{i.i.d.}{\sim} N(0, I_6)$ and $n = 1,000$. Let

$$\Pi_0 = \text{diag}(\mathbf{1}_{6-d}, \mathbf{0}_d) + \delta I_6 \text{ for } \delta \geq 0 \text{ and } d = 1, \ldots, 6 \ , \tag{22}$$

where $\mathbf{1}_{6-d}$ denotes a $(6-d) \times 1$ vector of ones and $\mathbf{0}_d$ denotes a $d \times 1$ vector of zeros. We examine the hypotheses (1) with $r = 5$, that is, we test whether Π_0 has full rank. The design of Π_0 implies H_0 is true if and only if $\delta = 0$. In particular, $\text{rank}(\Pi_0) = 6 - d$ under H_0, so $\text{rank}(\Pi_0) < r$ when $d \neq 1$ and $\text{rank}(\Pi_0) = r$ when $d = 1$. Thus, $d \neq 1$ represents the case when Π_0 is close to a matrix with rank strictly less than r, while $d = 1$ represents the regular case. From the above argument, it shall be expected that when $d \neq 1$, the Kleibergen and Paap (2006) test may over-reject H_0 when $\delta = 0$ or may be inefficient in detecting H_1 when $\delta > 0$. Moreover, the multiple testing method may be inefficient in detecting H_1 when $\delta > 0$. The value of δ represents how strong H_1 deviates away from H_0.

To implement the Kleibergen and Paap (2006) test and the multiple testing method, we estimate Π_0 by $\hat{\Pi}_n = \frac{1}{n} \sum_{i=1}^n W_i Z_i^\intercal$. See Appendix C for a review on the Kleibergen and Paap (2006) test. By the central limit theorem, the asymptotic distribution of $\hat{\Pi}_n$ is zero mean Gaussian with convergence rate $\sqrt{n}$ and all assumptions in Kleibergen and Paap (2006) are satisfied. Let the nominal level be 5%. The rejection rates, which are based on $10,000$ simulation replications, are plotted in Figures 1 and 2. We use KP-D to denote the Kleibergen and Paap (2006) test when directly applied and KP-M to denote the multiple testing method. First, as expected, the rejection rates of KP-M are no greater than the 5% nominal level when $\delta = 0$ and tend to one as δ increases for all cases. When $d = 1$, the null rejection rate is close to the 5% nominal level. When $d \neq 1$, however, the null rejection rates are far below the 5% nominal level. This suggests that KP-M may be conservative when $d \neq 1$. Indeed, the power curve shifts to right and more parts fall below the 5% nominal level as d increases. This hints a method of power improvement by dragging the curves to the left such that all of them are above the 5% nominal level. Similarly, as Figure 2 shows, KP-D has the same issue under the considered model. Note that the difference between the two methods in Figure 2 is negligible, despite the fact that KP-D is more powerful.

[3] Two main reasons for the focus are: the Kleibergen and Paap (2006) test is preferred in terms of assumptions and computation, and has the most citations (over $1,000$) among the existing rank tests according to Google Scholar.

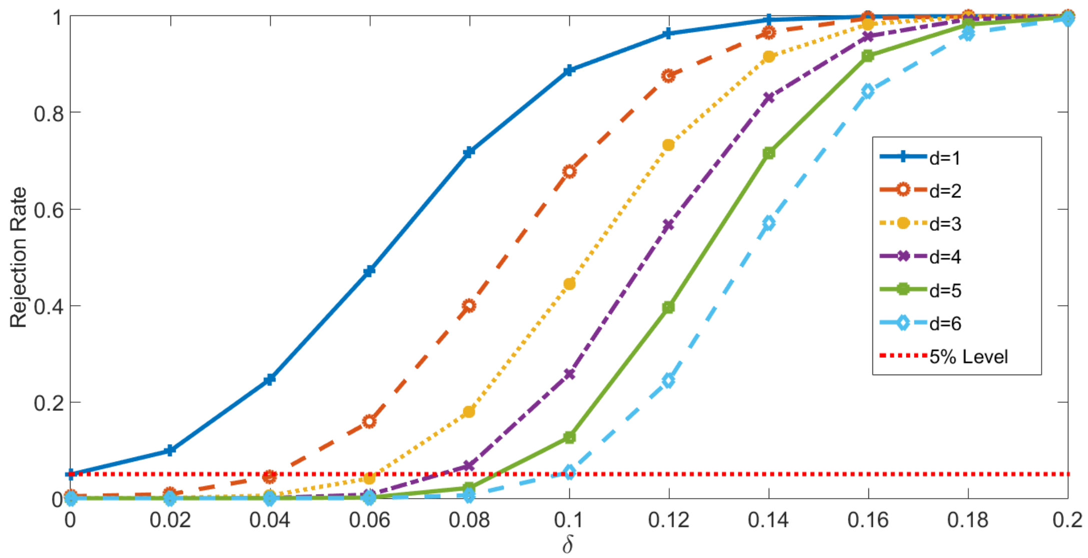

Figure 1: The rejection rate of the multiple testing method based on the Kleibergen and Paap (2006) test with 5% nominal level

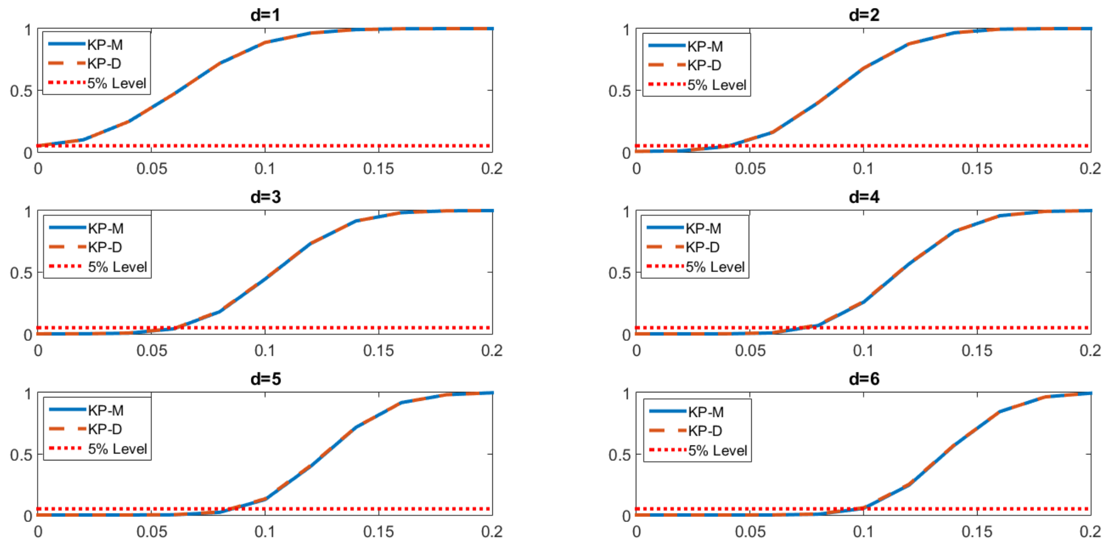

Figure 2: Comparison between the Kleibergen and Paap (2006) test and the multiple testing method based on it with 5% nominal level

3 Asymptotic Analysis

We can express the hypotheses (1) more tractably in terms of singular values. To see this, let $\sigma_1(\Pi_0) \geq \cdots \geq \sigma_k(\Pi_0) \geq 0$ be singular values of Π_0.[4] Then the rank of Π_0 is equal to the number of nonzero singular values of Π_0; see, for example, Problem 3.1.2 in Horn and Johnson (1991). It follows that the hypotheses (1) can be equivalently reformulated as

$$\mathrm{H}_0 : \sum_{j=r+1}^{k} \sigma_j^2(\Pi_0) = 0 \quad \text{v.s.} \quad \mathrm{H}_1 : \sum_{j=r+1}^{k} \sigma_j^2(\Pi_0) > 0 \ . \tag{23}$$

Given the reformulation in (23), it is convenient to study the differential properties of the map $\Pi_0 \mapsto \sum_{j=r+1}^{k} \sigma_j^2(\Pi_0)$. By leveraging the existing Delta method, we in turn establish the asymptotic distributions of the plug-in statistic $\sum_{j=r+1}^{k} \sigma_j^2(\hat{\Pi}_n)$ under the null for a given estimator $\hat{\Pi}_n$ of Π_0. Since the resulting asymptotic distributions are highly nonstandard, we resort to the resampling procedure developed by Fang and Santos (2015) and Chen and Fang (2015) in order to obtain critical values.

3.1 Differential Properties

For ease of exposition, define $\phi : \mathbf{M}^{m\times k} \to \mathbf{R}$ by

$$\phi(\Pi) \equiv \sum_{j=r+1}^{k} \sigma_j^2(\Pi) \ , \tag{24}$$

where we recall that $\sigma_j(\Pi)$ is the jth largest singular value of Π. To derive the differentiability of ϕ, it shall prove useful to establish the following representation.

Lemma 3.1. *Let $\mathbb{S}^{k\times q} \equiv \{U \in \mathbf{M}^{k\times q} : U^\intercal U = I_q\}$ for $q = 1, \ldots, k$. Then we have*

$$\phi(\Pi) = \min_{U\in\mathbb{S}^{k\times(k-r)}} \|\Pi U\|^2 \ . \tag{25}$$

Lemma 3.1 shows that $\phi(\Pi)$ can be represented as a quadratic minimum over the space of orthonormal matrices in $\mathbf{M}^{m\times(k-r)}$. The special case when $r = k-1$ – a test of Π having full rank – is a well known implication of the classical Courant-Fischer theorem, i.e., $\sigma_k^2(\Pi) = \min_{\|U\|=1} \|\Pi U\|^2$. Note that the minimum in (25) is achieved and hence well defined.

We are now in a position to analyze the differential properties of ϕ. It turns out that ϕ is not fully differentiable but belongs to a class of directionally differentiable maps. For completeness, we next introduce the appropriate notions of differentiability.

[4]Recall that $\sigma_1^2(\Pi_0), \ldots, \sigma_k^2(\Pi_0)$ are numerically identical to eigenvalues of $\Pi_0^\intercal \Pi_0$.

Definition 3.1. Let $\mathbf{M}^{m\times k}$ be equipped with the norm $\|\cdot\|$ and $\varphi : \mathbf{M}^{m\times k} \to \mathbf{R}$.

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

$$\lim_{n\to\infty} \frac{\varphi(\Pi + t_n M_n) - \varphi(\Pi)}{t_n} = \varphi'_{\Pi}(M) , \tag{26}$$

for all sequences $\{M_n\} \subset \mathbf{M}^{m\times k}$ and $\{t_n\} \subset \mathbf{R}_+$ such that $t_n \downarrow 0$, and $M_n \to M \in \mathbf{M}^{m\times k}$ as $n \to \infty$.

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

$$\lim_{n\to\infty} \frac{\varphi(\Pi + t_n M_n) - \varphi(\Pi) - t_n \phi'_{\Pi}(M_n)}{t_n^2} = \varphi''_{\Pi}(M) , \tag{27}$$

for all sequences $\{M_n\} \subset \mathbf{M}^{m\times k}$ and $\{t_n\} \subset \mathbf{R}^+$ such that $t_n \downarrow 0$, and $M_n \to M \in \mathbf{M}^{m\times k}$ as $n \to \infty$.

Compared with Hadamard full differentiability (van der Vaart, 1998) which requires continuity and linearity of the derivative, the directional derivative is generally nonlinear though necessarily continuous. In fact, linearity is the exact gap between these two notions of differentiability. Remarkably, the Delta method remains valid even if ϕ is only Hadamard directionally differentiable. We refer the readers to Shapiro (1990, 1991), Dümbgen (1993), and a recent review by Fang and Santos (2015) for further details. Unfortunately, as shall be proved, the asymptotic distribution of our statistic $\phi(\hat{\Pi}_n)$ implied by the Delta method is degenerate under the null, which creates substantial challenges for inference. This motivates the second order Hadamard directional differentiability. Compared with second order Hadamard full differentiability which requires a quadratic form of the derivative corresponding to a bilinear map, the directional derivative ϕ''_{θ} is generally nonquadratic though continuous. In fact, quadratic form structure is the exact gap between these two notions of differentiability. Similarly, the second order Delta method remains valid even if ϕ is only second order Hadamard directionally differentiable. We refer the readers to Shapiro (2000) and a recent review by Chen and Fang (2015) for further details.

The following proposition establishes the differentiability of ϕ.

Proposition 3.1. *Let $\phi : \mathbf{M}^{m\times k} \to \mathbf{R}$ be defined as in* (24).

(i) ϕ is first order Hadamard directionally differentiable at any $\Pi \in \mathbf{M}^{m\times k}$ with the

derivative $\phi'_{\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ *given by*

$$\phi'_{\Pi}(M) = \min_{U\in\Psi(\Pi)} 2\mathrm{tr}(U^{\intercal}\Pi^{\intercal}MU) \ , \tag{28}$$

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

(ii) ϕ *is second order Hadamard directionally differentiable at any* $\Pi \in \mathbf{M}^{m\times k}$ *satisfying* $\phi(\Pi) = 0$ *with the derivative* $\phi''_{\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ *given by*

$$\phi''_{\Pi}(M) = \min_{U\in\Psi(\Pi)} \min_{V\in\mathbf{M}^{k\times(k-r)}} \|MU + \Pi V\|^2 \ . \tag{29}$$

Proposition 3.1 implies that ϕ is Hadamard directionally differentiable at any $\Pi \in \mathbf{M}^{m\times k}$. In particular, when $\mathrm{rank}(\Pi) \leq r$, it exhibits a degenerate derivative, i.e., $\phi'_{\Pi}(\mathrm{M}) = 0$ for all $M \in \mathbf{M}^{m\times k}$. Moreover, Proposition 3.1 implies that ϕ is second order Hadamard directionally differentiable at any $\Pi \in \mathbf{M}^{m\times k}$ with $\mathrm{rank}(\Pi) \leq r$. In general, ϕ is not second order fully Hadamard differentiable at $\Pi \in \mathbf{M}^{m\times k}$ with $\mathrm{rank}(\Pi) \leq r$ unless $\mathrm{rank}(\Pi) = r$, see Lemma A.4. Thus, the accommodation of $\mathrm{rank}(\Pi) < r$ causes the irregularity of ϕ.

To conclude this section, we provide a simplified analytical expression for ϕ''_{Π}. Let $\Pi = P\Sigma Q^{\intercal}$ be a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$ and $Q \in \mathbb{S}^{k\times k}$, and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Let $r^* \equiv \mathrm{rank}(\Pi)$. Write $P = [P_1, P_2]$ and $Q = [Q_1, Q_2]$ for $P_1 \in \mathbf{M}^{m\times r^*}$ and $Q_1 \in \mathbf{M}^{k\times r^*}$, respectively. Thus, the columns of P_2 and Q_2 are the left-singular vectors and right-singular vectors of Π associated with the zero singular values, respectively. Then the following proposition gives a simplified analytical expression of ϕ''_{Π}.

Proposition 3.2. *Suppose* $r^* \leq r$ *and let* $\phi''_{\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ *be given as in Proposition 3.1. Then for* $M \in \mathbf{M}^{m\times k}$,

$$\phi''_{\Pi}(M) = \sum_{j=r-r^*+1}^{k-r^*} \sigma_j^2(P_2^{\intercal}MQ_2) \ . \tag{30}$$

Proposition 3.2 implies $\phi''_{\Pi}(M)$ is the sum of the $k - r$ smallest squared singular values of transformed matrix $P_2^{\intercal}MQ_2$. Observe that P_2 and Q_2 are from singular value decomposition, so calculation of the derivative requires no more than calculation of singular value decomposition as in the test statistic. As we will see later, this facilitates the computation of our test statistic and makes our test procedure attractive. Note P_2 and Q_2 can be chosen up to postmultiplication by $(m - r^*) \times (m - r^*)$ and $(k - r^*) \times (k - r^*)$ orthonormal matrices, respectively, but the term on the right hand side of (30) is invariant to the choice of P_2 and Q_2.

3.2 The Asymptotic Distributions

Given the established differentiability of ϕ and null first order derivative, the asymptotic distribution of $\phi(\hat{\Pi}_n)$ can be easily obtained by the second order Delta method (Shapiro, 2000), provided $\hat{\Pi}_n$ converges weakly. Towards this end, we impose the following assumption.

Assumption 3.1. *Let $\Pi_0 \in \mathbf{M}^{m\times k}$ and there are $\hat{\Pi}_n : \{X_i\}_{i=1}^n \to \mathbf{M}^{m\times k}$ such that $\tau_n(\hat{\Pi}_n - \Pi_0) \xrightarrow{L} \mathcal{M}$ for some $\tau_n \uparrow \infty$ and random matrix $\mathcal{M} \in \mathbf{M}^{m\times k}$.*

Assumption 3.1 imposes that the estimator $\hat{\Pi}_n$ for Π_0 admits a weak limit $\mathcal{M} \in \mathbf{M}^{m\times k}$ at a scalar rate τ_n. The estimator $\hat{\Pi}_n$ is defined as a function of the data $\{X_i\}_{i=1}^n$ into $\mathbf{M}^{m\times k}$, and the weak convergence "$\xrightarrow{L}$" is understood with respect to the joint law of $\{X_i\}_{i=1}^n$, which need not be i.i.d.. In particular, τ_n is allowed to be any parametric or nonparametric rate that covers all the above examples.

Let $\Pi_0 = P_0\Sigma_0 Q_0^\intercal$ be a singular value decomposition of Π_0, where $P_0 \in \mathbb{S}^{m\times m}$ and $Q_0 \in \mathbb{S}^{k\times k}$, and $\Sigma_0 \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Let $r_0 \equiv \operatorname{rank}(\Pi_0)$. Write $P_0 = [P_{0,1}, P_{0,2}]$ and $Q_0 = [Q_{0,1}, Q_{0,2}]$ for $P_{0,1} \in \mathbf{M}^{m\times r_0}$ and $Q_{0,1} \in \mathbf{M}^{k\times r_0}$, respectively. Thus, the columns of $P_{0,2}$ and $Q_{0,2}$ are the left-singular vectors and right-singular vectors of Π_0 associated with the zero singular values, respectively. The following proposition delivers the asymptotic distributions of $\phi(\hat{\Pi}_n)$.

