A General Theory of Rank Testing*

Majid M. Al-Sadoon

Universitat Pompeu Fabra & Barcelona GSE

February 1, 2015

Abstract

This paper demonstrates that all rank test statistics are functions of implicit null space estimators. The paper proposes a novel theory of null space estimation that allows for standard asymptotics, polynomial regressions, and cointegration asymptotics. The paper proves that the behaviour of rank test statistics is completely governed by the implicit null space estimators through a plug–in principle. This allows for a general theory of rank testing that simplifies the asymptotics of rank test statistics, clarifies the relationships between the various rank test statistics, makes full use of the numerical analysis literature, and motivates numerous new rank test statistics. A brief Monte Carlo study illustrates the results.

JEL Classification: C12, C13, C30.

Keywords: Rank testing, cointegration, plug–in principle, subspace estimation.

*An earlier version of this paper was titled, "The Stochastic Wald Test." Thanks are due to Sean Holly, M. Hashem Pesaran, Richard J. Smith, George Kapetanios, Rod McCrorie, Frank Kleibergen, Jesús Gonzalo, Vasco Carvalho, Kristoffer Nimark, Gábor Lugosi, Sergi Jiménez, Christian Brownlees, Barbara Rossi, Peter Boswijk, Herman van Dijk, Andrew Harvey, Luis Martins, Søren Johansen, anonymous referees, and to seminar participants at Universitat Pompeu Fabra and Universidad Carlos III de Madrid for helpful comments and suggestions. The author takes sole responsibility for any remaining errors. Research for this paper was supported by Spanish Ministry of Economy and Competitiveness project ECO2012-33247.

1 Introduction

Rank testing is ubiquitous in empirical data analysis. Examples of its applications include, demand systems (Lewbel, 1991), identification in GMM (Cragg & Donald, 1993; Wright, 2003; Arellano et al., 2012), identification robust inference (Kleibergen, 2005), cointegration (Dolado et al., 2001; Hubrich et al., 2001), model reduction (Velu et al., 1986; Camba-Mendez et al., 2003), Granger causality testing (Al-Sadoon, 2014), growth curve statistics (Reinsel & Velu, 1998), linear systems theory (Markovsky, 2012), and machine learning (Hastie et al., 2001). See Camba-Mendez & Kapetanios (2009) for a comprehensive review.

Much of this progress has taken place in spite of the difficulty of the asymptotics of these tests. Indeed the tests often involve the asymptotics of eigenvectors, matrix inverses, and other complicated objects. For example, Cragg & Donald (1996) have proven that some of the asymptotics of Gill & Lewbel (1992) are incorrect. The asymptotics are indeed so difficult that little is known about the behaviour of rank test statistics under local alternatives or under misspecification. The only paper in the standard asymptotics literature that touches on the issue of local power is Cragg & Donald (1997) and the only one that treats misspecification is Robin & Smith (2000). In both cases, the results relate to specific tests and do not generalize to all rank tests. In cointegration, the local power of the Johansen trace test is considered in a handful of papers surveyed in Hubrich et al. (2001). Caner (1998) and Cavaliere et al. (2010b) consider the effect of misspecifying the innovation process of a cointegrated VAR, while Aznar & Salvador (2002) and Cavaliere et al. (2014) consider the effect of misspecifying its lag order. Again, these results refer only to specific statistics and do not tell us how, for example, the Kleibergen & Paap (2006) might behave under the same local alternative or misspecification. Thus, there is no clear general principle that unifies and generalizes all of the results above.

Moreover, although this literature has proposed a number of tests of rank, it is not clear what the relationships are between some of these statistics. The statistics of Cragg & Donald (1996) and Cragg & Donald (1997) are known to be asymptotically equivalent. However, it is not clear how the statistics of Anderson (1951), Kleibergen & Paap (2006), and some of the statistics in Robin & Smith (1995) relate to these or to each other. It is also not clear how the Kleibergen & Paap (2006), Kleibergen & van Dijk (1994), and Johansen (1988) cointegration statistics relate to each other or to the standard asymptotics rank testing statistics. There is therefore a need for a general framework that encompasses all of these tests.

This literature has also developed in parallel to a much larger and older literature in numerical analysis, which has proposed a battery of algorithms for detecting rank (see the survey in Hansen (1998)). The econometric and statistical literature has made minimal use of these algorithms (see Appendix A). For example, the most popular rank revealing decomposition, the QR decomposition with pivoting, has never been used in a rank test. The Cholesky decomposition has also been absent from statistical tests as well as the multitude of variations on the algorithms commonly used in statistics and econometrics. It would be useful to have a theory general enough to allow the researcher to take any of these numerical algorithm and convert it into a statistical rank test.

This paper addresses all of the issues above by developing a general theory of rank testing that clarifies the asymptotics of the tests under the various alternatives (null, local, and global), the asymptotics under misspecification, the relationships between the various rank test statistics in the literature, all while taking full advantage of the numerical analysis literature. The approach is more general than that proposed by Reinsel & Velu (1998) and Massmann (2007), which nest some of the likelihood–based tests but not many of the Wald–type tests. This is accomplished in a number of steps.

First, the paper describes the general structure of all rank test statistics. The paper demonstrates that all rank test statistics are functions of implicit estimators of the null spaces of the matrix which is being tested.

Next, the paper conducts a detailed study of null space estimation. The approach is more general than that employed by Dufour & Valery (2011) in that it applies to general matrices rather than just the positive semi–definite ones and does not restrict itself to eigenprojections. The theory of null space estimation proposed in this paper has a number of novelties. First, it is identification–free in that it works with projection matrices rather than arbitrarily identified bases for the estimated subspaces. Second, it allows for the analysis of standard asymptotics as well as cointegration asymptotics. Finally, it allows the researcher to employ just about any algorithm proposed in the numerical analysis literature for the detection of rank. This theory is applicable not just in rank testing but should be useful for factor analysis, model reduction, and machine learning applications.

Finally, it is proven that the behaviour of rank test statistics is completely governed by the implicit null space estimators. A plug–in principle is shown to hold, whereby every rank

test statistic is asymptotically equivalent to an infeasible statistic that plugs in the population values for the estimated null spaces in the standard asymptotics. In the cointegration setting, the plug–in principle is also demonstrated although the infeasible statistic is no longer necessarily the population values of the null spaces. In the case of cointegration, the plug–in principle applies to the directions of fastest convergence to singularity of the estimator.

The new approach greatly simplifies the analysis of rank test statistics under the null hypothesis as well as local and global alternatives. Under the null hypothesis or the local alternative, one can simply ignore the fact that the null spaces are estimated and derive the asymptotics as if the appropriate null spaces were known. Under the global alternative, the paper finds conditions (conjectured to be generic) that ensure the feasible and infeasible statistics diverge at the same rate. When those conditions are not satisfied, one can still prove consistency by appealing to the properties of null space estimators.

The theory illuminates the continuity between standard asymptotics rank testing and cointegration rank testing. It is demonstrated that cointegration rank testing is nothing more than rescaled rank testing. It also clarifies some of the relationships between the various cointegration rank test statistics. It is proven, for example, that the Johansen (1988), Kleibergen & van Dijk (1994), and Kleibergen & Paap (2006) tests have the same local power.

The paper proposes a number of new statistics: (i) extensions of the likelihood ratio statistic of Anderson (1951) and the maximum eigenvalue statistic of Johansen (1991), (ii) statistics based on the QR and Cholesky decompositions, (iii) statistics based on the fixed–b theory proposed by Vogelsang (2001), Kiefer & Vogelsang (2002a), Kiefer & Vogelsang (2002b), and Kiefer & Vogelsang (2005), and (iv) some new statistics for testing common trends.

It is important to note two aspects of the plug–in principle that have been well known in the rank testing literature. First, as far back as Stock & Watson (1988) and as recently as Boswijk et al. (2015), researchers have relied on the idea that the cointegration vector, being super–consistent, is as good as known in determining the asymptotics of rank test statistics. This paper demonstrates that this idea does not hold in general (see Example 13) and proposes the necessary modifications. Second, the proofs of the asymptotics of rank test statistics under the null in standard asymptotics sometimes involved an implicit use of the plug–in principle (e.g. Cragg & Donald (1996) and Robin & Smith (2000)). However, the plug–in principle has not been recognized as an overarching principle that elucidates the asymptotics of rank

test statistics, not just under the null in standard asymptotics, but also the local and global alternatives as well as misspecification and more general asymptotics (e.g. cointegration).

We mention finally that some of the results of this paper were presented as part of the author's PhD thesis (Al-Sadoon, 2010). Specifically, Chapter 5 of the thesis included special cases of Lemma 2 (iv), Lemma 3 (i), and Corollary 1 of this paper as well as an early draft of Appendix A.

The paper is organized as follows. Section 2 develops the notation of the paper. Section 3 develops the theory of rank testing under standard asymptotics. Section 4 develops the theory of rank testing under general asymptotics. Section 5 provides Monte Carlo evidence. Section 6 concludes. Appendix A develops some of the necessary results from numerical analysis as well as containing some useful derivations. Appendix B consists of proofs.

2 Notation

Let $\mathbb{R}^{n\times m}$ be the set of real $n \times m$ matrices. Let $\mathbb{G}^{n\times m} \subset \mathbb{R}^{n\times m}$ be the set of matrices of full rank. Let $\mathbb{S}^m \subset \mathbb{R}^{m\times m}$ be the set of $m \times m$ symmetric matrices. Let $\mathbb{P}^m \subset \mathbb{S}^m$ be the sets of $m \times m$ positive semi–definite matrices and $\mathbb{P}^m_+ \subset \mathbb{P}^m$ the set of positive definite matrices. The ij–th element of B is denoted by $B_{(i,j)}$. We define $\text{vec}(B)$ as the vector formed by vertically stacking the columns of B and $\text{vech}(B)$ as the one formed by vertically stacking the elements below and including the diagonal elements of B. The mat operator is defined as the inverse to the vec operator (the particular range will be evident from the context). The Euclidean norm of B is defined as $\|B\| = (\text{vec}'(B)\text{vec}(B))^{1/2}$. The Mahalanobis norm is defined as $\|B\|_\Theta = (\text{vec}'(B)\Theta^{-1}\text{vec}(B))^{1/2}$ for $\Theta \in \mathbb{P}^{nm}_+$. The L^2 norm is defined as $\|B\|_2 = \max_{\|x\|=1} \|Bx\|$. The singular values of B are denoted by $\sigma_1(B) \geq \sigma_2(B) \geq \cdots \geq \sigma_m(B) \geq 0$. The condition number of B is defined as $\text{cond}(B) = \sigma_1(B)/\sigma_r(B)$, where $r = \text{rank}(B)$. When $B \in \mathbb{S}^m$, we denote the eigenvalues of B as $\lambda_1(B) \geq \lambda_2(B) \geq \cdots \geq \lambda_m(B)$. The duplication matrix D_m is the mapping $\text{vech}(B) \mapsto \text{vec}(B)$ over $B \in \mathbb{S}^m$. The Moore–Penrose inverse of $B \in \mathbb{R}^{n\times m}$ is denoted by $B^\dagger$. For any $B \in \mathbb{G}^{n\times m}$ with $n > m$, an orthogonal complement $B_\perp$ is any matrix in $\mathbb{G}^{n\times n-m}$ satisfying $B'_\perp B = 0$. The column space of $B \in \mathbb{R}^{n\times m}$ is denoted by $\text{span}(B)$. The

orthogonal projection onto span(B) is denoted by P_B.[1]

Finally, we say that a sequence of random matrices $X_T \in \mathbb{R}^{n\times m}$ indexed by T is bounded away from zero in probability if for all $\varepsilon > 0$, there exists a $\delta_\varepsilon > 0$ and a $T_\varepsilon \geq 0$ such that $P(\|X_T\| > \delta_\varepsilon) > 1 - \varepsilon$ for all $T \geq T_\varepsilon$. It is easy to show that $\|X_T\|^{-1}$ is bounded away from zero in probability if and only if $X_T = O_p(1)$. This suggests the notation $X_T = O_p^{-1}(1)$. The product of two $O_p^{-1}(1)$ sequences is again $O_p^{-1}(1)$ and $a_T\|X_T\| \xrightarrow{p} \infty$ for any non–random sequence $a_T \to \infty$. The deterministic version, $O^{-1}(1)$, is defined similarly.

3 Rank Testing Under Standard Asymptotics

In this section, we will study the general structure of all rank tests under standard asymptotics. We will show that their behaviour is completely governed by implicit null space estimators. The behaviour of null space estimators is studied in great detail. We then turn to its application to rank testing. Before we do that, however, we must fix a few ideas.

We will draw inference on matrices B in either $\mathbb{R}^{m\times m}$, $\mathbb{S}^m$, or $\mathbb{P}^m$. The particular parameter space will be evident from the context. For $0 \leq r < \min\{n, m\}$, we will be interested in testing the hypothesis

$$H_0(r) : B = B^*, \quad \text{rank}(B^*) = r$$

against the global alternative

$$H_1(r) : B = B^*, \quad \text{rank}(B^*) > r$$

as well as the local alternative

$$H_T(r) : B = B^* + D/\sqrt{T} \in \mathbb{R}^{n\times m}, \quad \text{rank}(B^*) = r.$$

The matrix D is assumed to be an element of the parameter space. $H_0(r)$ is considered the special case of $H_T(r)$ where $D = 0$.

The distinction between B and B^* is necessary under the local alternative as B will typically have a higher rank than B^* for a fixed T. For future reference, the reader may wish to keep in mind that B^* refers to a matrix of determined rank.

[1]The matrix analysis utilized in this paper derives mainly from Horn & Johnson (1985, 1991), Stewart & Sun (1990), and Golub & Van Loan (1996). Research for this paper also relied heavily on Dennis Bernstein's magnificent treatise (Bernstein, 2009) although only primary sources are cited in this paper.

3.1 Preliminary Examples

The next few examples, review the core statistics in the literature and form the basis of our investigation. The technical details are relegated to Appendix A.

Example 1 (Anderson (1951))**.** Suppose $\{\varepsilon_t : t \geq 1\}$ is an i.i.d. sequence of $N(0, \Sigma)$ random vectors with $\Sigma \in \mathbb{G}^{n\times n}$. Let $\{x_t : t \geq 1\}$ be a non–random sequence of vectors such that $\widehat{\Gamma} = T^{-1}\sum_{t=1}^{T} x_t x_t' \in \mathbb{G}^{m\times m}$ and $\lim_{T\to\infty}\widehat{\Gamma} = \Gamma \in \mathbb{G}^{m\times m}$. Now define

$$y_t = Bx_t + \varepsilon_t, \qquad t = 1, \ldots, T.$$

Let $\widehat{B}$ and $\widehat{\Sigma}$ be the maximum likelihood estimators of B and Σ and recall that the asymptotic variance of $\sqrt{T}\mathrm{vec}(\widehat{B} - B)$ is $\Omega = \Gamma^{-1} \otimes \Sigma$, which we estimate by $\widehat{\Omega} = \widehat{\Gamma}^{-1} \otimes \widehat{\Sigma}$. The likelihood ratio statistic for testing the hypothesis $H_0(r)$ against $H_1(r)$ derived by Anderson (1951) can be expressed as

$$T \sum_{i=1}^{\min\{n,m\}-r} \log\left(1 + \sigma_i^2\left(\widehat{N}_r'\widehat{B}\widehat{M}_r\right)\right),$$

where $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ act on $\widehat{B}$ by producing a diagonal matrix $\widehat{N}_r'\widehat{B}\widehat{M}_r$ with the smallest $\min\{n, m\} - r$ canonical correlations between y and x as the diagonal elements. These are the linear combinations of y and x that are least correlated (Reinsel & Velu, 1998). A few more lines of algebra furnish us with the following equivalent expression for the likelihood ratio statistic

$$LR\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) = T \sum_{i=1}^{\min\{n,m\}-r} \log\left(1 + \sigma_i^2\left((P_{\widehat{N}_r}\widehat{\Sigma}P_{\widehat{N}_r})^{\dagger/2}P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\widehat{\Gamma}^{-1}P_{\widehat{M}_r})^{\dagger/2}\right)\right).$$

Anderson (1951) also noticed that under $H_0(r)$, LR (we will drop the arguments when there is no danger of confusion) has the same limiting distribution as the statistic

$$A\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) = T\left\|(P_{\widehat{N}_r}\widehat{\Sigma}P_{\widehat{N}_r})^{\dagger/2}P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\widehat{\Gamma}^{-1}P_{\widehat{M}_r})^{\dagger/2}\right\|^2,$$

known as the trace statistic. In the context of cointegration (although it applies equally well here) Johansen (1991) proposed the maximum eigenvalue statistic

$$J\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) = T\left\|(P_{\widehat{N}_r}\widehat{\Sigma}P_{\widehat{N}_r})^{\dagger/2}P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\widehat{\Gamma}^{-1}P_{\widehat{M}_r})^{\dagger/2}\right\|_2^2.$$

These new expressions for LR, A, and J illustrate two important feature of rank test statistics. First, they are functions of matrices $\widehat{N}_r$ and $\widehat{M}_r$ that estimate the null spaces of B (we

give a precise definition later on). Second, they are functions of a standardized $P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}$. To see this more clearly, simply note that $\text{vec}\left((P_{\widehat{N}_r}\widehat{\Sigma}P_{\widehat{N}_r})^{\dagger/2}P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\widehat{\Gamma}^{-1}P_{\widehat{M}_r})^{\dagger/2}\right) = \left((P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\right)^{\dagger/2}\text{vec}(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r})$. □

Economic data is often heteroskedastic and/or autocorrelated. In that case, the asymptotic variance of $\widehat{B}$ is no longer of Kronecker product form and the limiting distribution of the statistics above are no longer pivotal. Robin & Smith (2000) prove this for the LR and A statistics and we will prove this for the J statistic later on. This has motivated a variety of alternative statistics, the first of which was the following.

Example 2 (Cragg & Donald (1996))**.** $H_0(r)$ holds if and only if the application of the LU decomposition with complete pivoting to B produces an upper triangular matrix whose lower right $(n-r)\times(m-r)$ block consists of zeros. Heuristically then, the same algorithm applied to $\widehat{B}$ should produces an $O_p(T^{-1/2})$ matrix that is asymptotically normal and centred at the $(n-r)\times(m-r)$ zero matrix. As the algorithm is performed by row and column operations, there exists matrices $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ such that $\widehat{N}_r'\widehat{B}\widehat{M}_r \in \mathbb{R}^{(n-r)\times(m-r)}$ is the matrix produced by the algorithm. Cragg & Donald (1996) then propose the statistic

$$T\text{vec}'(\widehat{N}_r'\widehat{B}\widehat{M}_r)\{(\widehat{M}_r\otimes\widehat{N}_r)'\widehat{\Omega}(\widehat{M}_r\otimes\widehat{N}_r)\}^{-1}\text{vec}(\widehat{N}_r'\widehat{B}\widehat{M}_r),$$

where $\widehat{\Omega}$ is an estimator of the asymptotic variance of $\widehat{B}$. With a bit of simple algebra, one can rewrite the statistic as

$$F\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}\right) = T\text{vec}'(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r})\{(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\}^{\dagger}\text{vec}(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}).$$

Note that $\widehat{N}_r$ and $\widehat{M}_r$ have the same functions as in Example 1. They estimate the null spaces of B. Note that $A\left(\widehat{B},\widehat{\Sigma},\widehat{\Gamma},P_{\widehat{N}_r},P_{\widehat{M}_r}\right) = F\left(\widehat{B},\widehat{\Gamma}^{-1}\otimes\widehat{\Sigma},P_{\widehat{N}_r},P_{\widehat{M}_r}\right)$ so that F is a generalization of A. □

The statistics proposed by Robin & Smith (1995), Robin & Smith (2000), and Kleibergen & Paap (2006) are all of the form F. We will show in the next subsection that the minimum distance statistic proposed by Cragg & Donald (1997) is also of the F form.

Example 2 suggests the following modifications to LR and J,

$$\begin{aligned}
&LRA\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}\right)\\
&\quad= T\sum_{i=1}^{\min\{n,m\}-r}\log\left(1+\sigma_i^2\left(\mathrm{mat}\left(((P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r}))^{\dagger/2}\mathrm{vec}(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r})\right)\right)\right)\\
&JA\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}\right)= T\left\|\mathrm{mat}\left(((P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r}))^{\dagger/2}\mathrm{vec}(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r})\right)\right\|_2^2.
\end{aligned}$$

These statistics are robust to heteroskedasticity and autocorrelation under correct specification and simplify to the LR and J respectively when $\widehat{\Omega}$ is of Kronecker product form.[2]

Example 3 (Donald et al. (2007))**.** Suppose $B\in\mathbb{S}^m$ in Example 1 and is estimated subject to the symmetry restriction so that $\widehat{B}\in\mathbb{S}^m$. Donald et al. (2007) apply the same idea as in Example 2, except that they preserve symmetry by applying identical operations to the rows and columns to produce $\widehat{M}_r'\widehat{B}\widehat{M}_r$. They then propose the test statistic

$$T\mathrm{vech}'(\widehat{M}_r'\widehat{B}\widehat{M}_r)\{D_{m-r}^{\dagger}(\widehat{M}_r\otimes\widehat{M}_r)'D_m\widehat{\Psi}D_m'(\widehat{M}_r\otimes\widehat{M}_r)D_{m-r}^{\dagger}{}'\}^{-1}\mathrm{vech}(\widehat{M}_r'\widehat{B}\widehat{M}_r),$$

where $\widehat{\Psi}$ is an estimator of the asymptotic covariance of $\mathrm{vech}(\widehat{B})$ and D_m is the duplication matrix. It can be shown that this statistic is exactly $F\left(\widehat{B},D_m\widehat{\Psi}D_m',P_{\widehat{M}_r},P_{\widehat{M}_r}\right)$. Camba-Mendez & Kapetanios (2005) construct a similar statistic.

When $B\in\mathbb{P}^m$, Donald et al. (2007) reason that $H_0(r)$ holds if and only if the sum of the $m-r$ smallest eigenvalues of B is zero and therefore propose the statistic $\sqrt{T}\sum_{i=r+1}^m\lambda_i(\widehat{B})$. If we collect the eigenvectors associated with the $m-r$ smallest eigenvalues in the matrix $\widehat{M}_r$, then we can express this statistic as $\sqrt{T}\mathrm{tr}(P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r})$. We may also standardize it as

$$t\left(\widehat{B},\widehat{\Psi},P_{\widehat{M}_r}\right)=\frac{\sqrt{T}\mathrm{tr}(P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r})}{\sqrt{\mathrm{vec}'(I_m)(P_{\widehat{M}_r}\otimes P_{\widehat{M}_r})D_m\widehat{\Psi}D_m'(P_{\widehat{M}_r}\otimes P_{\widehat{M}_r})\mathrm{vec}(I_m)}}.$$

Note that $t\left(\widehat{B},\frac{1}{m-r}(D_m'D_m)^{-1},P_{\widehat{M}_r}\right)$ yields the non–standardized statistic. We name this the t statistic because it is obtained from $\widehat{B}$ by straightforward multiplication, just like the t statistic for scalars. It is interesting to note that for general matrices, it is impossible to form t statistics that yield consistent tests of $H_0(r)$ against $H_1(r)$ when $r<\min\{n,m\}-1$ because

[2]One can avoid the computation of the projection matrices and the Moore–Penrose inverse in the LRA and JA statistics by expressing them as $T\sum_{i=1}^{\min\{n,m\}-r}\log\left(1+\sigma_i^2\left(\mathrm{mat}\left(((\widehat{M}_r\otimes\widehat{N}_r)\widehat{\Omega}(\widehat{M}_r\otimes\widehat{N}_r))^{-1/2}\mathrm{vec}(\widehat{N}_r'\widehat{B}\widehat{M}_r)\right)\right)\right)$ and $T\left\|\mathrm{mat}\left(((\widehat{M}_r\otimes\widehat{N}_r)\widehat{\Omega}(\widehat{M}_r\otimes\widehat{N}_r))^{-1/2}\mathrm{vec}(\widehat{N}_r'\widehat{B}\widehat{M}_r)\right)\right\|_2^2$ respectively.

the number of degrees of freedom in this case is greater than one.[3] t tests are possible for positive semi–definite matrices because the parameter space is restricted and positive semi–definite matrices behave much like scalars (Horn & Johnson, 1985, chapter 7).[4] It is important to note, however, that t statistics are not invariant to conformable rescaling of $\widehat{B}$ and $\widehat{\Omega}$, thus even a change of units can result in a different statistic.[5] □

It is perhaps no surprise that the statistics surveyed in the examples above should admit expressions in terms of $P_{\widehat{N}_r}$ and $P_{\widehat{M}_r}$. Under either $H_0(r)$ or $H_T(r)$, there are matrices $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ spanning the left and right null spaces of B^*. Since $H_0(r)$ or $H_T(r)$ are invariant to the choice of bases spanning $\mathrm{span}(N_r)$ and $\mathrm{span}(M_r)$, it is natural that the problem should depend on N_r and M_r only through the maximal invariants P_{N_r} and P_{M_r}, which in our case are estimated by $P_{\widehat{N}_r}$ and $P_{\widehat{M}_r}$. That is, it is natural that the statistics should inherit the invariance inherent in $H_0(r)$ or $H_T(r)$. The likelihood based theory of invariance in hypothesis testing is well developed (Ferguson (1967); Lehmann & Romano (2005)). However, a likelihood–based general theory of rank testing would fail to include far more tests than a Wald–like theory of rank testing that takes as its starting point a matrix estimator and (optionally) an estimator of its asymptotic variance. Therefore, this paper takes the Wald route and discusses the handful of non–nested rank tests (all of which are cointegration tests) in Section 4.

We will see that the asymptotic behaviour of rank test statistics in standard asymptotics is completely governed by the behaviour of $P_{\widehat{N}_r}$ and $P_{\widehat{M}_r}$. In particular, for every rank test statistic $T^\theta\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r})$, we will see that its asymptotics mimic those of an infeasible statistic $T^\theta\tau(\widehat{B},\widehat{\Omega},P_{N_r},P_{M_r})$, where N_r and M_r are population parameters.[6]

[3]For $r = \min\{n,m\} - 1$, we can formulate a t statistic based on the smallest singular value of $\widehat{B}$ as it can be obtained from $\widehat{B}$ by straightforward multiplication.

[4]Using t, we may also consider one–sided hypothesis tests, where the null hypothesis of positive definite B is tested against the alternative that the last $m-r$ eigenvalues of B are negative.

[5]Note that $t(I_2,I_3,I_2) = \sqrt{T}$ whereas $t(R'I_2R, D_2^\dagger(R\otimes R)'D_2I_3D_2'(R\otimes R)D_2^{\dagger\prime}, I_2) = \sqrt{T}\frac{a^2+1}{\sqrt{a^4+1}}$ when $R = \left[\begin{smallmatrix} a & 0 \\ 0 & 1\end{smallmatrix}\right]$.

