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

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

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

May 13, 2019

In this appendix, we present additional examples and results.

E Cointegration and Additional Examples

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

E.1 Inference on Cointegration Rank

Let $\{Y_t\}$ be a time series in $\mathbf{R}^k$ such that all its entries are unit root processes. For ease of exposition and to hight what is essential to our theory, we limit ourselves to processes without deterministic terms throughout. By the Granger representation theorem, the number h_0 of independent cointegrating vectors is precisely equal to $k - \text{rank}(\Omega_0)$ with Ω_0 the long run variance of ΔY_t.[1] Within this (nonparametric) system framework, one may be interested in testing

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

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

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

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

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

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

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

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

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

Nonparametric Cointegration Test

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

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

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

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

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

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

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

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

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

Cointegration Test in Error-Correction Models

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

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

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

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

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

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

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

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

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

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

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

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

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

Therefore, in order to invoke the Delta method, the only question that remains is: is our map ϕ_r as defined in (13) suitably differentiable with respect to these linear maps? The answer is affirmative, as shown by Proposition E.1 stated at the end of this subsection. In particular, if $\mathrm{rank}(\Pi_0) \leq r \equiv h$, then we have

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

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

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

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

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

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

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

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

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

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

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

Bootstrap Inference: The limiting distribution in (E.11) is highly nonstandard, and in particular depends on the true rank h_0. In order to apply our bootstrap procedure, we need to estimate both the "derivative" ϕ''_{r,Π_0} and the limit $\mathcal{M}$. Estimation of ϕ''_{r,Π_0} is no more special than what we have discussed in Section 3.3. For example, one may estimate ϕ''_{r,Π_0} by (28) with $\hat{r}_n$ given by (9). Since $\sqrt{n}\{\hat{\Pi}_n - \Pi_0\}$ converges in distribution by (E.11), we thus still have $\hat{r}_n \xrightarrow{p} r_0 = h_0$ provided $\kappa_n \to 0$ and $\sqrt{n}\kappa_n \to \infty$ by Lemma D.7. Condition (26) in turn follows from Lemma D.6. In fact, one can show along the lines in the proof of Lemma D.7 that it suffices to have $\kappa_n \to 0$ and $n\kappa_n \to \infty$.

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

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

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

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

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

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

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

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

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

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

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

Given an estimator $\hat{\phi}''_{r,n}$ satisfying (26) and the bootstrap estimator $\hat{\mathcal{M}}_n^*$ as in (E.17), we may finally estimate the limit in (E.11) by the conditional law (given the data) of

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

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

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

Proposition E.1. *Let $\phi_r : \mathbf{M}^{k\times k} \to \mathbf{R}$ be defined as in* (13) *with $m = k$ and $\Pi_0 \in \mathbf{M}^{k\times k}$ satisfy $\phi_r(\Pi_0) = 0$. Then, for $r_0 \equiv \mathrm{rank}(\Pi_0)$ and $T_n \equiv \mathrm{diag}(t_n\mathbf{1}_{r_0}, t_n^2\mathbf{1}_{k-r_0})$ with $t_n > 0$,*

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

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

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

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

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

$$\begin{aligned}
\min_{U\in\Psi(\Pi_0)_1^{t_n\Delta}} \|(\Pi_0 + M_n T_n B_0)U\| &\geq \min_{U\in\Psi(\Pi_0)_1^{t_n\Delta}} \|\Pi_0 U\| - \max_{U\in\mathbb{S}^{k\times d}} \|M_n T_n B_0 U\| \\
&\geq \frac{\sqrt{2}}{2} t_n \sigma^+_{\min}(\Pi_0)\Delta - t_n \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,1}Q_{0,1}^\intercal U\| - t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_{0,2}^\intercal U\| \\
&> t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_{0,2}^\intercal U\| \geq \min_{U\in\Psi(\Pi_0)} \|(\Pi_0 + M_n T_n B_0)U\| \\
&\geq \sqrt{\phi_r(\Pi_0 + M_n T_n B_0)}\ , \qquad \text{(E.20)}
\end{aligned}$$