Proposition 3.3. *Suppose Assumption 3.1 holds. Then we have*

$$\tau_n(\phi(\hat{\Pi}_n) - \phi(\Pi_0)) \xrightarrow{L} \min_{U\in\Psi(\Pi_0)} 2\operatorname{tr}(U^\intercal \Pi_0^\intercal \mathcal{M} U) \ , \tag{31}$$

and under H_0*,*

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

Proposition 3.3 implies that $\tau_n\phi(\hat{\Pi}_n)$ converges in distribution to a degenerate limit at 0 under H_0. This prevents us from making inference based on the first order framework (Chen and Fang, 2015). Proposition 3.3 also implies that $\tau_n^2\phi(\hat{\Pi}_n)$ converges in distribution to a generally nondegenerate limit under H_0. This enables us to make inference based on the second order framework. The limit is a nonlinear function of the weak limit $\mathcal{M}$ and remarkably nonstandard especially when $r_0 < r$. In general, an analytical (pivotal) distribution is not available. Note that $P_{0,2}$ and $Q_{0,2}$ are identified up to postmultiplication by $(m - r_0) \times (m - r_0)$ and $(k - r_0) \times (k - r_0)$ orthonormal matrices, respectively, but the term on the right hand side of (32) is invariant to the choice of $P_{2,0}$ and $Q_{2,0}$.

In order to see how our results apply to various settings, we now turn to examples introduced in Section 2.1. We shall focus on Examples 2.1 and 2.3 exclusively for conciseness; Examples 2.2 and 2.4-2.7 will be treated in Appendix B. In particular, Assumption 3.1 is not well satisfied in Example 2.2 since the convergence rates of $\hat{\Pi}_n$ are not homogenous across its columns, and we extend the result in Proposition 3.3 for it.

Example 2.1 (Continued). Suppose $\{W_i, Z_i\}_{i=1}^n$ is a sequence of data from Example 2.1. Let $\hat{\Pi}_n$ be the method of moment estimator

$$\hat{\Pi}_n = \frac{1}{n}\sum_{i=1}^{n} W_i Z_i^\intercal \ . \tag{33}$$

Under certain weak dependence and moment condition, the central limit theorem implies that Assumption 3.1 is satisfied with $\tau_n = \sqrt{n}$ and $\mathcal{M}$ being a zero mean Gaussian. When $r_0 < k-1$, the asymptotic distribution of $n\phi(\hat{\Pi}_n)$ can be highly nonstandard. ■

Example 2.3 (Continued). Suppose $\{Y_t\}_{t=1}^n$ is a sequence of data from Example 2.3. Let $\hat{\Pi}_n$ be a kernel HAC estimator

$$\hat{\Pi}_n = \sum_{j=-n+1}^{n-1} k(\frac{j}{b_n})\hat{\Gamma}_n(j) \ , \tag{34}$$

where $\hat{\Gamma}_n(j) \equiv \frac{1}{n}\sum_{t=1}^{n-j} \Delta Y_t \Delta Y_{t+j}^\intercal$ for $j \geq 0$, $\hat{\Gamma}_n(j) = \hat{\Gamma}_n(-j)^\intercal$ for $j < 0$, $k(\cdot)$ is a kernel function, and b_n is a bandwidth parameter. Under certain weak dependence and moment conditions, $\hat{\Pi}_n$ is asymptotically normal at the rate $\sqrt{n/b_n}$. For example, see Hannan (1970), Brillinger (1981), Priestley (1981) and Berkes et al. (2016). So, Assumption 3.1 is satisfied with $\tau_n = \sqrt{n/b_n}$ and $\mathcal{M}$ being a zero mean Gaussian. In testing for the existence of cointegration, when $r_0 < k-1$, the asymptotic distribution of $n\phi(\hat{\Pi}_n)/b_n$ can be highly nonstandard. ■

We now discuss the result of Proposition 3.3 when $r_0 = r$ and its relation to the literature. In this case, $P_{0,2}^\intercal \mathcal{M} Q_{0,2}$ has $k-r$ columns and $\sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^\intercal \mathcal{M} Q_{0,2})$ is equal to the Frobenius norm of $P_{0,2}^\intercal \mathcal{M} Q_{0,2}$. Thus, the asymptotic distribution in (32) becomes

$$\|P_{0,2}^\intercal \mathcal{M} Q_{0,2}\|^2 = \mathrm{vec}(P_{0,2}^\intercal \mathcal{M} Q_{0,2})^\intercal \mathrm{vec}(P_{0,2}^\intercal \mathcal{M} Q_{0,2}) \ . \tag{35}$$

When $\mathcal{M}$ is a zero mean Gaussian, the limit is a weighted sum of independent $\chi^2(1)$ random variables. Thus, Proposition 3.3 includes Robin and Smith (2000) as a special case. If, in addition, the covariance matrix of $\mathrm{vec}(P_{0,2}^\intercal \mathcal{M} Q_{0,2})$ is nonsingular, Kleibergen and Paap (2006) proved that a normalized version of $\tau_n^2\phi(\hat{\Pi}_n)$ has a $\chi^2((m-r)(k-r))$ asymptotic distribution under $\mathrm{H}_0^{(r)}$. The asymptotic distribution is not a χ^2-type distribution any more if $r_0 < r$. This suggests that the Robin and Smith (2000) test when directly applied to (1) may fail to control the asymptotic null rejection rate.

3.3 The Bootstrap

Given the nonstandard asymptotic distribution in Proposition 3.3, no analytical critical values can be employed for inference. We may resort to the standard bootstrap method (Efron, 1979) to consistently estimate the asymptotic distribution. Unfortunately, the consistency of this method fails due to the degeneracy of ϕ'_{Π_0} under the null (Chen and Fang, 2015). Moreover, the recentered bootstrap does not necessarily correct the inconsistency due to the nondifferentiability of ϕ. As such, we resort to the procedure developed by Chen and Fang (2015) for construction of critical values. See the discussion on m out of n bootstrap and subsampling in Remark 3.1.

Recall that the asymptotic distribution is a composition of $\mathcal{M}$ and ϕ''_{Π_0}. Our proposed procedure consists of first estimating $\mathcal{M}$ by bootstrap and then estimating ϕ''_{Π_0}. For the former, let $\hat{\Pi}_n^*$ denote a "bootstrapped version" of $\hat{\Pi}_n$, which is defined as a function of the data $\{X_i\}_{i=1}^n$ and random weights $\{W_i\}_{i=1}^n$ that are independent of $\{X_i\}_{i=1}^n$ into $\mathbf{M}^{m\times k}$. This general definition allows us to include special cases such as nonparametric, Bayesian, block, score, more generally multiplier and exchangeable bootstrap. To accommodate diverse resampling schemes, we simply impose the following high level condition.

Assumption 3.2. *(i)* $\hat{\Pi}_n^* : \{X_i, W_i\}_{i=1}^n \to \mathbf{M}^{m\times k}$ *with* $\{W_i\}_{i=1}^n$ *independent of* $\{X_i\}_{i=1}^n$*; (ii)* $\tau_n(\hat{\Pi}_n^* - \hat{\Pi}_n) \xrightarrow{L^*} \mathcal{M}$ *almost surely, where* $\xrightarrow{L^*}$ *denotes weak convergence with respect to the joint law of* $\{W_i\}_{i=1}^n$ *conditional on* $\{X_i\}_{i=1}^n$.

Assumption 3.2(i) defines the bootstrap analog $\hat{\Pi}_n^*$ of $\hat{\Pi}_n$, while Assumption 3.2(ii) simply imposes the consistency of the law of $\tau_n(\hat{\Pi}_n^* - \hat{\Pi}_n)$ conditional on the data $\{X_i\}_{i=1}^n$ for the law of $\mathcal{M}$, i.e., the bootstrap works for the estimator $\hat{\Pi}_n$.

Next we examine Assumption 3.2 in Examples 2.1 and 2.3; Examples 2.2 and 2.4-2.7 will be treated in Appendix B. In particular, Assumption 3.2 is not well satisfied in Example 2.2, and we extend the result in Theorem 3.1 for it.

Example 2.1 (Continued). Let $\{Z_i^*, W_i^*\}_{i=1}^n$ be obtained by nonparametric bootstrapping $\{Z_i, W_i\}_{i=1}^n$ when $\{Z_i, W_i\}_{i=1}^n$ is a sequence of i.i.d. data, and by block bootstrapping $\{Z_i, W_i\}_{i=1}^n$ when $\{Z_i, W_i\}_{i=1}^n$ is a sequence of dependent data. Under certain weak dependence and moment condition, Assumption 3.2 is satisfied with

$$\hat{\Pi}_n^* = \frac{1}{n}\sum_{i=1}^n W_i^* Z_i^{*\intercal} \; . \tag{36}$$

Multiplier and exchangeable bootstrap may also be employed for i.i.d. data. ■

Example 2.3 (Continued). Since $\hat{\Pi}_n$ only depends on $\{\Delta Y_t\}_{t=1}^n$, it suffices to resample $\{\Delta Y_t\}_{t=1}^n$. Note that $\{\Delta Y_t\}_{t=1}^n$ is stationary. Let $\{\Delta Y_t^*\}_{t=1}^n$ be obtained by block bootstrapping $\{\Delta Y_t\}_{t=1}^n$. Under certain weak dependence and moment condition, Assumption 3.2 is satisfied with

$$\hat{\Pi}_n^* = \sum_{j=-n+1}^{n-1} k(\frac{j}{b_n})\hat{\Gamma}_n^*(j) \ , \tag{37}$$

where $\hat{\Gamma}_n^*(j) \equiv \frac{1}{n}\sum_{t=1}^{n-j} \Delta Y_t^* \Delta Y_{t+j}^{*\intercal}$ for $j \geq 0$, $\hat{\Gamma}_n^*(j) = \hat{\Gamma}_n^*(-j)^\intercal$ for $j < 0$, $k(\cdot)$ and b_n are the same kernel function and bandwidth parameter. See Politis and Romano (1992, 1993) and Politis et al. (1992) for other bootstrap procedures. ■

There are two main methods for estimating ϕ''_{Π_0}: the structure-exploiting approach and the numerical differentiation approach. For the former, we describe how to estimate ϕ''_{Π_0} according to (30). Let $\hat{\Pi}_n = \hat{P}_n\hat{\Sigma}_n\hat{Q}_n^\intercal$ be a singular value decomposition of $\hat{\Pi}_n$, where $\hat{P}_n \in \mathbb{S}^{m\times m}$ and $\hat{Q}_n \in \mathbb{S}^{k\times k}$, and $\hat{\Sigma}_n \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Let $\hat{r}_n \equiv \min\{r, \#\{1 \leq j \leq k : \sigma_j(\hat{\Pi}_n) \geq \kappa_n\}\}$, where $\kappa_n \downarrow 0$ is a tuning parameter that is required to satisfy certain conditions below.[5] Write $\hat{P}_n = [\hat{P}_{1,n}, \hat{P}_{2,n}]$ and $\hat{Q}_n = [\hat{Q}_{1,n}, \hat{Q}_{2,n}]$ for $\hat{P}_{1,n} \in \mathbf{M}^{m\times\hat{r}_n}$ and $\hat{Q}_{1,n} \in \mathbf{M}^{k\times\hat{r}_n}$, respectively. By (30), we may estimate ϕ''_{Π_0} by

$$\hat{\phi}''_n(M) = \sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^\intercal M\hat{Q}_{2,n}) \ . \tag{38}$$

Note that $\hat{P}_{2,n}$ and $\hat{Q}_{2,n}$ can be chosen up to postmultiplication by $(m-\hat{r}_n)\times(m-\hat{r}_n)$ and $(k-\hat{r}_n)\times(k-\hat{r}_n)$ orthonormal matrices, respectively, but the term on the right hand side of (38) is invariant to the choice of $\hat{P}_{2,n}$ and $\hat{Q}_{2,n}$. For the latter, we estimate ϕ''_{Π_0} by

$$\hat{\phi}''_n(M) = \frac{\phi(\hat{\Pi}_n + \kappa_n M) - \phi(\hat{\Pi}_n)}{\kappa_n^2} \ . \tag{39}$$

Remark 3.1. In effect, m out of n bootstrap and subsampling amounts to estimating $\mathcal{M}$ based on subsamples (with and without replacement, respectively) and ϕ''_{Π_0} via the numerical differentiation approach, in which case the tuning parameters for choosing subsamples and estimation of the derivative coincide. Thus, our bootstrap procedure can be more efficient in two ways. First, our bootstrap procedure makes efficient use of the data in estimating $\mathcal{M}$, since it is based on full samples. Second, our bootstrap procedure also provides alternative method of estimating ϕ''_{Π_0} by exploiting more structural

[5] We use $\#A$ to denote the cardinality of a set A. One can theoretically ignore r in the expression of $\hat{r}_n$. However, taking minimum in the expression of $\hat{r}_n$ is a way of imposing the information under the null to ensure that the estimator in (38) is well defined and improve power.

information of the data .

Given a suitable condition on $\kappa_n \downarrow 0$, we are then able to prove that the law of the weak limit in (32) is consistently estimated by the law of $\hat{\phi}_n''(\tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\})$ conditional on the data. It in turn suggests employing the $1-\alpha$ quantile $\hat{c}_{1-\alpha}$ of $\hat{\phi}_n''(\tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\})$ conditional on the data:[6]

$$\hat{c}_{1-\alpha} \equiv \inf\{c \in \mathbf{R} : P_W(\hat{\phi}_n''(\tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\}) \leq c) \geq 1-\alpha\} \ . \tag{40}$$

Note that $\hat{c}_{1-\alpha}$ is generally infeasible in that it is constructed based on the "exact" distribution of $\hat{\phi}_n''(\tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\})$ conditional on the data. Nonetheless, it can be estimated by Monte Carlo simulation and the estimation error can be made arbitrarily small by choosing the number of bootstrap replications (Efron, 1979; Hall, 1992; Horowitz, 2001).

For each realization of $\tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\}$, the computation of $\hat{\phi}_n''(\tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\})$ requires no more than calculating singular value decompositions with $\hat{\phi}_n''$ in (38) and (39). When $\hat{\phi}_n''$ is given in (38), it is only necessary to calculate the singular value decomposition of $\hat{P}_{2,n}^{\intercal}\tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\}\hat{Q}_{2,n}$. When $\hat{\phi}_n''$ is given in (39), it is only necessary to calculate the singular value decomposition of $\hat{\Pi}_n + \kappa_n\tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\}$. Thus, the computation of simulated critical values is as simple as the computation of the test statistic. Comparisons between the estimators in (38) and (39) will be investigated in Monte Carlo studies.

The following theorem establishes that the test of rejecting H_0 when $\tau_n^2\phi(\hat{\Pi}_n) > \hat{c}_{1-\alpha}$ controls the asymptotic null rejection rate and is consistent.

Theorem 3.1. *Suppose Assumptions 3.1 and 3.2 hold. Let $\kappa_n \downarrow 0$ and $\tau_n\kappa_n \to \infty$. Let $\hat{c}_{1-\alpha}$ be given in* (40) *with $\hat{\phi}_n''$ in* (38) *or* (39)*. If the cdf of the limit distribution in* (32) *is continuous and strictly increasing at its $1-\alpha$ quantile for $\alpha \in (0,1)$, then under* H_0,

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

Furthermore, under H_1,

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

Theorem 3.1 implies that our tests have the asymptotic null rejection rate that is exactly equal to the nominal level, regardless of whether $r_0 = r$ or $r_0 < r$. This stems from the design of our bootstrap that estimates the asymptotic distribution pointwise in Π_0. In contrast to existing rank tests and the multiple testing method that may have the asymptotic null rejection rate strictly below the nominal level when $r_0 < r$, this distinct feature shall make our tests more powerful. In particular, when Π_0 is close to a matrix with rank strictly less than r, our tests shall be more powerful in detecting

[6] P_W denotes the probability with respect to the joint law of the random weights $\{W_i\}_{i=1}^n$.

H_1 than existing rank tests and the multiple testing method. In addition, in contrast to existing rank tests that may fail to control the asymptotic null rejection rate when $r_0 < r$, our tests control the asymptotic null rejection rate regardless of whether $r_0 = r$ or $r_0 < r$. Theorem 3.1 also implies that our tests are consistent.

Several simple, new and powerful tests are immediate from Theorem 3.1. First, applying Theorem 3.1 to Examples 2.1 yields tests for identification in linear IV models. Second, applying Theorem 3.1 to Examples 2.2 and 2.3 yields tests for the existence of stochastic trend and/or cointegration with or without VAR specification, respectively. Third, applying Theorem 3.1 to Examples 2.4 yields tests for the existence of common features.

We now discuss the quantile requirement on the limit distribution in (32) imposed in Theorem 3.1. A necessary condition for that requirement to hold is $P_{0,2}^{\intercal}\mathcal{M}Q_{0,2} \neq 0$ with positive probability, that is,

$$P(\mathcal{R}(\mathcal{M}) \cap \mathcal{N}(\Pi_0^{\intercal}) \neq \emptyset) > 0 \text{ and } P(\mathcal{R}(\mathcal{M}^{\intercal}) \cap \mathcal{N}(\Pi_0) \neq \emptyset) > 0 ,$$

where $\mathcal{R}(A)$ denotes the range of a matrix A and $\mathcal{N}(A)$ denotes the null space of a matrix A. When $\mathcal{M}$ is zero mean Gaussian and $r_0 = r$, the limit in (32) is a weighted sum of independent $\chi^2(1)$ random variables as shown in (35). This implies the limit distribution is continuous, unless the covariance matrix of $\text{vec}(P_{0,2}^{\intercal}\mathcal{M}Q_{0,2})$ is zero. Thus, in this special case, the sufficient and necessary condition for the requirement to hold is nonzero of the covariance matrix of $\text{vec}(P_{0,2}^{\intercal}\mathcal{M}Q_{0,2})$. In contrast, Kleibergen and Paap (2006) requires nonsingularity of the covariance matrix of $\text{vec}(P_{0,2}^{\intercal}\mathcal{M}Q_{0,2})$. In view of this, our tests rely on much weaker conditions than Kleibergen and Paap (2006).