[6]Previous versions of this paper termed the infeasible statistics classical statistics and the rank test statistics stochastic statistics. Thanks are due to Peter Boswijk for suggesting the change of terminology.

3.2 Estimating the Null Spaces

In this section, we will be concerned with estimating the null spaces of $B^* \in \mathbb{R}^{n\times m}$. We will maintain the assumption that $\widehat{B} \in \mathbb{G}^{n\times m}$. This is guaranteed if $\mathrm{vec}(\widehat{B})$ is a non–degenerate random vector (i.e. it has a continuous probability density function). It is also guaranteed if $\mathrm{vech}(\widehat{B})$ is non–degenerate when testing on $\mathbb{S}^m$ or $\mathbb{P}^m$.[7]

Definition 1 (Null Space Estimators)**.** For $0 \le r < \min\{n, m\}$, the random matrices $\widehat{N}_r \in \mathbb{R}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{R}^{m\times(m-r)}$ are respectively left and right null space estimators of a rank–r matrix in $\mathbb{R}^{n\times m}$ if they are both almost surely of full rank. When there is no possibility of confusion, we will refer to them simply as null space estimators. A matrix of full rank defines a unique column space, which in turn defines a unique orthogonal projection matrix (Rudin, 1986, Theorem 4.11), thus we will refer to $P_{\widehat{N}_r}$ and $P_{\widehat{M}_r}$ as null space estimators as well. □

The problem of estimating null spaces has a long history in the numerical analysis literature (Stewart, 1993; Golub & Van Loan, 1996; Hansen, 1998). The basic idea is illustrated in the following example.

Example 4. Suppose $\{\varepsilon_1, \varepsilon_2\} \subset (0, 1)$ and consider the set of matrices

$$\widehat{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \varepsilon_1 & 0 \\ 0 & 0 & \varepsilon_2 \end{bmatrix}, \quad \widehat{B}_2^{RRA} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \varepsilon_1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \widehat{B}_1^{RRA} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \widehat{B}_0^{RRA} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

When ε_2 is very small relative to ε_1 and 1, we may approximate $\widehat{B}$ by $\widehat{B}_2^{RRA}$ and estimate the null spaces by $\widehat{M}_2 = \widehat{N}_2 = [\,0 \quad 0 \quad 1\,]'$. When both ε_1 and ε_2 are small, we may approximate $\widehat{B}$ by $\widehat{B}_1^{RRA}$ and estimate the null spaces by $\widehat{M}_1 = \widehat{N}_1 = \left[\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]'$. Finally, the rank–0 approximation of $\widehat{B}$ is the zero matrix and we may estimate the null spaces by $\widehat{N}_0 = \widehat{M}_0 = I_3$.

Notice that the rank–2 approximation depends on the relative sizes of ε_1 and ε_2. If ε_1 is very small relative to ε_2 and 1, we estimate the null spaces by $\widehat{M}_2 = \widehat{N}_2 = [\,0 \quad 1 \quad 0\,]'$ instead.

[7] The set of matrices in $\mathbb{R}^{n\times m}$ of rank r is a submanifold of $\mathbb{R}^{nm}$ of dimension $nm - (n-r)(m-r)$ (Guillemin & Pollack, 1974, p. 27). It therefore has measure zero in $\mathbb{R}^{nm}$ for $r < \min\{n, m\}$ (Guillemin & Pollack, 1974, p. 45) and so the set of rank–deficient matrices in $\mathbb{R}^{n\times m}$ is of measure zero. By a similar argument, the set of matrices in $\mathbb{S}^m$ (resp. $\mathbb{P}^m$) of rank r is a submanifold (resp. intersection of submanifolds with boundaries) of $\mathbb{R}^{m(m+1)/2}$ of dimension $m(m+1)/2 - (m-r)(m-r+1)/2$ and therefore has measure zero in $\mathbb{R}^{m(m+1)/2}$ for $r < m$. So the set of rank–deficient matrices in $\mathbb{S}^m$ (resp. $\mathbb{P}^m$) is also of measure zero.

This implies that as $\varepsilon_1, \varepsilon_2 \to 0$, the null space estimators may fluctuate between $[\,0 \quad 0 \quad 1\,]'$ and $[\,0 \quad 1 \quad 0\,]'$ with no definite limit, although they will always be in the yz plane. This implies that null space estimators are nested.

If $\varepsilon_1 = 1$ and $\varepsilon_2 \to 0$, then the rank–1 approximation is not unique. One may choose either $\left[\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]'$ or $\left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right]'$ as the estimated null spaces. In either case, $P_{\widehat{N}_1} \widehat{B} P_{\widehat{M}_1}$ remains bounded away from zero as $\varepsilon_2 \to 0$.

Finally, as $\varepsilon_1, \varepsilon_2 \to 0$, it is clearly not possible for $\widehat{B}_0^{RRA}$ to approximate $\widehat{B}$ well in any meaningful sense and $P_{\widehat{N}_0} \widehat{B} P_{\widehat{M}_0}$ remains bounded away from zero as $\varepsilon_1, \varepsilon_2 \to 0$. $\square$

Thus, to estimate the null spaces of the rank–r matrix B^* based on $\widehat{B}$, we think of the latter as a perturbation of the former. If we can find a reduced rank approximation (RRA) $\widehat{B}_r^{RRA}$ of rank r that approximates $\widehat{B}$ well enough, then $\widehat{B}_r^{RRA}$ will approximate B^* well and we may obtain consistent estimates of the null spaces of B^* as the null spaces of $\widehat{B}_r^{RRA}$. Alternatively, if $i > r$ and $\widehat{B}_i^{RRA}$ continues to approximate $\widehat{B}$ well, then we may expect it to be consistent for B^* and we may expect the spans of the null space estimators to merge into the null spaces of B^*, although we cannot, in general, expect the null space estimators to converge. Finally, if $i < r$, then $\widehat{B}_i^{RRA}$ cannot possibly converge to B^* and a good RRA should ensure that $P_{\widehat{N}_i} \widehat{B} P_{\widehat{M}_i}$ is bounded away from zero.

There are essentially two types of RRAs: decomposition–based approximations and norm–based approximations. We discuss them briefly in turn. A more detailed discussion is relegated to Appendix A.

Definition 2 (Decomposition–based Approximations)**.** For $\widehat{B} \in \mathbb{G}^{n \times m}$, let

$$\widehat{B} = \widehat{U} \widehat{S} \widehat{V}', \tag{1}$$

where $\widehat{S} = \begin{bmatrix} \widehat{S}_{11} & \widehat{S}_{12} \\ 0 & \widehat{S}_{22} \end{bmatrix} \in \mathbb{R}^{n \times m}$ is upper triangular and $\widehat{U}$ and $\widehat{V}$ and their inverses are bounded. We further assume that if $\widehat{S}_{11} \in \mathbb{R}^{r \times r}$, then:

(i) There is a $K_1 > 0$, not dependent on $\widehat{B}$, such that $\sigma_r(\widehat{S}_{11}) \geq K_1 \sigma_r(\widehat{B})$.

(ii) There is a $K_2 > 0$, such that $\sigma_1(\widehat{S}_{22}) \leq K_2 \sigma_{r+1}(\widehat{B})$ and $K_2 = O(1)$ for any $\widehat{B}$ satisfying $\widehat{B} - \widehat{B}^* \to 0$ with $\widehat{B}^* = O(1)$, $\operatorname{rank}(\widehat{B}^*) = r$, and $\sigma_r(\widehat{B}^*) = O^{-1}(1)$.

The RRA suggested by this decomposition is then

$$\widehat{B}_r^{DBA} = \widehat{U} \begin{bmatrix} \widehat{S}_{11} & \widehat{S}_{12} \\ 0 & 0 \end{bmatrix} \widehat{V}'. \tag{2}$$

We refer to this RRA as a decomposition–based approximation (DBA). □

The idea behind DBAs is to apply elementary well–conditioned matrices (e.g. permutations, reflections, rotations, Gaussian elimination matrices) to $\widehat{B}$ to produce a triangular matrix that concentrates the effect of $\widehat{B}$ into the submatrix $\widehat{S}_{11}$ and leaves as little as possible in $\widehat{S}_{22}$. We will see in the process of proving Lemma 3 that condition (i) guarantees power of our rank tests, while (ii) guarantees approximability of $\widehat{B}$ as it approaches a rank–r matrix $\widehat{B}^*$, hence size. In this section $\widehat{B}^*$ will be fixed to B^*, however the definition is given more generally to allow for a sequence $\widehat{B}^*$ that varies with $\widehat{B}$ as this will be necessary for more general asymptotics. Existence and measurability of $\widehat{B}_r^{DBA}$ is guaranteed typically as a property of the algorithm used (e.g. LU and QR) but can also follow from analytic considerations (e.g. the spectral decomposition). Uniqueness may fail, although this will have no effect on our results as any solution will do.

The most important DBA is the singular value decomposition approximation, which takes $\widehat{U}$ and $\widehat{V}$ to be orthogonal and $\widehat{S}$ to be diagonal with non–negative elements in descending order. This RRA is utilized in Ratsimalahelo (2003), Kleibergen & Paap (2006), and Donald et al. (2007). We will refer to this RRA as the SVD approximation and denote it by $\widehat{B}_r^{SVD}$.

In the LU decomposition with complete pivoting, utilized by Cragg & Donald (1996), $\widehat{U}$ is the product of a well–conditioned lower triangular matrix and a permutation matrix, while $\widehat{V}$ is permutation matrix. Donald et al. (2007) use a similar LU algorithm with pivoting designed for symmetric matrices. We will refer to the collection of these RRAs by $\widehat{B}_r^{LU}$. Related to this is the block Gaussian elimination utilized in Kleibergen & van Dijk (1994), which is similar to the LU decomposition and satisfies the conditions of Definition 2 if the elimination matrix is nonsingular.[8] We will refer to these RRAs by $\widehat{B}_r^{BLU}$.

When $\widehat{B} \in \mathbb{S}^m$, we may also make use of the Spectral Decomposition Theorem, where $\widehat{S}$ is diagonal and $\widehat{U} = \widehat{V}$ is orthogonal. When the diagonal elements of $\widehat{S}$ are ordered by absolute value, we obtain $\widehat{B}_r^{SVD}$.[9] If the diagonal elements of $\widehat{S}$ are put in descending order we obtain

[8] Chapter 8 of Lucas (1996) provides an insightful discussion of this condition.

[9] This follows from the fact that the set of singular values of a symmetric matrix is the set of absolute values of

a different DBA, which we denote by $\widehat{B}_r^{EIG}$. This RRA satisfies the conditions of Definition 2 if $\widehat{B}$ approaches $\mathbb{P}^m$ in the limit. Donald et al. (2007) utilize both DBAs in their statistics.

There are also a number of DBAs that have never been used in rank testing statistics such as the QR and Cholesky decompositions. This is somewhat surprising as many of these DBAs are computationally efficient (Hansen, 1998).

Example 5. Suppose $\widehat{B} = \left[\begin{smallmatrix} 1 & 0.5 \\ -1 & 0.5 \end{smallmatrix}\right]$. The SVD of $\widehat{B}$ is given by $\left[\begin{smallmatrix} 0.71 & 0.71 \\ -0.71 & 0.71 \end{smallmatrix}\right]\left[\begin{smallmatrix} 1.41 & 0 \\ 0 & 0.71 \end{smallmatrix}\right]\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ so that $\widehat{B}_1^{SVD} = \left[\begin{smallmatrix} 1 & 0 \\ -1 & 0 \end{smallmatrix}\right]$. The LU decomposition with complete pivoting is $\left[\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 1 & 0.5 \\ 0 & 1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ so $\widehat{B}_1^{LU} = \left[\begin{smallmatrix} 1 & 0.5 \\ -1 & -0.5 \end{smallmatrix}\right]$. Finally, the QR decomposition with pivoting is $\left[\begin{smallmatrix} -0.71 & 0.71 \\ 0.71 & 0.71 \end{smallmatrix}\right]\left[\begin{smallmatrix} -1.41 & 0 \\ 0 & 0.71 \end{smallmatrix}\right]\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ so $\widehat{B}_1^{QR} = \left[\begin{smallmatrix} 1 & 0 \\ -1 & 0 \end{smallmatrix}\right]$. All three DBAs produce the same left null space estimators although the LU decomposition produces a different right null space estimators to the other two. □

Finally, we should note that not every matrix decomposition (1) satisfies conditions (i) and (ii) of Definition 2. For example, the Jordan canonical form for general matrices, the LU decomposition with partial or no pivoting, and the QR decomposition with no pivoting all fail to satisfy the conditions of Definition 2. The fact that all of the other DBAs surveyed above satisfy conditions (i) and (ii) is demonstrated in Appendix A.

We now present some general results that apply to all DBAs.

Lemma 1. Let $\widehat{B} \in \mathbb{G}^{n\times m}$, $B^* \in \mathbb{R}^{n\times m}$, $\text{rank}(B^*) = r$, and let $\widehat{N}_i$ and $\widehat{M}_i$ span the left and right null spaces of $\widehat{B}_i^{DBA}$ respectively.

(i) $\text{rank}(\widehat{B}_i^{DBA}) = i$.

(ii) As $\widehat{B} - B^* \to 0$, $\|\widehat{B} - \widehat{B}_i^{DBA}\| = O(\|\widehat{B} - B^*\|)$ for all $i \geq r$.

(iii) $(I_n - P_{\widehat{N}_r})P_{\widehat{N}_i} = 0$ and $(I_m - P_{\widehat{M}_r})P_{\widehat{M}_i} = 0$ for all $i \geq r$.

Lemma 1 (i) guarantees the existence of null space estimators $\widehat{N}_i \in \mathbb{G}^{n\times(n-i)}$ and $\widehat{M}_i \in \mathbb{G}^{m\times(m-i)}$ spanning the left and right null spaces of $\widehat{B}_i^{DBA}$. Lemma 1 (ii) proves that $\widehat{B}_i^{DBA}$ approximates $\widehat{B}$ well if i is greater than the rank of B^* and $\widehat{B}$ approaches B^*. Lemma 1 (iii) states that the estimated null spaces are nested in the sense that $\text{span}(\widehat{N}_i) \subset \text{span}(\widehat{N}_r)$ and $\text{span}(\widehat{M}_i) \subset \text{span}(\widehat{M}_r)$ whenever $i > r$. This result will allow us to prove dominance relationships between the various rank testing statistics.

the eigenvalues of the matrix, $\{\sigma_i(\widehat{B}) : i = 1, \ldots, m\} = \{|\lambda_i(\widehat{B})| : i = 1, \ldots, m\}$.

Definition 3 (Norm–based Approximations)**.** For $\widehat{B} \in \mathbb{G}^{n\times m}$, let

$$\widehat{B}_r^{CDA} \in \operatorname{argmin}\{\|\widehat{B} - A\|_\Theta : A \in \mathbb{R}^{n\times m}, \operatorname{rank}(A) \leq r\}, \tag{3}$$

where $\Theta \in \mathbb{P}_+^{nm}$. We term this the Cragg and Donald approximation (CDA), after Cragg & Donald (1997), who first proposed it in econometrics.[10] □

The idea behind the CDA is quite simply to find the closest rank–r matrix according to the Mahalanobis metric. Cragg & Donald (1997) restrict attention to the case where Θ is an estimator of the asymptotic covariance of $\widehat{B}$. We place no such restriction. Θ is to be understood simply as a weighting matrix. The existence of the CDA can be proven by standard methods (Cragg & Donald, 1997). Measurability also follows from standard methods (White, 1994, Theorem 2.11). Uniqueness, on the other hand, may not hold, although again this will have no bearing on our results.

The CDA nests a number of other RRAs as special cases. When Θ is the identity matrix we obtain the SVD approximation. More generally, when Θ is a Kronecker product of an $m \times m$ and an $n \times n$ matrix, we obtain the RRAs implicit in Bartlett (1947), Anderson (1951), and Izenman (1975). This RRA receives its most explicit statement in Robin & Smith (2000) and so we will refer to it as the Robin and Smith decomposition (RSD) approximation and denote it by $\widehat{B}_r^{RSD}$. The null space estimators in Example 1 are obtained from $\widehat{B}$ by an RSD with $\Theta = \widehat{\Gamma}^{-1} \otimes \widehat{\Sigma}$.

When Θ is not of Kronecker product form, there are no known analytical solutions. However, we detail a novel iterative scheme for obtaining the CDA in Appendix A, which works quite well in numerical experiments.

Example 6. Suppose $\widehat{B} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \varepsilon \end{smallmatrix}\right]$, $\Theta = I_4$, and $|\varepsilon| < 1$, then $\widehat{B}_1^{CDA} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$. Suppose now that $\Theta = \left[\begin{smallmatrix} I_3 & 0 \\ 0 & \delta^2 \end{smallmatrix}\right]$, with $\delta < \varepsilon$, then the higher relative weight on $\widehat{B}_{(2,2)}$ compensates for its smaller relative size in the approximation and so $\widehat{B}_1^{CDA} = \left[\begin{smallmatrix} 0 & 0 \\ 0 & \varepsilon \end{smallmatrix}\right]$. □

Next, we provide general results that apply to all CDAs.

Lemma 2. Let $\widehat{B} \in \mathbb{G}^{n\times m}$, $B^* \in \mathbb{R}^{n\times m}$, $\operatorname{rank}(B^*) = r$, $\Theta \in \mathbb{P}_+^{nm}$, and let $\widehat{N}_i$ and $\widehat{M}_i$ span the left and right null spaces of $\widehat{B}_i^{CDA}$ respectively.

(i) For all i, $\operatorname{rank}(\widehat{B}_i^{CDA}) = i$.

[10] A precursor to this RRA is the one proposed by Gabriel & Zamir (1979), although they take Θ to be diagonal.

Table 1: Reduced Rank Approximations Utilized in Rank Test Statistics.

Paper	Statistic	RRA	Paper	Statistic	RRA
Bartlett (1947)	F	RSD	Anderson (1951)	LR	RSD
Anderson (1951)	A	RSD	Johansen (1988)	F	RSD
Stock & Watson (1988)	NA†	EIG	Johansen (1991)	J	RSD
Kleibergen & van Dijk (1994)	F	BLU	Robin & Smith (1995)	A	BLU
Cragg & Donald (1996)	F	LU	Yang & Bewley (1996)	F	RSD
Yang & Bewley (1996)	J	RSD	Cragg & Donald (1997)	F	CDA
Bierens (1997)	J	RSD	Lucas (1997, 1998)	NA†	BLU
Caner (1998)	F	RSD	Quintos (1998)	F	RSD
Gonzalo & Pitarakis (1999)	κ	RSD	Lütkepohl & Saikkonen (1999)	F	RSD
Robin & Smith (2000)	κ	RSD	Nyblom & Harvey (2000)	t	EIG
Boswijk & Lucas (2002)	NA†	BLU	Kleibergen & Paap (2006)	F	SVD
Donald et al. (2007)	F	LU	Donald et al. (2007)	F	SVD
Donald et al. (2007)	F	CDA	Donald et al. (2007)	t	EIG
Avarucci & Velasco (2009)	F	SVD	Cavaliere et al. (2010a)	F	RSD
Cavaliere et al. (2010b)	F	RSD	Cavaliere et al. (2014)	F	RSD

†These statistics have more complicated expressions than LRA, F, JA, and t. They are discussed in Section 4.

(ii) If $\text{cond}(\Theta) = O(1)$ as $\widehat{B} - B^* \to 0$, then $\|\widehat{B} - \widehat{B}_i^{CDA}\| = O(\|\widehat{B} - B^*\|)$ for all $i \geq r$.

(iii) If $\text{cond}(\Theta) = O(1)$ as $\widehat{B} - B^* \to 0$, then $(I_n - P_{\widehat{N}_r})P_{\widehat{N}_i} = O(\|\widehat{B} - \widehat{B}_r^{CDA}\|)$ and $(I_m - P_{\widehat{M}_r})P_{\widehat{M}_i} = O(\|\widehat{B} - \widehat{B}_r^{CDA}\|)$ for all $i \geq r$.

(iv) For all i, $T\|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta^2 = F\left(\widehat{B}, \Theta, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right)$.

As before, Lemma 2 (i) implies that the null space estimators obtained from DBAs are well defined. Lemma 2 (ii) states that so long as i is greater than the rank of B^* and Θ does not give disproportionate weights in the approximation, $\widehat{B}_i^{CDA}$ is a good approximations for $\widehat{B}$ as $\widehat{B}$ approaches B^*. Lemma 2 (iii) proves that the null space estimators are asymptotically nested as $\widehat{B}$ approaches B^* in the sense that $\text{span}(\widehat{N}_i)$ merges into $\text{span}(\widehat{N}_r)$ and $\text{span}(\widehat{M}_i)$ merges into $\text{span}(\widehat{M}_r)$ as $\widehat{B}$ approaches B^* and $\text{rank}(B^*) = r < i$. This is a somewhat weaker result than Lemma 1 (iii) but will still allow us to prove asymptotic dominance results.[11] Finally, Lemma 2 (iv) tells us that the Cragg & Donald (1997) statistic is an F statistic. Markovsky & Van Huffel (2007) find a similar representation.

To summarize, Table 1 lists the implicit null space estimators in a selection of rank test

[11]In addition to its failure to satisfy nestedness of the subspace estimators, $(\widehat{B}_i^{CDA})_r^{CDA} \neq \widehat{B}_r^{CDA}$ for $i > r$ in general. The two known exceptions to this are the SVD and RSD as these can also be considered DBAs.

statistics in the literature. We are now able to extract the following lemma which describes the asymptotics of null space estimators.

Lemma 3. Let $\widehat{B}$ be an estimator of $B^* \in \mathbb{R}^{n\times m}$ such that $\widehat{B} \in \mathbb{G}^{n\times m}$, and $\sqrt{T}(\widehat{B} - B^*) = O_p(1)$. Let $\text{rank}(B^*) = r$ and let $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of B^* respectively. Let the RRAs $\{\widehat{B}_i^{RRA} : 0 \leq i < \min\{n, m\}\}$ be either DBAs or CDAs. In the latter case, we assume that $\text{cond}(\Theta) = O_p(1)$.

(i) $\sqrt{T}(\widehat{B} - \widehat{B}_r^{RRA})$, $\sqrt{T}(P_{\widehat{N}_r} - P_{N_r})$, and $\sqrt{T}(P_{\widehat{M}_r} - P_{M_r})$ are $O_p(1)$.

(ii) If $0 \leq i < r$ then $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$. If $n = m$ and $B^* \in \mathbb{P}^m$, then $P_{\widehat{M}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$ and $P_{\widehat{N}_i}\widehat{B}P_{\widehat{N}_i} = O_p^{-1}(1)$.

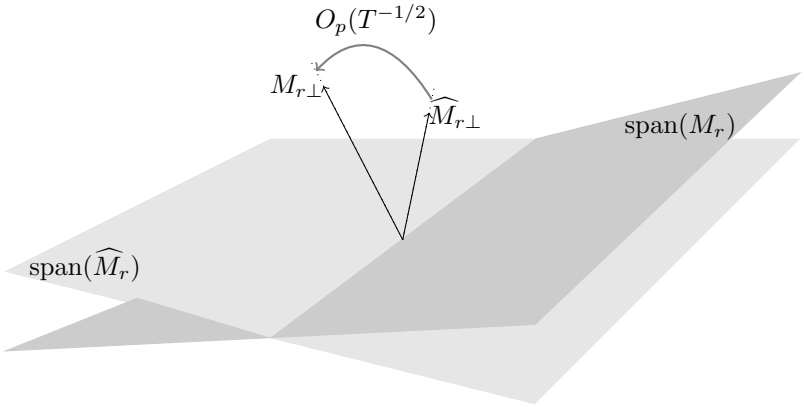
(iii) If $0 \leq i < r$ and $\widehat{B}_i^{RRA} - (B^*)_i^{RRA} = o_p(1)$, then $P_{\widehat{N}_i} - P_{N_i^*} = o_p(1)$ and $P_{\widehat{M}_i} - P_{M_i^*} = o_p(1)$, where N_i^* and M_i^* span the left and right null spaces of $(B^*)_i^{RRA}$.

Lemma 3 (i) establishes the rate of convergence of CDAs and DBAs. It also establishes the rate of convergence of the null space estimators. It is well known that the norm of the difference of two orthogonal projection matrices can be thought of as the distance between the two associated subspaces. In particular, the L^2 norm of this difference is the sine of the largest angle between the two subspaces, while the square of the Euclidean norm of this difference is the sum of the squares of the sines of all of the canonical angles between the two subspaces (Stewart & Sun, 1990; Gohberg et al., 2006). Thus, Lemma 3 (i) proves that the canonical angles between the estimated and population null spaces converge to zero at a rate of $\sqrt{T}$ (see Figure 1). Dufour & Valery (2011) obtain similar results for the special case of $B^*, \widehat{B} \in \mathbb{P}^m$, and $\widehat{B}_i^{EIG}$, whereas Lemma 3 (i) applies more generally to non–definite and non–symmetric matrices as well as any RRA.

Lemma 3 (ii) states that the null space estimators successfully capture non–vanishing components of $\widehat{B}$ when the rank is underestimated. In the case of positive semi–definite B^* the left and right null space estimators are equally capable of capturing non–vanishing components of $\widehat{B}$.

When the RRA is continuous at B^*, Lemma 3 (iii) states that the non–vanishing components of $\widehat{B}$ can be estimated consistently as $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i} \xrightarrow{p} P_{N_i}B^*P_{M_i} \neq 0$. Unfortunately, continuity of an RRA at a given matrix can fail for a variety of reasons including lack of uniqueness but also for reasons that have to do with the algorithm used for its computation

Figure 1: Convergence of a Two Dimensional Null Space Estimator in $\mathbb{R}^3$.



(e.g. multiple pivots in the LU and QR algorithms). However, continuity is known to be generic for the SVD, RSD, and EIG RRAs (Stewart & Sun, 1990; Markovsky, 2012) and it can also be shown to be generic for simple DBAs such as the LU, Cholesky, and QR RRAs. No results are available for the general CDA, although one might well conjecture that continuity is generic for all RRAs.

3.3 The Plug–in Principle

We are now in a position to see how null space estimation affects rank test statistics. Consider first the following set of assumptions, which will be useful for analysing the asymptotics of rank test statistics for general matrices.

Assumptions A. $B^* \in \mathbb{R}^{n\times m}$. $\widehat{B} \in \mathbb{R}^{n\times m}$ and $\widehat{\Omega} \in \mathbb{S}^{nm}$ are estimators indexed by T. Each $\mathrm{vec}(\widehat{B}) \in \mathbb{R}^{nm}$ is a non–degenerate random vector. $\widehat{\Omega} \in \mathbb{P}_+^{nm}$ almost surely. $\sqrt{T}(\widehat{B} - B^*)$, $\widehat{\Omega}$, and $\widehat{\Omega}^{-1}$ are $O_p(1)$.