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

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

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

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

$$\begin{aligned}
\min_{U\in\Psi(\Pi_0)_1^{t_n^{3/2}\Delta}\cap\Psi(\Pi_0)^{t_n\Delta}} \|(\Pi_0 + M_n T_n B_0)U\| &\geq \min_{U\in\Psi(\Pi_0)_1^{t_n^{3/2}\Delta}} \|\Pi_0 U\| - \max_{U\in\Psi(\Pi_0)^{t_n\Delta}} \|M_n T_n B_0 U\| \\
&\geq \frac{\sqrt{2}}{2} t_n^{3/2} \sigma^+_{\min}(\Pi_0)\Delta - t_n^2 \max_{V\in\Gamma^\Delta} \|M_{n,1}Q_{0,1}^\intercal V\| - t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_{0,2}^\intercal U\| \\
&> t_n^2 \max_{U\in\mathbb{S}^{k\times d}} \|M_{n,2}P_{0,2}^\intercal U\| \geq \min_{U\in\Psi(\Pi_0)} \|(\Pi_0 + M_n T_n B_0)U\| \\
&\geq \sqrt{\phi_r(\Pi_0 + M_n T_n B_0)}\ , \qquad \text{(E.22)}
\end{aligned}$$

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

$U \in \Psi(\Pi_0)$. In turn, by analogous arguments, we have, for all n sufficiently large,

$$\min_{U \in \Psi(\Pi_0)_1^{t_n^2 \Delta} \cap \Psi(\Pi_0)^{t_n^{3/2} \Delta}} \|(\Pi_0 + M_n T_n B_0)U\| > \sqrt{\phi_r(\Pi_0 + M_n T_n B_0)} \; . \tag{E.23}$$

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

$$\phi_r(\Pi_0 + M_n T_n B_0) = \min_{U \in \Psi(\Pi_0)^{t_n^2 \Delta}} \|(\Pi_0 + M_n T_n B_0)U\|^2 \; . \tag{E.24}$$

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

$$\begin{aligned} &\Big| \min_{U \in \Psi(\Pi_0)^{t_n^2 \Delta}} \|(\Pi_0 + M_n T_n B_0)U\|^2 - \min_{U \in \Psi(\Pi_0)^{t_n^2 \Delta}} \|(\Pi_0 + t_n M_1 + t_n^2 M_2)U\|^2 \Big| \\ &\leq (O(t_n^2) + O(t_n^2)) \max_{U \in \Psi(\Pi_0)^{t_n^2 \Delta}} \|(t_n(M_{1,n} Q_{0,1}^{\intercal} - M_1) + t_n^2(M_{2,n} P_{0,2}^{\intercal} - M_2))U\| \; , \end{aligned} \tag{E.25}$$

where the inequality follows by the formula $a^2 - b^2 = (a+b)(a-b)$, the Lipschitz inequality of the min operator, the triangle inequality, and the facts that $\min_{U \in \Psi(\Pi_0)^{t_n^2 \Delta}} \|(\Pi_0 + M_n T_n B_0)U\| = O(t_n^2)$ and $\min_{U \in \Psi(\Pi_0)^{t_n^2 \Delta}} \|(\Pi_0 + M T_n B_0)U\| = O(t_n^2)$. For the second term on the right hand side of (E.25), we have

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

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

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

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

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

where the second equality follows by the facts that $\Pi_0 U = 0$ and $M_1 U = 0$ for $U \in$

$\Psi(\Pi_0)$, and $\|V\| \leq \Delta$ for all $V \in \Gamma^\Delta$. By analogous arguments in (A.15), we have

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

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

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

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

Proposition E.2. *Suppose that there is an estimator* $\hat{\Pi}_n : \{X_i\}_{i=1}^n \to \mathbf{M}^{k\times k}$ *for* $\Pi_0 \in \mathbf{M}^{k\times k}$ *such that* $\{\hat{\Pi}_n - \Pi_0\}B_0^{-1}D_nB_0 \xrightarrow{L} \mathcal{M}$ *for some* $\tau_n \uparrow \infty$ *and random matrix* $\mathcal{M} \in \mathbf{M}^{k\times k}$*, where* $D_n \equiv diag(\tau_n \mathbf{1}_{r_0}, \tau_n^2 \mathbf{1}_{k-r_0})$*. If* $\operatorname{rank}(\Pi_0) \leq r$*, then we have*