Remark 3.2. The requirement on the limit distribution in (32) imposed in Theorem 3.1 may not be satisfied in testing for perfect multicollinearity in Example 2.1. When $\hat{\Pi}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^{\intercal}$, then the limit in (32) is degenerate at zero, which can be best seen from (32) since $\mathcal{M}Q_{0,2} = 0$. Heuristically, if the smallest singular value of Π_0 is zero, then $\lambda^{\intercal} Z_i$ is constantly zero for some constant $\lambda \in \mathbb{S}^k$ and the smallest singular value of $\hat{\Pi}_n$ is constantly zero. Nevertheless, one can easily prove that the properties of size control and consistency continue to hold. ■

4 Simulations and Applications

In this section, we first conduct Monte Carlo studies to examine the finite sample performance of our tests, and show how existing rank tests when directly applied to (1) and the multiple testing method may be conservative. We then apply our tests to study identification in stochastic discount factor models (Jagannathan and Wang, 1996). Lastly, we demonstrate how our tests can improve the accuracy of the sequential testing procedure for rank determination.

4.1 Simulation Studies

We start with the performance of our tests for the problem in Section 2.2. To implement our tests, we use the same estimator $\hat{\Pi}_n$ as in Section 2.2 and the same nominal level 5%. The rejection rates, which are based on 10,000 simulation replications with 500 nonparametric i.i.d. bootstrap replications for each Monte Carlo, are plotted in Figure 3. Clearly, Assumptions 3.1 and 3.2 are satisfied. The result is based on the derivative estimator in (38) with $\kappa_n = n^{-1/4}$, while the result for the derivative estimator in (38) with $\kappa_n = n^{-1/3}$ is similar and available upon request. For ease of comparison, we combine Figures 1 and 3 to yield Figure 4, where CF denotes our tests and KP-M is defined in Section 2.1. In contrast to KP-M, the null rejection rates of CF are close to the 5% nominal level for all d as shown in Figure 3. As expected from Theorem 3.1, CF are more powerful than KP-M uniformly over $d \neq 1$ and all $\delta > 0$ as shown in Figure 4. In particular, in contrast to KP-M, all power curves of CF lie above the 5% nominal level line. Note the power curves do not coincide since the data generating process (DGP) is varied for different d. Figure 4 also shows that the greater the value of d is, the greater the power improvement is. In addition, when $d = 1$, CF are as powerful as KP-M. Thus, these findings confirm that KP-M are too conservative, and CF provide power improvement over KP-M. Given Figure 2, the comparison between CF and KP-D is the same.

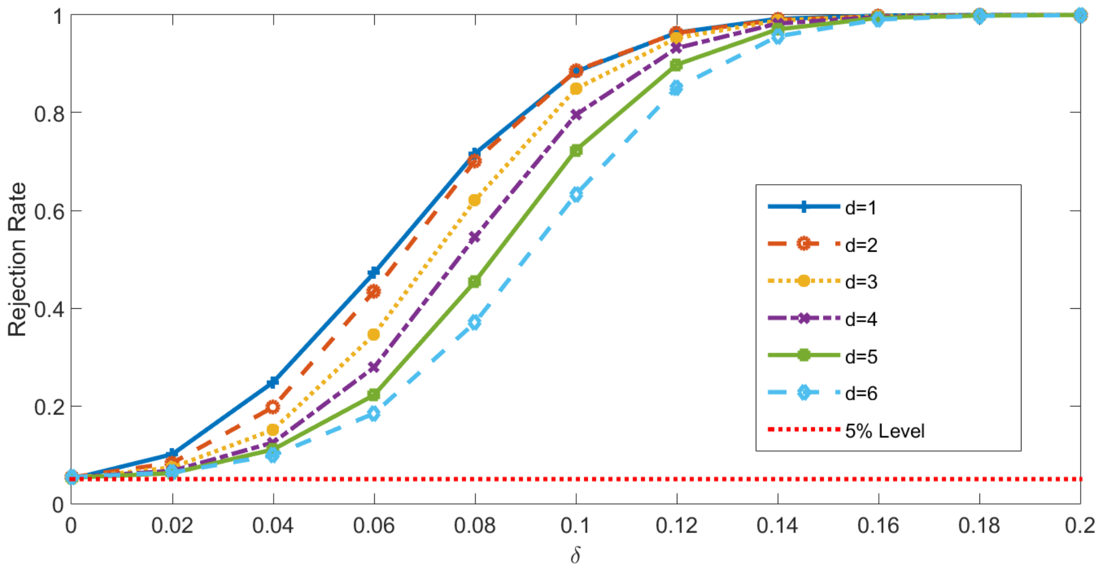

Figure 3: The rejection rate of our tests with 5% nominal level

We next investigate the finite sample performance of our tests, the Kleibergen and

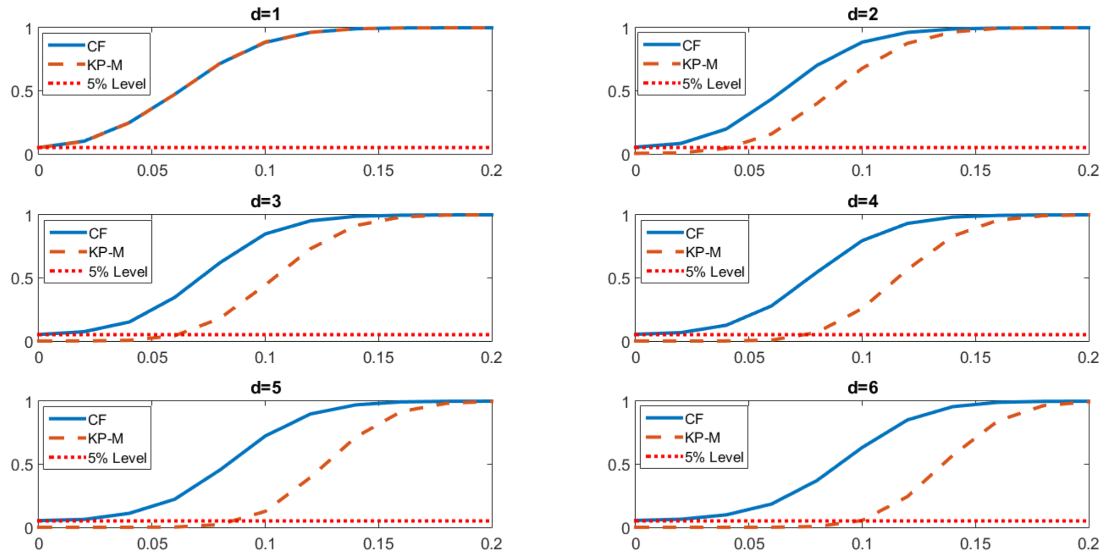

Figure 4: Comparison between our tests and the multiple testing method based on the Kleibergen and Paap (2006) test with 5% nominal level

Paap (2006) test when directly applied, and the multiple testing method in more complicated DGPs with heteroskedasticity, dependence and different sample sizes. We consider two types of DGPs. For the first DGP (DGP1), we assume

$$Z_t^\intercal = W_t^\intercal \Pi_0 + W_{1,t} u_t^\intercal \text{ with } u_t = v_t - \frac{1}{4}\mathbf{1}_4\mathbf{1}_4^\intercal v_{t-1}, t = 1, \ldots, T,$$

where $v_t \overset{i.i.d.}{\sim} N(0, I_4)$, $W_t \overset{i.i.d.}{\sim} N(0, I_4)$ and $W_{1,t}$ is the first element of W_t. Note the errors now are heteroskedastic and autocorrelated. Let

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

For the second DGP (DGP2), following Kleibergen and Paap (2006) we assume

$$R_t = \Pi_0 F_t + \varepsilon_t \text{ with } \varepsilon_t = v_t + \Gamma v_{t-1}, t = 1, \ldots, T,$$

where $v_t \overset{i.i.d.}{\sim} N(0, \Sigma_v)$ and $F_t \overset{i.i.d.}{\sim} N(0, \Sigma_F)$ with $\Gamma \in \mathbf{M}^{10\times 10}$, $\Sigma_v \in \mathbf{M}^{10\times 10}$ and $\Sigma_F \in \mathbf{M}^{4\times 4}$ given in Appendix D. Let

$$\Pi_0 = \beta\alpha^\intercal + \rho \Pi_1 \text{ for } \rho \geq 0 \ ,$$

where $\alpha \in \mathbf{R}^4$, $\beta \in \mathbf{R}^{10}$ and $\Pi_1 \in \mathbf{M}^{10\times 4}$ are given in Appendix D. These values are estimates based on the real data used in Section 4.2. In view of this, we use DGP2 to mimic possible scenarios in Section 4.2 as in Kleibergen and Paap (2006).

We examine the hypotheses (1) with $r = 2$ and $r = 3$ for DGP1, and the hypotheses

(1) with $r = 3$ for DGP2. The design of Π_0 implies that H_0 is true if and only if $\delta = 0$ for both cases. In particular, for DGP1 $r_0 = 2$ under H_0, and for DGP2 $r_0 = 1$ under H_0. So $r = 3$ for both DGPs represents the case when Π_0 is close to a matrix with rank strictly less than r, while $r = 2$ for DGP1 represents the regular case. Given the findings in Figure 4, for the hypotheses with $r = 3$ for both DGPs, it shall be expected that our tests are more powerful than the Kleibergen and Paap (2006) test when directly applied and the multiple testing method.

To implement all tests, we estimate Π_0 by $\hat{\Pi}_T = \frac{1}{T}\sum_{t=1}^{T} W_t Z_t^{\intercal}$ for DGP1 and by $\hat{\Pi}_T = \sum_{t=1}^{T} R_t F_t^{\intercal}\,(\sum_{t=1}^{T} F_t F_t^{\intercal})^{-1}$ for DGP2. It is clear that the asymptotic distribution $\mathcal{M}$ of $\hat{\Pi}_T$ is zero mean Gaussian with convergence rate $\sqrt{T}$, so Assumption 3.1 is satisfied. As the data exhibits first order autocorrelation, we adopt the simple block bootstrap (Lahiri, 2003) to resample the data with block size $b = 2$ for implementing our tests. For derivative estimation in (38) and (39), we set the tuning parameter $\kappa_T = T^{-1/4}$ and $T^{-1/3}$. It is also clear that all assumptions in Kleibergen and Paap (2006) are satisfied. We use HACC matrix estimator with one lag (West, 1997) for the long run covariance matrix estimator. See Appendix C for a review on the Kleibergen and Paap (2006) test.

We let $\rho = 0, 0.1, \cdots, 0.5$ for DGP1 and $\rho = 0, 0.01, \cdots, 0.1$ for DGP2, where ρ represents how strong H_1 deviates away from H_0. We consider $T = 50, 100, 300, 1000$ for DGP1 and $T = 330$ for DGP2. The rejection rates, which are based on $5,000$ simulation replications with 500 bootstrap replications, are reported in Tables 1-3. We use CF1 and CF2 to denote our tests using derivative estimator in (38) and (39), respectively, and KP-D and KP-M to denote the Kleibergen and Paap (2006) test when directly applied and the multiple testing method, respectively. The nominal level is 5% throughout.

The main findings are summarized as follows. First, CF1 exhibits good finite sample performance for all cases, even when $T = 50$. Interestingly, as Tables 1 and 3 show, the rejection rates of CF1 for $r = 2$ under DGP1 and $r = 3$ under DGP2 are invariant to the choice of κ_T in most of cases. The rejection rates of CF1 for $r = 3$ under DGP1 are not quite sensitive to the choice of κ_T. Second, the performance of CF2 is not as satisfactory as that of CF1 in small samples. In particular, CF2 is over rejected for all cases with $\rho = 0$ when $T = 50$ or 100. This indicates that good performance of CF2 may require a larger T than CF1 does. This may be explained by the fact that the structural method (CF1) exploits more information of the derivative. For large T, CF2 seems to be more powerful than CF1 under DGP1 when $T = 300$ or $1,000$, while CF1 seems to be more powerful than CF2 under DGP2. We leave a thorough comparison between these two methods of derivative estimation for future study. Third, the performance of KP-M and KP-D is less satisfactory than our tests. As Table 1 shows, KP-M and KP-D over-reject the null for $r = 2$ under DGP1 with $\rho = 0$ when $T = 50$ or 100. This indicates that good performance of KP-M and KP-D may require a large T. On other other hand, as Tables 2 and 3 show, KP-M and KP-D under-reject the null for $r = 3$ under DGP1 and DGP2 with $\rho = 0$. This is consistent with the finding in Figure 1. Moreover, as expected, CF1 and CF2 are uniformly more powerful than KP-M and KP-D as shown in Tables 2 and 3.[7] In addition, in our designed simulation, the rejection rates of KP-M and KP-D are similar with insignificant difference, although the latter is slightly more powerful.

4.2 Testing for Identification in SDF Models

Following Jagannathan and Wang (1996), the stochastic discount factor (SDF) model based on the conditional capital asset pricing model is specified as

$$E[R_{t+1}F_{t+1}^{\intercal}\gamma_0|\mathcal{I}_t] = \mathbf{1}_m \ , \tag{41}$$

where $R_t \in \mathbf{R}^m$ is a vector of returns on m assets at time t, $F_t \in \mathbf{R}^k$ is a vector of common factors at time t, $\mathcal{I}_t$ is the information set at time t, and $\gamma_0 \in \mathbf{R}^k$ is a vector of risk premia. The risk premia γ_0 can be estimated by the generalized method of moments (Hansen, 1982), see, for example, Jagannathan et al. (2002). The GMM estimator of γ_0 is consistent if

$$E[R_{t+1}F_{t+1}^{\intercal}|\mathcal{I}_t] \tag{42}$$

is of full rank at time t, see, for example, Hansen (1982) and Newey and McFadden (1994). Therefore, it is of importance to test for the full rank of (42) to indicate whether γ_0 is identifiable.

When the conditional expectation of $R_{t+1}F_{t+1}^{\intercal}$ does not depend on $\mathcal{I}_t$ and R_t satisfies a linear factor model

$$R_t = \Pi_0 F_t + \varepsilon_t \tag{43}$$

with $E[F_t\varepsilon_t] = 0$ and $E[F_tF_t^{\intercal}]$ being nonsingular, then testing for the full rank of (42) is equivalent to testing for the full rank of Π_0. Following Kleibergen and Paap (2006), instead of testing for the full rank of (42), we opt to test whether Π_0 is of full rank. Thus, this amounts to examining the hypotheses (1) with $r = k-1$. We cannot restrict ourself to examine the hypotheses (2) since it is unrealistic to assume $r_0 \geq k-1$ unless $k = 1$.

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

[7] In Table 2, the rejection rates of KP-M and KP-D under the alternatives are size adjusted ones.

Table 1: Rejection rates for $r = 2$ under DGP1

	CF1		CF2		KP	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	KP-M	KP-D
$\rho = 0$						
$T = 50$	0.0380	0.0374	0.2764	0.2078	0.1276	0.1658
$T = 100$	0.0402	0.0402	0.1958	0.1232	0.0930	0.0952
$T = 300$	0.0450	0.0450	0.1218	0.0512	0.0606	0.0606
$T = 1000$	0.0472	0.0472	0.0752	0.0368	0.0526	0.0526
$\rho = 0.1$						
$T = 50$	0.0812	0.0812	0.3520	0.2676	0.0720	0.0828
$T = 100$	0.1210	0.1210	0.3356	0.2314	0.1262	0.1316
$T = 300$	0.3458	0.3458	0.5144	0.3600	0.3961	0.3962
$T = 1000$	0.8976	0.8976	0.9238	0.8784	0.9784	0.9152
$\rho = 0.2$						
$T = 50$	0.2248	0.2248	0.5714	0.4880	0.1904	0.2078
$T = 100$	0.4254	0.4254	0.6950	0.5750	0.4410	0.4520
$T = 300$	0.9350	0.9350	0.9694	0.9366	0.9526	0.9526
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\rho = 0.3$						
$T = 50$	0.4776	0.4776	0.7852	0.7208	0.3682	0.4142
$T = 100$	0.8044	0.8044	0.9348	0.8906	0.7964	0.8102
$T = 300$	0.9992	0.9992	0.9980	1.0000	1.0000	1.0000
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\rho = 0.4$						
$T = 50$	0.7220	0.7220	0.9236	0.8896	0.5380	0.6212
$T = 100$	0.9618	0.9618	0.9954	0.9832	0.9456	0.9586
$T = 300$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\rho = 0.5$						
$T = 50$	0.8872	0.8872	0.9786	0.9658	0.6696	0.7846
$T = 100$	0.9960	0.9960	0.9994	0.9992	0.9840	0.9946
$T = 300$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 2: Rejection rates for $r = 3$ under DGP1

	$\rho = 0$					
	CF1		CF2		KP	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	KP-M	KP-D
$T = 50$	0.0594	0.0388	0.1410	0.1248	0.0072	0.0156
$T = 100$	0.0556	0.0328	0.1156	0.0944	0.0066	0.0110
$T = 300$	0.0486	0.0324	0.0766	0.0564	0.0050	0.0062
$T = 1000$	0.0550	0.0484	0.0656	0.0544	0.0044	0.0056
	$\rho = 0.1$					
	CF1		CF2		KP	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	KP-M	KP-D
$T = 50$	0.0874	0.0534	0.1770	0.1600	0.0168	0.0270
$T = 100$	0.1114	0.0626	0.1936	0.1624	0.0270	0.0344
$T = 300$	0.2926	0.1562	0.3628	0.3068	0.0926	0.0994
$T = 1000$	0.8070	0.5948	0.8226	0.7730	0.5396	0.5428
	$\rho = 0.2$					
	CF1		CF2		KP	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	KP-M	KP-D
$T = 50$	0.1804	0.1182	0.3334	0.3030	0.0698	0.0910
$T = 100$	0.3162	0.2060	0.4882	0.4342	0.1692	0.1806
$T = 300$	0.7774	0.6724	0.8872	0.8426	0.6644	0.6666
$T = 1000$	0.9988	0.9986	0.9960	0.9994	0.9982	0.9982
	$\rho = 0.3$					
	CF1		CF2		KP	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	KP-M	KP-D
$T = 50$	0.3254	0.2538	0.5566	0.5166	0.1906	0.2432
$T = 100$	0.5678	0.4986	0.7886	0.7414	0.4856	0.4962
$T = 300$	0.9602	0.9576	0.9940	0.9874	0.9602	0.9602
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\rho = 0.4$					
	CF1		CF2		KP	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	KP-M	KP-D
$T = 50$	0.4916	0.4460	0.7488	0.7110	0.3544	0.4434
$T = 100$	0.7758	0.7626	0.9422	0.9120	0.7552	0.7656
$T = 300$	0.9972	0.9972	0.9998	0.9996	0.9974	0.9974
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\rho = 0.5$					
	CF1		CF2		KP	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	KP-M	KP-D
$T = 50$	0.6432	0.6288	0.8798	0.8442	0.5182	0.6290
$T = 100$	0.9146	0.9138	0.9880	0.9766	0.9016	0.9116
$T = 300$	0.9998	0.9998	1.0000	1.0000	0.9998	0.9998
$T = 1000$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 3: Rejection rates for $r = 3$ under DGP2 when $T = 330$

	CF1		CF2		KP	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	KP-M	KP-D
$\rho = 0.00$	0.0514	0.0514	0.0468	0.0406	0.0006	0.0008
$\rho = 0.01$	0.2834	0.2834	0.1770	0.1104	0.0460	0.0482
$\rho = 0.02$	0.4228	0.4228	0.1648	0.0864	0.0956	0.1018
$\rho = 0.03$	0.5850	0.5850	0.2192	0.1242	0.2044	0.2166
$\rho = 0.04$	0.7526	0.7526	0.3268	0.2388	0.3562	0.3768
$\rho = 0.05$	0.8706	0.8706	0.4944	0.4010	0.5314	0.5598
$\rho = 0.06$	0.9500	0.9500	0.6622	0.5796	0.6898	0.7294
$\rho = 0.07$	0.9822	0.9606	0.8064	0.7388	0.7994	0.8464
$\rho = 0.08$	0.9932	0.9852	0.9032	0.8628	0.8748	0.9276
$\rho = 0.09$	0.9982	0.9936	0.9582	0.9368	0.9144	0.9670
$\rho = 0.10$	0.9998	0.9984	0.9842	0.9754	0.9306	0.9852

labor income growth. We estimate Π_0 by

$$\hat{\Pi}_T = \sum_{t=1}^{T} R_t F_t^{\intercal} (\sum_{t=1}^{T} F_t F_t^{\intercal})^{-1} . \tag{44}$$

As demonstrated in Kleibergen and Paap (2006), the data on returns R_t exhibits first order autocorrelation. To compute the test statistics of Kleibergen and Paap (2006) test, we use HACC matrix estimator with one lag (West, 1997) for the long run covariance matrix estimator. To implement our tests, we adopt the simple block bootstrap (Lahiri, 2003) to resample the data with block size $b = 1, 2, 3, 4$. For derivative estimation in (39) and (38), we set the tuning parameter $\kappa_T = T^{-1/4}$ and $T^{-1/3}$.