Assumptions A arises in the context of Examples 1 and 2. They also arise in much more general settings.

Example 7. In standard GMM or ML estimation, we have an estimator $\widehat{B}$, satisfying $\sqrt{T}(\widehat{B} - B) \xrightarrow{d} N(0, \Phi)$. Usually, $\widehat{\Omega}$ is a consistent estimator for Φ. However, Assumptions A also allows $\widehat{\Omega}$ to converge in probability to a different matrix than Φ, allowing for misspecification analysis *à la* White (1994). Note in particular that, just as in Robin & Smith (2000), Φ need not be positive definite. □

Example 8. Consider Example 2, where $\widehat{\Omega}$ is a nonparametric kernel–based estimator of the type considered in den Haan & Levin (1997) and Cushing & McGravey (1999). These estimators require the specification of a bandwidth that diverges to infinity but at a slower rate than the sample size. A recent literature has considered allowing the bandwidth to grow proportionally to the sample size. In that case, $\widehat{\Omega}$ fails to converge in probability although t and F statistics are asymptotically pivotal (Kiefer et al., 2000; Vogelsang, 2001; Kiefer & Vogelsang, 2002b,a, 2005). These fixed–bandwidth asymptotics are also allowed under Assumptions A. They are often called "fixed–b" to distinguish them from the "small–b" theory of Example 7. We illustrate these asymptotics in Section 5. □

We will also want to prove results for symmetric matrices, in which case, we will rely on the following set of assumptions.

Assumptions B. $B^* \in \mathbb{S}^m$. $\widehat{B} \in \mathbb{S}^{m\times m}$ and $\widehat{\Psi} \in \mathbb{S}^{m(m+1)/2}$ are estimators indexed by T. Each $\text{vech}(\widehat{B}) \in \mathbb{R}^{m(m+1)/2}$ is a non–degenerate random vector. $\widehat{\Psi} \in \mathbb{P}_+^{m(m+1)/2}$ almost surely. $\sqrt{T}(\widehat{B} - B^*)$, $\widehat{\Psi}$, and $\widehat{\Psi}^{-1}$ are $O_p(1)$. In this context, we will set $\Omega = D_m \Psi D_m'$ and $\widehat{\Omega} = D_m \widehat{\Psi} D_m'$.

Assumptions B were satisfied in the context of Example 3. They also arise naturally in the variety of contexts considered in Examples 7 and 8 (Donald et al., 2007).

Under Assumptions A and B, we will prove that all rank test statistics satisfy the following plug–in principle.

Definition 4 (The Plug–in Principle in Standard Asymptotics)**.** Suppose $\widehat{B} \in \mathbb{R}^{n\times m}$ and $\widehat{\Omega} \in \mathbb{R}^{nm\times nm}$ are estimators indexed by T and let $B^* \in \mathbb{R}^{n\times m}$. For a given $0 \le r < \min\{n, m\}$ and RRA scheme, let $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $\widehat{B}_r^{RRA}$. The weak plug–in principle for rank test statistics is said to hold for the rank test statistic $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$ relative to the null spaces of B^* if

(i) Under either $H_0(r)$ or $H_T(r)$, $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) - T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r}) = O_p(T^{-1/2})$, where $N_r \in \mathbb{G}^{n\times n-r}$ and $M_r \in \mathbb{G}^{m\times m-r}$ span the left and right null spaces of B^*.

(ii) Under $H_1(r)$, then $|\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})| = O_p^{-1}(1)$ if $|\tau(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r})| = O_p^{-1}(1)$, where $N_r \in \mathbb{G}^{n\times n-r}$ and $M_r \in \mathbb{G}^{m\times m-r}$ span the left and right null spaces of $(B^*)_r^{RRA}$.

It is said to satisfy the strong plug–in principle relative to the null spaces of B^* if additionally

(iii) Under $H_1(r)$, $\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) - \tau(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r}) = o_p(1)$, where $N_r \in \mathbb{G}^{n\times n-r}$ and $M_r \in \mathbb{G}^{m\times m-r}$ span the left and right null spaces of $(B^*)_r^{RRA}$. □

Condition (i) requires that the feasible and infeasible statistics differ from each other by no more than $O_p(T^{-1/2})$ under $H_0(r)$ and $H_T(r)$. This is much stronger than asymptotic equivalence in large sample statistics, which requires only that the two have the same limiting distribution (Lehmann & Romano, 2005, p. 577). We will see, however, that it is easily satisfied. Condition (ii) ensures that a rank test has power against $H_1(r)$ if the associated infeasible test has power. Condition (iii) strengthens (ii) in that it requires the feasible and infeasible statistics to diverge at the same rate under the global alternative.

The variant of the plug–in principle we will prove applies to the class of rank test statistics of the form $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) = T^\theta \kappa\left(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}, (P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\right)$, where κ satisfies the following set of assumptions.

Assumptions K. $\mathcal{P} \subseteq \mathcal{X} \subseteq \mathbb{R}^{n\times m}$. $\mathcal{P}$ is closed and convex. $\mathcal{Y} \subseteq \mathbb{R}^{nm\times nm}$. $\kappa : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ satisfies:

(i) There exist measurable functions L_1 and L_2 such that for all $X, \widehat{X} \in \mathcal{X}$ and $Y, \widehat{Y} \in \mathcal{Y}$,

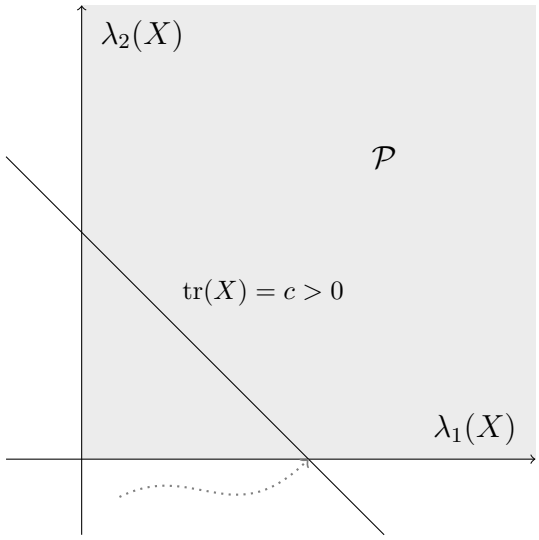
$$|\kappa(\widehat{X}, \widehat{Y}) - \kappa(X, Y)| \leq L_1(\widehat{X}, X, \widehat{Y}, Y)\|\widehat{X} - X\| + L_2(\widehat{X}, X, \widehat{Y}, Y)\|\widehat{Y} - Y\|.$$

For $\theta > 0$, $L_1(\widehat{X}, X, \widehat{Y}, Y) = O(\|X\|^{2\theta-1})$ and $L_2(\widehat{X}, X, \widehat{Y}, Y) = O(\|X\|^{2\theta})$ if $\|Y^\dagger\| = O(1)$ as $\|\widehat{X} - X\| + \|\widehat{Y} - Y\| \to 0$.

(ii) For every $C_1 > 0$ and $C_2 > 0$ there exists a $C > 0$ such that for all $X \in \mathcal{P}$ and $Y \in \mathcal{Y}$ with $\mathrm{vec}(X) \in \mathrm{span}(Y)$, $\|X\| \geq C_1$ and $\|Y\| \leq C_2$ imply that $|\kappa(X, Y)| \geq C$.

The Lipschitz inequality in condition (i) of Assumptions K allows the weak plug–in principle to hold under the null and local alternatives. It also allows the strong plug–in principle to hold under the global alternative. The boundedness condition in (ii) allows the feasible and infeasible rank tests to have power against $H_1(r)$. Note that boundedness of κ away from zero is only ensured on $\mathcal{P} \times \mathcal{Y}$ rather than the potentially larger set $\mathcal{X} \times \mathcal{Y}$. This is to allow for the t test, which has power against positive semi–definite matrices of rank higher than r but not against general matrices of rank higher than r. That is, the t test has power under $H_1(r)$ not just because $P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r} = O_p^{-1}(1)$ but also because $P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r}$ approaches $\mathbb{P}^m$. Figure 2 illustrates the case of $m = 2$ where $P_{\widehat{M}_0}\widehat{B}P_{\widehat{M}_0} = \widehat{B}$ is initially non–definite but converges along

Figure 2: The $\mathcal{P}$ Set of the t Statistic.



the dotted curve to a rank–1 matrix in $\mathcal{P}$ and stays bounded away from the origin. Another possibility is for convergence to proceed into the interior of $\mathcal{P}$ and away from the origin, when $\mathrm{rank}(B^*) = 2$.

It is easily verified that the class of all rank test statistics where Assumptions K are satisfied includes all of the statistics we have considered so far (see Table 2) and many more.

Table 2: Examples of Rank Testing Statistics in the Literature.

τ	θ	$\mathcal{P}$	$\mathcal{X}$	$\mathcal{Y}$	$\kappa(X,Y)$
LR	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^m \otimes \mathbb{P}^n$	$\sum_{i=1}^{\min\{n,m\}-r} \log(1+\sigma_i^2(\mathrm{mat}(Y^{\dagger/2}\mathrm{vec}(X))))$
LRA	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^{nm}$	$\sum_{i=1}^{\min\{n,m\}-r} \log(1+\sigma_i^2(\mathrm{mat}(Y^{\dagger/2}\mathrm{vec}(X))))$
A	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^m \otimes \mathbb{P}^n$	$\lVert Y^{\dagger/2}\mathrm{vec}(X)\rVert^2$
F	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^{nm}$	$\lVert Y^{\dagger/2}\mathrm{vec}(X)\rVert^2$
J	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^m \otimes \mathbb{P}^n$	$\lVert \mathrm{mat}(Y^{\dagger/2}\mathrm{vec}(X))\rVert_2^2$
JA	1	$\mathbb{R}^{n\times m}$	$\mathbb{R}^{n\times m}$	$\mathbb{P}^{nm}$	$\lVert \mathrm{mat}(Y^{\dagger/2}\mathrm{vec}(X))\rVert_2^2$
t	$\frac{1}{2}$	$\mathbb{P}^m$	$\mathbb{S}^{m\times m}$	$\mathbb{P}^{m(m+1)/2}$	$\frac{\mathrm{tr}(X)}{\mathrm{vec}'(I_m)Y\mathrm{vec}(I_m)}$

Example 9. A class of statistics asymptotically equivalent to the one that satisfies Assumptions K is the class of statistics proposed by Robin & Smith (2000), where $\kappa(X,Y) = \lVert Y^{\dagger/2}\mathrm{vec}(X)\rVert^2 + O(\lVert Y^{\dagger/2}\mathrm{vec}(X)\rVert^3)$ as $Y^{\dagger/2}\mathrm{vec}(X) \to 0$. Here, $\theta = 1$, $\mathcal{P} = \mathcal{X} = \mathbb{R}^{n\times m}$, and $\mathcal{Y} = \mathbb{P}^{nm}$. The advantage of this class is that it yields χ^2 asymptotic distributions under the usual conditions. □

Example 10. We can take any vector norm $\varphi : \mathbb{R}^{nm} \to \mathbb{R}$ and formulate a rank test statistic with $\theta = 1$ and $\kappa(X,Y) = \varphi^2(Y^{\dagger/2}\mathrm{vec}(X))$. Here, $\theta = 1$, $\mathcal{P} = \mathcal{X} = \mathbb{R}^{n\times m}$, and $\mathcal{Y} = \mathbb{P}^{nm}$. Rank

test statistics in this class are either bounded in probability for every norm φ or divergent for every φ.[12] The asymptotic distributions based on this choice of τ will not generally be standard (an example is the limit of J in Example 1). However, under the usual assumptions, it will be asymptotically pivotal and therefore obtainable either analytically or by simulation. $\square$

We can now state and prove the first main theorem of the paper.

Theorem 1. Suppose Assumptions K hold along with either Assumptions A or B. Suppose the null space estimators $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ are obtained by either a DBA or a CDA with $\text{cond}(\Theta) = O_p(1)$. Suppose the following inclusions hold almost surely

$$P_{N_r}\widehat{B}P_{M_r} \in \mathcal{X} \qquad (P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r}) \in \mathcal{Y}$$

$$P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r} \in \mathcal{X} \qquad (P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}) \in \mathcal{Y},$$

$\inf_{X\in\mathcal{P}} \|P_{N_r}\widehat{B}P_{M_r} - X\| = o_p(1)$, and $\inf_{X\in\mathcal{P}} \|P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r} - X\| = o_p(1)$. Then $T^\theta\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) = T^\theta\kappa\left(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}, (P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\right)$ satisfies the weak plug–in principle for rank test statistics. If, additionally, the RRA is continuous at B^*, then the statistic satisfies the strong plug–in principle.

A number of comments are in order.

First, it follows from Theorem 1 that all of the rank testing statistics of the standard asymptotics literature (and the many more we have listed above) satisfy the weak plug–in principle and when the underlying RRA is continuous at the population matrix (a feature we have conjectured to be generic for all RRAs), also satisfy the strong plug–in principle.

Second, it also follows from Theorem 1 that statistics that utilize different null space estimators differ from each other by $O_p(T^{-1/2})$ under $H_0(r)$ and $H_T(r)$. Thus, the Cragg & Donald (1996), Cragg & Donald (1997), and Kleibergen & Paap (2006) statistics do not only have the same limiting distribution under $H_0(r)$ and $H_T(r)$, but they are in fact asymptotically the same statistic. When $\widehat{\Omega}$ is of Kronecker product form, we may add to the list the statistics of Anderson (1951), Robin & Smith (1995), and Robin & Smith (2000). Thus, tests based on these statistics will exhibit the same asymptotic size and local power. In the symmetric case, we have additionally that all three F statistics proposed by Donald et al. (2007) are asymptotically the same statistic with the associated tests having equivalent asymptotic size and local power.

[12] All norms on a finite dimensional vector spaces are equivalent (Horn & Johnson, 1985, Corollary 5.4.5).

Third, as there are no first–order asymptotic differences between statistics that use different null space estimators, the researcher must rely on either small sample performance or numerical convenience in choosing the right test. In the latter case, we note that the CDA with non–Kronecker product weighting matrix is the most computational expensive of the RRAs considered in this paper. Next are the RSD, SVD, and EIG RRAs, which although much faster than the CDA, are not the most efficient computationally. The fastest available algorithms are the LU and QR algorithms (Hansen, 1998; Golub & Van Loan, 1996). Therefore, these latter algorithms are the recommended algorithms for high intensity computations such as the bootstrap.

Fourth, an immediate corollary of Theorem 1 is that the test for identification proposed by Wright (2003) does not have to be conducted using the Cragg & Donald (1997) statistic but can instead be done using the much simpler to compute Cragg & Donald (1996) or Kleibergen & Paap (2006) statistics. The same statistic can also be avoided in the rank estimator proposed by Cragg & Donald (1997).

Fifth, the weak plug–in principle simplifies the asymptotics of rank test statistics tremendously. It allows us to immediately see the asymptotic distribution under $H_0(r)$ and $H_T(r)$ – we simply derive the asymptotic distribution as if the population null spaces were known. It also allows us to obtain the asymptotics under $H_1(r)$ and misspecification. The strong plug–in principle, in turn, allows us (under possibly generic conditions) to obtain precise estimates of the rates of divergence of the statistics under $H_1(r)$. See Examples 15 and 16 for a Monte Carlo illustration of the weak and strong plug–in principles in standard asymptotics.

Sixth, the $\sqrt{T}$ consistency of the null space estimators is sufficient but not necessary for the plug–in principle to hold. If we are willing to relax condition (i) of the plug–in principle from $O_p(T^{-1/2})$ to $o_p(1)$, then only one of the null space estimators need be $\sqrt{T}$–consistent if the other is consistent (the proof of Theorem 1 makes this quite evident). However, one cannot do away with $\sqrt{T}$–consistency altogether as we demonstrate in the following example.

Example 11. Consider the non–standardized F statistic $F\left(\widehat{B}, I_{nm}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right)$ under Assumptions A. Even if the null space estimators are consistent under $H_0(r)$, if their convergence rate is slow enough, then the F statistic will diverge. Take for example $B = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and null space estimators $\widehat{N}_1 = \left[\begin{smallmatrix} \sin(\rho_T) \\ \cos(\rho_T) \end{smallmatrix}\right]$ and $\widehat{M}_1 = \left[\begin{smallmatrix} \sin(\nu_T) \\ \cos(\nu_T) \end{smallmatrix}\right]$, where $\rho_T, \nu_T \to 0$ as $T \to \infty$. Then $F\left(\widehat{B}, I_{nm}, P_{\widehat{N}_1}, P_{\widehat{M}_1}\right) = (\sqrt{T}\sin(\rho_T)\sin(\nu_T)\widehat{B}_{(1,1)} + O_p(1))^2$. But $|\sqrt{T}\sin(\rho_T)\sin(\nu_T)\widehat{B}_{(1,1)}| \geq$

$|\widehat{B}_{(1,1)}|\sqrt{T}|\rho_T\nu_T|/2$ for small enough ρ_T and ν_T. Thus if ρ_T and ν_T converges slowly enough that $\sqrt{T}|\rho_T\nu_T| \to \infty$ (e.g. $\rho_T = \nu_T = T^{-1/5}$), then the F statistic diverges to infinity. $\square$

Finally, we may relax the condition that $\widehat{\Omega}^{-1} = O_p(1)$ (resp. $\widehat{\Psi}^{-1} = O_p(1)$) under Assumptions A (resp. B). Here, there are two cases to consider: reducible singularity, which can be treated by rescaling (this is taken up in the next section), and irreducible singularity, which requires regularization (the terminology is due to Dufour & Valery (2011)). In the latter case, we may pursue the approach of Moore (1977) if we can ensure that $(P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r})$ satisfies the conditions of Andrews (1987). If not, we will need to substitute $Y^\dagger$ in Table 2 with one of the regularized inverses proposed by Lütkepohl & Burda (1997) or Dufour & Valery (2011). Recently, Duplinskiy (2014) has proposed avoiding regularization altogether and simply bootstrapping the non–standardized test statistics.

If the plug–in principle holds, one need only determine the asymptotics of the infeasible statistic in order to determine the asymptotics of the associated rank test statistic. We explore this in the following corollaries.

Corollary 1. Suppose Assumptions K and A hold and suppose we have null space estimators $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M_r} \in \mathbb{G}^{m\times(m-r)}$ obtained by either a DBA or a CDA with $\text{cond}(\Theta) = O_p(1)$. Under $H_0(r)$ or $H_T(r)$, let $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of B^*. Then if $T^\theta\kappa\left(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r}\right) \xrightarrow{d} \zeta$, then $T^\theta\kappa\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M_r}}\right) \xrightarrow{d} \zeta$. In particular, if

$$\left(\sqrt{T}\text{vec}(N_r'\widehat{B}M_r), (M_r \otimes N_r)'\widehat{\Omega}(M_r \otimes N_r)\right) \xrightarrow{d} (\xi_r, \Omega_r),$$

then we have

$$\begin{aligned} LRA\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M_r}}\right) &\xrightarrow{d} \|\xi_r\|^2_{\Omega_r} \\ F\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M_r}}\right) &\xrightarrow{d} \|\xi_r\|^2_{\Omega_r} \\ JA\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M_r}}\right) &\xrightarrow{d} \|\text{mat}(\Omega_r^{-1/2}\xi_r)\|^2_2. \end{aligned}$$

It follows from Corollary 1 that, in the context of Example 7, where $\sqrt{T}(\widehat{B} - B) \xrightarrow{d} N(0, \Phi)$ and $\widehat{\Omega}$ converges to a constant positive definite matrix, then LRA and F converge in distribution to the same quadratic form in $(n-r)(m-r)$ normal random variables, while JA converges to the L^2 norm of a random matrix with normal entries. Under correct specification (i.e. $\widehat{\Omega} \xrightarrow{p}$

Φ) and $H_0(r)$, $F \xrightarrow{d} \chi^2((n-r)(m-r))$ and $JA \xrightarrow{d} \|Z\|_2^2$, where $\mathrm{vec}(Z) \sim N(0, I_{(n-r)(m-r)})$. Under correct specification and $H_T(r)$, $F \xrightarrow{d} \chi^2\left((n-r)(m-r), \|N_r' D M_r\|^2_{(M_r \otimes N_r)'\Phi(M_r \otimes N_r)}\right)$ and $JA \xrightarrow{d} \left\|Z + \mathrm{mat}((M_r \otimes N_r)'\Phi(M_r \otimes N_r))^{\dagger/2}\mathrm{vec}(N_r' D M_r)\right\|_2^2$, with Z as before. For the limiting distribution of F under $H_0(r)$ and $H_T(r)$ and incorrect specification (i.e. $\widehat{\Omega} \xrightarrow{p} \Omega \neq \Phi$), the reader is referred to Lemma 8.2 of White (1994). The limiting distributions of JA under $H_0(r)$ and $H_T(r)$ and incorrect specification is non–standard and does not appear to simplify further than what is stated in the result above. Under fixed–b asymptotics $\widehat{\Omega}$ does not converge in probability. However, Corollary 1 continues to hold and the reader is referred to the literature cited in Example 8 for the limiting distributions (see also Examples 15 and 16).

Corollary 1 generalizes the misspecification results of Robin & Smith (2000), who consider the asymptotics of only the A statistic. It also generalizes the local power result of Cragg & Donald (1997), who consider only the F statistic that employs the CDA null space estimator. Finally, it allows for more general functional forms of κ than previously used in the literature.

Corollary 2. Suppose Assumptions K and B hold and suppose we have a null space estimator $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ obtained by either a DBA or a CDA with $\mathrm{cond}(\Theta) = O_p(1)$. Under $H_0(r)$ or $H_T(r)$, let $M_r \in \mathbb{G}^{m\times(m-r)}$ span the null space of B^*. Then if $T^\theta \kappa\left(\widehat{B}, \widehat{\Omega}, P_{N_r}, P_{M_r}\right) \xrightarrow{d} \zeta$, then $T^\theta \kappa\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \zeta$. In particular, if

$$\left(\sqrt{T}\mathrm{vech}(M_r'\widehat{B}M_r), D^\dagger_{m-r}(M_r \otimes M_r)'\widehat{\Omega}(M_r \otimes M_r)D^\dagger_{m-r}{}'\right) \xrightarrow{d} (\xi_r, \Omega_r),$$

then we have

$$\begin{aligned} LRA\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r}\right) &\xrightarrow{d} \|\xi_r\|^2_{\Omega_r} \\ F\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r}\right) &\xrightarrow{d} \|\xi_r\|^2_{\Omega_r} \\ JA\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r}\right) &\xrightarrow{d} \|\mathrm{mat}(D_{m-r}\Omega_r^{-1/2}\xi_r)\|_2^2 \end{aligned}$$

and if M_r is chosen to have orthogonal columns then

$$t\left(\widehat{B}, \widehat{\Psi}, P_{\widehat{M}_r}\right) \xrightarrow{d} \frac{\mathrm{tr}(\mathrm{mat}(D_{m-r}\xi_r))}{(\mathrm{vec}'(I_{m-r})D_{m-r}\Omega_r D'_{m-r}\mathrm{vec}(I_{m-r}))^{1/2}}.$$

It follows from Corollary 2 that if $\mathrm{vech}(\widehat{B}-B) \xrightarrow{d} N(0,\Phi)$ and $\widehat{\Omega}$ converges in probability to a positive definite matrix, then LRA and F converge in distribution to the same quadratic form in $(m-r)(m-r+1)/2$ normal random variables, JA converges to the L^2 norm of a random matrix with normal entries, and t converges to a normal random variable. Under correct specification (i.e. $\widehat{\Psi} \xrightarrow{p} \Phi$) and $H_0(r)$, $F \xrightarrow{d} \chi^2((m-r)(m-r+1)/2)$, $JA \xrightarrow{d} \|Z\|_2^2$, where $Z = Z'$ almost

surely and $\text{vech}(Z) \sim N(0, I_{(m-r)(m-r+1)/2})$, and $t \xrightarrow{d} N(0,1)$. Under correct specification and $H_T(r)$, $F \xrightarrow{d} \chi^2\left((m-r)(m-r+1)/2, \|\text{vech}(M_r' D M_r)\|^2_{D^\dagger_{m-r}(M_r \otimes M_r)'\Phi(M_r \otimes M_r)D^\dagger_{m-r}{}'}\right)$, $JA \xrightarrow{d} \left\| Z + \text{mat}(D_{m-r}((D^\dagger_{m-r}(M_r \otimes M_r)'\Phi(M_r \otimes M_r)D^\dagger_{m-r}{}')^{-1/2}\text{vech}(M_r' D M_r))) \right\|_2^2$, with Z the same as before, and $t \xrightarrow{d} N\left(\frac{\text{tr}(M_r' D M_r)}{\text{vec}'(I_{m-r})(M_r \otimes M_r)'\Phi(M_r \otimes M_r)\text{vec}(I_{m-r})}, 1\right)$. For the limiting distribution of F and t under $H_0(r)$ and $H_T(r)$ and incorrect specification (i.e. $\Omega \neq \Phi$), the reader is referred again to Lemma 8.2 of White (1994). The limiting distributions of JA under $H_0(r)$ and $H_T(r)$ and incorrect specification is, again, not amenable to further simplification. For fixed–b asymptotics, the reader is referred to the literature cited in Example 8 (see also Example 17).

Donald et al. (2007) proved the $H_0(r)$ and $H_1(r)$ results for F and t in the case of normality and correct specification. Thus Corollary 2 extends their results in the direction of local power, misspecification, fixed–b asymptotics, and more general functional forms of the rank test statistics.

Some rank test statistics exhibit interesting dominance relationships. Under either Assumptions A or B, either $H_0(r)$ or $H_T(r)$, and $i \geq r$, the rank test statistics LR, F, and J have the property that

$$T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_i}, P_{\widehat{M}_i}) \leq T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) + O_p(\|P_{\widehat{N}_i}(I_n - P_{\widehat{N}_r})\| + \|P_{\widehat{M}_i}(I_m - P_{\widehat{M}_r})\|).^{13}$$

The non–standardized t statistic also has this property when $\widehat{B}$ is restricted to $\mathbb{P}^m$. Then, when the null spaces are estimated by DBA, Lemma 1 (iii) implies that $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_i}, P_{\widehat{M}_i}) \leq T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$. If, on the other hand, the null spaces are estimated by CDA, Lemma 2 (iii) implies that $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_i}, P_{\widehat{M}_i}) \leq T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) + O_p(T^{-1/2})$. In both cases, we have that under $H_0(r)$ or $H_T(r)$, every subsequence of $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_i}, P_{\widehat{M}_i})$ that converges in distribution is asymptotically distributed as a random variable stochastically dominated by the limiting distribution of $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$. This generalizes the results of Cragg & Donald (1997) and Donald et al. (2007), which establish the limiting stochastic dominance for F statistics with particular choices of the null space estimators and $\widehat{\Omega}$.[14] Cragg & Donald (1993) show that under the assumptions of Example 7, when $\widehat{\Omega}$ is of Kronecker product form, then F has an asymptotic distribution stochastically dominated by the $\chi^2((n-i)(m-i))$

[13] This can be seen by explicitly working out the plug–in principle for these statistics.