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

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

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

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

E.2 Additional Examples

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

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

(2009) establish a lower bound for γ_0, i.e., $\gamma_0 \geq \text{rank}(\Pi_0)$ with

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

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

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

with p_{X_1,X_2} the probability mass function (pmf) of (X_1, X_2) and $p_{Z_2|X_1}$ the conditional pmf of Z_2 given X_1. Under additional conditions, Kasahara and Shimotsu (2009) show in fact $\gamma_0 = \text{rank}(\Pi_0)$. We focus on discrete variables and two periods for ease of exposition, but the results extend to more general cases by Remark 2(iv) and Propositions 3 and 8 in Kasahara and Shimotsu (2009). The number of types is crucial for the specification of mixture distributions, and yet inference on γ_0 without parametric assumptions on the component distributions has been understood to be challenging (Kasahara and Shimotsu, 2009, 2014; Bonhomme et al., 2016). By further restricting each component distribution to have independent marginals, Kasahara and Shimotsu (2014) and Bonhomme et al. (2016) accomplish nonparametric estimation of γ_0 based on the rank test of Kleibergen and Paap (2006). Interestingly, Xiao (2018) derives a similar nonparametric identification result for the number of equilibria in incomplete information games and obtains a consistent estimator based on the rank test of Robin and Smith (2000). ■

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Then under appropriate restrictions, the rank of $\mathrm{Var}(F)$ is equal to the rank of

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

References

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

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.

Berkowitz, J. and Diebold, F. X. (1998). Bootstrapping multivariate spectra. *The Review of Economics and Statistics*, **80** 664–666.

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

Bonhomme, S., Jochmans, K. and Robin, J.-M. (2016). Non-parametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **78** 211–229.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory*, vol. 36. Siam.

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

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

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.

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.

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

Hannan, E. J. (1970). *Multiple Time Series*. Wiley.

Harris, D. (1997). Principal components analysis of cointegrated time series. *Econometric Theory*, **13** 529–557.

Hayashi, F. (2000). *Econometrics*. Princeton University Press, New Jersey.

Hubrich, K., Lütkepohl, H. and Saikkonen, P. (2001). A review of systems cointegration tests. *Econometric Reviews*, **20** 247–318.

Jentsch, C. and Kreiss, J.-P. (2010). The multiple hybrid bootstrap—resampling multivariate linear processes. *Journal of Multivariate Analysis*, **101** 2320–2345.

Jentsch, C. and Politis, D. N. (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *The Annals of Statistics*, **43** 1117–1140.

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

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

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

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

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

Kasahara, H. and Shimotsu, K. (2014). Non-parametric identification and estimation of the number of components in multivariate mixtures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76** 97–111.

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

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, **54** 159–178.

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

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

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

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

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

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

Nyblom, J. and Harvey, A. (2000). Tests of common stochastic trends. *Econometric Theory*, **16** 176–199.

Phillips, P. C., Sun, Y. and Jin, S. (2006). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review*, **47** 837–894.

Politis, D. (2001). On nonparametric function estimation with infinite-order flat-top kernels. In *Probability and Statistical Models with Applications*. Chapman and Hall/CRC, 484–498.

Politis, D. N. (2011). Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, **27** 703–744.

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

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

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

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

Snell, A. (1999). Testing for r versus $r-1$ cointegrating vectors. *Journal of Econometrics*, **88** 151–191.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American statistical Association*, **83** 1097–1107.

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

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

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

van Giersbergen, N. P. (1996). Bootstrapping the trace statistic in VAR models: Monte Carlo results and applications. *Oxford Bulletin of Economics and Statistics*, **58** 391–408.

Xiao, R. (2018). Identification and estimation of incomplete information games with multiple equilibria. *Journal of Econometrics*, **203** 328 – 343.