The results, which are based on $1,000$ bootstrap replications, are reported in Table 4. We use CF1 and CF2 to denote our tests using derivative estimator in (38) and (39), respectively, and KP-D and KP-M to denote the Kleibergen and Paap (2006) test when directly applied and the multiple testing method based on it, respectively. As Panel A of Table 4 indicates, all our tests fail to reject the non-full rank of Π_0 with 5% nominal level, which is consistent with the finding in Kleibergen and Paap (2006). However, the p values of our tests are uniformly smaller than 15% with some smaller than 10%, while the p values of the two conventional tests are larger than 90%. This implies that our tests reject the non-full rank of Π_0 in some cases at the 10% level, while the conventional tests never reject the non-full rank of Π_0 at any conventional significance level. In this sense, the evidence for non-identification of γ_0 from our tests is very weak, while the evidence from the conventional tests is very strong. Given the drawback of the conventional tests, the conclusion from our tests is more reliable.

Table 4: p values for different tests

Panel A: our tests				
	CF1		CF2	
	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$	$\kappa_T = T^{-1/4}$	$\kappa_T = T^{-1/3}$
$b = 1$	0.079	0.079	0.118	0.121
$b = 2$	0.094	0.094	0.113	0.119
$b = 3$	0.103	0.103	0.128	0.140
$b = 4$	0.082	0.082	0.137	0.138

Panel B: conventional tests	
KP-M	KP-D
0.9063	0.9063

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

4.3 Rank Determination

Testing for the hypotheses (1) only tells whether r_0 satisfies the inequality or not. In many cases, however, we still want to know what r_0 is. In addition to employing the multiple testing method to test for inequality of cointegration rank, Johansen (1995, Chapter 12) also proposed a sequential testing procedure to determine the rank of cointegration in VAR models (see, for instance, Example 2.2). More examples that concern the true rank of a matrix can be found in Examples 2.5-2.7. In this section, we demonstrate how our tests can improve the accuracy of the sequential testing procedure for rank determination.

We first characterize the sequential testing procedure for rank determination in our general framework following Johansen (1995, Chapter 12). For $\alpha \in (0, 1)$, let $\psi_n^{(r)}$ be a test for the hypotheses (1) or (2) such that $\lim_{n\to\infty} P(\psi_n^{(r)} = 1) = \alpha$ when $r_0 = r$, and $\lim_{n\to\infty} P(\psi_n^{(r)} = 1) = 1$ when $r_0 > r$. For example, it can be any one of existing rank tests or our tests. The sequential testing procedure starts with $q = 0$ and carries out $\psi_n^{(q)}$ with progressively larger q. The rank estimator $\hat{r}_n^*$ is defined as the threshold value q^* when $\psi_n^{(q^*)}$ does not reject the null hypothesis for the first time, and $\hat{r}_n^* = k$ if such q^* does not exist. Formally, $\hat{r}_n^* = k$ if $\psi_n^{(q)} = 1$ for all $0 \le q \le k - 1$ and otherwise

$$\hat{r}_n^* = \min\{0 \le q \le k - 1 : \psi_n^{(q)} = 0\} . \tag{45}$$

Remark 4.1. Clearly, $\hat{r}_n^* > r$ is equivalent to $\psi_n^{(q)} = 1$ for all $0 \le q \le r$. Thus, for given existing rank tests $\{\psi_n^{(q)}\}_{q=1}^r$, rejecting H_0 by the multiple testing method based on $\{\psi_n^{(q)}\}_{q=1}^r$ is equivalent to $\hat{r}_n^* > r$ where $\hat{r}_n^*$ is based on $\{\psi_n^{(q)}\}_{q=1}^r$. In fact, Kleibergen and Paap (2006) relied on this relation for $r = k - 1$ to test for identification of the risk premia parameters in stochastic discount factor models. ■

The following theorem establishes that $\hat{r}_n^*$ is a good estimator for r_0.

Theorem 4.1. *For $\alpha \in (0,1)$, let $\psi_n^{(r)}$ be a test for the hypotheses* (1) *or* (2) *such that $\lim_{n\to\infty} P(\psi_n^{(r)} = 1) = \alpha$ when $r_0 = r$, and $\lim_{n\to\infty} P(\psi_n^{(r)} = 1) = 1$ when $r_0 > r$. Then $\lim_{n\to\infty} P(\hat{r}_n^* < r_0) = 0$,*

$$\lim_{n\to\infty} P(\hat{r}_n^* = r_0) = 1 - \alpha \text{ if } r_0 < k \text{ and } 1 \text{ if } r_0 = k \ ,$$

and

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

Theorem 4.1 implies that the true rank is correctly chosen with probability no smaller than $1-\alpha$ asymptotically, a smaller rank is chosen with probability going to zero, and a larger rank is chosen with probability no larger than α asymptotically. In short, $\{\hat{r}_n^*\}$ provides a confidence set for r_0 with asymptotic coverage probability no smaller than $1-\alpha$. Interestingly, Theorem 4.1 does not rely on the behavior of $\psi_n^{(q)}$ when $q > r_0$, since the sequential testing procedure carries out $\psi_n^{(q)}$ progressively from $q = 0$ to larger q and terminates before $q = r_0$ with probability no smaller than $1-\alpha$ asymptotically. That is, efficient rank determination does not require the ability of detecting whether $\text{rank}(\Pi_0)$ is strictly less than a hypothesized value. This explains why the hypotheses (2) has become prevalent.

However, the procedure crucially depends on the behavior of $\psi_n^{(q)}$ when $q < r_0$, that is, the power of detecting whether $\text{rank}(\Pi_0)$ is strictly greater than hypothesized values. In particular, the probability of ensuring a no smaller rank crucially depends on the probability of accepting $r_0 > q$ for $q = 0, \ldots, r_0 - 1$, which is the power of $\psi_n^{(q)}$ for $q = 0, \ldots, r_0 - 1$. This suggests that our tests may be leveraged for accuracy improvement in the sequential testing procedure for rank determination, provided the improved power property of our tests over existing rank tests as shown in Sections 2.2 and 4.1.

To show how the sequential testing procedure based on our tests can be more accurate than that based on existing rank tests, we focus on the case of the Kleibergen and Paap (2006) test and present some simulation evidence. We use the same DGP given in (21) and (22) with $\delta = 0.1$ and 0.12. The design of Π_0 implies that $r_0 = 6$ for both δ's and all $d = 1, \ldots, 6$. The Kleibergen and Paap (2006) test and our tests are implemented as in Section 2.1 and 4.1. The probability distributions of $\hat{r}_n^*$, which are based on $5,000$ simulation replications are reported in Figures 5 and 6. We use CF to denote the sequential testing procedure based on our tests and KP to denote the one based on the Kleibergen and Paap (2006) test. The result is based on $\kappa_n = n^{-1/4}$ and the derivative estimator in (38). The result for $\kappa_n = n^{-1/3}$ is similar and is available upon request. As shown in both figures, CF yields more accurate rank estimators than KP uniformly over $d = 1, \ldots, 6$ for both δ's. In particular, KP tends to underestimate the true rank when d

increases. The coverage probability of the resulting rank estimator is 5.46% when $d = 6$ and $\delta = 0.1$, and 25.3% when $d = 6$ and $\delta = 0.12$. The coverage probabilities of CF's rank estimator are greater than those of KP's rank estimator.

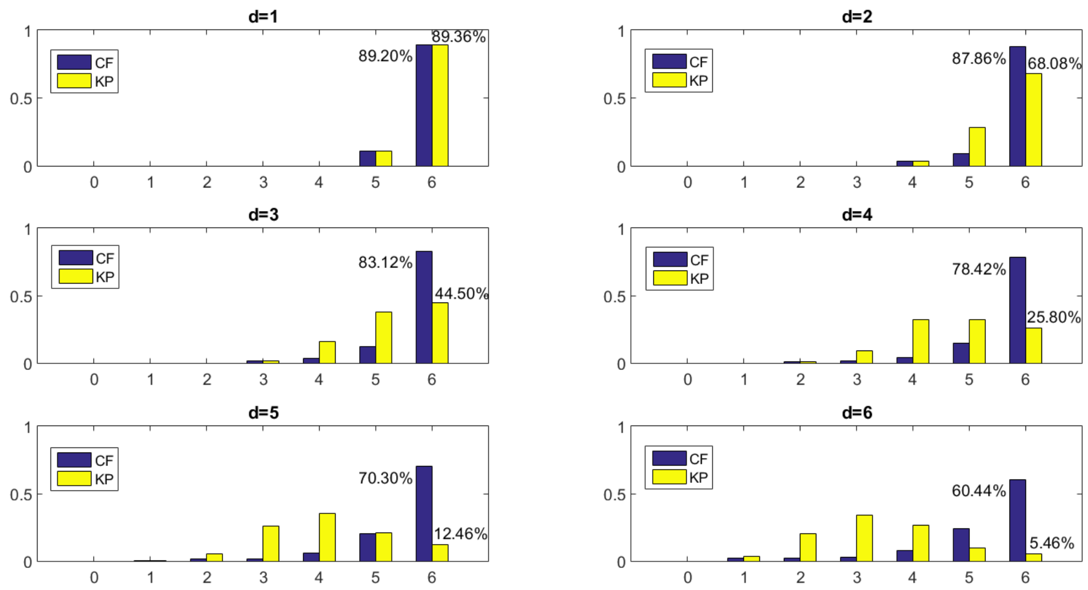

Figure 5: Comparison between the sequential testing procedures based on our tests and the Kleibergen and Paap (2006) test with $\alpha = 5\%$ and $\delta = 0.1$

Remark 4.2. To obtain a consistent estimator for r_0, Cragg and Donald (1997) and Robin and Smith (2000) make an adjustment dependent on n to the nominal level α. The consistency of $\hat{r}_n^*$ can be obtained when the adjusted nominal level $\alpha_n \to 0$ as $n \to \infty$ and satisfies certain rate requirement. In fact, the estimator $\hat{r}_n$ used in (38) provides a simple and consistent estimator for r_0, see Lemma A.6. ■

5 Conclusion

In this paper, we developed a more powerful method for examining a "no greater than" inequality of the rank of a matrix and a more accurate procedure for rank determination in a general setup. We proved that our tests have the asymptotic null rejection rate that is exactly equal to the nominal level regardless of whether the rank is less than or equal to the hypothesized value. Our simulation results showed that our tests are often more powerful than the multiple testing method, and improve the accuracy of the sequential testing procedure for rank determination. We illustrated our methods in several examples, including testing for identification and testing for the existence of stochastic trend and/or cointegration, to show the wide applicability of our methods.

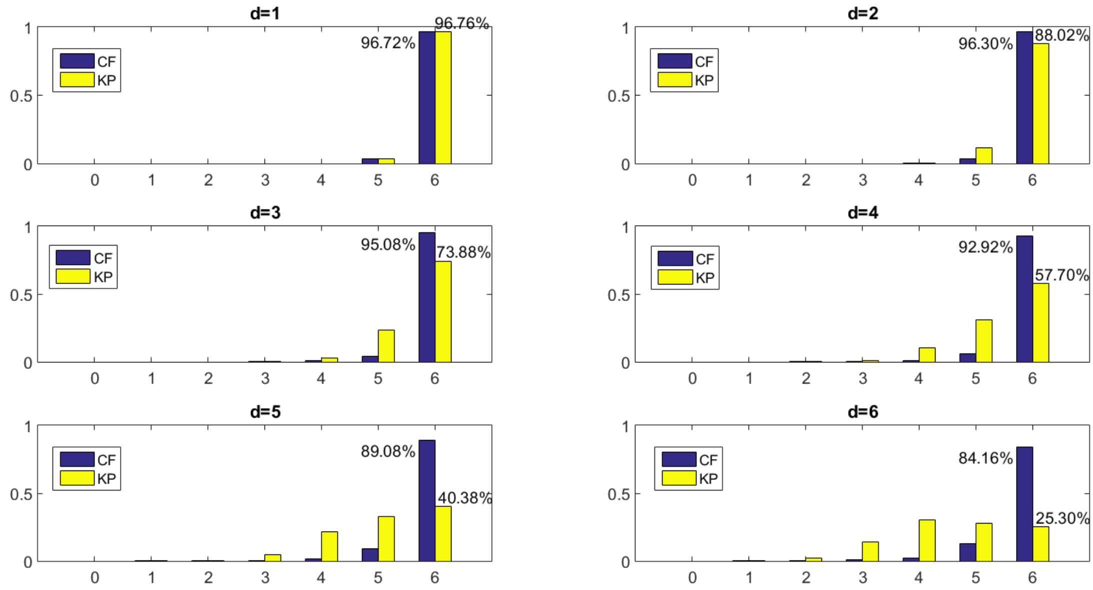

Figure 6: Comparison between the sequential testing procedures based on our tests and the Kleibergen and Paap (2006) test with $\alpha = 5\%$ and $\delta = 0.12$

Appendix A Proofs of Main Results

The following list includes notation and definitions that will be used in the appendix.

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

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

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

where the infimum is taken over all $u_1, \ldots, u_d \in \mathbf{R}^k$ that are orthonormal. Let $U \equiv [u_1, \ldots, u_d]$. Clearly, $U \in \mathbb{S}^{k\times d}$. By (A.1) and the definition of Frobenius norm, we further have

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

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

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

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

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

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

$$\begin{aligned}\sup_{U \in \mathbb{S}^{k\times d}} |g_n(U) - g(U)| &= \sup_{U \in \mathbb{S}^{k\times d}} |2\operatorname{tr}((\Pi U)^\intercal (M_n - M)U) + t_n \|M_n U\|^2| \\ &\le \sup_{U \in \mathbb{S}^{k\times d}} \{2\|\Pi U\| \|(M_n - M)U\| + t_n \|M_n U\|^2\} \ ,\end{aligned} \tag{A.3}$$

where the inequality follows by the triangle inequality and the Cauchy-Schwarz inequality for the trace operator. For the right hand side of (A.3), we have

$$\begin{aligned}&\sup_{U \in \mathbb{S}^{k\times d}} \{2\|\Pi U\| \|(M_n - M)U\| + t_n \|M_n U\|^2\} \\ &\le \sup_{U \in \mathbb{S}^{k\times d}} \{2\|\Pi\| \|U\| \|M_n - M\| \|U\| + t_n \|M_n\|^2 \|U\|^2\} = o(1) \ ,\end{aligned} \tag{A.4}$$

where we exploited the sub-multiplicativity of Frobenius norm and the fact that $\|U\| = \sqrt{d}$ and that $M_n \to M$ as well as $t_n \downarrow 0$ as $n \to \infty$. We thus conclude from (A.3) and (A.4) that $g_n \to g$ uniformly in $C(\mathbb{S}^{k\times d})$, or equivalently ϕ_1 is first order Hadamard directionally differentiable at Π with derivative $\phi'_{1,\Pi} : \mathbf{M}^{m\times k} \to C(\mathbb{S}^{k\times d})$ given by

$$\phi'_{1,\Pi}(M)(U) = 2\operatorname{tr}((\Pi U)^\intercal M U) \ . \tag{A.5}$$

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

$\phi'_{2,f} : C(\mathbb{S}^{k\times d}) \to \mathbf{R}$ given by

$$\phi'_{2,f}(h) = \min_{U\in\Psi(f)} h(U) , \tag{A.6}$$

where, by abuse of notation, $\Psi(f) \equiv \arg\min_{U\in\mathbb{S}^{k\times d}} f(U)$. Combining (A.5), (A.6) and the chain rule (Shapiro, 1990, Proposition 3.6), we may now conclude that $\phi : \mathbf{M}^{m\times k} \to \mathbf{R}$ is first order Hadamard directionally differentiable at any $\Pi \in \mathbf{M}^{m\times k}$ with the derivative $\phi'_\Pi : \mathbf{M}^{m\times k} \to \mathbf{R}$ given by