[14] Donald et al. (2007) establish a further stochastic dominance result for t, which is not nested here. Their result depends on stronger assumptions on the limiting distribution of $\widehat{B}$ than we are allowing ourselves here.

distribution. Cragg & Donald (1997) have noted this does not seem to generalize to general forms of $\widehat{\Omega}$.

The above dominance relationships can be used to deduce the fact that $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_i}, P_{\widehat{M}_i}) = O_p(1)$ under $H_0(r)$ or $H_T(r)$ when $i \geq r$. More generally, if κ in Theorem 1 satisfies $\kappa(X, Y) = O(\|X\|^{2\theta})$ whenever $Y^\dagger = O(1)$ as $\|X\| \to 0$, then we have that $T^\theta \tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_i}, P_{\widehat{M}_i}) = O_p(1)$ under either Assumptions A or B. All of the statistics in Table 2 satisfy this condition. This then allows us to estimate the rank of B using the index

$$I_\tau(i) = T^\theta \tau\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_i}, P_{\widehat{M}_i}\right) + f(T)g(i),$$

where g is strictly increasing, $f(T) \to \infty$ as $T \to \infty$, and $f(T)/T^\theta \to 0$. The intuition here that $T^\theta \tau\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_i}, P_{\widehat{M}_i}\right)$ can be thought of as measures of fit of the RRA as it is bounded for $i \geq \text{rank}(B)$ and unbounded for $i < \text{rank}(B)$. As r increases, however, the parametrization of the RRA increases and so a cost is added to prevent over–parametrization. Using standard model selection techniques and under the assumptions we have made so far on κ, it can be shown that $\text{argmin}\{I_\tau(i) : i = 0, \ldots, \min\{n, m\}\}$ is a consistent estimator of $\text{rank}(B)$ (see e.g. Cragg & Donald (1997)).

4 Rank Testing Under General Asymptotics

Cointegration presents some truly fascinating anomalies for rank testing. In the next few examples we will show that the framework of the last section cannot be applied verbatim. The examples will, however, point to the necessary generalization.

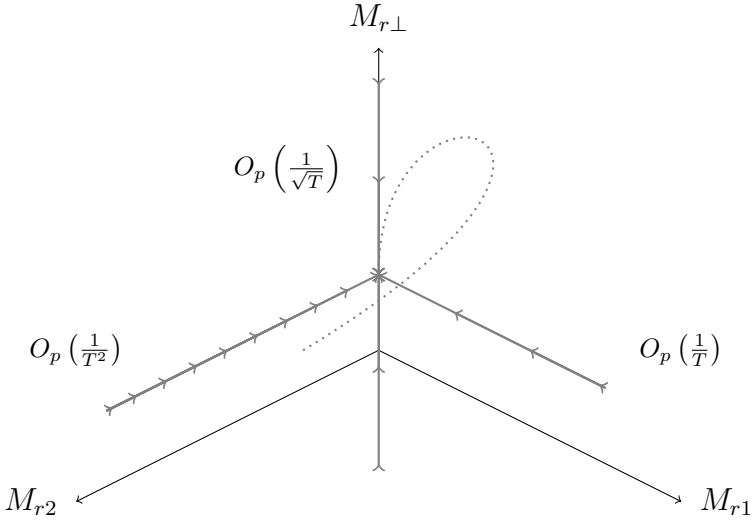
4.1 Preliminary Examples

Example 12 (Johansen (1988))**.** Let $\{\varepsilon_t : t \geq 1\}$ be i.i.d. $N(0, \Sigma)$, $\Sigma \in \mathbb{P}_+^m$, $y_0 = 0$, and

$$\Delta y_t = B y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T.$$

We assume that the roots of the characteristic polynomial of the system are either outside the unit circle or else at 1. Assume for the moment that the model generates data of order of integration no higher than 1 (see Theorem 4.2 of Johansen (1995a) for the conditions). Then $r = \text{rank}(B) < m$ is the number of cointegration relationships. Let $N_r \in \mathbb{G}^{m \times (m-r)}$ and

Figure 3: Convergence of a Column of $\widehat{B}$ in Example 12.



$M_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of B. Let $\widehat{B}$ be the maximum likelihood estimator of B and let $\widehat{\Omega} = \widehat{\Gamma}^{-1} \otimes \widehat{\Sigma}$, where $\widehat{\Sigma}$ and $\widehat{\Gamma}$ are as in Example 1.

It is easy to check that Johansen's (1988) trace statistic is $A\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right)$, where the null space estimators are the RSD estimators, and has the same limiting distribution as that of $A\left(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{N_r}, P_{M_r}\right)$. This suggests that the plug–in principle holds for cointegration. Unfortunately, however, $\widehat{\Omega}$ converges to a singular matrix so Assumptions A fail. In particular, $\widehat{\Omega}(M_r \otimes I_n)$ converges in probability to zero. On closer inspection, however, we find that $\widehat{\Omega}$'s rate of convergence along its asymptotic null space is exactly equal to $\widehat{B}$'s rate of convergence along its asymptotic right null space. That is, $\widehat{\Omega}(M_r \otimes I_n)$ and $\widehat{B}M_r$ are each $O_p(T^{-1})$. This counterbalancing of the accelerated rates of convergence is of crucial importance in the theory of cointegration rank testing.

Now suppose that the model generates data of order of integration no higher than 2 (see Theorem 4.6 of Johansen (1995a) for the conditions). Then Johansen (1995b) finds that $M_r = [\ M_{r1} \quad M_{r2}\ ]$, where $\widehat{\Omega}(M_{r1} \otimes I_n)$ and $\widehat{B}M_{r1}$ are $O_p(T^{-1})$ and $\widehat{\Omega}(M_{r2} \otimes I_n)$ and $\widehat{B}M_{r2}$ are $O_p(T^{-2})$. Thus, there may be heterogenous rates of accelerated convergence that need to be taken into account. A hypothetical column of $\widehat{B}$ might converge along the dotted curve in Figure 3. □

The phenomenon illustrated in Example 12 is well known in cointegration (Johansen, 1995a) and in regressions with polynomial trends (Hamilton, 1994, Chapter 16). $\widehat{B}$ and $\widehat{\Omega}$

shrink to zero along certain direction at exactly offsetting rates. Thus, all that is required to evaluate the asymptotics of the infeasible rank test statistic is to rescale $\widehat{B}$ and $\widehat{\Omega}$ by the appropriate power of T along the appropriate directions. For this reason, Dufour & Valery (2011) refer to the limiting singularity of $\widehat{\Omega}$ as reducible singularity. We will show that the conformably rescaled feasible statistics continues to mimic the infeasible statistic, thus proving the plug–in principle for cointegration. This will require a deeper analysis of the behaviour of null space estimators in the context of accelerated and possibly heterogeneous rates of convergence.

Example 13 (Nyblom & Harvey (2000))**.** Let $\{(\varepsilon_t', u_t')' : t \geq 1\}$ be a $2m$–dimensional sequence of i.i.d. $N\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}\Sigma & 0\\0 & B\end{smallmatrix}\right]\right)$ random vectors, $\Sigma \in \mathbb{G}^{m\times m}$, $x_0 = 0$, and

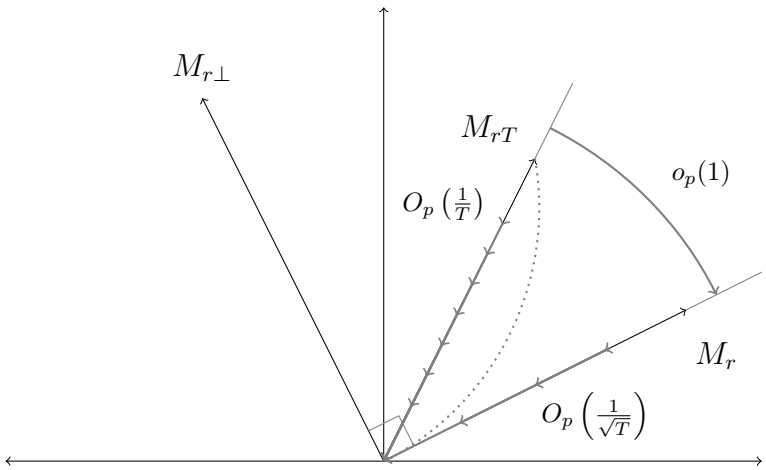
$$\begin{aligned} y_t &= x_t + \varepsilon_t \\ x_t &= x_{t-1} + u_t \end{aligned} \qquad t = 1, \ldots, T.$$

Then the rank of $B \in \mathbb{R}^{m\times m}$ determines the number of stochastic trends in the model. Let $M_r \in \mathbb{G}^{m\times(m-r)}$ span the null space of B. Let $\overline{y} = T^{-1}\sum_{t=1}^{T} y_t$, $\widehat{\Sigma} = T^{-2}\sum_{t=1}^{T}(y_t - \overline{y})(y_t - \overline{y})'$, and $\widehat{\Gamma} = T^{-4}\sum_{t=1}^{T}\sum_{s=1}^{t}(y_s - \overline{y})\sum_{s=1}^{t}(y_s - \overline{y})'$. We will work with $\widehat{B} = \widehat{\Sigma}^{-1/2}\widehat{\Gamma}\widehat{\Sigma}^{-1/2}$.

Nyblom and Harvey show that $\widehat{B}$ converges in distribution to a random matrix whose null space is exactly the span of M_r. As $[\ M_{r\perp} \quad \sqrt{T}M_r\ ]'\,\widehat{B}\,[\ M_{r\perp} \quad \sqrt{T}M_r\ ]$ converges in distribution to an almost surely positive definite matrix, they propose the test statistic $\mathrm{tr}(TP_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r})$, where $\widehat{M}_r$ is the eigenvalue null space estimator of B. Experience would then suggest that this statistic should mimic $\mathrm{tr}(TP_{M_r}\widehat{B}P_{M_r})$. Surprisingly, however, this is not the case. It would seem then that the plug–in principle fails.

In fact, the plug–in principle still holds but for a different matrix than M_r. One can check that $\widehat{B}M_r$ converges at a rate slower than $O_p(T^{-1})$. On the other hand, the Poincaré Separation Theorem (Horn & Johnson, 1985, Corollary 4.3.16) implies that along the eigenvectors associated with the smallest $m-r$ eigenvalues of $\widehat{B}$ our estimator is actually $O_p(T^{-1})$. That is, normalizing and collecting these eigenvectors in $M_{rT} \in \mathbb{G}^{m\times(m-r)}$, we have that $\widehat{B}M_{rT} = O_p(T^{-1})$. Therefore we find the surprising fact that M_r fails to capture the appropriate rate of convergence of $\widehat{B}$ to singularity and there are other directions along which $\widehat{B}$ converges faster (a hypothetical column of $\widehat{B}$ might converge along the dotted curve in Figure 4, although it does not converge along $M_{r\perp}$). Another, simpler, example of this is

Figure 4: Convergence of a Column of $\widehat{B}$ in Example 13.



$M_{rT} = \sqrt{T}\widehat{\Sigma}^{1/2}(I_m - M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma})M_r$, which is bounded in probability and satisfies $\widehat{B}M_{rT} = O_p(T^{-1})$. The algebraic intuition behind this choice is that it performs a Gaussian elimination of the troublesome (because of its slow convergence) off diagonal block $M'_{r\perp}\widehat{\Gamma}M_r$ from $[\ M_{r\perp} \quad M_r\ ]'\widehat{\Gamma}[\ M_{r\perp} \quad M_r\ ]$. In both cases, $\widehat{B}$ converges faster along M_{rT} than it does along M_r, even though $P_{M_{rT}}$ converges to P_{M_r}.[15] The crucial point to note here is that any reasonable subspace estimators will detect M_{rT} rather than M_r. Thus, the plug–in principle continues to hold, albeit for M_{rT} rather than M_r and the limiting distribution of $\mathrm{tr}(TP_{M_{rT}}\widehat{B}P_{M_{rT}})$ is precisely the limiting distribution of the Nyblom and Harvey statistic.[16] □

Example 13 suggests two additional features of cointegration. First, $\widehat{B}$ need not be consistent for B, so we must allow for statistics of varying rates of convergence. Second, rescaling should be allowed to occur along possibly random and T–varying directions.

Examples 12 and 13 show that rank test statistics have a slightly different scaling factor than in standard asymptotics. In standard asymptotics, we scaled by T^{θ}, where θ was determined by the curvature of the test statistic at the origin. Here, we must take into account accelerated rates of convergence to the origin and so we scale by $T^{2\gamma\theta}$, where θ is determined

[15]From the perspective of matrix perturbation theory (Stewart & Sun, 1990), there is nothing surprising about this at all. By the Almost Sure Representation Theorem (Lehmann & Romano, 2005, Theorem 11.2.19), we may think of $\widehat{B}$ as a perturbation of its probability limit, which has the null space $\mathrm{span}(M_r)$. This null space can be perturbed arbitrarily, as an invariant subspace, and does not have to remain constant as in the setting of Example 12. Thus, the real surprise is that the phenomenon of Example 13 does not occur more frequently.

[16]The technical details of these asymptotics can be found in the appendix.

by the curvature of τ and γ is the minimal rate of convergence along N_{rT} and M_{rT}. In Example 12, $\gamma = \frac{1}{2}$, while in Example 13, $\gamma = 1$. In short, the appropriate rank test statistics in general asymptotics are of the form $T^{2\gamma\theta}\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$.

Using these statistics, we will be interested in testing $H_0(r)$ against $H_1(r)$ and $H_T(r)$, which is now defined as

$$H_T(r) : B = B^* + T^{-\gamma-\omega}D \in \mathbb{R}^{n\times m}, \quad \text{rank}(B^*) = r,$$

where the faster rate of convergence imposed by $\omega \geq 0$ ensures that the local alternative does not stray too far away from the null because $\widehat{B}$ can converge at a faster rate than $T^{-\gamma}$ along certain rows and/or columns. In Example 12 the appropriate ω is $\frac{1}{2}$ for the $I(1)$ case and $\frac{3}{2}$ for the $I(2)$ case. In Example 13, the appropriate ω is 1.

4.2 Estimating the Null Spaces

For an RRA to produce good null space estimates based on a matrix $\widehat{B}$ that is converging at different rates along different (potentially non–constant) linear combinations of the rows and/or columns towards a matrix of reduced rank, it must (i) be a good approximation to $\widehat{B}$ and (ii) match the accelerated rates of convergence of $\widehat{B}$ along its columns and/or rows. We will first prove that the DBA and CDA are capable of delivering these two properties, then demonstrate the implications for subspace estimation.

Lemma 4. Let $\widehat{B} \in \mathbb{G}^{n\times m}$. Let $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ and suppose there are sequences $N_{rT} \in \mathbb{G}^{n\times(n-r)}$ and $M_{rT} \in \mathbb{G}^{m\times(m-r)}$, whose singular values are bounded away from zero and $P_{N_{rT}} \to P_{N_r}$ and $P_{N_{rT}} \to P_{N_r}$ as $\widehat{B}_T - \widehat{B}^*_T \to 0$, where $\widehat{B}_T = [\ N_{r\perp} \quad N_{rT}\ ]'\widehat{B}[\ M_{r\perp} \quad M_{rT}\ ]$ and $\widehat{B}^*_T = \begin{bmatrix} N'_{r\perp}\widehat{B}M_{r\perp} & 0 \\ 0 & 0 \end{bmatrix}$. Assume, moreover, that $\widehat{B}^*_T = O(1)$ and $\sigma_r(\widehat{B}^*_T) = \sigma_r(N'_{r\perp}\widehat{B}M_{r\perp})$ is bounded away from zero as $\widehat{B}_T - \widehat{B}^*_T \to 0$. Then as $\widehat{B}_T - \widehat{B}^*_T \to 0$ and for all $i \geq r$, $\widehat{B} - \widehat{B}^{DBA}_i$, $\widehat{B}^{DBA}_i M_{rT}$, and $N'_{rT}\widehat{B}^{DBA}_i$ are $O(\|\widehat{B}_T - \widehat{B}^*_T\|)$.

Lemma 4 generalizes Lemma 1 (ii) in two important respects. First, it does not rely on $\widehat{B}$ being a perturbation of a fixed reduced rank matrix. $\widehat{B}$ is only required to converge to zero along N_{rT} and M_{rT} but its component $N'_{r\perp}\widehat{B}M_{r\perp}$ is not even required to converge. This allows for the scenario we witnessed in Example 13, where $N_{r\perp}\widehat{B}M_{r\perp}$ converges in distribution but not in probability. In this setting, the lower bound on $\sigma_r(N_{r\perp}\widehat{B}M_{r\perp})$ is necessary in order to

ensure $\widehat{B}$ approaches a sequence of rank–r matrices whose null spaces are well approximated by N_{rT} and M_{rT}. When $\widehat{B}$ converges to a fixed rank–r matrix, as in Example 12 or Section 3, the lower bound is redundant. Second, Lemma 4 allows $\widehat{B}$ to converge along N_{rT} and M_{rT} at arbitrary rates. This allows for the accelerated rates of convergence we saw in Examples 12 and 13. In that regard, the condition that N_{rT} and M_{rT} have singular values bounded away from zero is important in order to to ensure that they specify proper directions along which $\widehat{B}$ goes to zero. Without it, $\widehat{B}M_{rT} \to 0$ could occur not because $\widehat{B}$ is converging to zero along M_{rT} but because certain line combinations of M_{rT} itself are shrinking.

In summary, Lemma 4 specializes to Lemma 1 (ii) if $N_{rT} = N_r$ and $M_{rT} = M_r$ span the left and right null spaces of B^* respectively and $\widehat{B} \to B^*$, in which case the lower bound on $\sigma_r(N_{r\perp}\widehat{B}M_{r\perp})$ is redundant.

The CDA also continues to perform well if Θ matches the accelerated rates of convergence along N_{rT} and M_{rT}. The intuition here is that the fast–converging components of $\widehat{B}$ should receive more weight in the approximation $\widehat{B}_r^{CDA}$. Otherwise, the CDA may miss important features of $\widehat{B}$.

Lemma 5. Let $\widehat{B}$, $N_{r\perp}$, $M_{r\perp}$, N_{rT}, M_{rT}, $\widehat{B}_T$, and $\widehat{B}_T^*$ be as in Lemma 4. Let $\Theta \in \mathbb{P}_+^{nm}$, let $\Theta_T = Z_T'\Theta Z_T$, where $Z_T = [\ M_{r\perp} \quad M_{rT}\ ] \otimes [\ N_{r\perp} \quad N_{rT}\ ]$, and assume $\text{cond}(\Theta_T) = O(1)$ as $\widehat{B}_T - \widehat{B}_T^* \to 0$. Then as $\widehat{B}_T - \widehat{B}_T^* \to 0$ and for all $i \geq r$, $\widehat{B} - \widehat{B}_i^{CDA}$, $\widehat{B}_i^{CDA}M_{rT}$, $N_{rT}'\widehat{B}_i^{CDA}$, $(I_n - P_{\widehat{N}_r})P_{\widehat{N}_i}$, and $(I_m - P_{\widehat{M}_r})P_{\widehat{M}_i}$ are $O(\|\widehat{B}_T - \widehat{B}_T^*\|)$, where $\widehat{N}_i$ and $\widehat{M}_i$ span the left and right null spaces of $\widehat{B}_i^{CDA}$ respectively.

Based on these results, we may now generalize Lemma 3 as follows.

Lemma 6. Let $\widehat{B}$ be an estimator indexed by T such that $\widehat{B} \in \mathbb{G}^{n\times m}$ almost surely and $\widehat{B} = O_p(1)$. Let $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ and suppose there exists sequences of possibly random matrices $N_{rT} \in \mathbb{G}^{n\times(n-r)}$ and $M_{rT} \in \mathbb{G}^{m\times(m-r)}$, whose singular values are bounded away from zero in probability, $P_{N_{rT}} \xrightarrow{p} P_{N_r}$ and $P_{N_{rT}} \xrightarrow{p} P_{N_r}$, and, for $\gamma > 0$,

$$\sigma_r(N_{r\perp}'\widehat{B}M_{r\perp}) = O_p^{-1}(1), \quad T^\gamma N_{rT}'\widehat{B} = O_p(1), \quad T^\gamma \widehat{B}M_{rT} = O_p(1), \quad T^\gamma N_{rT}'\widehat{B}M_{rT} = O_p(1).$$

Let the RRAs $\{\widehat{B}_i^{RRA} : 0 \leq i < \min\{n, m\}\}$ be either DBAs or CDAs. In the latter case, we assume that $\text{cond}(\Theta_T) = O_p(1)$, where $\Theta_T = Z_T'\Theta Z_T$ and $Z_T = [\ M_{r\perp} \quad M_{rT}\ ] \otimes [\ N_{r\perp} \quad N_{rT}\ ]$. Finally, set $\widehat{B}^* = [\ N_{r\perp} \quad N_{rT}\ ]^{-1\prime} \begin{bmatrix} N_{r\perp}'\widehat{B}M_{r\perp} & 0 \\ 0 & 0 \end{bmatrix} [\ M_{r\perp} \quad M_{rT}\ ]^{-1}$.

(i) $T^{\gamma}(\widehat{B}-\widehat{B}_r^{RRA})$, $T^{\gamma}(\widehat{B}-\widehat{B}^*)$, $T^{\gamma}(P_{\widehat{N}_r}-P_{N_{rT}})[\ N_{r\perp} \quad N_{rT}\ ]$, and $T^{\gamma}(P_{\widehat{M}_r}-P_{M_{rT}})[\ M_{r\perp} \quad M_{rT}\ ]$ are $O_p(1)$.

(ii) If $0 \leq i < r$ and the null spaces are estimated by the DBA, then $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$. If, on the other hand, the null spaces are estimated by CDA, then $P_{\widehat{N}_{iT}}\widehat{B}_T P_{\widehat{M}_{iT}} = O_p^{-1}(1)$ and $[\ N_{r\perp} \quad N_{rT}\ ]' P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}[\ M_{r\perp} \quad M_{rT}\ ] = O_p^{-1}(1)$. Here $\widehat{N}_{iT} = [\ N_{r\perp} \quad N_{rT}\ ]^{-1}\widehat{N}_i$, $\widehat{M}_{iT} = [\ M_{r\perp} \quad M_{rT}\ ]^{-1}\widehat{M}_i$, and $\widehat{B}_T = [\ N_{r\perp} \quad N_{rT}\ ]'\widehat{B}[\ M_{r\perp} \quad M_{rT}\ ]$.

If $n = m$, $N_{r\perp} = M_{r\perp}$, $N_{rT} = M_{rT}$, and $\inf_{X\in\mathbb{P}^r} \|M_{r\perp}'\widehat{B}M_{r\perp} - X\| = o_p(1)$, then $P_{\widehat{M}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$ and $P_{\widehat{N}_i}\widehat{B}P_{\widehat{N}_i} = O_p^{-1}(1)$ for the DBA and $P_{\widehat{M}_{iT}}\widehat{B}_T P_{\widehat{M}_{iT}} = O_p^{-1}(1)$ and $P_{\widehat{N}_{iT}}\widehat{B}_T P_{\widehat{N}_{iT}} = O_p^{-1}(1)$ for the CDA.

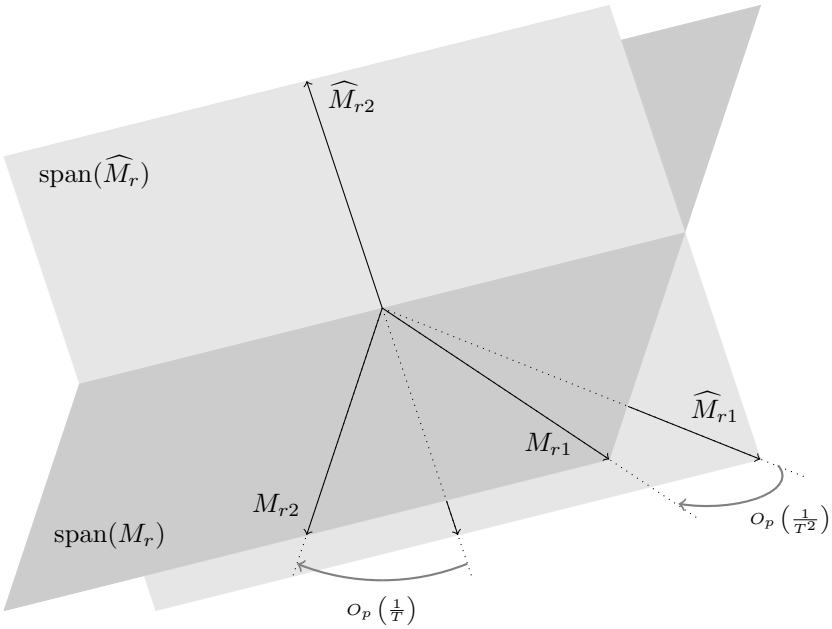
(iii) If, for $0 \leq i < r$, $\widehat{B}_i^{RRA} - (\widehat{B}^*)_i^{RRA} = o_p(1)$, then $P_{\widehat{N}_i} - P_{\widehat{N}_i^*} = o_p(1)$ and $P_{\widehat{M}_i} - P_{\widehat{M}_i^*} = o_p(1)$, where $\widehat{N}_i^*$ and $\widehat{M}_i^*$ span the left and right null spaces of $(\widehat{B}^*)_i^{RRA}$.

Lemma 6 specializes to Lemma 3 when $\gamma = \frac{1}{2}$, $N_{rT} = N_r$, $M_{rT} = M_r$, and $\sqrt{T}(\widehat{B} - B^*) = O_p(1)$. In the $I(1)$ case of Example 12, $\gamma = \frac{1}{2}$, $N_{rT} = N_r$, and $M_{rT} = T^{1/2}M_r$, while in the $I(2)$ case $M_{rT} = [\ T^{1/2}M_{r1} \quad T^{3/2}M_{r2}\ ]$. In Example 13, on the other hand, we have $\gamma = 1$ and $N_{rT} = M_{rT}$ can be chosen as $\sqrt{T}\widehat{\Sigma}^{1/2}(I_m - M_{r\perp}(M_{r\perp}'\widehat{\Gamma}M_{r\perp})^{-1}M_{r\perp}'\widehat{\Gamma})M_r$.

Lemma 6 (i) may seem peculiar compared to its counterpart in Lemma 3 (i). In fact, it is more general in two important respects. First, it allows for the estimated subspaces to converge at a potentially faster rate than the overall convergence rate of $\widehat{B}$. This is the analogue to super–consistency of point estimates in cointegration analysis. Here subspace estimates are super–consistent. For example, the left null space estimator in Example 12 converges at a rate of $\sqrt{T}$ while the right null space estimator in the $I(1)$ case converges at a rate of T as $T(P_{\widehat{M}_r} - P_{M_r}) = O_p(1)$. The estimated null spaces converge at a rate of T in Example 13 even though $\widehat{B}$ is inconsistent. Second, it allows the subspaces of span$(\widehat{M}_r)$ to converge at different rates. In the $I(2)$ case of Example 12, $T(P_{\widehat{M}_r}M_{r1} - M_{r1}) = O_p(1)$ and $T^2(P_{\widehat{M}_r}M_{r2} - M_{r2}) = O_p(1)$. Thus, it is possible to decompose $\widehat{M}_r$ as $[\ \widehat{M}_{r1} \quad \widehat{M}_{r2}\ ]$ with $T(P_{\widehat{M}_{r1}} - P_{M_{r1}}) = O_p(1)$ and $T^2(P_{\widehat{M}_{r2}} - P_{M_{r2}}) = O_p(1)$ (see Figure 5). The rates given in Lemma 6 (i) are therefore more parsed descriptions of the rates of convergence of the estimated subspaces.

Lemma 6 (ii) is different than Lemma 3 (ii) for the CDA in that we are only ensured the weaker result that $P_{\widehat{N}_{iT}}\widehat{B}_T P_{\widehat{M}_{iT}} = O_p^{-1}(1)$, which implies the even weaker result that $[\ N_{r\perp} \quad N_{rT}\ ]' P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}[\ M_{r\perp} \quad M_{rT}\ ] = O_p^{-1}(1)$. That is, we are only ensured that a rescaled

Figure 5: Heterogeneous Rates of Null Space Convergence in $\mathbb{R}^3$.



version of $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}$ is bounded away from zero in probability. When N_{rT} and M_{rT} are bounded in probability we still have that $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}$ is bounded away from zero in probability. When N_{rT} and M_{rT} are unbounded, we will see that the CDA is still capable of delivering power in rank tests.

Lemma 6 (iii) differs from its counterpart in Lemma 3 in that the comparison is to a sequence $\widehat{B}^*$ rather than the probability limit of $\widehat{B}$, which may not be defined (e.g. Example 13). For this reason, the strong plug–in principle for general asymptotics will apply to a more general non–vanishing component of $\widehat{B}$ than it did for a standard asymptotics.

4.3 The Plug–In Principle

The following assumptions generalize Assumptions A and B to the settings of Examples 12 and 13 as well as polynomial regressions.

Assumptions C. $B^* \in \mathbb{R}^{n\times m}$. $\widehat{B} \in \mathbb{R}^{n\times m}$ and $\widehat{\Omega} \in \mathbb{S}^{nm}$ are estimators indexed by T. Each $\text{vec}(\widehat{B}) \in \mathbb{R}^{nm}$ is a non–degenerate random vector and $\widehat{B} = O_p(1)$. $\widehat{\Omega} \in \mathbb{P}_+^{nm}$ almost surely. If $N_q \in \mathbb{G}^{n\times(n-q)}$ and $M_q \in \mathbb{G}^{m\times(m-q)}$ span the left and right null spaces of B^*, there exists sequences of possibly random matrices $N_{qT} \in \mathbb{G}^{n\times(n-q)}$ and $M_{qT} \in \mathbb{G}^{m\times(m-q)}$, whose singular values are bounded away from zero in probability, $P_{N_{qT}} \xrightarrow{p} P_{N_q}$ and $P_{M_{qT}} \xrightarrow{p} P_{M_q}$, and, for

$\gamma > 0$,

$$\sigma_q(N'_{q\perp}\widehat{B}M_{q\perp}) = O_p^{-1}(1), \qquad T^\gamma N'_{qT}\widehat{B} = O_p(1), \qquad T^\gamma \widehat{B}M_{qT} = O_p(1),$$
$$T^\gamma N'_{qT}\widehat{B}M_{qT} = O_p(1), \qquad Z'_T\widehat{\Omega}Z_T = O_p(1), \qquad (Z'_T\widehat{\Omega}Z_T)^{-1} = O_p(1),$$

where $Z_T = [\ M_{q\perp} \quad M_{qT}\ ] \otimes [\ N_{q\perp} \quad N_{qT}\ ]$.

The symmetric analogue is given by the following set of assumptions.

Assumptions D. $B^* \in \mathbb{S}^m$. $\widehat{B} \in \mathbb{S}^m$ and $\widehat{\Psi} \in \mathbb{S}^{m(m+1)/2}$ are estimators indexed by T. Each $\mathrm{vec}(\widehat{B}) \in \mathbb{R}^{mm}$ is a non–degenerate random vector and $\widehat{B} = O_p(1)$. $\widehat{\Psi} \in \mathbb{P}_+^{m(m+1)/2}$ almost surely. If $M_q \in \mathbb{G}^{m\times(m-q)}$ spans the null space of B^*, there exists a sequence of possibly random matrices $M_{qT} \in \mathbb{G}^{m\times(m-q)}$, whose singular values are bounded away from zero in probability, $P_{M_{qT}} \xrightarrow{p} P_{M_q}$, and, for $\gamma > 0$,

$$\sigma_q(M'_{q\perp}\widehat{B}M_{q\perp}) = O_p^{-1}(1), \qquad T^\gamma \widehat{B}M_{qT} = O_p(1), \qquad T^\gamma M'_{qT}\widehat{B}M_{qT} = O_p(1),$$
$$D_m^\dagger Z'_T D_m \widehat{\Psi} D'_m Z_T D_m^{\dagger\prime} = O_p(1), \qquad (D_m^\dagger Z'_T D_m \widehat{\Psi} D'_m Z_T D_m^{\dagger\prime})^{-1} = O_p(1),$$

where $Z_T = [\ M_{q\perp} \quad M_{qT}\ ] \otimes [\ M_{q\perp} \quad M_{qT}\ ]$. In this context, we will set $\widehat{\Omega} = D_m\widehat{\Psi}D'_m$.

Assumptions C and D specialize to Assumptions A and B when $\sqrt{T}(\widehat{B} - B^*) = O_p(1)$, $\gamma = \frac{1}{2}$, and N_{qT} and M_{qT} are constant. The assumptions of Lemma 6 are implied by both Assumptions C and D. Thus, we can find estimates of the limiting null spaces of $\widehat{B}$ by either the DBA or CDA (assuming that Θ satisfies the necessary condition) and obtain the rates of convergence $T^\gamma(P_{\widehat{N}_r} - P_{N_{rT}})[\ N_{r\perp} \quad N_{rT}\ ] = O_p(1)$ and $T^\gamma(P_{\widehat{M}_r} - P_{M_{rT}})[\ M_{r\perp} \quad M_{rT}\ ] = O_p(1)$ under either $H_0(r)$ or $H_T(r)$. Under $H_1(r)$, we have $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$ in the case the DBA and $P_{\widehat{N}_{rT}}\widehat{B}_T P_{\widehat{M}_{rT}} = O_p^{-1}(1)$ and $[\ N_{r\perp} \quad N_{rT}\ ]P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}[\ M_{r\perp} \quad M_{rT}\ ] = O_p^{-1}(1)$ in the case of the CDA.

Definition 5 (The Plug–in Principle in General Asymptotics)**.** Suppose $\widehat{B} \in \mathbb{R}^{n\times m}$ and $\widehat{\Omega} \in \mathbb{R}^{nm\times nm}$ are estimators indexed by T and let $B^* \in \mathbb{R}^{n\times m}$. Let $\widehat{B}^* \in \mathbb{R}^{n\times m}$ be a random sequence indexed by T whose null spaces converge in probability to the null spaces of B^* and let $N_{qT} \in \mathbb{G}^{n\times(n-q)}$ and $M_{qT} \in \mathbb{G}^{m\times(m-q)}$ span the left and right null spaces of $\widehat{B}^*$ respectively. For a given $0 \le r < \min\{n, m\}$ and RRA scheme, let $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $\widehat{B}_r^{RRA}$. The weak plug–in principle for rank test statistics is said to hold for the rank test statistic $T^{2\gamma\theta}\tau(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$ relative to the null spaces of $\widehat{B}^*$ if

(i) Under either $H_0(r)$ or $H_T(r)$, $T^{2\gamma\theta}\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r})-T^{2\gamma\theta}\tau(\widehat{B},\widehat{\Omega},P_{N_{rT}},P_{M_{rT}})=O_p(T^{-\gamma})$.

(ii) Under $H_1(r)$, then $|\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r})|=O_p^{-1}(1)$ if $|\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r^*},P_{\widehat{M}_r^*})|=O_p^{-1}(1)$, where $\widehat{N}_r^*\in\mathbb{G}^{n\times n-r}$ and $\widehat{M}_r^*\in\mathbb{G}^{m\times m-r}$ span the left and right null spaces of $(\widehat{B}^*)_r^{RRA}$.

It is said to satisfy the strong plug–in principle relative to the null spaces of $\widehat{B}^*$ if additionally

(iii) Under $H_1(r)$, $\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r})-\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r^*},P_{\widehat{M}_r^*})=o_p(1)$, where $\widehat{N}_r^*\in\mathbb{G}^{n\times n-r}$ and $\widehat{M}_r^*\in\mathbb{G}^{m\times m-r}$ span the left and right null spaces of $(\widehat{B}^*)_r^{RRA}$. □

Definition 5 generalizes Definition 4. When $\widehat{B}^*$ is fixed at B^* and N_{qT} and M_{qT} are both fixed and span the null spaces of B^*, the general asymptotics plug–in principle implies the standard asymptotics plug–in principle.

Note that the choice of the symbol γ is deliberate as it will turn out to be exactly the γ that appears in Lemmas 4 – 6. That is, the quality of the approximation implicit in the plug–in principle under $H_0(r)$ and $H_T(r)$ depends on the minimum rate of convergence of $\widehat{B}$ along N_{rT} and M_{rT}. This was $\frac{1}{2}$ under standard asymptotics and is different here due to the more general asymptotics.

The generalized set of assumptions and the generalized notion of the plug–in principle together allow us to generalize Theorem 1.

Theorem 2. Suppose Assumptions K hold along with either Assumptions C or D. Suppose the null space estimators $\widehat{N}_r\in\mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r\in\mathbb{G}^{m\times(m-r)}$ are obtained by either a DBA or a CDA with $\text{cond}(\Theta_T)=O_p(1)$. Let $\widehat{B}^*=[\ N_{q\perp}\quad N_{qT}\ ]^{-1\prime}\begin{bmatrix}N_{q\perp}'\widehat{B}M_{q\perp} & 0\\ 0 & 0\end{bmatrix}[\ M_{q\perp}\quad M_{qT}\ ]^{-1}$ and let $\widehat{N}_r^*\in\mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r^*\in\mathbb{G}^{m\times(m-r)}$ span its left and right null spaces respectively. Suppose the following inclusions hold almost surely

$$P_{\widehat{N}_r^*}\widehat{B}P_{\widehat{M}_r^*}\in\mathcal{X}\qquad\qquad (P_{\widehat{M}_r^*}\otimes P_{\widehat{N}_r^*})\widehat{\Omega}(P_{\widehat{M}_r^*}\otimes P_{\widehat{N}_r^*})\in\mathcal{Y}$$
$$P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}\in\mathcal{X}\qquad\qquad (P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\in\mathcal{Y}$$

And suppose either of the following two conditions hold

(i) For every $X\in\mathcal{X}$, $Y\in\mathcal{Y}$, $N\in\mathbb{G}^{n\times n}$, and $M\in\mathbb{G}^{m\times m}$, $\widetilde{X}=P_{N^{-1}X}N'XMP_{M^{-1}X'}\in\mathcal{X}$, $\widetilde{Y}=(P_{M^{-1}X'}M\otimes P_{N^{-1}X}N)'Y(P_{M^{-1}X'}M\otimes P_{N^{-1}X}N)\in\mathcal{Y}$, and $\kappa(\widetilde{X},\widetilde{Y})=\kappa(X,Y)$. Moreover, $\inf_{X\in\mathcal{P}}\|P_{\widehat{N}_{rT}^*}\widehat{B}_TP_{\widehat{M}_{rT}^*}-X\|=o_p(1)$ and $\inf_{X\in\mathcal{P}}\|P_{\widehat{N}_{rT}}\widehat{B}_TP_{\widehat{M}_{rT}}-X\|=o_p(1)$, where $\widehat{N}_{rT}^*=[\ N_{q\perp}\quad N_{qT}\ ]^{-1}\widehat{N}_r^*$, $\widehat{M}_{rT}^*=[\ M_{q\perp}\quad M_{qT}\ ]^{-1}\widehat{M}_r^*$, $\widehat{N}_{rT}=[\ N_{q\perp}\quad N_{qT}\ ]^{-1}\widehat{N}_r$, $\widehat{M}_{rT}=[\ M_{q\perp}\quad M_{qT}\ ]^{-1}\widehat{M}_r$, $\widehat{B}_T=[\ N_{q\perp}\quad N_{qT}\ ]'\widehat{B}[\ M_{q\perp}\quad M_{qT}\ ]$.

(ii) $Z_T = O_p(1)$, $\inf_{X\in\mathcal{P}} \|P_{\widehat{N}_r^*}\widehat{B}P_{\widehat{M}_r^*} - X\| = o_p(1)$, and $\inf_{X\in\mathcal{P}} \|P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r} - X\| = o_p(1)$.

Then $T^{2\gamma\theta}\tau(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}) = T^{2\gamma\theta}\kappa\left(P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r},(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r}\otimes P_{\widehat{N}_r})\right)$ satisfies the weak plug–in principle relative to the sequence $\widehat{B}^*$. If, additionally, the distance between $\widehat{B}^*$ and the points of discontinuity of the rank–r RRA is bounded away from zero in probability, then the statistic satisfies the strong plug–in principle relative to $\widehat{B}^*$.

Theorem 2 is strictly more general than Theorem 1. When Assumptions C and D specialize to Assumptions A and B respectively, Theorem 1 is a special case of Theorem 2 (ii).

Condition (i) of Theorem 2 is an invariance condition that allows the plug–in principle to hold in the context of Example 12, where both $\widehat{B}$ and $\widehat{\Omega}$ must be rescaled conformably in order to evaluate the asymptotics. The set of transformations in this condition may seem peculiar. However, they are simple manifestations of the invariance of all of the statistics we have considered so far (except for t) with respect to the group of transformations

$$(\widehat{B},\widehat{\Omega},\widehat{N}_r,\widehat{M}_r) \mapsto (N'\widehat{B}M,(M\otimes N)'\widehat{\Omega}(M\otimes N),N^{-1}\widehat{N}_r,M^{-1}\widehat{M}_r),$$

where $N\in\mathbb{G}^{n\times n}$ and $M\in\mathbb{G}^{m\times m}$.[17] Thus, the set of transformations $(X,Y)\mapsto(\widetilde{X},\widetilde{Y})$ with respect to which κ is invariant defines a group.[18]

Condition (ii), on the other hand, allows the plug–in principle to hold in the context of Example 13, where Z_T is bounded in probability and the invariance conditions in (i) do not hold. This condition also allows the plug–in principle to hold in standard asymptotics.

The plug–in principle applied to the context of Example 12 allows one to simply plug–in the limiting null space of $\widehat{B}$. The plug–in principle applies regardless of the order of integration of the process and not only to the F statistic but also to the JA and LRA statistics. It also applies in the contexts of added lags and arbitrary deterministic terms such as polynomial trends and dummies. The plug–in principle in Example 13 applies relative to a random sequence rather than a constant one. In particular, one cannot plug–in the null space of B^*. That is because under either $H_0(r)$ or $H_T(r)$, $P_{\widehat{M}_r} - P_{M_r} = O_p(T^{-1/2})$, which is too slow for the plug–in principle to work. One can, however, plug in $M_{rT} = \sqrt{T}\widehat{\Sigma}^{1/2}(I_m - M_{r\perp}(M_{r\perp}'\widehat{\Gamma}M_{r\perp})^{-1}M_{r\perp}'\widehat{\Gamma})M_r$ because $P_{\widehat{M}_r} - P_{M_{rT}} = O_p(T^{-1})$. See Example 19 for a Monte Carlo illustration.

[17] The t statistic is not invariant to this group of transformations but it is invariant to the subgroup of transformations where both N and M are non–zero scalars.

[18] See Ferguson (1967) and Lehmann & Romano (2005) for more on invariance in hypothesis testing.

A large class of statistics is nested under Theorem 2, including all of the standard asymptotics statistics of the literature as well as the majority of the cointegration rank statistics in the literature. In particular, it nests all of the statistics included in Table 1 except for the ones superscripted by the symbol †. Those, along with recent statistics by Hallin et al. (2012) and Boswijk et al. (2015), are of the form $T^{\theta}\tau(\{y_t : t = 1, \ldots, T\}, P_{\widehat{M}_r})$. Thus, they explicitly depend on a null space estimator and their dependence on the data is more complicated than what we have considered in this paper. However, it is evident from the proofs of the asymptotics of these results that these statistics are asymptotically equivalent to infeasible versions $T^{\theta}\tau(\{y_t : t = 1, \ldots, T\}, P_{M_r})$. Thus, the plug–in principle continues to hold for these statistics as well.

It follows from Theorem 2 that the Johansen (1988), Kleibergen & van Dijk (1994), and Kleibergen & Paap (2006) statistics, which differ from each other only in their implicit null space estimators, differ from each other by $O_p(T^{-1/2})$ under $H_0(r)$ and $H_T(r)$. Thus, the choice among these will have to depend on either Monte Carlo performance or numerical expedience as noted in Section 3.

It is important at this point to warn against an often repeated mistake in the cointegration literature, that any estimator of the cointegration space will do in working out the asymptotics of cointegration rank tests. This is clearly not the case of Example 13, where plugging in the true null spaces produces incorrect asymptotics (see Example 19). It is also not true in the context of Example 12, where the null space estimator must match the rate of convergence of the matrix itself. Example 11 can easily be modified to a cointegration example where the statistic diverges under $H_0(r)$.

The simplification to asymptotic analysis afforded by Theorem 2 is noteworthy. It allows the researcher to obtain the asymptotics not only for the different alternatives but also under misspecification. We summarize in the following corollaries.

Corollary 3. Suppose Assumptions K and C hold and suppose we have null space estimators $\widehat{N}_r \in \mathbb{G}^{n\times n-r}$ and $\widehat{M}_r \in \mathbb{G}^{m\times m-r}$ obtained by either a DBA or a CDA with $\text{cond}(\Theta_T) = O_p(1)$. Under $H_0(r)$ or $H_T(r)$, if $T^{2\gamma\theta}\kappa\left(\widehat{B}, \widehat{\Omega}, P_{N_{rT}}, P_{M_{rT}}\right) \xrightarrow{d} \zeta$, then $T^{2\gamma\theta}\kappa\left(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r}\right) \xrightarrow{d} \zeta$. In particular, if

$$\left(T^{\gamma}\text{vec}(N_{rT}'\widehat{B}M_{rT}), (M_{rT} \otimes N_{rT})'\widehat{\Omega}(M_{rT} \otimes N_{rT})\right) \xrightarrow{d} (\xi_r, \Omega_r),$$

then we have

$$T^{2\gamma-1}LRA\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\xi_r\|^2_{\Omega_r}$$
$$T^{2\gamma-1}F\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\xi_r\|^2_{\Omega_r}$$
$$T^{2\gamma-1}JA\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\mathrm{mat}(\Omega_r^{-1/2}\xi_r)\|^2_2.$$

Under $H_0(r)$, ξ_r and Ω_r in Corollary 3 are typically functionals of a Brownian motion and deterministic terms (if deterministic trends are included), while under $H_T(r)$ they are typically of the Ornstein–Uhlenbeck form (Hubrich et al., 2001). The limiting behaviour under $H_0(r)$ and $H_T(r)$ of all of the statistics in Johansen (1988), Johansen (1991), Kleibergen & van Dijk (1994), Yang & Bewley (1996), Quintos (1998), Gonzalo & Pitarakis (1999), Lütkepohl & Saikkonen (1999), Kleibergen & Paap (2006), Avarucci & Velasco (2009), and Cavaliere et al. (2010a) follow from Corollary 3. These results assume correct specification, so the limiting distributions above are nuisance–parameter–free. In the case of misspecification, the limiting distributions may not be free of nuisance parameters. It follows from Corollary 3 that the F statistics proposed by Johansen (1988), Kleibergen & van Dijk (1994), and Kleibergen & Paap (2006) have the exact same behaviour under the misspecification conditions of Caner (1998) (infinite variance shocks), Cavaliere et al. (2010b) (heteroskedastic shocks), and Aznar & Salvador (2002) and Cavaliere et al. (2014) (misspecified lag length).

Corollary 4. Suppose Assumptions K and D hold and suppose we have a null space estimator $\widehat{M}_r\in\mathbb{G}^{m\times m-r}$ obtained by either a DBA or a CDA with $\mathrm{cond}(\Theta_T)=O_p(1)$. Under $H_0(r)$ or $H_T(r)$, if $T^{2\gamma\theta}\kappa\left(\widehat{B},\widehat{\Omega},P_{N_r},P_{M_r}\right)\xrightarrow{d}\zeta$, then $T^{2\gamma\theta}\kappa\left(\widehat{B},\widehat{\Omega},P_{\widehat{N}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\zeta$. In particular, if

$$\left(T^{\gamma}\mathrm{vech}(M_{rT}'\widehat{B}M_{rT}),D^{\dagger}_{m-r}(M_{rT}\otimes M_{rT})'\widehat{\Omega}(M_{rT}\otimes M_{rT})D^{\dagger}_{m-r}{}'\right)\xrightarrow{d}(\xi_r,\Omega_r),$$

then we have

$$T^{2\gamma-1}LRA\left(\widehat{B},\widehat{\Omega},P_{\widehat{M}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\xi_r\|^2_{\Omega_r}$$
$$T^{2\gamma-1}F\left(\widehat{B},\widehat{\Omega},P_{\widehat{M}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\xi_r\|^2_{\Omega_r}$$
$$T^{2\gamma-1}JA\left(\widehat{B},\widehat{\Omega},P_{\widehat{M}_r},P_{\widehat{M}_r}\right)\xrightarrow{d}\|\mathrm{mat}(D_{m-r}\Omega_r^{-1/2}\xi_r)\|^2_2$$

and if $M_{rT}\xrightarrow{p}M_r$, a non–random matrix of orthonormal columns, then

$$T^{\gamma-1/2}t\left(\widehat{B},\widehat{\Omega},P_{\widehat{M}_r}\right)\xrightarrow{d}\frac{\mathrm{tr}(\mathrm{mat}(D_{m-r}\xi_r))}{(\mathrm{vec}'(I_{m-r})D_{m-r}\Omega_rD'_{m-r}\mathrm{vec}(I_{m-r}))^{1/2}}.$$

Similar observations apply to Corollary 4 as do to Corollary 3. The asymptotic distribution of the Bierens (1997) statistic under $H_0(r)$ and $H_T(r)$ follow from Corollary 4 as do the asymptotics of Nyblom & Harvey (2000) (see Example 19). As an application, we explicitly derive the asymptotics of the Nyblom & Harvey (2000) statistic under $H_T(r)$.

Example 14. Consider the setting of Example 13. Under $H_T(r)$, if $M_r \in \mathbb{G}^{m\times(m-r)}$ consists of orthonormal columns and spans the null space of B^*, the Nyblom & Harvey (2000) statistic converges in distribution to $\mathrm{tr}(C_{22} - C_{12}'C_{11}^{-1}C_{12})$, where

$$\begin{aligned}
C_{11} &= \int_0^1 \left[\int_0^u W_1^*(s)ds\right]\left[\int_0^u W_1^*(s)ds\right]' du \\
C_{12} &= \int_0^1 \left[\int_0^u W_1^*(s)ds\right] K'(u)du \\
C_{22} &= \int_0^1 K(u)K'(u)du \\
W_1^*(u) &= W_1(u) - \int_0^1 W_1(s)ds \\
K(u) &= W_2(u) - uW_2(1) + (M_r'\Sigma M_r)^{-1/2}(M_r'DM_r)^{1/2}\int_0^u W_3^*(s)ds,
\end{aligned}$$

and $(W_1', W_2', W_3')'$ is a standard Brownian motion whose three components have dimensions r, $m-r$, and $m-r$ respectively. The limiting distribution reduces to the one reported by Cappuccio & Lubian (2009) in the univariate case. As the distribution is invariant to the transformation $(K, W_3) \mapsto (MK, NW_3)$ whenever M and N are orthogonal matrices, it depends on D only through $\{\sigma_i((M_r'\Sigma M_r)^{-1/2}(M_r'DM_r)^{1/2})\}$, the singular values of the local alternative's signal–to–noise ratio along the null spaces of B^*.

Based on these asymptotics, we can formulate others statistics for testing rank in the context of Example 13. One example is $T^2\sum_{i=r+1}^m \lambda_i^2(\widehat{B})$, which is equivalent to $TF(\widehat{B}, I_{nm}, P_{\widehat{M_r}}, P_{\widehat{M_r}})$ with the null spaces estimated from the EIG or SVD RRAs and converges in distribution to $\mathrm{tr}((C_{22} - C_{12}'C_{11}^{-1}C_{12})^2)$. This statistic is illustrated in Example 19. Another example is $T^2\lambda_{r+1}^2(\widehat{B})$, which is equivalent to $TJ(\widehat{B}, I_n, I_m, P_{\widehat{M_r}}, P_{\widehat{M_r}})$ and converges in distribution to $\|C_{22} - C_{12}'C_{11}^{-1}C_{12}\|_2^2$. □

Finally, we note that the stochastic dominance relationships discussed at the end of subsection 3.3 continue to hold as do the rank estimation techniques, with $2\gamma\theta$ taking the place of θ everywhere. See Hubrich et al. (2001) for more details.

5 Monte Carlo

This section illustrates the theory presented in previous sections through a series of Monte Carlo experiments organized in examples. It is in no way intended as a study of the small sample performance of the various tests (that is a topic for a separate study). We follow Donald et al. (2007) in reporting PP plots, which plot nominal size, p, on the vertical axis against the observed rejection rate, $\alpha(p)$. In each case, we report PP plots for a number of statistics, including the infeasible statistic to which the plug–in principle applies. For computational efficiency, we have used the statistics in their simplified forms that avoid the computation of Moore–Penrose inverses and projection matrices (see footnote 2). The Matlab code for generating these plots is available at `www.econ.upf.edu/~alsadoon/`.