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

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

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

$$\begin{aligned}|\phi(\Pi + t_nM_n) - \phi(\Pi + t_nM)| \le |&\min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM_n)U\| - \min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM)U\|| \\ &\times (\min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM_n)U\| + \min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM)U\|) ,\end{aligned} \tag{A.7}$$

where the inequality follows by the formula $a^2 - b^2 = (a + b)(a - b)$. For the first term on the right hand side of (A.7), we have

$$|\min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM_n)U\| - \min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM)U\|| \le t_n\sqrt{d}\|M_n - M\| = o(t_n) , \tag{A.8}$$

where the inequality follows by the Lipschitz continuity of the infimum operator, the triangle inequality, the sub-multiplicativity of Frobenius norm and $\|U\| = \sqrt{d}$ for $U \in \mathbb{S}^{k\times d}$. For the second term on the right hand side of (A.7), we have

$$\begin{aligned}\min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM_n)U\| + \min_{U\in\mathbb{S}^{k\times d}} \|(\Pi + t_nM)U\| &\le \|(\Pi + t_nM_n)U^*\| \\ + \|(\Pi + t_nM)U^*\| \le t_n\|M_n\|\|U^*\| + t_n\|M\|\|U^*\| &= O(t_n) ,\end{aligned} \tag{A.9}$$

where the first inequality follows by letting U^* be an element from $\Psi(\Pi)$, and the second inequality follows by $\Pi U^* = 0$, the sub-multiplicativity of Frobenius norm and the fact that $\|U^*\| = \sqrt{d}$ and that $M_n \to M$ as $n \to \infty$. Combining (A.7)-(A.9), we thus obtain

$$|\phi(\Pi + t_nM_n) - \phi(\Pi + t_nM)| = o(t_n^2) . \tag{A.10}$$

Next, for $\epsilon > 0$, let $\Psi(\Pi)^\epsilon \equiv \{U \in \mathbb{S}^{k\times d} : \min_{U'\in\Psi(\Pi)} \|U' - U\| \le \epsilon\}$ and $\Psi(\Pi)^\epsilon_1 \equiv \{U \in \mathbb{S}^{k\times d} : \min_{U'\in\Psi(\Pi)} \|U' - U\| \ge \epsilon\}$. In what follows we consider the nontrivial case $\Pi \ne 0$ and $M \ne 0$. In this case, $\Psi(\Pi) \subsetneqq \mathbb{S}^{k\times d}$ in view of Proposition 1.3.4 in Tao (2012) and hence $\Psi(\Pi)_1^\epsilon \neq \emptyset$ for ϵ sufficiently small. Let $\sigma_{\min}^+(\Pi)$ denote the smallest positive singular value of Π which exists since $\Pi \neq 0$, and $\Delta \equiv 3\sqrt{2}[\sigma_{\min}^+(\Pi)]^{-1}\max_{U\in\mathbb{S}^{k\times d}}\|MU\| > 0$ since $M \neq 0$. Then it follows that for all n sufficiently large

$$\begin{aligned}\min_{U\in\Psi(\Pi)_1^{t_n\Delta}}\|(\Pi+t_nM)U\| &\geq \min_{U\in\Psi(\Pi)_1^{t_n\Delta}}\|\Pi U\| - t_n\max_{U\in\mathbb{S}^{k\times d}}\|MU\| \\ &\geq \frac{\sqrt{2}}{2}t_n\sigma_{\min}^+(\Pi)\Delta - t_n\max_{U\in\mathbb{S}^{k\times d}}\|MU\| > t_n\max_{U\in\mathbb{S}^{k\times d}}\|MU\| \\ &\geq \min_{U\in\Psi(\Pi)}\|(\Pi+t_nM)U\| \geq \sqrt{\phi(\Pi+t_nM)}\ ,\end{aligned} \tag{A.11}$$

where the first inequality follows by the Lipschitz continuity of the infimum operator, the triangle inequality and the fact that $\Psi(\Pi)_1^{t_n\Delta} \subset S^{k\times d}$, the second inequality follows by Lemma A.1, the third inequality follows by the definition of Δ, and the fourth inequality holds by the fact that $\Pi U = 0$ for $U \in \Psi(\Pi)$. By (A.11), we thus obtain that for all n sufficiently large

$$\phi(\Pi+t_nM) = \min_{U\in\Psi(\Pi)^{t_n\Delta}}\|(\Pi+t_nM)U\|^2\ . \tag{A.12}$$

Now, for fixed $U \in \Psi(\Pi)$, $\Delta > 0$ and $t \in \mathbf{R}$, let $\Gamma^\Delta \equiv \{V \in \mathbf{M}^{k\times d} : \|V\| \leq \Delta\}$ and $\Gamma_U^\Delta(t) \equiv \{V \in \Gamma^\Delta : U + tV \in \mathbb{S}^{k\times d}\} = \{V \in \Gamma^\Delta : V^\intercal U + U^\intercal V = -tV^\intercal V\}$. Define a correspondence $\varphi : \mathbf{R} \twoheadrightarrow \mathbb{S}^{k\times d}\times\Gamma^\Delta$ by $\varphi(t) = \{(U,V) : U \in \Psi(\Pi), V \in \Gamma_U^\Delta(t)\}$. Then the right hand side of (A.12) can be written as

$$\begin{aligned}\min_{U\in\Psi(\Pi)^{t_n\Delta}}\|(\Pi+t_nM)U\|^2 &= \min_{(U,V)\in\varphi(t_n)}\|(\Pi+t_nM)(U+t_nV)\|^2 \\ &= t_n^2\min_{(U,V)\in\varphi(t_n)}\|\Pi V + MU\|^2 + o(t_n^2)\ ,\end{aligned} \tag{A.13}$$

where the second equality follows by the fact that $\Pi U = 0$ for all $U \in \Psi(\Pi)$ and $\|MV\| \leq \|M\|\Delta$ for all $V \in \Gamma^\Delta$. By Lemma A.2, $\varphi(t)$ is continuous at $t = 0$. Moreover, φ is obviously compact-valued. We may then obtain by Theorem 17.31 in Aliprantis and Border (2006) that

$$\begin{aligned}\min_{(U,V)\in\varphi(t_n)}\|\Pi V + MU\|^2 &= \min_{(U,V)\in\varphi(0)}\|\Pi V + MU\|^2 + o(1) \\ &= \min_{U\in\Psi(\Pi)}\min_{V\in\mathbf{M}^{k\times d}}\|\Pi V + MU\|^2 + o(1)\ ,\end{aligned} \tag{A.14}$$

where the second equality holds by letting Δ sufficiently large in view of Lemma A.3. Combining (A.10), (A.12), (A.13) and (A.14) then yields part (ii) of the proposition. ■

Proof of Proposition 3.2 Recall that $d = k - r$ and let $d^* \equiv k - r^*$. Noting that the column vectors in Q_2 form a orthonormal basis for the null space of Π_0, we may rewrite

$\Psi(\Pi)$ as $\Psi(\Pi) = \{Q_2V : V \in \mathbb{S}^{d^*\times d}\}$. This together with the projection theorem implies

$$\phi''_{\Pi}(M) = \min_{V\in\mathbb{S}^{d^*\times d}} \|(I - \Pi(\Pi^\intercal\Pi)^-\Pi^\intercal)MQ_2V\|^2 \ , \tag{A.15}$$

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

$$\begin{aligned}(I - \Pi(\Pi^\intercal\Pi)^-\Pi^\intercal)P &= P - P\Sigma Q^\intercal(Q\Sigma^\intercal P^\intercal P\Sigma Q^\intercal)^- Q\Sigma^\intercal P^\intercal P \\ = P - P\Sigma Q^\intercal Q(\Sigma^\intercal P^\intercal P\Sigma)^- Q^\intercal Q\Sigma^\intercal P^\intercal P &= P - P\Sigma(\Sigma^\intercal\Sigma)^-\Sigma^\intercal = [0, P_2] \ ,\end{aligned} \tag{A.16}$$

where the second equality exploited Theorem 20.5.6 in Harville (2008), the third equality follows from P and Q being orthonormal, and the fourth equality is obtained by carrying out the Moore-Penrose inverse by Exercise 2.7.4 in Magnus and Neudecker (2007) and noting that Σ is diagonal. In view of (A.16), we have

$$\begin{aligned}\min_{V\in\mathbb{S}^{d^*\times d}} \|(I - \Pi(\Pi^\intercal\Pi)^-\Pi^\intercal)MQ_2V\|^2 &= \min_{V\in\mathbb{S}^{d^*\times d}} \|[0, P_2]P^\intercal MQ_2V\|^2 \\ = \min_{V\in\mathbb{S}^{d^*\times d}} \|P_2P_2^\intercal MQ_2V\|^2 = \min_{V\in\mathbb{S}^{d^*\times d}} \|P_2^\intercal MQ_2V\|^2 &= \sum_{j=r-r^*+1}^{k-r^*} \sigma_j^2(P_2^\intercal MQ_2) \ ,\end{aligned} \tag{A.17}$$

where the third equality follows from $P_2^\intercal P_2 = I_{m-r^*}$ and the final equality follows from Lemma 3.1. Combining (A.15) and (A.17) concludes the proof of the proposition. ■

Proof of Proposition 3.3: The first and second results are straightforward application of Theorem 2.1 in Fang and Santos (2015) and Chen and Fang (2015) by noting that $\phi'_{\Pi_0} = 0$ under H_0, respectively. In particular, Assumptions 2.1(i)-(ii) are satisfied in view of Proposition 3.1 and Assumption 2.2 is satisfied by Assumption 3.1. ■

Proof of Theorem 3.1: By Lemma A.5 and the maintained assumptions, each of the two derivative estimators are consistent for ϕ''_{Π_0} in the sense that they satisfy Assumption 3.4 in Chen and Fang (2015). This, together with Lemma A.2 in Chen and Fang (2015), Assumption 3.2, Proposition 3.3, and the cdf of the limit distribution being strictly increasing at its $1-\alpha$ quantile $c_{1-\alpha}$, implies that $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$, following exactly the same proof of Corollary 3.2 in Fang and Santos (2015). Then under H_0, the first claim of the theorem follows from combining Proposition 3.3, Slutsky's lemma, $c_{1-\alpha}$ being a continuity point of the limit distribution and the portmanteau theorem.

For the second part of the theorem, let $\mathbb{G}_n^* \equiv \tau_n\{\hat{\Pi}_n^* - \hat{\Pi}_n\}$. By the definition of $\hat{c}_{1-\alpha}$, if $P_W(\hat{\phi}''_n(\mathbb{G}_n^*) \le \tau_n^2\phi(\hat{\Pi}_n)) \ge 1-\alpha$, then we must have $\hat{c}_{1-\alpha} \le \tau_n^2\phi(\hat{\Pi}_n)$ and hence

$$P(\tau_n^2\phi(\hat{\Pi}_n) \ge \hat{c}_{1-\alpha}) \ge P_X(P_W(\hat{\phi}''_n(\mathbb{G}_n^*) \le \tau_n^2\phi(\hat{\Pi}_n)) \ge 1-\alpha) \ . \tag{A.18}$$

We shall show that the right side of (A.18) tends to one as $n \to \infty$ for each of the two derivative estimators. First, consider the numerical estimator (39). Note that

$$
\begin{aligned}
P_W(\hat{\phi}''_n(\mathbb{G}^*_n) \leq \tau_n^2\phi(\hat{\Pi}_n)) &= P_W(\frac{\phi(\hat{\Pi}_n + \kappa_n\tau_n\{\hat{\Pi}^*_n - \hat{\Pi}_n\}) - \phi(\hat{\Pi}_n)}{\kappa_n^2} \leq \tau_n^2\phi(\hat{\Pi}_n)) \\
&\geq P_W(\frac{\phi(\hat{\Pi}_n + \kappa_n\tau_n\{\hat{\Pi}^*_n - \hat{\Pi}_n\})}{\kappa_n^2} \leq \tau_n^2\phi(\hat{\Pi}_n)) \\
&= P_W(\phi(\hat{\Pi}_n + \kappa_n\tau_n\{\hat{\Pi}^*_n - \hat{\Pi}_n\}) \leq (\kappa_n\tau_n)^2\phi(\hat{\Pi}_n)) \; . \qquad \text{(A.19)}
\end{aligned}
$$

Since $\hat{\Pi}_n \xrightarrow{p} \Pi_0$ and ϕ is continuous at Π_0, the continuous mapping theorem implies that: under H_1,

$$\phi(\hat{\Pi}_n) \xrightarrow{p} \phi(\Pi_0) > 0 \; . \qquad \text{(A.20)}$$

By Assumptions 3.1 and 3.2, together with the assumption that $\kappa_n = o(1)$ as $n \to \infty$ and continuity of ϕ, we have $\phi(\hat{\Pi}_n + \kappa_n\tau_n\{\hat{\Pi}^*_n - \hat{\Pi}_n\}) = O_{P_W}(1)$ with probability approaching one. Consequently, by $\kappa_n\tau_n \to \infty$, with probability approaching one,

$$P_W(\phi(\hat{\Pi}_n + \kappa_n\tau_n\{\hat{\Pi}^*_n - \hat{\Pi}_n\}) \leq (\kappa_n\tau_n)^2\phi(\hat{\Pi}_n)) \to 1 > 1 - \alpha \; . \qquad \text{(A.21)}$$

By the dominated convergence theorem, we may conclude from results (A.19), (A.20) and (A.21) that

$$P_X(P_W(\hat{\phi}''_n(\mathbb{G}^*_n) \leq \tau_n^2\phi(\hat{\Pi}_n)) \geq 1 - \alpha) \to 1 \; . \qquad \text{(A.22)}$$

This implies the second claim of the theorem holds when $\hat{\phi}''_n$ is the numerical derivative estimator. Second, consider the derivative estimator (38). Recall that $\hat{d}_n = k - \hat{r}_n$ and $d = k - r$. By Lemma 3.1, we have

$$
\begin{aligned}
P_W(\hat{\phi}''_n(\mathbb{G}^*_n)) \leq \tau_n^2\phi(\hat{\Pi}_n) &= P_W(\min_{U \in \mathbb{S}^{\hat{d}_n \times d}} \|\hat{P}^\intercal_{2,n}\tau_n\{\hat{\Pi}^*_n - \hat{\Pi}_n\}\hat{Q}_{2,n}U\|^2 \leq \tau_n^2\phi(\hat{\Pi}_n)) \\
&\geq P_W(\|\tau_n\{\hat{\Pi}^*_n - \hat{\Pi}_n\}\|^2 mkd \leq \tau_n^2\phi(\hat{\Pi}_n)) \; ,
\end{aligned}
$$

where the second inequality exploited $\|\hat{P}^\intercal_{2,n}\|^2\|\hat{Q}_{2,n}\|^2 \leq mk$ and $\|U\|^2 = d$. The second claim of the theorem then follows by analogous arguments as above. ■

Proof of Theorem 4.1: We prove the results for three different cases: when $r_0 = k$, when $1 \leq r_0 \leq k - 1$ and when $r_0 = 0$. It suffices to show the first two results. First, we show the second result. When $r_0 = k$, we have

$$\lim_{n\to\infty} P(\hat{r}^*_n = r_0) = \lim_{n\to\infty} P(\psi^{(0)}_n = 1, \cdots, \psi^{(k-1)}_n = 1) \geq 1 - \sum_{q=0}^{k-1}(1 - \lim_{n\to\infty} P(\psi^{(q)}_n = 1)) = 1 \; ,$$

where the inequality follows by the Boole's inequality. When $1 \leq r_0 \leq k-1$, we have

$$\lim_{n\to\infty} P(\hat{r}_n^* = r_0) = \lim_{n\to\infty} P(\psi_n^{(0)} = 1, \cdots, \psi_n^{(r_0-1)} = 1, \psi_n^{(r_0)} = 0)$$
$$\leq 1 - \lim_{n\to\infty} P(\psi_n^{(r_0)} = 1) = 1 - \alpha \ , \tag{A.23}$$

where the inequality follows by the fact that $P(A) \leq P(B)$ for $A \subset B$, and

$$\lim_{n\to\infty} P(\hat{r}_n^* = r_0) = \lim_{n\to\infty} P(\psi_n^{(0)} = 1, \cdots, \psi_n^{(r_0-1)} = 1, \psi_n^{(r_0)} = 0)$$
$$\geq 1 - \sum_{q=0}^{r_0-1}(1 - \lim_{n\to\infty} P(\psi_n^{(q)} = 1)) - \lim_{n\to\infty} P(\psi_n^{(r_0)} = 1) = 1 - \alpha \ , \tag{A.24}$$

where the inequality follows by the Boole's inequality. Combining results (A.23) and (A.24) gives the result when $1 \leq r_0 \leq k-1$. When $r_0 = 0$, we have

$$\lim_{n\to\infty} P(\hat{r}_n^* = r_0) = \lim_{n\to\infty} P(\psi_n^{(0)} = 0) = 1 - \lim_{n\to\infty} P(\psi_n^{(0)} = 1) = 1 - \alpha \ .$$

Next, we show the first result. When $r_0 = k$, we have

$$\lim_{n\to\infty} P(\hat{r}_n^* < r_0) \leq \sum_{q=0}^{k-1}(1 - \lim_{n\to\infty} P(\psi_n^{(q)} = 1)) = 0 \ ,$$

where the inequality holds by the Boole's inequality. When $1 \leq r_0 \leq k-1$, we have

$$\lim_{n\to\infty} P(\hat{r}_n^* < r_0) \leq \sum_{q=0}^{r_0-1}(1 - \lim_{n\to\infty} P(\psi_n^{(q)} = 1)) = 0 \ ,$$

where the inequality holds by the Boole's inequality. When $r_0 = 0$, obviously $P(\hat{r}_n^* < r_0) = 0$. This completes the proof of the theorem. ■

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

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

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

$$\|\Pi U\| = \|P\Sigma Q^\intercal U\| = \|\Sigma U_Q\| \geq \sigma_{\min}^+(\Pi)\|U_Q^{(1)}\| \ , \tag{A.25}$$