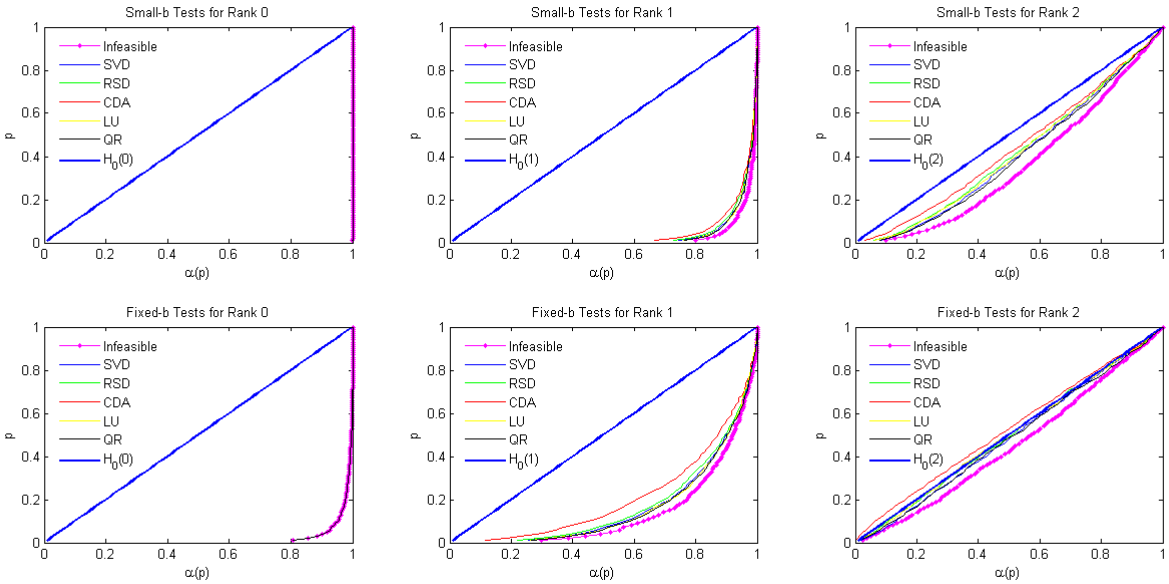
5.1 Standard Asymptotics

Example 15 (The Plug–in Principle for General Matrices)**.** Let $\{(x_t', u_t')' : t = 1, \ldots, 50\}$ be i.i.d. $N(0, I_8)$. Let $\{\varepsilon_t : t = 0, \ldots, T\}$ be a stationary process satisfying $\varepsilon_t = 0.5\varepsilon_{t-1} + u_t$. Let $B = \begin{bmatrix} 0.75 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ and $y_t = Bx_t + \varepsilon_t$ for $t = 1, \ldots, 50$. Our sample consists of $\{(y_t', x_t')' : t = 1, \ldots, 50\}$. The small size of the dataset and the moderate autocorrelation is at about the range where first order asymptotics usually begin to break down for an F statistic with 4 degrees of freedom. Thus, the specification is chosen to be challenging. We estimate B by OLS and the asymptotic variance of $\widehat{B}$ nonparametrically using a Bartlett kernel with the traditional bandwidth $\lfloor 4(50/100)^{1/4} \rfloor = 3$ for the small–b case and bandwidth T for the fixed–b case. The number of replications is set to 2000. For each replication, we computed F statistics based on the SVD, RSD, CDA, LU, and QR RRAs. For the CDA, we used Θ equal to the small–b estimate of the asymptotic variance matrix of $\widehat{B}$.

The limiting distribution of the rank–2 small–b F statistic is $\chi^2(4)$. The limiting distribution of the rank–2 fixed–b F statistic is $W'(1)\left(2\int_0^1 (W(s) - sW(1))(W(s) - sW(1))'ds\right)^{-1} W(1)$, where W is a standard Brownian motion of dimension 4 (Kiefer & Vogelsang, 2002a).

The strong plug–in principle is clearly in effect in this example. The third column of Figure 6 demonstrates that the feasible and infeasible statistics are practically the same under $H_0(r)$. The first and second columns, on the other hand, illustrate the strong plug–in principle under $H_1(r)$. Figure 6 also demonstrates the superior performance of fixed–b tests in terms of size.

Figure 6: PP Plots for the F tests of Example 15



This is consistent with the fixed–b theory and Monte Carlo evidence in the literature (Kiefer et al., 2000; Kiefer & Vogelsang, 2002a,b, 2005).

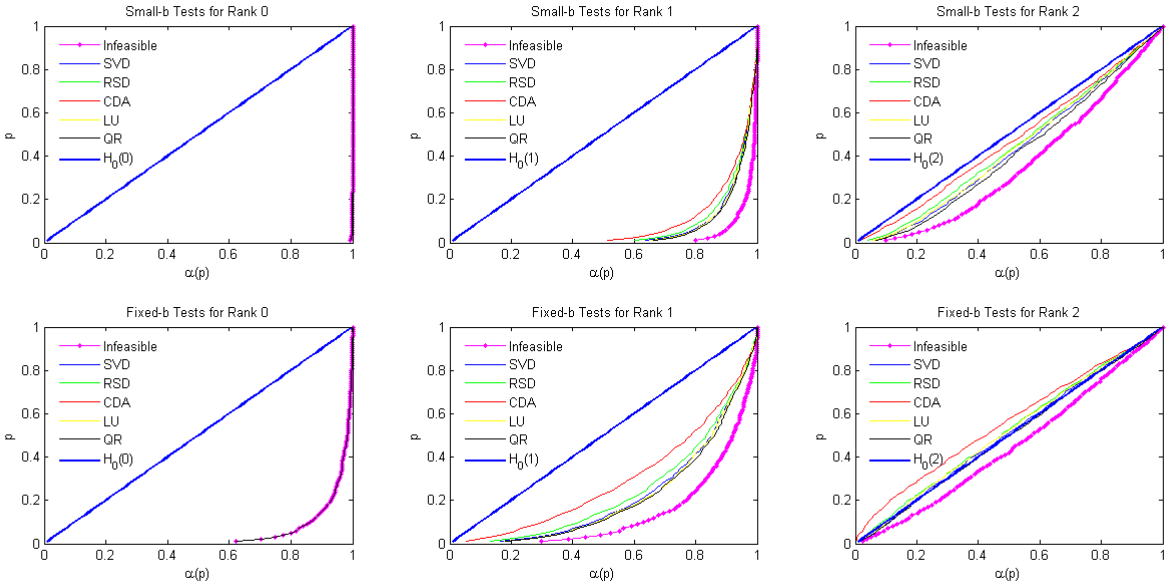
Note that the infeasible statistics have a tendency to over–rejects relative to the feasible statistics. This makes sense because the null space estimators implicit in the infeasible statistics are passive and do not adapt to $\widehat{B}$. The null space estimators implicit in the feasible statistics, in contrast, actively seek to annihilate $\widehat{B}$. □

Example 16 (Discontinuity of RRAs)**.** Suppose that instead of the B of Example 15, we have $B = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Then there are multiple rank–1 SVD approximations, the LU and QR algorithms run into multiple pivots, and as a result, these RRAs will have a discontinuity at B. This implies that the rank test statistics for rank–1 will not necessarily diverge at the same rate. In fact, they can be seen to diverge at heterogenous rates in the second column of Figure 7. Thus, the strong plug–in principle fails in this example, although the weak plug–in principle continues to hold. □

Example 17 (The Plug–in Principle for Symmetric Matrices)**.** Suppose now that we have the same setting above, with $B = \begin{bmatrix} 0.75 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ and we estimate it by OLS subject to the restriction of symmetry. Thus $\widehat{B}$ is symmetric and $\text{vech}(\widehat{B})$ has a positive definite asymptotic variance, which was again estimated non–parametrically with bandwidths 3 and 50.

For each replication, we computed F and t statistics as in Example 3 with the same RRAs

Figure 7: PP Plots for the F tests of Example 16



as above except that we used the LU decomposition utilized in Donald et al. (2007) as it is designed for symmetric matrices. The statistic based on the QR null space estimator takes the right null space estimator $\widehat{M}_r$ and uses $P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r}$ to formulate the rank test statistics. That tests based on this null space estimator should have power follows from Lemma 3 (ii) because $B \in \mathbb{P}^m$. There is no guarantee that this QR test for symmetric matrices would have power against a non–definite symmetric matrix.

The limiting distribution of the rank–2 small–b F statistic is $\chi^2(3)$. The limiting distribution of the rank–2 fixed–b F statistic has the same functional form as in Example 15, except the dimension of the underlying Brownian motion here is 3. The limiting distribution of the rank–2 small–b t statistic is $N(0,1)$. The limiting distribution of the rank–2 fixed–b t statistic is $W(1)\Big/\sqrt{2\int_0^1 (W(s)-sW(1))^2 ds}$, where W is a standard Brownian motion (Kiefer & Vogelsang, 2002a).

The results are ploted in Figure 8 and 9. Clearly, the tests exhibit similar behaviour to what we have seen in Example 15. □

5.2 General Asymptotics

Example 18 (The Cointegrated VAR Model)**.** Consider a variant of the model given in Example 12 with $B = \begin{bmatrix} -0.25 & 0 & 0 & 0 \\ 0 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, $y_0 = 0$, and $T = 50$. This model generates $I(1)$ data. If we proceed as in Example 12, estimating the model by OLS and formulating F and JA

Figure 8: PP Plots for the F tests of Example 17

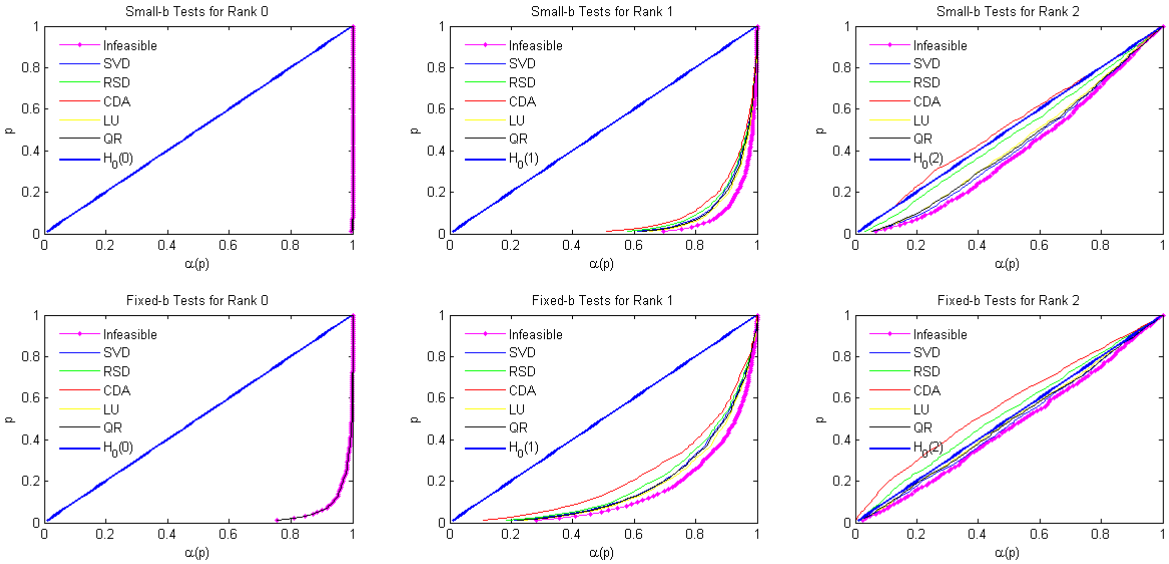
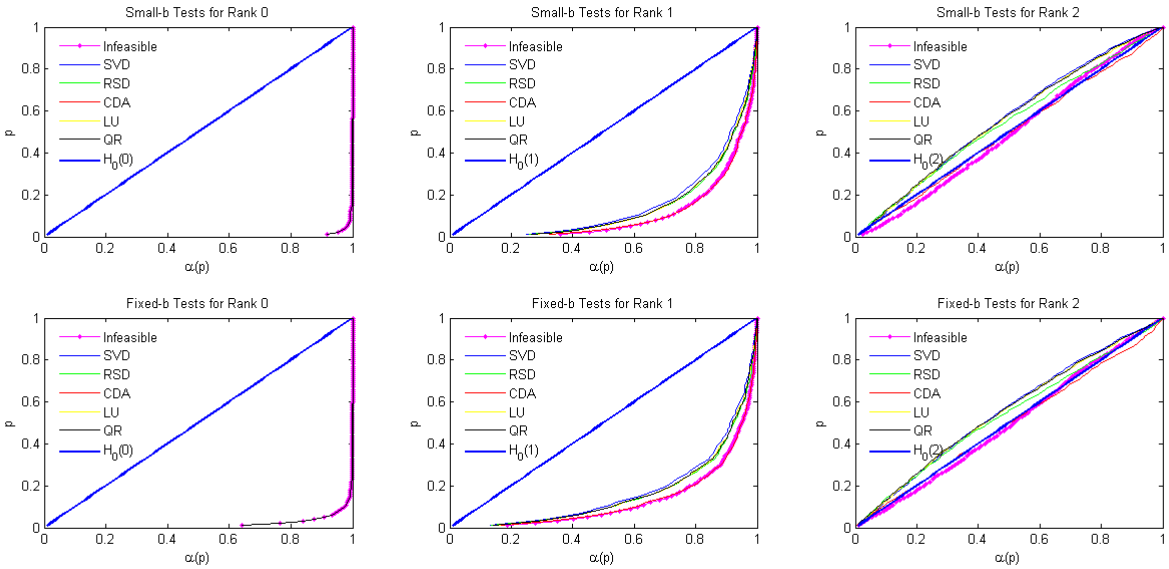


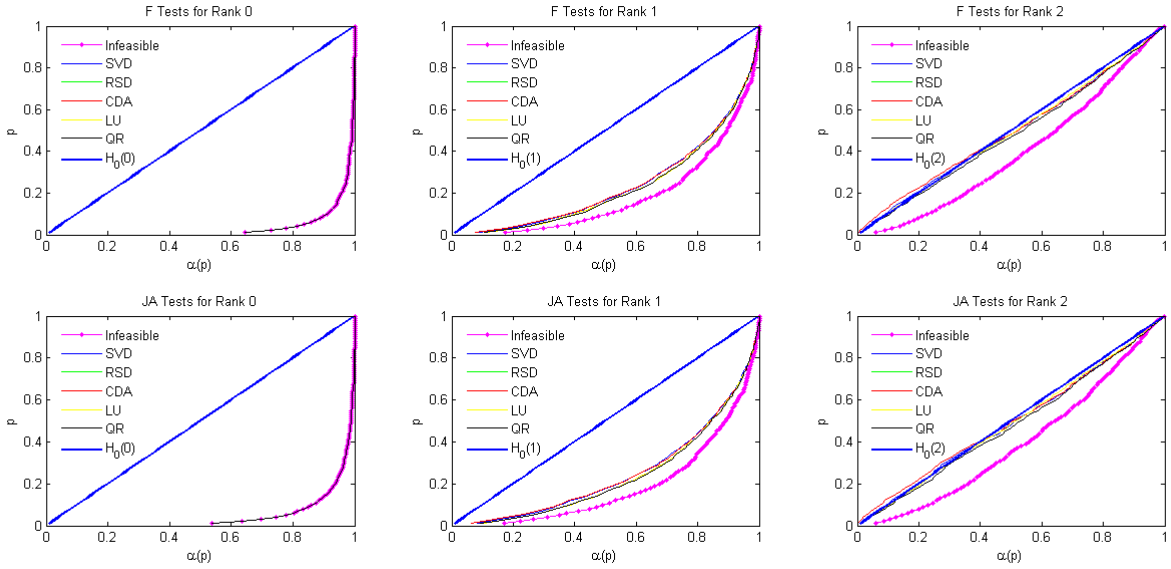
Figure 9: PP Plots for the t tests of Example 17



statistics, the strong plug–in principle is expected to hold. The F statistic for rank 2 converges in distribution to a generalized Dickey Fuller limiting distribution with 2 degrees of freedom (Johansen, 1988), while the JA statistics converges in distribution to the maximum eigenvalue distribution with 2 degrees of freedom (Johansen, 1991). Both statistics evaluated at rank 0 and 1 diverge at a rate determined by the strong plug–in principle. Figure 7 confirms these results for a variety of null space estimators. □

Example 19 (Common Stochastic Trends)**.** Consider the model given in Example 13 with $\Sigma = I_4$, $\Gamma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, $x_0 = 0$, and $T = 50$. Here we look at the t statistic proposed by

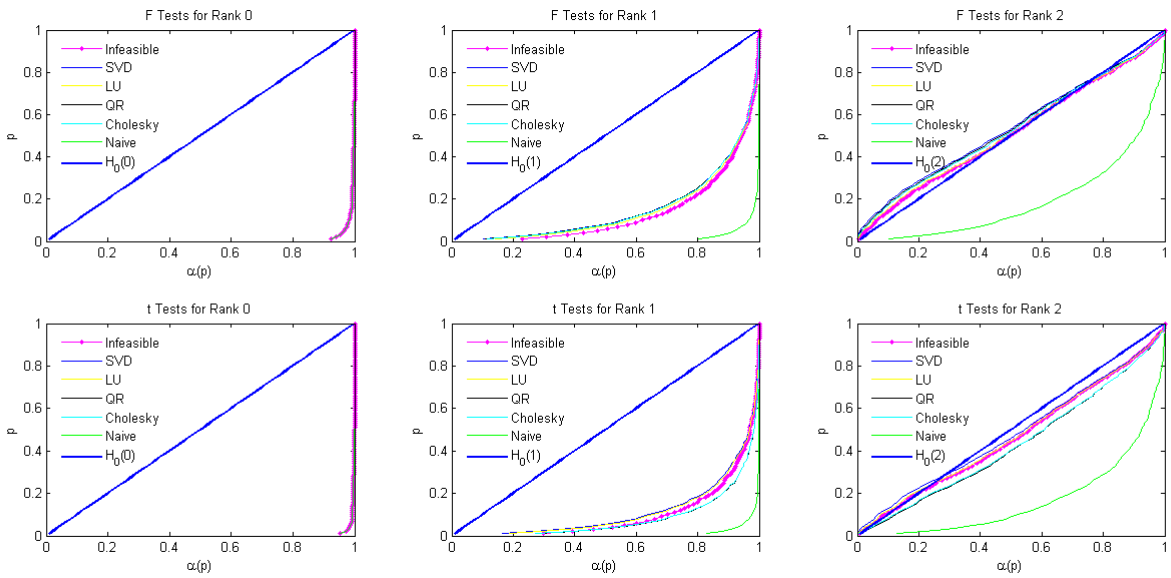
Figure 10: PP Plots for the F and JA tests of Example 18



Nyblom & Harvey (2000) as well as the F statistic proposed in Example 14. For each of these, we also consider the naive statistic that plugs in $M_r' = \left[\begin{smallmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{smallmatrix}\right]$ rather than the correct choice of M_{rT} discussed in Example 13.

The results are given in Figure 11. Clearly, the naive statistic perform quite differently to all other statistics. The figure also displays results for the Cholesky RRA, which can only be used here because $\widehat{B} \in \mathbb{P}^4$. $\square$

Figure 11: PP Plots for the F and t tests of Example 19



6 Conclusion

This paper has demonstrated that all rank test statistics are functions of implicit null space estimators. The paper presented a novel theory of null space estimation that allows for standard asymptotics, polynomial regressions, and cointegration asymptotics. The paper has proven that the behaviour of rank test statistics is completely governed by the implicit null space estimators through a plug–in principle. This has allowed for a general theory of rank testing that simplifies the asymptotics of rank test statistics, clarifies the relationships between the various rank test statistics, makes full use of the numerical analysis literature, and motivates numerous new rank test statistics.

We briefly mention some possible venues for future research. First, as a number of statistics in the literature have been shown to be asymptotically equivalent, the next natural step is to study small sample performance and higher order asymptotics. Second, as this paper has presented a theory of subspace estimation, the natural next step is to consider inference on subspaces. Third, the paper has presented a number of estimators of rank that ought to be tested in a simulation study. Fourth, this paper has considered the case of fixed n and m and it would be useful to extend the theory for application to high–dimensional data.

Appendix A: Reduced Rank Approximations

In this section, we develop the properties of some of the most popular RRAs in the literature. We assume throughout that $\widehat{B} \in \mathbb{G}^{n\times m}$, $\widehat{\Sigma} \in \mathbb{P}_+^n$, $\widehat{\Gamma} \in \mathbb{P}_+^m$, $\widehat{\Omega} \in \mathbb{P}_+^{nm}$, and $\widehat{\Psi} \in \mathbb{P}_+^{m(m+1)/2}$.

The Singular Value Decomposition. The most important reduced rank approximation is the singular value decomposition (SVD) approximation.[19] The SVD of $\widehat{B}$ is of the form $\widehat{B} = \widehat{U}\widehat{S}\widehat{V}'$, where $\widehat{U} \in \mathbb{R}^{n\times n}$ and $\widehat{V} \in \mathbb{R}^{m\times m}$ are orthogonal matrices and $\widehat{S}$ is diagonal with diagonal elements $\sigma_1(\widehat{B}) \geq \sigma_2(\widehat{B}) \geq \cdots \geq \sigma_{\min\{n,m\}}(\widehat{B}) > 0$. We have

$$\widehat{B} = \begin{bmatrix} \widehat{U}_{\cdot 1} & \widehat{U}_{\cdot 2} \end{bmatrix} \begin{bmatrix} \widehat{S}_1 & 0 \\ 0 & \widehat{S}_2 \end{bmatrix} \begin{bmatrix} \widehat{V}_{\cdot 1}' \\ \widehat{V}_{\cdot 2}' \end{bmatrix} = \widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}' + \widehat{U}_{\cdot 2}\widehat{S}_2\widehat{V}_{\cdot 2}',$$

[19] Interestingly, the singular value decomposition has a long history in applied mathematics as a rank revealing decomposition (Stewart, 1993) and yet it was one of the very last decompositions to be used in a rank test (Ratsimalahelo, 2003; Kleibergen & Paap, 2006; Donald et al., 2007).

where $\widehat{S}_1 \in \mathbb{R}^{r\times r}$. Then, as is well known $\widehat{B}_r^{SVD} = \widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}'$ is closest in Euclidian distance to $\widehat{B}$ among all matrices of rank r. In particular,

$$\sum_{i=r+1}^{\min\{n,m\}} \sigma_i^2(\widehat{B}) = \|\widehat{B} - \widehat{B}_r^{SVD}\|^2 \leq \|\widehat{B} - A\|^2$$

whenever $\text{rank}(A) = r$ (Horn & Johnson, 1985, Example 7.4.1). $\widehat{B}_r^{SVD}$ is unique if and only if $\sigma_r(\widehat{B}) \neq \sigma_{r+1}(\widehat{B})$ (Markovsky, 2012, Theorem 2.23). Finally, the above suggests the null space estimators $\widehat{N}_r = \widehat{U}_{\cdot 2}$ and $\widehat{M}_r = \widehat{V}_{\cdot 2}$. This implies that $\widehat{N}_r'\widehat{B}\widehat{M}_r = \widehat{S}_2$. Ratsimalahelo (2003) and Kleibergen & Paap (2006) utilize these null space estimators to formulate the statistic

$$T\text{vec}'(\widehat{N}_r'\widehat{B}\widehat{M}_r)\{(\widehat{M}_r \otimes \widehat{N}_r)'\widehat{\Omega}(\widehat{M}_r \otimes \widehat{N}_r)\}^{-1}\text{vec}(\widehat{N}_r'\widehat{B}\widehat{M}_r).$$

Using the fact that for any $H \in \mathbb{R}^{p\times k}$ and $G \in \mathbb{R}^{p\times p}$ satisfying $H'GH \in \mathbb{G}^{k\times k}$,

$$(P_H G P_H)^\dagger = H(H'GH)^{-1}H = P_H(P_H G P_H)^\dagger P_H, \tag{4}$$

it follows that the Ratsimalahelo–Kleibergen–Paap statistic is equal to $F(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$. The identity (4) is verified by checking the four conditions of the Moore–Penrose inverse (Stewart & Sun, 1990, Theorem III.1.1).

When $\widehat{B} \in \mathbb{S}^m$, $\widehat{U}'\widehat{V}$ is a diagonal matrix with either $+1$ or -1 on the diagonal. In particular, every column of $\widehat{N}_r$ differs from a column in $\widehat{M}_r$ by at most a sign and vice versa so that $P_{\widehat{N}_r} = P_{\widehat{M}_r}$. Donald et al. (2007) use this to construct the statistic

$$T\text{vech}'(\widehat{M}_r'\widehat{B}\widehat{M}_r)\{D_{m-r}^\dagger(\widehat{M}_r \otimes \widehat{M}_r)'D_m\widehat{\Psi}D_m'(\widehat{M}_r \otimes \widehat{M}_r)D_{m-r}^{\dagger\prime}\}^{-1}\text{vech}(\widehat{M}_r'\widehat{B}\widehat{M}_r),$$

Noting that $\text{vech}(\widehat{M}_r'\widehat{B}\widehat{M}_r) = D_{m-r}^\dagger(\widehat{M}_r \otimes \widehat{M}_r)'D_m\text{vech}(\widehat{B})$ and using (4) again, this statistic may be written as

$$T\text{vec}'(\widehat{B})D_m'P_{(\widehat{M}_r\otimes\widehat{M}_r)D_{m-r}^{\dagger\prime}}\{P_{(\widehat{M}_r\otimes\widehat{M}_r)D_{m-r}^{\dagger\prime}}D_m\widehat{\Psi}D_m'P_{(\widehat{M}_r\otimes\widehat{M}_r)D_{m-r}^{\dagger\prime}}\}^\dagger P_{(\widehat{M}_r\otimes\widehat{M}_r)D_{m-r}^{\dagger\prime}}D_m\text{vec}(\widehat{B}),$$

This can be simplified by writing

$$\begin{aligned}
P_{(\widehat{M}_r\otimes\widehat{M}_r)D_{m-r}^{\dagger\prime}} &= (\widehat{M} \otimes \widehat{M}_r)D_{m-r}^{\dagger\prime}\left(D_{m-r}^\dagger(\widehat{M}_r \otimes \widehat{M}_r)'(\widehat{M} \otimes \widehat{M}_r)D_{m-r}^{\dagger\prime}\right)^{-1} D_{m-r}^\dagger(\widehat{M}_r \otimes \widehat{M}_r)' \\
&= (\widehat{M}_r \otimes \widehat{M}_r)D_{m-r}\left(D_{m-r}'(\widehat{M}_r'\widehat{M}_r \otimes \widehat{M}_r'\widehat{M})D_{m-r}\right)^{-1} D_{m-r}'(\widehat{M}_r \otimes \widehat{M}_r)' \\
&= (\widehat{M}_r \otimes \widehat{M}_r)D_{m-r}D_{m-r}^\dagger\left((\widehat{M}_r'\widehat{M}_r)^{-1} \otimes (\widehat{M}_r'\widehat{M})^{-1}\right) D_{m-r}^{\dagger\prime}D_{m-r}'(\widehat{M}_r \otimes \widehat{M}_r)' \\
&= (\widehat{M}_r \otimes \widehat{M}_r)D_{m-r}D_{m-r}^\dagger\left((\widehat{M}_r'\widehat{M}_r)^{-1} \otimes (\widehat{M}_r'\widehat{M})^{-1}\right)(\widehat{M}_r \otimes \widehat{M}_r)' \\
&= D_m D_m^\dagger(\widehat{M}_r \otimes \widehat{M}_r)\left((\widehat{M}_r'\widehat{M}_r)^{-1} \otimes (\widehat{M}_r'\widehat{M})^{-1}\right)(\widehat{M}_r \otimes \widehat{M}_r)' \\
&= D_m D_m^\dagger(P_{\widehat{M}_r} \otimes P_{\widehat{M}_r})
\end{aligned}$$

where the third equality follows from Theorem 3.13 (d) of Magnus & Neudecker (1999), while the fourth and fifth follow from Theorem 3.12 (b) and Theorem 3.9 (a) of Magnus & Neudecker (1999). Substituting into the statistic and noting that $D_m D_m^{\dagger}(P_{\widehat{M}_r} \otimes P_{\widehat{M}_r})D_m = (P_{\widehat{M}_r} \otimes P_{\widehat{M}_r})D_m$ by Theorem 3.13 (a) of Magnus & Neudecker (1999), we have that the Donald et al. (2007) statistic is exactly $F(\widehat{B}, D_m\widehat{\Psi}D_m', P_{\widehat{M}_r}, P_{\widehat{M}_r})$.

The Robin–Smith Decomposition. The RSD of $\widehat{B}$ takes matrices $\Xi \in \mathbb{P}_+^n$ and $\Upsilon \in \mathbb{P}_+^m$ and obtains $\widehat{B} = \widehat{U}\widehat{S}\widehat{V}'$, where $\widehat{U} \in \mathbb{G}^{n\times n}$, $\widehat{V} \in \mathbb{G}^{m\times m}$, and $\widehat{S}$ satisfy:

(i) The columns of $\widehat{U}^{-1\prime}$ are generalized eigenvectors of $(\widehat{B}\Upsilon^{-1}\widehat{B}', \Xi)$.

(ii) The columns of $\widehat{V}^{-1\prime}$ are generalized eigenvectors of $(\widehat{B}'\Xi^{-1}\widehat{B}, \Upsilon)$.

(iii) $\widehat{S}$ is diagonal with diagonal entries, $\mu_1(\widehat{B}) \geq \mu_2(\widehat{B}) \geq \cdots \geq \mu_{\min\{n,m\}}(\widehat{B}) > 0$.[20]

The RSD arises naturally in a number of contexts. In canonical correlation analysis $\widehat{B}$ is a sample covariance matrix of two random vectors with sample covariance matrices Ξ and Υ, the columns found in (i) and (ii) define the coefficients of the sample canonical variates, while (iii) lists the sample canonical correlations. In reduced rank regression, on the other hand, we take $\widehat{B}$ to be the OLS estimator in the regression of y on x, while Ξ is the sample variance of the residuals, and Υ is the inverse of the sample second moment of x (Reinsel & Velu, 1998).

The RSD is easily derived from the SVD: if $\widehat{U}_0\widehat{S}_0\widehat{V}_0'$ is the SVD of $\Xi^{-\frac{1}{2}}\widehat{B}\Upsilon^{-\frac{1}{2}}$ then $\widehat{B} = \widehat{U}\widehat{S}\widehat{V}'$ with $\widehat{U} = \Xi^{-\frac{1}{2}}\widehat{U}_0$, $\widehat{S} = \widehat{S}_0$, and $\widehat{V} = \Upsilon^{-\frac{1}{2}}\widehat{V}_0$ and it is easily checked that $\widehat{U}$, $\widehat{S}$, and $\widehat{V}$ satisfy (i) – (iii) above, with $\mu_i(\widehat{B}) = \sigma_i(\Xi^{-\frac{1}{2}}\widehat{B}\Upsilon^{-\frac{1}{2}})$. Clearly the RSD reduces to the SVD when $\Xi = I_n$ and $\Upsilon = I_m$.

Now just as we did in the SVD case, write $\widehat{B} = \widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}' + \widehat{U}_{\cdot 2}\widehat{S}_2\widehat{V}_{\cdot 2}'$, where $\widehat{S}_1 \in \mathbb{R}^{r\times r}$ and set $\widehat{B}_r^{RSD} = \widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}'$. We may also write $\widehat{B}_r^{RSD} = \Xi^{1/2}(\Xi^{-1/2}\widehat{B}\Upsilon^{-1/2})_r^{SVD}\Upsilon^{1/2}$, which minimizes $\|\widehat{B} - A\|_{\Upsilon\otimes\Xi} = \|\Xi^{-1/2}(\widehat{B} - A)\Upsilon^{-1/2}\|$ with respect to all matrices A of rank r. Clearly $\widehat{B}_r^{RSD}$ is unique if and only if $\mu_r(\widehat{B}) \neq \mu_{r+1}(\widehat{B})$ (see Theorem 2.29 of Markovsky (2012)). We may estimate the null spaces by setting $\widehat{N}_r$ to the last $n-r$ columns of $\widehat{U}^{-1}$ and $\widehat{M}_r$ to the last $m-r$ columns of $\widehat{V}^{-1}$. This implies that $\widehat{N}_r'\widehat{B}\widehat{M}_r = \widehat{S}_2$.

By property (i) of the RSD, $\widehat{N}_r\Xi\widehat{N}_r = I_{n-r}$, while property (ii) implies that $\widehat{M}_r\Upsilon\widehat{M}_r = I_{m-r}$. This, and the fact $\sigma_i^2(A) = \lambda_i(AA') = \lambda_i(A'A)$ for any matrix A (Horn & Johnson,

[20] The RSD is also a special case of the generalized singular value decomposition of Van Loan (1976), which is also used as a rank revealing decomposition (Hansen, 1998).

1991, p. 135), implies that

$$
\begin{aligned}
\mu_{i+r}(\widehat{B}) &= \sigma_i^2(\widehat{S}_2) \\
&= \sigma_i\left(\widehat{N}_r'\widehat{B}\widehat{M}_r\right) \\
&= \sigma_i^2\left((\widehat{N}_r\Xi\widehat{N}_r)^{-1/2}\widehat{N}_r'\widehat{B}\widehat{M}_r(\widehat{M}_r'\Upsilon\widehat{M}_r)^{-1/2}\right) \\
&= \lambda_i\left((\widehat{N}_r'\Xi\widehat{N}_r)^{-1/2}\widehat{N}_r'\widehat{B}\widehat{M}_r(\widehat{M}_r'\Upsilon\widehat{M}_r)^{-1}\widehat{M}_r'\widehat{B}'\widehat{N}_r(\widehat{N}_r\Xi\widehat{N}_r)^{-1/2}\right) \\
&= \lambda_i\left((\widehat{N}_r'\Xi\widehat{N}_r)^{-1/2}\widehat{N}_r'\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\Upsilon P_{\widehat{M}_r})^{\dagger}P_{\widehat{M}_r}\widehat{B}'\widehat{N}_r(\widehat{N}_r\Xi\widehat{N}_r)^{-1/2}\right), \qquad \text{using (4)} \\
&= \lambda_i\left((P_{\widehat{M}_r}\Upsilon P_{\widehat{M}_r})^{\dagger/2}P_{\widehat{M}_r}\widehat{B}'\widehat{N}_r(\widehat{N}_r\Xi\widehat{N}_r)^{-1}\widehat{N}_r'\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\Upsilon P_{\widehat{M}_r})^{\dagger/2}\right) \\
&= \lambda_i\left((P_{\widehat{M}_r}\Upsilon P_{\widehat{M}_r})^{\dagger/2}P_{\widehat{M}_r}\widehat{B}'P_{\widehat{N}_r}(P_{\widehat{N}_r}\Xi P_{\widehat{N}_r})^{\dagger}P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\Upsilon P_{\widehat{M}_r})^{\dagger/2}\right), \qquad \text{using (4)} \\
&= \sigma_i^2\left((P_{\widehat{N}_r}\Xi P_{\widehat{N}_r})^{\dagger/2}P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}(P_{\widehat{M}_r}\Upsilon P_{\widehat{M}_r})^{\dagger/2}\right)
\end{aligned}
$$

The Anderson (1951) likelihood ratio statistic takes $\Xi = \widehat{\Sigma}$ and $\Upsilon = \widehat{\Gamma}^{-1}$ to obtain $\widehat{N}_r$ and $\widehat{M}_r$ and is given by $LR(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) = T\sum_{i=r+1}^{\min\{n,m\}} \log\left(1 + \mu_i^2(\widehat{B})\right)$. Similarly, $A(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) = T\sum_{i=r+1}^{\min\{n,m\}} \mu_i^2(\widehat{B})$ and $J(\widehat{B}, \widehat{\Sigma}, \widehat{\Gamma}, P_{\widehat{N}_r}, P_{\widehat{M}_r}) = T\mu_{r+1}^2(\widehat{B})$.

The Cragg–Donald Approximation. This RRA is discussed at length in Section 3.2, here we will mention only the computational details.

If $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ span the left and right null spaces of $\widehat{B}_r^{CDA}$ respectively, then Lemma 2 (iv) proves that the Cragg & Donald (1997) statistic is $F(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$. An often cited criticism of the CDA in the econometrics literature is that it is difficult to compute for general forms of Θ. The statistics literature, on the other hand, has often resorted to manipulating first order conditions to yield iterative solutions of the RRAs (see e.g. p. 33 and p. 63 of Reinsel & Velu (1998) and Gabriel & Zamir (1979)). In the process of proving Lemma 2 (iv) (see equation (6)) we find that

$$
\mathrm{vec}(\widehat{B}_r^{CDA}) = (\widehat{M}_r \otimes \widehat{N}_r)_\perp\{(\widehat{M}_r \otimes \widehat{N}_r)_\perp'\Theta^{-1}(\widehat{M}_r \otimes \widehat{N}_r)_\perp\}^{-1}(\widehat{M}_r \otimes \widehat{N}_r)_\perp'\Theta^{-1}\mathrm{vec}(\widehat{B}),
$$

We may therefore iterate this equation as outlined in the following algorithm, which is used in all of the Monte Carlo experiments of this paper.

Algorithm 1 (Cragg–Donald Approximation)**.** Initialize $\widehat{B}_r^{CDA}$ as any rank–r RRA of $\widehat{B}$ and set $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ to span the left and right null spaces of $\widehat{B}_r^{CDA}$ respectively. Iterate the following steps until $\widehat{B}_r^{CDA}$ converges:

(i) Obtain

$$\mathrm{vec}(\widetilde{B}) = (\widehat{M}_r \otimes \widehat{N}_r)_\perp \{(\widehat{M}_r \otimes \widehat{N}_r)'_\perp \Theta^{-1} (\widehat{M}_r \otimes \widehat{N}_r)_\perp\}^{-1} (\widehat{M}_r \otimes \widehat{N}_r)'_\perp \Theta^{-1} \mathrm{vec}(\widehat{B})$$

(ii) Set $\widehat{B}_r^{CDA}$ to any rank–r RRA of $\widetilde{B}$ and $\widehat{N}_r \in \mathbb{G}^{n\times(n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m\times(m-r)}$ to span the left and right null spaces of $\widehat{B}_r^{CDA}$ respectively. □

The initial choice of $\widehat{B}_r^{CDA}$ can be important, particularly when Θ is ill conditioned. In this case, the initial choice may throw away information that Θ would have picked up. For example, with the SVD as the initial RRA it is impossible to obtain the correct CDA in Example 6 when $\delta < \varepsilon$. The algorithm employed in this paper obtains an RRA of $\mathrm{mat}(\Theta^{-1/2}\mathrm{vec}(\widehat{B}))$, which appears to work quite well. The author also tried using an RSD with the Kronecker product approximation of Θ proposed in Robin & Smith (1995) as well as the method proposed by Van Loan & Pitsianis (1993) but these still were not able to provide the correct CDA in Example 6.

The LU Decomposition with Complete Pivoting. This decomposition arises from Gaussian elimination in linear system and is used in Cragg & Donald (1996) to construct an F statistic. The algorithm constructs permutation matrices $\widehat{\Pi}_1$ and $\widehat{\Pi}_2$ such that $\widehat{\Pi}_1 \widehat{B} \widehat{\Pi}_2 = \widehat{L}\widehat{S}$, where $\widehat{L} \in \mathbb{G}^{n\times n}$ is lower triangular with 1's along its diagonal and all subdiagonal elements are smaller than 1 in absolute value and $\widehat{S} \in \mathbb{R}^{n\times m}$ is upper triangular with $|\widehat{S}_{(i,i)}| \geq |\widehat{S}_{(i,j)}|$ for all $j \geq i$ (Golub & Van Loan, 1996, Theorem 3.4.2). Thus $\widehat{L}$ is bounded and so is it's inverse (Higham, 1987, Theorem 6.1). The rank–r approximation is then given by $\widehat{B}_r^{LU} = \widehat{\Pi}_1' \widehat{L} \begin{bmatrix} \widehat{S}_{11} & \widehat{S}_{12} \\ 0 & 0 \end{bmatrix} \widehat{\Pi}_2'$, where $\widehat{S}_{11} \in \mathbb{G}^{r\times r}$.[21] We may take $\widehat{N}_r' = [\ 0 \quad I_{n-r}\ ]\widehat{L}^{-1}\widehat{\Pi}_1$ and $\widehat{M}_r = \widehat{\Pi}_2 \begin{bmatrix} \widehat{S}_{11}^{-1}\widehat{S}_{12} \\ I_{m-r} \end{bmatrix}$. Following the same steps as we used in the representation of the Ratsimalahelo–Kleibergen–Paap statistic, the Cragg & Donald (1996) statistic is precisely $F(\widehat{B}, \widehat{\Omega}, P_{\widehat{N}_r}, P_{\widehat{M}_r})$.

It remains to show that this algorithm satisfies conditions (i) and (ii) of Definition 2.[22] Let $\widehat{S}^{(r)}$ be the result of the r–th permutation and Gaussian elimination of the algorithm.

[21] Cragg & Donald (1996) obtain their estimates of the null spaces by running the LU algorithm only up to the r–th step. This is exactly equivalent to our formulation because subsequent steps of the LU algorithm have no effect on $\widehat{B}_r^{LU}$. See also problem 3.2.2 of Golub & Van Loan (1996).

[22] The LU decomposition with complete pivoting is not commonly used in the numerical analysis literature to detect rank because it is neither as efficient as the SVD nor as computationally attractive as the QR decomposition. The bounds for the LU decomposition are given here because they cannot be found elsewhere.

Then $\widehat{S} = \widehat{S}^{(\min\{n,m\}-1)}$. Now $|\widehat{S}_{(1,1)}| = |\widehat{S}^{(1)}_{(1,1)}| = \max_{i,j} |\widehat{B}_{(i,j)}| \geq \sigma_1(\widehat{B})/\sqrt{nm}$. Likewise, $|\widehat{S}_{(2,2)}| = |\widehat{S}^{(2)}_{(2,2)}| = \max_{i>1} |\widehat{S}^{(1)}_{(i,j)}| \geq \sigma_1(\widehat{S}^{(1)}_{(2:n,1:m)})/\sqrt{(n-1)(m-1)} \geq \sigma_1(\widehat{S}^{(1)}_{(2:n,1:m)})/\sqrt{nm} \geq \sigma_2(\widehat{S}^{(1)})/\sqrt{nm}$, where the last inequality follows from Corollary 3.1.3 of Horn & Johnson (1991).[23] Since the smallest singular value of the r–th step Gaussian elimination matrix is bounded below by $(1+n)^{-1}$,[24] we have that $\sigma_2(\widehat{S}^{(1)}) \geq \sigma_2(\widehat{B})/(1+n)$ (Horn & Johnson, 1991, Theorem 3.3.16). Therefore $|\widehat{S}_{(2,2)}| \geq \frac{\sigma_2(\widehat{B})}{(1+n)\sqrt{nm}}$. Following the same logic, we find that $|\widehat{S}_{(r,r)}| \geq \frac{\sigma_r(\widehat{B})}{(n+1)^{r-1}\sqrt{nm}}$ for $r = 1, \ldots, \min\{n,m\}$.

To prove the other inequality, first note that $\widehat{S}_{(i,j)} = O(\widehat{S}_{(i,i)})$ for $i \leq j$ by construction. We also have that $\widehat{S}_{(j,j)} = O(\widehat{S}_{(i,i)})$ for $i \leq j$ (Wilkinson, 1961, equation 5.3). It follows that $\widehat{S}_{(i,j)} = O(\widehat{S}_{(k,k)})$ for $k \leq i \leq j$. Therefore, we will have proven the inequality if we can show that $\widehat{S}_{(r+1,r+1)} = O(\sigma_{r+1}(\widehat{B}))$ as $\widehat{B}$ converges to a rank–r matrix. The case $r = 0$ follows from the fact that $|\widehat{S}_{11}| \leq \sigma_1(\widehat{B})$. Therefore, let $r > 0$ and consider the inequality

$$|\widehat{S}_{(r+1,r+1)}| \leq \frac{\sqrt{2(r+1)}\sigma_1(\widehat{S})}{\sigma_r(\widehat{S}_{(1:r,1:r)})}\sigma_{r+1}(\widehat{S}_{(1:r+1,1:r+1)}), \quad r = 1, \ldots, \min\{n,m\}-1,$$

a proof of which can be found in Chandrasekaran & Ipsen (1994) p. 601–602. Now $\sigma_r(\widehat{S}_{(1:r,1:r)}) \geq \frac{3\min_{1\leq i\leq r}|\widehat{S}_{(i,i)}|}{\sqrt{4^r+6r-1}} \geq \frac{3\sigma_r(\widehat{B})}{(n+1)^{r-1}\sqrt{nm(4^r+6r-1)}}$, where the first inequality follows from Theorem 6.1 of Higham (1987) and the second inequality follows from our analysis above. On the other hand, since $\widehat{S}$ is obtained from $\widehat{B}$ by multiplying it with $\min\{n,m\}-1$ Gaussian elimination matrices, $\sigma_1(\widehat{S}) \leq (n+1)^{\min\{n,m\}-1}\sigma_1(\widehat{B})$. Finally, applying Corollary 3.1.3 of Horn & Johnson (1991) again, $\sigma_{r+1}(\widehat{S}_{(1:r+1,1:r+1)}) \leq \sigma_{r+1}(\widehat{S}) \leq (n+1)^{\min\{n,m\}-1}\sigma_{r+1}(\widehat{B})$. Putting this all together we obtain, $|\widehat{S}_{(r+1,r+1)}| \leq \frac{1}{3}\sqrt{2(r+1)nm(4^r+6r-1)}(n+1)^{r+2\min\{n,m\}-3}\sigma_{r+1}(\widehat{B})/\sigma_r(\widehat{B})$.

Donald et al. (2007) utilize the Bunch–Parlett LU algorithm designed for symmetric matrices (Golub & Van Loan, 1996, Section 4.4). The algorithm can be shown to satisfy the conditions for a DBA by similar arguments to those used above. Using these null spaces in $F(\widehat{B}, \widehat{\Omega}, P_{\widehat{M}_r}, P_{\widehat{M}_r})$ we obtain the Donald et al. (2007) LU statistic.

Other LU RRAs based on different pivoting strategies can be found in Hansen (1998). It

[23] For $A \in \mathbb{R}^{n\times m}$, $1 \leq i \leq j \leq n$ and $1 \leq k \leq l \leq m$, the matrix $A_{(i:j,k:l)}$ denotes the submatrix of A consisting of the rows i to j and columns k to l.

[24] The r–th step Gaussian elimination matrix in the algorithm is of the form $I_n + \theta e_r'$, where $\theta = (0, \ldots, 0, \theta_{r+1}, \ldots, \theta_n)'$, with $|\theta_j| \leq 1$ and e_r the r–th column of I_n (Golub & Van Loan, 1996, p. 95). It follows that $\|(I_n + \theta e_r')x\| \leq \|x\| + \|\theta\||x_r| \leq (1+\|\theta\|)\|x\| \leq (1+\sqrt{n-r})\|x\|$ and so $\sigma_1(I_n + \theta e_r') \leq 1 + \sqrt{n-r} \leq 1+n$. Since $(I_n + \theta e_r')^{-1} = I_n - \theta e_r'$ is also a Gaussian elimination matrix, it follows that $\sigma_n(I_n + \theta e_r') = \frac{1}{\sigma_1(I_n - \theta e_r')} \geq \frac{1}{1+n}$.

is important to note, however, that the LU algorithm with no pivoting and the LU algorithm with partial pivoting are not rank–revealing. For example, if $\widehat{B} = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right]$, then both algorithms produce $\widehat{S} = \widehat{B}$. Thus, both algorithms fail to push the content of $\widehat{B}$ into the upper left corner block of $\widehat{S}$ and to leave the bottom block empty. That is, they fail to satisfy the bounds in Definition 2.

The Block LU Decomposition. A related RRA to the LU is the BLU. Whereas the LU decomposition arises from the Gaussian elimination algorithm, the BLU arises from Gaussian elimination of a block of matrices. If $\widehat{B} \in \mathbb{R}^{n\times m}$ is partitioned as $\left[\begin{smallmatrix} \widehat{B}_{11} & \widehat{B}_{12} \\ \widehat{B}_{21} & \widehat{B}_{22} \end{smallmatrix}\right]$, where $\widehat{B}_{11} \in \mathbb{R}^{r\times r}$, $\sigma_r(\widehat{B}_{11}) = O^{-1}(1)$ as $\widehat{B} - \widehat{B}^* \to 0$, $\text{rank}(\widehat{B}^*) = r$, and $\sigma_r(\widehat{B}^*) = O^{-1}(1)$, then $\widehat{B}_{22} - \widehat{B}_{21}\widehat{B}_{11}^{-1}\widehat{B}_{12} = O(\sigma_{r+1}(\widehat{B}))$ as $\widehat{B} - \widehat{B}^* \to 0$. To see this, write

$$\widehat{S} = \begin{bmatrix} \widehat{B}_{11} & 0 \\ 0 & \widehat{B}_{22} - \widehat{B}_{21}\widehat{B}_{11}^{-1}\widehat{B}_{12} \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ -\widehat{B}_{21}\widehat{B}_{11}^{-1} & I_{n-r} \end{bmatrix} \widehat{B} \begin{bmatrix} I_r & -\widehat{B}_{11}^{-1}\widehat{B}_{12} \\ 0 & I_{m-r} \end{bmatrix} = \widehat{U}^{-1}\widehat{B}\widehat{V}^{-1\prime}.$$

The singular values of $\widehat{S}$ are the union of the singular values of $\widehat{B}_{11}$ and the singular values of $\widehat{B}_{22} - \widehat{B}_{21}\widehat{B}_{11}^{-1}\widehat{B}_{12}$. On the other hand, the i–th singular values of the right hand side expression are bounded above by $O(1)\sigma_i(\widehat{B})$ (Horn & Johnson, 1991, Theorem 3.3.16 (d)). Since the singular values of $\widehat{B}_{11}$ are bounded away from zero and $\sigma_{r+1}(\widehat{B}) \to 0$, it follows that $\sigma_i(\widehat{B}_{22} - \widehat{B}_{21}\widehat{B}_{11}^{-1}\widehat{B}_{12}) \to 0$ and $\sigma_i(\widehat{B}_{22} - \widehat{B}_{21}\widehat{B}_{11}^{-1}\widehat{B}_{12}) = O(\sigma_{r+i}(\widehat{B}))$. Thus conditions (i) and (ii) of Definition 2 are satisfied when $\widehat{B} - \widehat{B}^* \to 0$. We can define $\widehat{B}_r^{BLU} = \left[\begin{smallmatrix} \widehat{B}_{11} & \widehat{B}_{12} \\ \widehat{B}_{21} & \widehat{B}_{21}\widehat{B}_{11}^{-1}\widehat{B}_{12} \end{smallmatrix}\right]$. The null spaces may then be estimated as $\widehat{N}_r' = [\ -\widehat{B}_{21}\widehat{B}_{11}^{-1} \quad I_{n-r}\ ]$ and $\widehat{M}_r = \left[\begin{smallmatrix} -\widehat{B}_{11}^{-1}\widehat{B}_{12} \\ I_{m-r} \end{smallmatrix}\right]$. Kleibergen & van Dijk (1994) and Robin & Smith (1995) construct F statistics using such null space estimators.

The QR Decomposition with Pivoting. This decomposition arises from extensions to the Gram–Schmidt orthogonalization algorithm and, to the author's knowledge, has never been used in a rank test. The algorithm, which can be found in section 5.4.1 of Golub & Van Loan (1996), constructs a permutation matrix $\widehat{V}$ and an orthogonal matrix $\widehat{U}$ such that $\widehat{B}\widehat{V} = \widehat{U}\widehat{S}$. Chandrasekaran & Ipsen (1994) prove that if $\widehat{S}$ is partitioned as $\left[\begin{smallmatrix} \widehat{S}_{11} & \widehat{S}_{12} \\ 0 & \widehat{S}_{22} \end{smallmatrix}\right]$, with $\widehat{S}_{11} \in \mathbb{R}^{r\times r}$, then $\sigma_r(\widehat{S}_{11}) \geq (r\max\{n,m\})^{-1/2}2^{-r}\sigma_r(\widehat{B})$.[25] The second inequality can be derived exactly as

[25] Chandrasekaran & Ipsen (1994) report their bounds for the case $n = m$. The bounds given above result from applying the Chandrasekaran & Ipsen bounds to the completed matrices $[\ \widehat{B} \quad 0\ ]$ when $n > m$ and $\left[\begin{smallmatrix} \widehat{B} \\ 0 \end{smallmatrix}\right]$ when $n < m$ and noting that the added rows or columns have no effect on $\widehat{S}_{11}$ or the singular values of $\widehat{S}_{22}$ and $\widehat{B}$.

in the discussion of the LU decomposition above, on noting that, by construction, $\widehat{S}_{(i,j)} \leq \widehat{S}_{(k,k)}$ whenever $k \leq i \leq j$. There are numerous other rank–revealing QR algorithms in the literature (see Hansen (1998) for a survey). The QR algorithm we have presented is often preferred to the other RRAs as it is quicker to compute.