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

$$\|U_Q^{(2)}\|^2 = \sum_{j=1}^{d} \sigma_j^2(U_Q^{(2)}) \leq \sum_{j=1}^{d} \sigma_j(U_Q^{(2)}) = \mathrm{tr}([I_d, \mathbf{0}_{r-r^*}]\Sigma_U^{(2)}) \ , \tag{A.26}$$

where the inequality follows by the fact that $\sigma_j(U_Q^{(2)}) \in [0, 1]$ as singular values of $U_Q^{(2)}$ due to $U_Q^{(2)\intercal}U_Q^{(2)} + U_Q^{(1)\intercal}U_Q^{(1)} = I_d$, and the second equality follows by noting that the diagonal entries of $\Sigma_U^{(2)}$ are singular values of $U_Q^{(2)}$. Since $\|U_Q^{(1)}\|^2 + \|U_Q^{(2)}\|^2 = \|U_Q\|^2 = d$, thus combining (A.25) and (A.26) yields that for $U \in \mathbb{S}^{k\times d}$,

$$\|\Pi U\| \geq \sigma^+_{\min}(\Pi)\sqrt{d - \mathrm{tr}([I_d, \mathbf{0}_{r-r^*}]\Sigma_U^{(2)})} \ . \tag{A.27}$$

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

$$2(d - \mathrm{tr}([I_d, \mathbf{0}_{d-r^*}]\Sigma_U^{(2)})) = \|U_Q^{(1)}\|^2 + \|\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r^*}]^\intercal\|^2 \ . \tag{A.28}$$

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

$$\|U_Q^{(1)}\|^2 + \|\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r^*}]^\intercal\|^2 = \|Q_1U_Q^{(1)} + Q_2P_U^{(2)}(\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r^*}]^\intercal)Q_U^{(2)\intercal}\|^2 \ . \tag{A.29}$$

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

$$|Q_1U_Q^{(1)} + Q_2P_U^{(2)}(\Sigma_2 - [I_d, \mathbf{0}_{r-r^*}]^\intercal)Q_U^{(2)\intercal}\|^2 = \|U - Q_2P_U^{(2)}[I_d, \mathbf{0}_{r-r^*}]^\intercal Q_U^{(2)\intercal}\|^2 \ . \tag{A.30}$$

Clearly, $Q_2P_U^{(2)}[I_d, \mathbf{0}_{r-r^*}]^\intercal Q_U^{(2)\intercal} \in \Psi(\Pi)$, so combining (A.28)- (A.30) yields that for $U \in \mathbb{S}^{k\times d}$,

$$2(d - \mathrm{tr}([I_d, \mathbf{0}_{r-r^*}]\Sigma_U^{(2)})) \geq \min_{U' \in \Psi(\Pi)} \|U - U'\|^2 \ . \tag{A.31}$$

Since $\Pi \neq 0$, then $\Psi(\Pi)_1^\epsilon \neq \emptyset$ for all sufficiently small $\epsilon > 0$ by Proposition 1.3.4 in Tao (2012). Fix such an ϵ. By the definition of $\Psi(\Pi)_1^\epsilon$, combining (A.27) and (A.31) yields that for all $U \in \Psi(\Pi)_1^\epsilon$,

$$\|\Pi U\| \geq \frac{\sqrt{2}}{2}\sigma_{\min}^{+}(\Pi) \min_{U' \in \Psi(\Pi)} \|U - U'\| \geq \frac{\sqrt{2}}{2}\sigma_{\min}^{+}(\Pi)\epsilon \ . \tag{A.32}$$

Then the lemma follows by applying minimum over $\Psi(\Pi)_1^\epsilon$ to both sides of (A.32) and noting that the result continues to hold for all sufficiently small $\epsilon > 0$. ■

Lemma A.2. *Let the correspondence φ be as in the proof of Proposition 3.1. Then $\varphi(t)$ is continuous at $t = 0$.*

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

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

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

$$V_{1n}^\intercal U_0 + U_0^\intercal V_{1n} = (V_0 - \frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n})^\intercal U_0 + U_0^\intercal (V_0 - \frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n}) = -t_n V_{1n}^\intercal V_{1n} \ . \tag{A.33}$$

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

$$\|V_{1n} - V_0\| = \|\frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n}\| \leq \frac{t_n}{2}\sqrt{d}\Delta^2 \ . \tag{A.34}$$

In case ii), since $U_0 \in \Psi(\Pi)$, $\lambda_n^2 V_0 \in \bar{\varphi}(0)$ and $\lambda_n V_{2n} = \lambda_n^2 f(V_{2n})$, then by analogous calculations as in (A.33), we have

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

This together with $\lambda_n V_{2n} \in \Gamma^\Delta$ due to $\lambda_n \in (0, 1)$ and $V_{2n} \in \Gamma^\Delta$ implies that $\lambda_n V_{2n} \in \bar{\varphi}(t_n)$. Moreover, since $\lambda_n V_{2n} = \lambda_n^2 f(V_{2n})$, then by analogous calculations as in (A.34), we have

$$\|\lambda_n V_{2n} - V_0\| \leq \|\lambda_n^2 f(V_{2n}) - \lambda_n^2 V_0\| + |\lambda_n^2 - 1|\|V_0\| \leq \frac{t_n}{2}\sqrt{d}\Delta^2 + |\lambda_n^2 - 1|\Delta \ , \tag{A.35}$$

where the first inequality follows the triangle inequality and the second inequality follows since $\lambda_n \in (0,1)$. Now, for each $n \in \mathbf{N}$, define V_n^* to be V_{1n} if case (i) happens and $\lambda_n V_{2n}$ otherwise. Let $\delta_n \equiv 1$ if case (i) happens and $\delta_n \equiv \lambda_n$ otherwise. Then $V_n^* \in \Gamma_{U_0}^{\Delta}(t_n)$ for all $n \in \mathbf{N}$, and combination of (A.34) and (A.35) yields

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

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

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

Lemma A.3. *Suppose* $\Pi \in \mathbf{M}^{m\times k}$ *with* $\Pi \neq 0$ *and rank*$(\Pi) \le r$*, and* $M \in \mathbf{M}^{m\times k}$ *with* $M \neq 0$*. Let* $\Psi(\Pi)$ *given in the proof of Proposition 3.1. For* $U \in \Psi(\Pi)$ *and* $\Delta > 0$*, let* $\Gamma_U^{\Delta}(0)$ *be as in the proof of Proposition 3.1. Recall that* $d = k - r$*. When* Δ *is sufficiently large, then for all* $U \in \Psi(\Pi)$*,*

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

Proof: Recall that $\Pi = P\Sigma Q^{\intercal}$ is a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$ and $Q \in \mathbb{S}^{k\times k}$ are orthonormal, and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Recall that $r^* = \text{rank}(\Pi) < r$. We may write $\Sigma = [\Sigma_1, 0]$ such that $\Sigma_1 \in \mathbf{M}^{m\times r^*}$ is of full rank with $r^* < r$. It follows that

$$\min_{V \in \mathbf{M}^{k\times d}} \|\Pi V + MU\|^2 = \min_{V \in \mathbf{M}^{r^*\times d}} \|[P\Sigma_1 V + MU\|^2 \ . \tag{A.36}$$

By the projection theorem, the minimum on the right hand side of (A.36) is attained at some point, say $V_1^* \in \mathbf{M}^{r^*\times d}$. Moreover, V_1^* is uniformly bounded over $U \in \Psi(\Pi)$. Let $V^* \equiv Q[V_1^{*\intercal}, 0]^{\intercal} \in \mathbf{M}^{k\times d}$, then the minimum on the left hand side of (A.36) is attained at V^*. Recall that $Q = [Q_1, Q_2]$, where $Q_1 \in \mathbf{M}^{k\times r^*}$. Then $V^* = Q_1 V_1^* \in \Gamma_U^{\Delta}(0)$ for all $U \in \Psi(\Pi)$, when Δ is sufficiently large. It implies that the minimum on the right hand side of (A.36) is attained within $\Gamma_U^{\Delta}(0)$ as well for all $U \in \Psi(\Pi)$, when Δ is sufficiently large. This implies that when Δ is sufficiently large,

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

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

Lemma A.4. *Suppose rank*$(\Pi) \leq r$ *and let* $\phi''_\Pi : \mathbf{M}^{m\times k} \to \mathbf{R}$ *be given in Proposition 3.1. If rank*$(\Pi) = r$, *there exists a bilinear map* $\Phi''_\Pi : \mathbf{M}^{m\times k} \times \mathbf{M}^{m\times k} \to \mathbf{R}$ *such that* $\phi''_\Pi(M) = \Phi''_\Pi(M, M)$ *for all* $M \in \mathbf{M}^{m\times k}$*; if rank*$(\Pi) < r$, *such a* Φ''_Π *does not exist.*

PROOF: Recall that $\Pi = P\Sigma Q^\intercal$ is a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$ and $Q \in \mathbb{S}^{k\times k}$ are orthonormal, and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Recall that $d = k - r$. If $\text{rank}(\Pi) = r$, then Proposition 3.2 and Lemma 3.1 imply

$$\phi''_\Pi(M) = \min_{V\in\mathbb{S}^{d\times d}} \|P_2^\intercal M Q_2 V\|^2 = \|P_2^\intercal M Q_2\|^2 ,$$

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

Next, suppose that $\text{rank}(\Pi) < r_0$ and assume that there exists a bilinear map Φ''_Π corresponding to ϕ''_Π. In turn, bilinearity of Φ''_Π implies that

$$\phi''_\Pi(M_1) + \phi''_\Pi(M_2) = \frac{\phi''_\Pi(M_1 + M_2) + \phi''_\Pi(M_1 - M_2)}{2} \tag{A.37}$$

for all $M_1 \in \mathbf{M}^{m\times k}$ and $M_2 \in \mathbf{M}^{m\times k}$. Recall that $r^* = \text{rank}(\Pi)$. If $M = P_2 H Q_2^\intercal$ for some $H \in \mathbf{M}^{(m-r^*)\times(k-r^*)}$, then Proposition 3.2 and Lemma 3.1 imply

$$\phi''_\Pi(M) = \sigma^2_{r-r^*+1}(H) + \cdots + \sigma^2_{k-r^*}(H) . \tag{A.38}$$

Now, let $H_1 \in \mathbf{M}^{(m-r^*)\times(k-r^*)}$ be diagonal with the (j, j)th entry equal to 1 for $j = 1, \ldots, k - r^*$ and $H_2 \in \mathbf{M}^{(m-r^*)\times(k-r^*)}$ be diagonal with the (j, j)th entry equal to -1 for $j = 1$ and 1 for $j = 2, \ldots, k - r^*$. Set $M_i = P_2 H_i Q_2^\intercal$ for $i = 1, 2$, the result in (A.38) implies $\phi''_\Pi(M_1) = \phi''_\Pi(M_2) = k - r$, $\phi''_\Pi(M_1 + M_2) = 4(k - r) - 4$ and $\phi''_\Pi(M_1 - M_2) = 0$. It follows that

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

which contradicts the result (A.37). Thus, the second result of the lemma follows. ■

Lemma A.5. *Suppose Assumption 3.1 holds,* $\kappa_n \downarrow 0$ *and* $\tau_n\kappa_n \to \infty$*. Let* $\hat{\phi}''_n$ *be constructed as in* (39) *or* (38)*. Then we have under* H_0*,*

$$\hat{\phi}''_n(M_n) \xrightarrow{p} \phi''_{\Pi_0}(M)$$

whenever $M_n \to M$ *as* $n \to \infty$ *for* $\{M_n\} \subset \mathbf{M}^{m\times k}$ *and* $M \in \mathbf{M}^{m\times k}$*.*

Proof: When $\hat{\phi}''_n$ is constructed as in (39), the result of the lemma follows by Proposition 3.1 of Chen and Fang (2015). Next we consider the derivative estimator (38). Recall that $d = k - r$ and let $\hat{d}_n \equiv k - \hat{r}_n$. By Lemma 3.1, we have

$$|\hat{\phi}''_n(M_n) - \hat{\phi}''_n(M)| \le |\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^{\intercal} M_n \hat{Q}_{2,n} U\| - \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n} U\|| \\ \times (\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^{\intercal} M_n \hat{Q}_{2,n} U\| + \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n} U\|) \ , \tag{A.39}$$

where the inequality follows by the formula $(a^2 - b^2) = (a+b)(a-b)$. For the first term on the right hand side of (A.39), we have

$$|\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^{\intercal} M_n \hat{Q}_{2,n} U\| - \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n} U\|| \le \sqrt{kmd}\|M_n - M\| = o_p(1) \ , \tag{A.40}$$

where the inequality follows by the Lipschitz continuity of the infimum operator, the triangle inequality and $\|\hat{P}_{2,n}\| \le \sqrt{m}$, $\|\hat{Q}_{2,n}\| \le \sqrt{k}$ and $\|U\| = \sqrt{r}$ for all $U \in \mathbb{S}^{\hat{d}_n\times d}$, and the equality follows since $M_n \to M$. For the second term on the right hand side of (A.39), we have

$$\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^{\intercal} M_n \hat{Q}_{2,n} U\| + \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \|\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n} U\| \le \sqrt{kmd}\|M_n\| + \sqrt{kmd}\|M\| \ , \tag{A.41}$$

where the inequality follows by the sub-multiplicability of the Frobenius norm, $\|\hat{P}_{2,n}\| \le \sqrt{m}$, $\|\hat{Q}_{2,n}\| \le \sqrt{k}$ and $\|U\| = \sqrt{r}$ for all $U \in \mathbb{S}^{\hat{d}_n\times d}$. Combining (A.39)-(A.41), then we obtain

$$|\hat{\phi}''_n(M_n) - \hat{\phi}''_n(M)| = o_p(1) \ . \tag{A.42}$$

Recall that $\phi''_{\Pi_0}(M) = \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^{\intercal} M Q_{0,2})$. By (A.42), Lemma 3.1 and A.6, it suffices to show that given $\hat{r}_n = r_0$,

$$|\sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2(\hat{P}_{2,n}^{\intercal} M \hat{Q}_{2,n}) - \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(P_{0,2}^{\intercal} M Q_{0,2})| = o_p(1) \ . \tag{A.43}$$

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

$$[\hat{\Pi}_n^{\intercal}\hat{\Pi}_n - \Pi_0^{\intercal}\Pi_0 - (\sigma_{r_0+j}^2(\hat{\Pi}_n) - \sigma_{r_0+j}^2(\Pi_0))I_k + \Pi_0^{\intercal}\Pi_0 - \sigma_{r_0+j}^2(\Pi_0)I_k]Q_0\hat{u}_j \\ = [\hat{\Pi}_n^{\intercal}\hat{\Pi}_n - \sigma_{r_0+j}^2(\hat{\Pi}_n)I_k]\hat{q}_j = 0 \ . \tag{A.44}$$

Noting that $\|\hat{\Pi}_n^{\intercal}\hat{\Pi}_n - \Pi_0^{\intercal}\Pi_0\| = o_p(1)$ and $|\sigma_{r_0+j}^2(\hat{\Pi}_n) - \sigma_{r_0+j}^2(\Pi_0)| = o_p(1)$ by the continuous mapping theorem, the Weyl inequality (Tao, 2012, Exercise 1.3.22(iv)) and

Assumption 3.1, we then conclude from (A.44) that

$$o_p(1) = [\Pi_0^\intercal \Pi_0 - \sigma_{r_0+j}^2(\Pi_0) I_k] Q_0 \hat{u}_j = Q_0 \Sigma_0^\intercal \Sigma_0 \hat{u}_j \ , \tag{A.45}$$

where we exploited the singular value decomposition $\Pi_0 = P_0 \Sigma_0 Q_0^\intercal$, and the fact that $\sigma_{r_0+j}^2(\Pi_0) = 0$. Since the first r_0 diagonal elements of the diagonal matrix $\Sigma_0^\intercal \Sigma_0$ are positive and Q_0 being nonsingular, we may conclude from result (A.45) that the first r_0 elements of $\hat{u}_j$ are $o_p(1)$ and moreover by the definition of $\hat{q}_j$ that for some random $U_2 \in \mathbb{S}^{(k-r_0)\times(k-r_0)}$,

$$\hat{Q}_{2,n} = Q_{0,2} U_2 + o_p(1) \ , \tag{A.46}$$

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

$$\hat{P}_{2,n} = P_{0,2} V_2 + o_p(1) \ . \tag{A.47}$$

Combining results (A.46) and (A.47) and the continuous mapping theorem yields that given $\hat{r}_n = r_0$,

$$\|\hat{P}_{2,n}^\intercal M \hat{Q}_{2,n} - V_2^\intercal P_{0,2}^\intercal M Q_{0,2} U_2\| = o_p(1) \ . \tag{A.48}$$

Thus, (A.43) is obtained by (A.48), the continuous mapping theorem and the fact that the singular values of $V_2^\intercal P_{0,2}^\intercal M Q_{0,2} U_2$ are equal to those of $P_{0,2}^\intercal M Q_{0,2}$. This completes the proof of the lemma. ■

Lemma A.6. *Suppose Assumption 3.1 holds, $\kappa_n \downarrow 0$ and $\tau_n \kappa_n \to \infty$. Let $\hat{r}_n = \min\{r, \#\{1 \le j \le k : \sigma_j(\hat{\Pi}_n) \ge \kappa_n\}\}$. Then we have under* H_0*,*

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

Proof: Noting that $\hat{r}_n > r_0$ implies $\sigma_{r_0+1}(\hat{\Pi}_n) \ge \kappa_n$ and that $\sigma_{r_0+1}(\Pi_0) = 0$, we then have

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n > r_0) &\le \limsup_{n\to\infty} P(|\sigma_{r_0+1}(\hat{\Pi}_n) - \sigma_{r_0+1}(\Pi_0)| \ge \kappa_n) \\ &\le \limsup_{n\to\infty} P(\|\tau_n(\hat{\Pi}_n - \Pi_0)\| \ge \tau_n \kappa_n) = 0 \ ,\end{aligned} \tag{A.49}$$

where the first inequality follows by $P(A) \le P(B)$ for $A \subset B$, the second inequality follows by the Weyl inequality (Tao, 2012, Exercise 1.3.22(iv)), and the equality follows by Assumption 3.1 and $\tau_n \kappa_n \to \infty$. Noting that $\hat{r}_n < r_0$ implies $\sigma_{r_0}(\hat{\Pi}_n) < \kappa_n$, we then have

$$\limsup_{n\to\infty} P(\hat{r}_n < r_0) \le \limsup_{n\to\infty} P(|\sigma_{r_0}(\hat{\Pi}_n) - \sigma_{r_0}(\Pi_0)| > -\kappa_n + \sigma_{r_0}(\Pi_0))$$

$$\leq \limsup_{n\to\infty} P(\|\tau_n(\hat{\Pi}_n - \Pi_0)\| \geq \tau_n \sigma_{r_0}(\Pi_0)(1 - \kappa_n/\sigma_{r_0}(\Pi_0)) = 0 \ , \tag{A.50}$$

where the first inequality follows by $P(A) \leq P(B)$ for $A \subset B$, the second inequality follows by the Weyl inequality (Tao, 2012, Exercise 1.3.22(iv)), and the equality follows by Assumption 3.1, $\sigma_{r_0}(\Pi_0) > 0$, $\tau_n \to \infty$ and $\kappa_n \downarrow 0$. Combining (A.49) and (A.50) yields

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

This completes the proof of the lemma by noting that $\lim_{n\to\infty} P(\hat{r}_n = r_0) = 1 - \lim_{n\to\infty} P(\hat{r}_n \neq r_0) = 1$. ■

Appendix B Results for Examples 2.1-2.7

Example 2.2 (Continued). Suppose $\{Y_t\}_{t=1}^n$ is a sequence of data from Example 2.2. Let $\hat{\Pi}_n$ be the least squares estimator

$$\hat{\Pi}_n = \frac{1}{n}\sum_{t=2}^{n} \Delta Y_t Y_{t-1}^\intercal (\frac{1}{n}\sum_{t=2}^{n} Y_{t-1}Y_{t-1}^\intercal)^{-1} \ . \tag{B.1}$$

Let $D_n \equiv \operatorname{diag}(\sqrt{n}\mathbf{1}_{r_0}, n\mathbf{1}_{k-r_0})$ and $B_0 \equiv [Q_{0,1}, P_{0,2}]^\intercal$, where r_0, $Q_{0,1}$ and $P_{0,2}$ are given in Proposition 3.3. By Lemma A.2 of Liao and Phillips (2015), if Φ_0 has eigenvalues on or inside the unit circle, then

$$(\hat{\Pi}_n - \Pi_0)B_0^{-1}D_nB_0 \xrightarrow{L} \mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2 \ , \tag{B.2}$$

where $\mathcal{M}_1 \in \mathbf{M}^{k\times k}$ with $\operatorname{vec}(\mathcal{M}_1) \sim N(0, \Sigma \otimes (Q_{0,1}\Sigma_1^{-1}Q_{0,1}^\intercal))$ and $\Sigma_1 \equiv \operatorname{Var}(Q_{0,1}^\intercal Y_t)$, and $\mathcal{M}_2 \in \mathbf{M}^{k\times k}$ with

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

and $B_k(t)$ is a $k\times 1$ Brownian motion defined on the unit interval with identity covariance matrix at time t. Given that Assumption 3.1 is not satisfied since the rates in D_n are not homogenous unless $r_0 = 0$ or $r_0 = k$, we extend Proposition 3.3 to accommodate this case. Next we focus on the nontrivial case of testing for the existence of stochastic trend. By Proposition B.2, the asymptotic distribution of $n^2\phi(\hat{\Pi}_n)$ under H_0 is given by

$$\sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2(\Sigma_{r_0}^{1/2}\int_0^1 dB_{k-r_0}(t)B_{k-r_0}(t)^\intercal(\int_0^1 B_{k-r_0}(t)B_{k-r_0}(t)^\intercal dt)^{-1}\Sigma_{r_0}^{-1/2}P_{0,2}^\intercal Q_{0,2}) \ , \tag{B.3}$$

where $\Sigma_{r_0} = P_{0,2}^\intercal \Sigma P_{0,2}$ and $Q_{0,2}$ is given in Proposition 3.3. When $r_0 < k-1$, the asymptotic distribution can be highly nonstandard. Note that $P_{0,2}$ and $Q_{0,2}$ are identified up to postmultiplication by $(k - r_0) \times (k - r_0)$ orthonormal matrices, so the weak limits in (B.2) and (B.3) are invariant to the choice of $P_{2,0}$ and $Q_{2,0}$.

Another distinct feature of this example is that $\mathcal{M}$ depends on Π_0, in particular, on r_0. This presents a challenge for estimating $\mathcal{M}$ by bootstrap. We propose a residual based bootstrap following Swensen (2006) and Cavaliere et al. (2012). To this end, we need a consistent estimator for r_0, that can be obtained by various methods, for example, the estimator $\hat{r}_n$ used in (38). We propose the following bootstrap algorithm.

1. Given the consistent estimator $\hat{r}_n$ of r_0, calculate the reduced rank estimate $\hat{\Pi}_{r,n}$ and the corresponding residuals $\hat{u}_{r,t}$, for example, following Johansen (1991). Let $\hat{u}^c_{r,t} \equiv \hat{u}_{r,t} - n^{-1}\sum_{t=1}^n \hat{u}_{r,t}$, i.e., $\hat{u}^c_{r,t}$ are recentered residuals of $\hat{u}_{r,t}$.
2. Check that $\det|(1-z)I_k - \hat{\Pi}_{r,n}z|$ has $k - \hat{r}_n$ roots equal to one and all other roots outside the unit circle. If so, proceed to the next step.
3. Construct the bootstrap sample $\{Y^*_t\}_{t=1}^n$ recursively from (7) with the initial value Y_0, $\Pi_0 = \hat{\Pi}_{r,n}$, and u^*_t being generated from $\{\hat{u}^c_{r,t}\}_{t=1}^n$ by the nonparametric bootstrap. Calculate the least squares estimator

$$\hat{\Pi}^*_n = \frac{1}{n}\sum_{t=2}^n \Delta Y^*_t Y^{*\intercal}_{t-1}\Big(\frac{1}{n}\sum_{t=2}^n Y^*_{t-1}Y^{*\intercal}_{t-1}\Big)^{-1} . \tag{B.4}$$

Let $\hat{B}_n$ is the analog of B_0 and $\hat{D}_n$ is the analog of D_n by letting $\Pi_0 = \hat{\Pi}_{r,n}$. It then can be proved that

$$(\hat{\Pi}^*_n - \hat{\Pi}_{r,n})\hat{B}_n^{-1}\hat{D}_n\hat{B}_n \xrightarrow{L^*} \mathcal{M} \tag{B.5}$$

almost surely, where $\xrightarrow{L^*}$ denotes the weak convergence conditional on the data. That is, the law of the weak limit $\mathcal{M}$ is consistently estimated by the proposed bootstrap. Note that Assumption 3.2 is not satisfied.

Given that Assumptions 3.1 and 3.2 are not satisfied, we extend Theorem 3.1 to accommodate this case. Let $\kappa_n \downarrow 0$, $n\kappa_n \to \infty$, and $\hat{\phi}''_n$ be given in (38). We note that the same argument in the proof of Theorem 3.2 of Fang and Santos (2015) and Theorem 3.3 of Chen and Fang (2015) can be applied to prove that the law of the weak limit in (B.3) is consistently estimated by the law of

$$\hat{\phi}''_n((\hat{\Pi}^*_n - \hat{\Pi}_{r,n})\hat{B}_n^{-1}\hat{D}_n\hat{B}_n) \tag{B.6}$$

conditional on the data. Let $\hat{c}_{1-\alpha}$ be the $1-\alpha$ quantile of (B.6) conditional on the data. Then the same argument in the proof of Theorem 3.1 can be applied to prove that the test of rejecting H_0 when $n^2\phi(\hat{\Pi}_n) > \hat{c}_{1-\alpha}$ controls the asymptotic null rejection rate and is consistent. ■

Example 2.4-2.7 (Continued). The analysis here is similar to Example 2.1. Suppose the data is generated in Examples 2.4-2.7. In Example 2.4, let $\hat{\Pi}_n$ be the least squares estimator of Γ_0 from regressing Y_t on Z_t and W_t based on (11). In Examples 2.5-2.7, let $\hat{\Pi}_n$ be the method of moment estimators based on (14), (16) and (18), respectively. Then, under certain weak dependence and moment condition, Assumption 3.1 is satisfied by all of four examples with $\tau_n = \sqrt{n}$ and $\mathcal{M}$ bing a zero mean Gaussian. Specifically, in Example 2.4 the Gaussian limit follows by the standard result of linear regression, and Examples 2.5-2.7 the Gaussian limit follows by the central limit theorem.

Let the resampled data be generated by the nonparametric bootstrap when the original data is a sequnce of i.i.d. data, and by a block bootstrap when the original data is a sequence of dependent data. Then, under certain weak dependence and moment condition, in Example 2.4 Assumption 3.2 is satisfied with $\hat{\Pi}_n^*$ being be the least squares estimator of Γ_0 from regressing Y_t^* on Z_t^* and W_t^* based on (11), and in Examples 2.5-2.7 Assumption 3.2 is satisfied with $\hat{\Pi}_n^*$ being the method of moment estimators based on (14), (16) and (18), respectively. ■

Proposition B.1. *Let $\phi : \mathbf{M}^{k\times k} \to \mathbf{R}$ be defined as in (24). For $\Pi \in \mathbf{M}^{k\times k}$ satisfying $\phi(\Pi) = 0$, let r^*, P_2, Q_1 and Q_2 be given in Proposition 3.2. Let $B^* \equiv [Q_1, P_2]^\intercal$. Then for $\Pi \in \mathbf{M}^{k\times k}$ satisfying $\phi(\Pi) = 0$, we have*

$$\lim_{n\to\infty} \frac{\phi(\Pi + M_n T_n^* B^*)}{t_n^4} = \sum_{j=r-r^*+1}^{k-r^*} \sigma_j^2(P_2^\intercal M Q_2) \quad \text{with } T_n^* \equiv \operatorname{diag}(t_n \mathbf{1}_{r^*}, t_n^2 \mathbf{1}_{k-r^*}) ,$$

for all sequences $\{M_n\} \subset \mathbf{M}^{k\times k}$ and $\{t_n\} \subset \mathbf{R}^+$ such that $t_n \downarrow 0$, $M_n B^* \to M \in \mathbf{M}^{m\times k}$ as $n \to \infty$.

PROOF: Let $\{M_n\} \subset \mathbf{M}^{k\times k}$ be such that $M_n B^* \to M \in \mathbf{M}^{k\times k}$ and $t_n \downarrow 0$ as $n \to \infty$. Write $M_n = [M_{n,1}, M_{n,2}]$ such that $M_{n,1} \in \mathbf{M}^{k\times r^*}$, and $M = M_1 + M_2$ with $M_{n,1}Q_1^\intercal \to M_1$ and $M_{n,2}P_2^\intercal \to M_2$. Clearly, $M_1 U = 0$ for all $U \in \Psi(\Pi)$. Recall that $d = k - r$. For $\epsilon > 0$, let $\Psi(\Pi)^\epsilon$ and $\Psi(\Pi)_1^\epsilon$ be given in the proof of Proposition 3.1. In what follows we consider the nontrivial case with $\Pi \neq 0$ and $M_2 \neq 0$. In this case, $\Psi(\Pi) \subsetneq \mathbb{S}^{k\times d}$ in view of Proposition 1.3.4 in Tao (2012) and hence $\Psi(\Pi)_1^\epsilon \neq \emptyset$ for ϵ sufficiently small. Let $\sigma_{\min}^+(\Pi)$ be the smallest positive singular value of Π, which exists since $\Pi \neq 0$. Let $\Delta \equiv 5\sqrt{2}[\sigma_{\min}^+(\Pi)]^{-1}(\max_{U\in\mathbb{S}^{k\times d}} \|M_2 U\| + \max_{U\in\mathbb{S}^{k\times d}} \|M_1 U\|) > 0$, which holds since $M_2 \neq 0$. Then it follows that for all n sufficiently large

$$\begin{aligned}
&\min_{U\in\Psi(\Pi)_1^{t_n\Delta}} \|(\Pi + M_n T_n^* B^*)U\| \geq \min_{U\in\Psi(\Pi)_1^{t_n\Delta}} \|\Pi U\| - \max_{U\in\mathbb{S}^{k\times d}} \|M_n T_n^* B^* U\| \\
&\geq \frac{\sqrt{2}}{2} t_n \sigma_{\min}^+(\Pi)\Delta - t_n \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,1}Q_1^\intercal U\| - t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_2^\intercal U\| \\
&> t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_2^\intercal U\| \geq \min_{U\in\Psi(\Pi)} \|(\Pi + M_n T_n^* B^*)U\| \geq \sqrt{\phi(\Pi + M_n T_n^* B^*)} , \quad \text{(B.7)}
\end{aligned}$$

where the first inequality follows by the Lipschitz continuity of the infimum operator, the triangle inequality and the fact that $\Psi(\Pi)_1^{t_n\Delta} \subset \mathbb{S}^{k\times d}$, the second inequality follows by Lemma A.1 and the triangle inequality, the third inequality follows by the definition of Δ, $t_n \downarrow 0$, $M_{n,1}Q_1^\intercal \to M_1$ and $M_{n,2}P_2^\intercal \to M_2$ as $n \to \infty$, the fourth inequality holds by the fact that $\Pi U = 0$ and $Q_1^\intercal U = 0$ for $U \in \Psi(\Pi)$, and the last inequality follows by Lemma 3.1. Let Γ^Δ and the correspondence $\varphi : \mathbf{R} \twoheadrightarrow \mathbb{S}^{k\times d} \times \Gamma^\Delta$ be given in the proof of Proposition 3.1. Then it follows that

$$\begin{aligned}\max_{U\in\Psi(\Pi)^{t_n\Delta}} \|M_n T_n^* B^* U\| &\le t_n \max_{(U,V)\in\varphi(t_n)} \|(M_{n,1}Q_1^\intercal)(U + t_n V)\| + t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_2^\intercal U\| \\ &\le t_n^2 \max_{V\in\Gamma^\Delta} \|M_{n,1}Q_1^\intercal V\| + t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_2^\intercal U\| \ , \end{aligned} \tag{B.8}$$

where the first inequality follows by the triangle inequality and the fact that $\Psi(\Pi)^{t_n\Delta} \subset \mathbb{S}^{k\times d}$, and the second inequality follows by the fact that $Q_1^\intercal U = 0$ for $U \in \Psi(\Pi)$ and $\varphi(t_n) \subset \Psi(\Pi) \times \Gamma^\Delta$. By analogous arguments in (B.7), we have for all n sufficiently large

$$\begin{aligned}&\min_{U\in\Psi(\Pi)_1^{t_n^{3/2}\Delta}\cap\Psi(\Pi)^{t_n\Delta}} \|(\Pi + M_n T_n^* B^*)U\| \ge \min_{U\in\Psi(\Pi)_1^{t_n^{3/2}\Delta}} \|\Pi U\| - \max_{U\in\Psi(\Pi)^{t_n\Delta}} \|M_n T_n^* B^* U\| \\ &\ge \frac{\sqrt{2}}{2} t_n^{3/2} \sigma_{\min}^+(\Pi)\Delta - t_n^2 \max_{V\in\Gamma^\Delta} \|M_{n,1}Q_1^\intercal V\| - t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_2^\intercal U\| \\ &> t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_2^\intercal U\| \ge \min_{U\in\Psi(\Pi)} \|(\Pi + M_n T_n^* B^*)U\| \ge \sqrt{\phi(\Pi + M_n T_n^* B^*)} \ , \end{aligned} \tag{B.9}$$

where the first inequality follows by the Lipschitz continuity of the infimum operator, the triangle inequality and the fact that $\Psi(\Pi)_1^{t_n^{3/2}\Delta} \cap \Psi(\Pi)^{t_n\Delta} \subset \Psi(\Pi)_1^{t_n^{3/2}\Delta}$ and $\Psi(\Pi)_1^{t_n^{3/2}\Delta} \cap \Psi(\Pi)^{t_n\Delta} \subset \Psi(\Pi)^{t_n\Delta}$, the second inequality follows by (B.8) and Lemma A.1, the third inequality follows by the definition of Δ and Γ^Δ, $t_n \downarrow 0$, $M_{n,1}Q_1^\intercal \to M_1$ and $M_{n,2}P_2^\intercal \to M_2$ as $n \to \infty$, the fourth inequality holds by the fact that $\Pi U = 0$ and $Q_1^\intercal U = 0$ for $U \in \Psi(\Pi)$, and the last inequality follows by Lemma 3.1. By analogous arguments in (B.9), we have for all n sufficiently large

$$\min_{U\in\Psi(\Pi)_1^{t_n^2\Delta}\cap\Psi(\Pi)^{t_n^{3/2}\Delta}} \|(\Pi + M_n T_n^* B^*)U\| > \sqrt{\phi(\Pi + M_n T_n^* B^*)} \ . \tag{B.10}$$

Combining (B.7), (B.9), (B.10) and Lemma 3.1, we thus obtain that for all n sufficiently large

$$\phi(\Pi + M_n T_n^* B^*) = \min_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(\Pi + M_n T_n^* B^*)U\|^2 \ . \tag{B.11}$$

Now, for the right hand side of (B.11), we have

$$
\begin{aligned}
&|\min_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(\Pi + M_nT_n^*B^*)U\|^2 - \min_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(\Pi + t_nM_1 + t_n^2M_2)U\|^2| \\
&\quad \le (O(t_n^2) + O(t_n^2)) \max_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(t_n(M_{1,n}Q_1^\intercal - M_1) + t_n^2(M_{2,n}P_2^\intercal - M_2))U\| \ , \qquad \text{(B.12)}
\end{aligned}
$$

where the inequality follows by the formula $a^2-b^2 = (a+b)(a-b)$, the Lipschitz inequality of the infimum operator, the triangle inequality, and the fact that $\min_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(\Pi + M_nT_n^*B^*)U\| = O(t_n^2)$ and $\min_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(\Pi+MT_n^*B^*)U\| = O(t_n^2)$. For the second term on the right hand side of (B.12), we have

$$
\begin{aligned}
&\max_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(t_n(M_{1,n}Q_1^\intercal - M_1) + t_n^2(M_{2,n}P_2^\intercal - M_2))U\| \\
&\le t_n \max_{(U,V)\in\varphi(t_n^2)} \|(M_{n,1}Q_1^\intercal - M_1)(U + t_n^2V)\| + t_n^2 \max_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(M_{n,2}P_2^\intercal - M_2)U\| \\
&\le \max_{V\in\Gamma^\Delta} t_n^3\|(M_{n,1}Q_1^\intercal - M_1)V\| + t_n^2 \max_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(M_{n,2}P_2^\intercal - M_2)U\| = o(t_n^2) \ , \qquad \text{(B.13)}
\end{aligned}
$$

where the first inequality follows by the triangle inequality and the definition of $\varphi(t_n^2)$, the second inequality follows by the fact that $Q_1^\intercal U = 0$ and $M_1U = 0$ for $U \in \Psi(\Pi)$ and $\varphi(t_n^2) \subset \Psi(\Pi) \times \Gamma^\Delta$, and the equality follows by applying the sub-multiplicativity of Frobenius norm and the fact that $M_{n,1}Q_1^\intercal \to M_1$ and $M_{n,2}P_2^\intercal \to M_2$ as $n \to \infty$. Combining (B.11), (B.12) and (B.13), we then obtain

$$
\phi(\Pi + M_nT_n^*B^*) = \min_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(\Pi + t_nM_1 + t_n^2M_2)U\|^2 + o(t_n^4) \ . \qquad \text{(B.14)}
$$

Next, the first term on the right hand side of (B.14) can be written as

$$
\begin{aligned}
\min_{U\in\Psi(\Pi)^{t_n^2\Delta}} \|(\Pi + t_nM_1 + t_n^2M_2)U\|^2 &= \min_{(U,V)\in\varphi(t_n^2)} \|(\Pi + t_nM_1 + t_n^2M_2)(U + t_n^2V)\|^2 \\
&= t_n^4 \min_{(U,V)\in\varphi(t_n^2)} \|\Pi V + MU\|^2 + o(t_n^4) \ , \qquad \text{(B.15)}
\end{aligned}
$$

where the second equality follows by the fact that $\Pi U = 0$ and $M_1U = 0$ for $U \in \Psi(\Pi)$ and $\|V\| \le \Delta$ for all $V \in \Gamma^\Delta$. By analogous arguments in (A.14), we have

$$
\min_{(U,V)\in\varphi(t_n^2)} \|\Pi V + MU\|^2 = \min_{U\in\Psi(\Pi)} \min_{V\in\mathbf{M}^{k\times d}} \|\Pi V + MU\|^2 + o(1) \ . \qquad \text{(B.16)}
$$

Combining (B.14), (B.15) and (B.16), we may conclude that

$$
\lim_{n\to\infty} \frac{\phi(\Pi + M_nT_n^*B^*)}{t_n^4} = \min_{U\in\Psi(\Pi)} \min_{V\in\mathbf{M}^{k\times d}} \|\Pi V + MU\| = \sum_{j=r-r^*+1}^{k-r^*} \sigma_j^2(P_2^\intercal MQ_2) \ , \qquad \text{(B.17)}
$$

where the second equality follows by Proposition 3.2. This completes the proof of the lemma. ■

Proposition B.2. *Suppose* $\Pi_0 \in \mathbf{M}^{k\times k}$, *and let* r_0, $Q_{0,1}$ *and* $P_{0,2}$ *be given in Proposition 3.3. Suppose there are* $\hat{\Pi}_n : \{X_i\}_{i=1}^n \to \mathbf{M}^{k\times k}$ *such that* $(\hat{\Pi}_n - \Pi_0)B_0^{-1}D_nB_0 \xrightarrow{L} \mathcal{M}$ *for some* $\tau_n \uparrow \infty$ *and random matrix* $\mathcal{M} \in \mathbf{M}^{k\times k}$, *where* $D_n \equiv diag(\tau_n \mathbf{1}_{r_0}, \tau_n^2 \mathbf{1}_{k-r_0})$ *and* $B_0 \equiv [Q_{0,1}, P_{0,2}]^\intercal$. *Then we have under* H_0,

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

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

$$g_n(M) \equiv \tau_n^4 \phi(\Pi_0 + MD_n^{-1}B_0) \ . \tag{B.18}$$

By Proposition B.1, $g_n(M_n) \to \sum_{j=r-r^*+1}^{k-r^*} \sigma_j^2(P_2^\intercal M Q_2)$ whenever $M_n B^* \to M$. Note that $\tau_n^4\phi(\hat{\Pi}_n) = g_n((\hat{\Pi}_n - \Pi_0)B_0^{-1}D_n)$, then the result of the proposition follows by Theorem 1.11.1(i) in van der Vaart and Wellner (1996). ■

Appendix C Kleibergen and Paap (2006)'s Test

For ease of reference, we review the rank test by Kleibergen and Paap (2006). Let $\hat{\Pi}_n \in \mathbf{M}^{m\times k}$ be an estimator for $\Pi_0 \in \mathbf{M}^{m\times k}$ that satisfies Assumption 3.1 with $\tau_n = \sqrt{n}$ and $\mathrm{vec}(\mathcal{M}) \sim N(0, \Omega)$ for some positive semidefinite matrix Ω. Let $\hat{\Omega}_n$ be a consistent estimator of Ω. Let $\hat{\Pi}_n = \hat{P}_n \hat{\Sigma}_n \hat{Q}_n^\intercal$ be a singular value decomposition of $\hat{\Pi}_n$, where $\hat{P}_n \in \mathbb{S}^{m\times m}$ and $\hat{Q}_n \in \mathbb{S}^{k\times k}$, and $\hat{\Sigma}_n \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Write $\hat{P}_n = [\hat{A}_n, \hat{B}_n]$ and $\hat{Q}_n = [\hat{C}_n, \hat{D}_n]$ for $\hat{A}_n \in \mathbf{M}^{m\times r}$ and $\hat{C}_n \in \mathbf{M}^{k\times r}$, and let $\hat{S}_n$ be the right bottom $(m-r)\times(k-r)$ block submatrix of $\hat{\Sigma}_n$. Then the test statistic for the hypotheses (2) is given by

$$\mathrm{rk}(r) = n\mathrm{vec}(\hat{S}_n)^\intercal[(\hat{D}_n \otimes \hat{B}_n)^\intercal \hat{\Omega}_n (\hat{D}_n \otimes \hat{B}_n)]^{-1}\mathrm{vec}(\hat{S}_n) \ , \tag{C.1}$$

where $\otimes$ denotes the kronecker product. Thus, the rank test with the nominal level $\alpha \in (0,1)$ rejects the null $\mathrm{H}_0^{(r)}$ in the hypotheses (2) whenever $\mathrm{rk}(r) > \chi^2((m-r)(k-r), 1-\alpha)$. Note that $\hat{B}_n$ and $\hat{D}_n$ can be chosen up to postmultiplication by $(m-r)\times(m-r)$ and $(k-r)\times(k-r)$ orthonormal matrices, respectively, but $\mathrm{rk}(r)$ is invariant to the choice of $\hat{B}_n$ and $\hat{D}_n$.

In order to examine the asymptotic behavior of the rank test when $\mathrm{rank}(\Pi_0) < r$, we consider the case with $\Pi_0 = \mathbf{0}_{2\times 2}$, Ω is positive definite and $r = 1$. Let $\mathcal{M} = \mathcal{P}\mathcal{W}\mathcal{Q}$ be a singular value decomposition of $\mathcal{M}$, where $\mathcal{P} \in \mathbb{S}^{2\times 2}$ and $\mathcal{Q} \in \mathbb{S}^{2\times 2}$, and $\mathcal{W} \in \mathbf{M}^{2\times 2}$ is diagonal with diagonal entries in descending order. Write $\mathcal{P} = [\mathcal{P}_1, \mathcal{P}_2]$ and $\mathcal{Q} = [\mathcal{Q}_1, \mathcal{Q}_2]$

for $\mathcal{P}_1 \in \mathbf{M}^{2\times 1}$ and $\mathcal{Q}_1 \in \mathbf{M}^{2\times 1}$, and let $\mathcal{S}$ be (2,2)th entry of $\mathcal{W}$. Then by Lemma C.1, the asymptotic distribution of rk(1) is given by

$$\text{rk}(1) \xrightarrow{L} \frac{\mathcal{S}^2}{(\mathcal{Q}_2 \otimes \mathcal{P}_2)^\intercal \Omega (\mathcal{Q}_2 \otimes \mathcal{P}_2)} \,. \tag{C.2}$$

Note that $\mathcal{P}_2$ and $\mathcal{Q}_2$ can be chosen up to a sign, respectively, but the asymptotic distribution is invariant to the choice of $\mathcal{P}_2$ and $\mathcal{Q}_2$.

We now plot the distribution function of the weak limit in (C.2) by simulation. We consider two values of Ω:

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

The distribution functions based on 100,000 simulation replications are plotted in Figure 7. The weak limit when $\Omega = \Omega_1$ is first order dominated by the $\chi^2(1)$ random variable, and the weak limit when $\Omega = \Omega_2$ first order dominates the $\chi^2(1)$ random variable. This implies that directly applying the test to (1) will under-reject the null when $\Omega = \Omega_1$, and will over-reject the null when $\Omega = \Omega_2$.

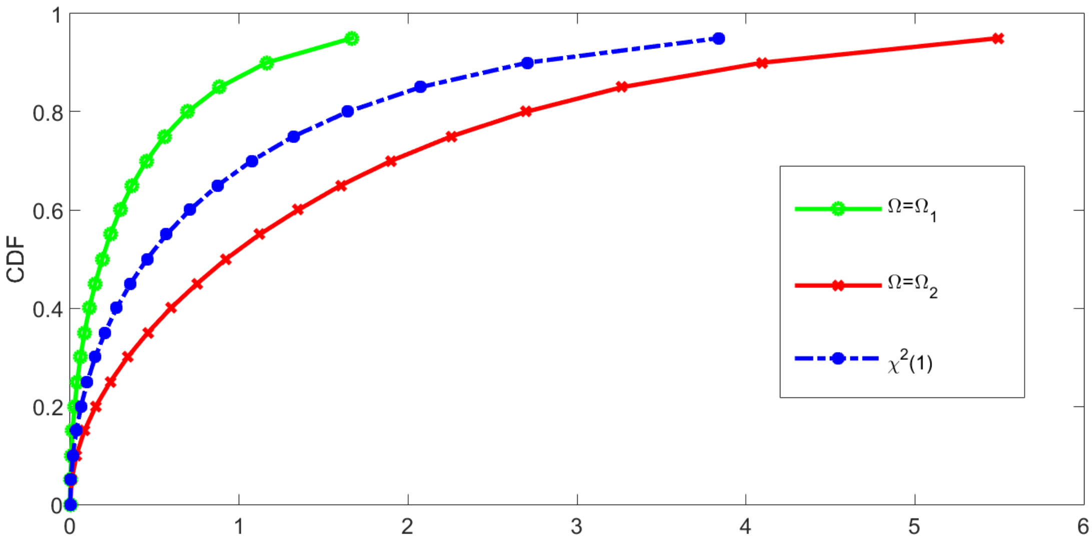

Figure 7: The distribution function of the weak limit of rk(1) when $\Pi_0 = \mathbf{0}_{2\times 2}$

Lemma C.1. *Let* rk(r) *be given in* (C.1)*. Suppose* $\Pi_0 = \mathbf{0}_{2\times 2}$ *and* Ω *is positive definite. Then the asymptotic distribution of* rk(1) *is given in* (C.2)*.*

Proof: For $x \in \mathbf{R}$, let $\text{sgn}(x) \equiv 1\{x \geq 0\} - 1\{x < 0\}$. Note that $\hat{D}_n$ and $\mathcal{Q}_2$ are the eigenvalue of $n\hat{\Pi}_n^\intercal \hat{\Pi}_n$ and $\mathcal{M}^\intercal \mathcal{M}$ associated with the smallest eigenvalue. By analogous arguments in Lemma 4.3 of Bosq (2000), we have

$$\|\operatorname{sgn}(\hat{D}_n^\intercal \mathcal{Q}_2)\hat{D}_n - \mathcal{Q}_2\| \leq \frac{2\sqrt{2}}{\sigma_1^2(\mathcal{M}) - \sigma_2^2(\mathcal{M})} \|n\hat{\Pi}_n^\intercal \hat{\Pi}_n - \mathcal{M}^\intercal \mathcal{M}\| \ . \tag{C.3}$$

Similarly, we have

$$\|\operatorname{sgn}(\hat{B}_n^\intercal \mathcal{P}_2)\hat{B}_n - \mathcal{P}_2\| \leq \frac{2\sqrt{2}}{\sigma_1^2(\mathcal{M}) - \sigma_2^2(\mathcal{M})} \|n\hat{\Pi}_n \hat{\Pi}_n^\intercal - \mathcal{M} \mathcal{M}^\intercal\| \ . \tag{C.4}$$

Note that $\sqrt{n}\hat{S}_n = \sigma_2(\sqrt{n}\hat{\Pi}_n)$ and $\mathcal{S} = \sigma_2(\mathcal{M})$. By the fact that singular values are continuous, (C.3), (C.4) and the continuous mapping theorem, we thus obtain that

$$(\sqrt{n}\hat{S}_n, \operatorname{sgn}(\hat{B}_n^\intercal \mathcal{P}_2)\hat{B}_n^\intercal, \operatorname{sgn}(\hat{D}_n^\intercal \mathcal{Q}_2)\hat{D}_n^\intercal) \xrightarrow{L} (\mathcal{S}, \mathcal{P}_2^\intercal, \mathcal{Q}_2^\intercal) \ . \tag{C.5}$$

Note that rk(1) does not change by replacing $\hat{B}_n$ and $\hat{D}_n$ with $\operatorname{sgn}(\hat{B}_n^\intercal \mathcal{P}_2)\hat{B}_n$ and $\operatorname{sgn}(\hat{D}_n^\intercal \mathcal{Q}_2)\hat{D}_n$, respectively, so the result of the lemma follows by (C.5) together with the continuous mapping theorem. ■

Appendix D Parameters in Section 4.1

The values of parameters for DGP2 in the simulation studies in Section 4.1 are as follows:

- The value of Σ_F is specified as the sample correlation matrix of $\{F_t\}_{t=1}^T$, where $\{F_t\}_{t=1}^T$ is the real data in Section 4.2;
- The values of α and β are specified as $\alpha = (0.0813, -0.0271, -0.6203, -0.0460)^\intercal$ and $\beta = (-0.3411, -0.1277, -0.3838, -0.5312, -0.2728, -0.3527, -0.2188, -0.2934, -0.2035, -0.3427)^\intercal$;
- The value of Π_1 is specified as $\Pi_1 = \bar{\Pi}_T - \beta\alpha^\intercal$, where $\bar{\Pi}_T = \sum_{t=1}^T R_t F_t^\intercal (\sum_{t=1}^T F_t F_t^\intercal)^{-1}$ with $\{F_t, R_t\}_{t=1}^T$ being the real data in Section 4.2;
- The value of Γ is specified as

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

- The value of Σ_v is specified as

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

References

Al-Sadoon, M. M. (2015). A general theory of rank testing. Working paper.

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

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

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley New York.

Aoki, M. (1990). *State Space Modeling of Time Series*. 2nd ed. Springer.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70** 191–221.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, **25** 52–60.

Barnett, W. A. and Serletis, A. (2008). Measuring consumer preferences and estimating demand systems. In *Quantifying Consumer Preferences*. 1–35.

Berkes, I., Horváth, L. and Rice, G. (2016). On the asymptotic normality of kernel estimators of the long run covariance of functional time series. *Journal of Multivariate Analysis*, **144** 150–175.

Bierens, H. (1997). Nonparametric cointegration analysis. *Journal of Econometrics*, **77** 379–404.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Springer.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory.* Holden-Day.

Bura, E. and Yang, J. (2011). Dimension estimation in sufficient dimension reduction: a unifying approach. *Journal of Multivariate Analysis*, **102** 130–142.

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

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

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

Chen, Q. and Fang, Z. (2015). Inference on functionals under first order degeneracy. Working paper.

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

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

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

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

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

Dovonon, P. and Renault, E. (2013). Testing for common conditionally heteroskedastic factors. *Econometrica*, **81** 2561–2586.

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

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

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

Engle, R. F. and Kozicki, S. (1993). Testing for common features. *Journal of Business & Economic Statistics*, **11** 369–380.

ENGLE, R. F. and SUSMEL, R. (1993). Common volatility in international equity markets. *Journal of Business & Economic Statistics*, **11** 167–176.

FANG, Z. and SANTOS, A. (2015). Inference on directionally differentiable functions. Working paper.

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

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

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

GORMAN, W. M. (1981). Some Engel curves. In *Essays in the Theory and Measurement of Consumer Behaviour: In Honour of Sir Richard Stone*. Cambridge University Press, 7–29.

GRANAS, A. and DUGUNDJI, J. (2003). *Fixed Point Theory.* Springer.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer.

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

HANNAN, E. (1970). *Multiple Time Series.* Wiley.

HANSEN, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50** 1029–1054.

HARVILLE, D. A. (2008). *Matrix Algebra From a Statistician's Perspective.* Springer.

HORN, R. A. and JOHNSON, C. R. (1991). *Topics in Matrix Analysis.* Cambridge University Press.

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

JAGANNATHAN, R., SKOULAKIS, G. and SKOULAKIS, Z. (2002). Generalized methods of moments: Applications in finance. *Journal of Business & Economic Statistics*, **20** 470–481.

JAGANNATHAN, R. and WANG, Z. (1996). The conditional CAPM and the cross-section of expected returns. *The Journal of Finance*, **51** 3–53.

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

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

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

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

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. **40** 694–726.

Lewbel, A. (1991). The rank of demand systems: Theory and nonparametric estimation. *Econometrica*, **59** 711–730.

Lewbel, A. (2006). Engel curves. In *The New Palgrave Dictionary of Economics*, vol. 2.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86** 316–327.

Liao, Z. and Phillips, P. (2015). Automated estimation of vector error correction models. *Econometric Theory*, **31** 581–646.

Magnus, J. R. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. 3rd ed. John Wiley & Sons.

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

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, vol. 4. 2111–2245.

Politis, D. and Romano, J. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *The Annals of Statistics* 1985–2007.

Politis, D. and Romano, J. (1993). Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis*, **47** 301–328.

Politis, D., Romano, J. and Lai, T. (1992). Bootstrap confidence bands for spectra and cross-spectra. *IEEE Transactions on Signal Processing*, **40** 1206–1215.

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

Priestley, M. (1981). *Spectral Analysis and Time Series.* Adademic Press.

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

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

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13** 341–360.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, **39** 577–591.

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

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

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

Shintani, M. (2001). A simple cointegrating rank test without vector autoregression. *Journal of Econometrics*, **105** 337–362.

Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models1. *Econometrica*, **74** 1699–1714.

Tao, T. (2012). *Topics in Random Matrix Theory.* American Mathematical Society.

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

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

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