Eigenvalue Decomposition. If a square matrix is rank deficient, then it has an eigenvalue at zero. By the continuity of the eigenvalues of a matrix with respect to its elements (Horn & Johnson, 1985, Appendix D), a matrix should be close to rank deficiency if it has eigenvalues close to zero. Unfortunately, there is little else to infer from the eigenvalues of a general matrix. An $n \times n$ matrix can have all its eigenvalues at zero but have rank–$(n-1)$ (e.g. the Jordan canonical nilpotent matrix). A matrix can also have its eigenvalues uniformly bounded away from zero and still be ill–conditioned (Stewart & Sun, 1990, Exercise 8 of Section I.4). Eigenvalues are, moreover, non–differentiable in the entries of the matrix (Stewart & Sun, 1990, Example I.3.2). Thus, eigenvalues are not recommended for rank detection in general.

Exceptions to this rule occur when the limit of $\widehat{B}$ has a special structure that allows one to infer the behaviour of the eigenvalues under the various alternatives. This is the case for the Stock & Watson (1988) statistic as the limit of $\widehat{B}$ is known to have either all eigenvalues at 1 under the null or else a determined number of eigenvalues with real part less than one under the alternative.

Another exception to the rule is the case of symmetric matrices, which have well behaved eigenvalues by Weyl's Theorem (Stewart & Sun, 1990, Corollary IV.4.9). In particular, if $\widehat{B} \in \mathbb{R}^{m \times m}$ is symmetric, then the SVD of $\widehat{B}$ has $|\lambda_1(\widehat{B})|, \ldots, |\lambda_m(\widehat{B})|$ (possibly reordered) along the diagonal of $\widehat{S}$. Thus the SVD null space estimator is obtained by collecting the eigenvectors associated with the $m-r$ eigenvalues of $\widehat{B}$ that are closest to zero. We have already seen how Donald et al. (2007) construct an F statistic using this RRA.

If $\widehat{B}$ converges to a positive semi–definite matrix, we may utilize the spectral decomposition instead. Here $\widehat{B} = \widehat{U}\widehat{S}\widehat{U}'$ has $\lambda_1(\widehat{B}) \geq \cdots \geq \lambda_m(\widehat{B})$ along the diagonal of $\widehat{S}$. We may then estimate the null space by collecting the eigenvectors associated with $\lambda_{m-r+1}(\widehat{B}), \ldots, \lambda_m(\widehat{B})$ in the columns of $\widehat{M}_r$. Then $\sqrt{T}\mathrm{tr}(\widehat{M}_r'\widehat{B}\widehat{M}_r) = \sqrt{T}\sum_{i=r+1}^{m} \lambda_i(\widehat{B})$, the rank test statistic

proposed by Donald et al. (2007), is exactly $t\left(\widehat{B}, \frac{1}{m-r}(D_m'D_m)^{-1}, P_{\widehat{M}_r}\right)$. To see this, write

$$\begin{aligned}
t\left(\widehat{B}, \frac{1}{m-r}(D_m'D_m)^{-1}, P_{\widehat{M}_r}\right) &= \frac{\sqrt{T}\mathrm{tr}(\widehat{M}_r'\widehat{B}\widehat{M}_r)}{\sqrt{\frac{1}{m-r}\mathrm{vec}'(I_m)(P_{\widehat{M}_r}\otimes P_{\widehat{M}_r})D_mD_m^{\dagger}(P_{\widehat{M}_r}\otimes P_{\widehat{M}_r})\mathrm{vec}(I_m)}} \\
&= \frac{\sqrt{T}\mathrm{tr}(\widehat{M}_r'\widehat{B}\widehat{M}_r)}{\sqrt{\frac{1}{m-r}\mathrm{vec}'(I_m)(P_{\widehat{M}_r}\otimes P_{\widehat{M}_r})D_mD_m^{\dagger}\mathrm{vec}(I_m)}} \\
&= \frac{\sqrt{T}\sum_{i=r+1}^{m}\lambda_i(\widehat{B})}{\sqrt{\frac{1}{m-r}\mathrm{tr}(\mathrm{mat}((P_{\widehat{M}_r}\otimes P_{\widehat{M}_r})\mathrm{vec}(I_m)))}} \\
&= \frac{\sqrt{T}\sum_{i=r+1}^{m}\lambda_i(\widehat{B})}{\sqrt{\frac{1}{m-r}\mathrm{tr}(P_{\widehat{M}_r})}} \\
&= \sqrt{T}\sum_{i=r+1}^{m}\lambda_i(\widehat{B})
\end{aligned}$$

The second equality follows from Theorem 3.12 (b) and Theorem 3.9 (a) of Magnus & Neudecker (1999) and the third from the fact that $D_mD_m^{\dagger}\mathrm{vec}(I_m) = D_mD_m^{\dagger}D_m\mathrm{vech}(I_m) = D_m\mathrm{vech}(I_m) = \mathrm{vec}(I_m)$.

More generally, Donald et al. (2007) suggest normalizing using an estimate of the asymptotic variance of $\widehat{B}$.

$$t\left(\widehat{B}, \widehat{\Psi}, P_{\widehat{M}_r}\right) = \frac{\sqrt{T}\mathrm{tr}(\widehat{M}_r\widehat{M}_r'\widehat{B})}{\sqrt{\mathrm{vec}'(I_m)(P_{\widehat{M}_r}\otimes P_{\widehat{M}_r})D_m\widehat{\Psi}D_m'(P_{\widehat{M}_r}\otimes P_{\widehat{M}_r})\mathrm{vec}(I_m)}}.$$

Cholesky Decomposition. If $\widehat{B} \in \mathbb{P}_+^m$, we may employ a rank revealing Cholesky decomposition as well. Following (Higham, 1990), we can write $\widetilde{V}'\widehat{B}\widetilde{V} = \widetilde{S}'\widetilde{S}$, where $\widehat{B}^{1/2}\widetilde{V} = \widetilde{U}\widetilde{S}$ is the QR decomposition (with pivoting) of $\widehat{B}^{1/2}$. Now if $\widetilde{S}$ is partitioned as $\left[\begin{smallmatrix}\widetilde{S}_{11} & \widetilde{S}_{12}\\ 0 & \widetilde{S}_{22}\end{smallmatrix}\right]$, with $\widetilde{S}_{11} \in \mathbb{R}^{r\times r}$, then using the Chandrasekaran & Ipsen (1994) formula, we have that $\sigma_r(\widetilde{S}_{11}) \geq (rm)^{-1/2}2^{-r}\sigma_r^{1/2}(\widehat{B})$. The fact that $\sigma_1(\widetilde{S}_{22}) = O(\sigma_{r+1}^{1/2}(\widehat{B}))$ as $\widehat{B}$ converges to a rank–r matrix follows as indicated in the discussion of the QR algorithm. The decomposition implicit in the Cholesky RRA of $\widehat{B}$ is then given by $\widehat{S} = \mathrm{diag}^2(\widetilde{S})$ and $\widehat{U} = \widehat{V} = \widetilde{V}\widetilde{S}'\mathrm{diag}^{-1}(\widetilde{S})$.[26] The fact that this decomposition satisfies the conditions for an RRA follows from the results of Higham (1987). Just as in the QR case, the Cholesky RRA has never been used in a rank test. For an estimate of $\widehat{M}_r$, one may simply take $\widetilde{V}\widetilde{S}^{-1}\mathrm{diag}(\widetilde{S})\left[\begin{smallmatrix}0\\ I_{m-r}\end{smallmatrix}\right]$.

[26]For a $A \in \mathbb{R}^{m\times m}$, $\mathrm{diag}(A) \in \mathbb{R}^{m\times m}$ is defined by $\mathrm{diag}(A)_{(i,j)} = A_{(i,j)}$ when $i = j$ and zero otherwise.

Appendix B: Proofs

Proof of Lemma 1. (i) $\text{rank}(\widehat{B}_i^{DBA}) > i$ is impossible since the $\widehat{U}^{-1}\widehat{B}_i^{DBA}\widehat{V}^{-1\prime}$ has a rank of at most i by construction. If $\text{rank}(\widehat{B}_i^{DBA}) < i$, then $\text{rank}([\ \widehat{S}_{11} \quad \widehat{S}_{12}\ ]) < i$, which is impossible since $\sigma_i(\widehat{S}_{11}) \geq K_1\sigma_i(\widehat{B}) > 0$ by Definition 2 (i) so $\widehat{S}_{11}$ is non–singular.

(ii) Set $\widehat{B}^* = B^*$. Since its r–th singular value is fixed and positive, it follows from Definition 2 (ii) that $\|\widehat{B} - \widehat{B}_i^{DBA}\| \leq \|\widehat{U}\|\|\widehat{V}\|\|\widehat{S}_{22}\| = O(\sigma_{r+1}(\widehat{B}))$ as $\widehat{B} - B^* \to 0$ since U and V are bounded. The result then follows from the fact that $\sigma_{r+1}(\widehat{B}) = O(\|\widehat{B} - B^*\|)$ as $\widehat{B} - B^* \to 0$ (Horn & Johnson, 1985, Example 7.4.1).

(iii) Let $\widehat{U} = [\ \widehat{u}_1 \ \cdots \ \widehat{u}_n\ ]$. $\widehat{N}_i$ is orthogonal to $\text{span}(\widehat{B}_i^{DA}) = \text{span}\,([\ \widehat{u}_1 \ \cdots \ \widehat{u}_i\ ])$. It is therefore orthogonal to $\text{span}\,([\ u_1 \ \cdots \ \widehat{u}_r\ ]) = \text{span}(\widehat{B}_r^{DA})$. Thus, $\text{span}(\widehat{N}_i) \subset \text{span}(\widehat{N}_r)$. The nestedness of the right null spaces is proven similarly. □

Proof of Lemma 2. (i) Suppose $\text{rank}(\widehat{B}_i^{CDA}) < i$. Then for arbitrary $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and $h \in \mathbb{R}$, $\text{rank}(\widehat{B}_i^{CDA} + hxy') \leq i$ (Horn & Johnson, 1985, Result 0.4.5). By the definition of $\widehat{B}_i^{CDA}$,

$$\begin{aligned}\|\widehat{B} - \widehat{B}_i^{CDA} - hxy'\|_\Theta^2 &= \|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta^2 - 2\text{vec}'(\widehat{B} - \widehat{B}_i^{CDA})\Theta^{-1}(y \otimes x)h + \|xy'\|_\Theta^2 h^2 \\ &\geq \|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta^2.\end{aligned}$$

The left hand side is quadratic in h and achieves a minimum at $h = 0$, it follows that its derivative with respect to h at $h = 0$ must be zero and so

$$\text{vec}'(\widehat{B} - \widehat{B}_i^{CDA})\Theta^{-1}(y \otimes x) = 0.$$

Since x and y are arbitrary, $\widehat{B} = \widehat{B}_i^{CDA}$, which is impossible because $\text{rank}(\widehat{B}) > \text{rank}(\widehat{B}_i^{CDA})$.

(ii) It follows from the basic theory of positive definite matrices and the definition of the CDA that

$$\|\widehat{B} - \widehat{B}_i^{CDA}\| \leq \lambda_1^{1/2}(\Theta)\|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta \leq \lambda_1^{1/2}(\Theta)\|\widehat{B} - B^*\|_\Theta \leq \left(\frac{\lambda_1(\Theta)}{\lambda_{nm}(\Theta)}\right)^{1/2}\|\widehat{B} - B^*\|$$

for all $i \geq r$. The result follows from the fact that $\text{cond}(\Theta) = \lambda_1(\Theta)/\lambda_{nm}(\Theta)$.

(iii) Lemma III.3.5 of Stewart & Sun (1990) proves that

$$\begin{aligned}\max\{\|(I_n - P_{\widehat{N}_r})P_{\widehat{N}_i}\|_2, \|(I_m - P_{\widehat{M}_r})P_{\widehat{M}_i}\|_2\} &\leq \|(\widehat{B}_r^{CDA})^\dagger\|_2\|\widehat{B}_r^{CDA} - \widehat{B}_i^{CDA}\|_2 \\ &\leq \|(\widehat{B}_r^{CDA})^\dagger\|_2\left(\|\widehat{B} - \widehat{B}_r^{CDA}\|_2 + \|\widehat{B} - \widehat{B}_i^{CDA}\|_2\right) \\ &\leq 2\|(\widehat{B}_r^{CDA})^\dagger\|_2\|\widehat{B} - \widehat{B}_r^{CDA}\|_2.\end{aligned}$$

The result then follows from the fact that $(\widehat{B}_r^{CDA})^\dagger \to (B^*)^\dagger$ as $\widehat{B}_r^{CDA} \to B^*$ because $\text{rank}(\widehat{B}_r^{CDA}) = \text{rank}(B^*) = r$ (Stewart & Sun, 1990, p. 146).

(iv) For $h \in \mathbb{R}^{r^2}$ the matrix, $\widehat{B}_i^{CDA} + \text{mat}((\widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp})h)$ has a rank of at most i. From the definition of the CDA, we know that

$$\begin{aligned}\|\widehat{B} - \widehat{B}_i^{CDA} - \text{mat}((\widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp})h)\|_\Theta^2 &= \|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta^2 - 2\text{vec}'(\widehat{B} - \widehat{B}_i^{CDA})\Theta^{-1}(\widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp})h \\ &\quad + h'(\widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp})'\Theta^{-1}(\widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp})h \\ &\geq \|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta^2\end{aligned}$$

and it follows just as in (i) that

$$\text{vec}'(\widehat{B} - \widehat{B}_i^{CDA})\Theta^{-1}(\widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp}) = 0.$$

By the same logic we can show that

$$\begin{aligned}\text{vec}'(\widehat{B} - \widehat{B}_i^{CDA})\Theta^{-1}(\widehat{M}_i \otimes \widehat{N}_{i\perp}) &= 0, \\ \text{vec}'(\widehat{B} - \widehat{B}_i^{CDA})\Theta^{-1}(\widehat{M}_{i\perp} \otimes \widehat{N}_i) &= 0.\end{aligned}$$

Since $(\widehat{M}_i \otimes \widehat{N}_i)_\perp = \begin{bmatrix} \widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp} & \widehat{M}_i \otimes \widehat{N}_{i\perp} & \widehat{M}_{i\perp} \otimes \widehat{N}_i \end{bmatrix}$, we can combine the three equations above to arrive at,

$$\text{vec}'(\widehat{B} - \widehat{B}_i^{CDA})\Theta^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp = 0. \tag{5}$$

It follows that $\widehat{B}_i^{CDA}$ satisfies the equation

$$\begin{bmatrix} (\widehat{M}_i \otimes \widehat{N}_i)'_\perp \Theta^{-1/2} \\ (\widehat{M}_i \otimes \widehat{N}_i)'\Theta^{1/2} \end{bmatrix} \Theta^{-1/2}\text{vec}(\widehat{B}_i^{CDA}) = \begin{bmatrix} (\widehat{M}_i \otimes \widehat{N}_i)'_\perp \Theta^{-1/2} \\ 0 \end{bmatrix} \Theta^{-1/2}\text{vec}(\widehat{B}).$$

The matrix on the left hand side consists of two blocks of full rank that are orthogonal to each other. It is therefore invertible and we have the unique solution $\text{vec}(\widehat{B}_i) = \widehat{P}_i\text{vec}(\widehat{B})$, where

$$\widehat{P}_i = (\widehat{M}_i \otimes \widehat{N}_i)_\perp\{(\widehat{M}_i \otimes \widehat{N}_i)'_\perp \Theta^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp\}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)'_\perp \Theta^{-1}. \tag{6}$$

Now using the well known identity

$$I_n = G^{\frac{1}{2}}H(H'GH)^{-1}H'G^{\frac{1}{2}} + G^{-\frac{1}{2}}H_\perp(H'_\perp G^{-1}H_\perp)^{-1}H'_\perp G^{-\frac{1}{2}}, \tag{7}$$

for $G \in \mathbb{P}_+^n$ and $H \in \mathbb{G}^{n\times m}$, we have that

$$\begin{aligned}\Theta^{-1} &= \Theta^{-\frac{1}{2}} I_{nm} \Theta^{-\frac{1}{2}} \\ &= (\widehat{M}_i \otimes \widehat{N}_i)\{(\widehat{M}_i \otimes \widehat{N}_i)'\Theta(\widehat{M}_i \otimes \widehat{N}_i)\}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)' \\ &\quad + \Theta^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp\{(\widehat{M}_i \otimes \widehat{N}_i)'_\perp \Theta^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp\}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)'_\perp \Theta^{-1}. \qquad (8)\end{aligned}$$

Substituting (5) and (8) into $T\|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta^2$ proves (iv). □

Proof of Lemma 3. (i) The rate of convergence of $\widehat{B}_r^{RRA}$ follows from Lemma 1 (ii) and Lemma 2 (ii). The rate of convergence of the subspace estimators follows from the inequality

$$\max\{\|P_{\widehat{N}_r} - P_{N_r}\|, \|P_{\widehat{M}_r} - P_{M_r}\|\} \le K\|\widehat{B}_r^{RRA} - B^*\|, \qquad (9)$$

for some K that depends only on B^* (Gohberg et al., 2006, Theorem 13.5.1).

(ii) We will need the following lemma.

Lemma 7. Let $\widehat{B}, \widehat{B}^* \in \mathbb{G}^{n\times m}$, $\operatorname{rank}(\widehat{B}^*) = r$, and $\Theta \in \mathbb{P}_+^{nm}$. We assume that $\widehat{B}^* = O(1)$ and $\sigma_r(\widehat{B}^*) = O^{-1}(1)$ as $\widehat{B} - \widehat{B}^* \to 0$. Let the RRAs $\{\widehat{B}_i^{RRA} : 0 \le i < \min\{n, m\}\}$ be either DBAs or CDAs. In the latter case, we assume that $\operatorname{cond}(\Theta) = O(1)$ as $\widehat{B} - \widehat{B}^* \to 0$. Finally, let $\widehat{N}_i$ and $\widehat{M}_i$ span the left and right null spaces of $\widehat{B}_i^{RRA}$ respectively. Then there exists an oblique projection matrix $\widehat{P}_i$ such that:

(i) For $0 \le i < \min\{n, m\}$, $\operatorname{vec}(\widehat{B}_i) = \widehat{P}_i \operatorname{vec}(\widehat{B})$.

(ii) For $0 \le i < \min\{n, m\}$, $(I_{nm} - \widehat{P}_i)P_{\widehat{M}_i \otimes \widehat{N}_i} = I_{nm} - \widehat{P}_i$.

(iii) For $0 \le i \le r$, $\widehat{P}_i = O(1)$.

Proof of Lemma 7. First, we consider the DBA. Let $\widehat{U} = [\ \widehat{U}_{\cdot 1} \quad \widehat{U}_{\cdot 2}\ ]$ and $\widehat{S} = \begin{bmatrix} \widehat{S}_{1\cdot} \\ \widehat{S}_{2\cdot} \end{bmatrix}$ be partitioned conformably, with $\widehat{S}_{2\cdot} \in \mathbb{G}^{(n-i)\times m}$. Define the oblique projection matrices $\widehat{Q}_i = \widehat{U}_{\cdot 2}(\widehat{U}'_{\cdot 1\perp}\widehat{U}_{\cdot 2})^{-1}\widehat{U}'_{\cdot 1\perp}$ and $\widehat{W}_i = \widehat{V}'^{-1}\widehat{S}'_{1\cdot\perp}(\widehat{S}_{2\cdot}\widehat{S}'_{1\cdot\perp})^\dagger \widehat{S}_{2\cdot}\widehat{V}'$. Then clearly $\widehat{B} - \widehat{B}_i^{DBA} = \widehat{Q}_i\widehat{B} = \widehat{B}\widehat{W}_i = \widehat{Q}_i\widehat{B}\widehat{W}_i$. We may therefore, define $I_{nm} - \widehat{P}_i = \widehat{W}'_i \otimes \widehat{Q}_i$. Thus (i) and (ii) follow from the fact that the null space estimators may be chosen as $\widehat{N}_i = \widehat{U}_{\cdot 1\perp}$ and $\widehat{M}_i = \widehat{V}'^{-1}\widehat{S}'_{1\cdot\perp}$. We prove (iii) by showing that both $\widehat{Q}_i$ and $\widehat{W}_i$ are bounded. $\widehat{Q}_i$ is the product of $\widehat{U}_{\cdot 2}$, a submatrix of $\widehat{U}$, and $(\widehat{U}'_{\cdot 1\perp}\widehat{U}_{\cdot 2})^{-1}\widehat{U}'_{\cdot 1\perp}$, a submatrix of $\widehat{U}^{-1}$. Since both $\widehat{U}$ and $\widehat{U}^{-1}$ are bounded, $\widehat{Q}_i$ must be bounded. Next, we may choose $\widehat{S}'_{1\cdot\perp} = \begin{bmatrix} \widehat{S}_{11}^{-1}\widehat{S}_{12} \\ -I_{m-i} \end{bmatrix}$ so that $\widehat{W}_i = \widehat{V}'^{-1}\begin{bmatrix} 0 & \widehat{S}_{11}^{-1}\widehat{S}_{12}\widehat{S}_{22}^\dagger\widehat{S}_{22} \\ 0 & -\widehat{S}_{22}^\dagger\widehat{S}_{22} \end{bmatrix}\widehat{V}'$. $\widehat{V}$ and its inverse are bounded by assumption. $\widehat{S}_{22}^\dagger\widehat{S}_{22}$ is an orthogonal projection matrix

(Stewart & Sun, 1990, Theorem III.1.3) and therefore bounded. $\widehat{S}_{12}$ is a submatrix of $\widehat{S}$, which is bounded. Finally, $\|\widehat{S}_{11}^{-1}\|_2 = \sigma_i^{-1}(\widehat{S}_{11}) \leq \frac{1}{K_1\sigma_i(\widehat{B})}$, which is bounded as $\widehat{B} - \widehat{B}^* \to 0$.

Next, consider the CDA. (i) follows from equation (6). From (i) and (7), we find that

$$I_{nm} - \widehat{P}_i = \Theta(\widehat{M}_i \otimes \widehat{N}_i)((\widehat{M}_i \otimes \widehat{N}_i)'\Theta(\widehat{M}_i \otimes \widehat{N}_i))^{-1}(\widehat{M}_i \otimes \widehat{N}_i)', \tag{10}$$

which clearly satisfies $(I_{nm} - \widehat{P}_i)P_{\widehat{M}_i \otimes \widehat{N}_i} = I_{nm} - \widehat{P}_i$. (iii) follows from the fact that

$$\begin{aligned}
\|\widehat{P}_i\|_2 &= \|\Theta^{1/2}\Theta^{-1/2}\widehat{P}_i\Theta^{1/2}\Theta^{-1/2}\|_2 \\
&\leq \|\Theta^{1/2}\|_2\|\Theta^{-1/2}(\widehat{M}_i \otimes \widehat{N}_i)_\perp\{(\widehat{M}_i \otimes \widehat{N}_i)'_\perp\Theta^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp\}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)'_\perp\Theta^{-1/2}\|_2\|\Theta^{-1/2}\|_2 \\
&= (\mathrm{cond}(\Theta))^{1/2}.
\end{aligned}$$

The middle term on the right hand side of the inequality has an L^2 norm of 1 because it is an orthogonal projection. □

Lemma 7 (i) states that $\mathrm{vec}(\widehat{B}_i^{RRA})$ is obtained from $\mathrm{vec}(\widehat{B})$ by an oblique projection. Lemma 7 (ii) states that $\mathrm{vec}(\widehat{B}_i^{RRA})$ and $\mathrm{vec}(\widehat{B})$ can only be different if they differ along $\widehat{M}_i \otimes \widehat{N}_i$. That is, the RRA is obtained by removing from $\mathrm{vec}(\widehat{B})$ components along the estimated null spaces. Lemma 7 (iii) ensures that the oblique projection is bounded if the rank is underestimated.

Now note that $\|\widehat{B}_i^{RRA} - \widehat{B}\|^2 \geq \sigma_{i+1}^2(\widehat{B})$ by (Horn & Johnson, 1985, Example 7.4.1). If $i < r$, the right hand side converges in probability to $\sigma_{i+1}^2(B^*) > 0$ and so $\|\widehat{B}_i^{RRA} - \widehat{B}\|^2$ is $O_p^{-1}(1)$. Next, $\|\widehat{B}_i^{RRA} - \widehat{B}\|^2 = \|\mathrm{vec}(\widehat{B}_i^{RRA} - \widehat{B})\|^2 = \|(I_{nm} - \widehat{P}_i)\mathrm{vec}\widehat{B}\|^2$ by Lemma 7 (i). By Lemma 7 (ii), $\|(I_{nm} - \widehat{P}_i)\mathrm{vec}(\widehat{B})\| = \|(I_{nm} - \widehat{P}_i)(P_{\widehat{M}_i} \otimes P_{\widehat{N}_i})\mathrm{vec}(\widehat{B})\| \leq \|I_{nm} - \widehat{P}_i\|\|P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}\|$. Finally, Lemma 7 (iii) implies that $\widehat{P}_i = O_p(1)$ so that $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}$ is $O_p^{-1}(1)$.

If $n = m$ and $B^* \in \mathbb{P}^m$, the fact that $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$ implies that $\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$. The consistency of $\widehat{B}$ then implies that $B^*P_{\widehat{M}_i} = O_p^{-1}(1)$, which implies that $(B^*)^{1/2}P_{\widehat{M}_i} = O_p^{-1}(1)$ and therefore $P_{\widehat{M}_i}B^*P_{\widehat{M}_i} = O_p^{-1}(1)$. Appealing to consistency again gives us that $P_{\widehat{M}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$. The analogous result for $P_{\widehat{N}_i}\widehat{B}P_{\widehat{N}_i}$ follows from a similar argument.

(iii) The result follows from inequality (9) applied to $(B^*)_i^{RRA}$. □

Proof of Theorem 1. Substitute $\widehat{X} = P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}$, $X = P_{N_r}\widehat{B}P_{M_r}$, $\widehat{Y} = (P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})\widehat{\Omega}(P_{\widehat{M}_r} \otimes P_{\widehat{N}_r})$, and $Y = (P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r})$.

Then, under $H_0(r)$ or $H_T(r)$, Lemma 3 (i) implies

$$\begin{aligned}
\widehat{X} - X &= (P_{\widehat{N}_r} - P_{N_r})\widehat{B}(P_{\widehat{M}_r} - P_{M_r}) + P_{N_r}\widehat{B}(P_{\widehat{M}_r} - P_{M_r}) + (P_{\widehat{N}_r} - P_{N_r})\widehat{B}P_{M_r} = O_p(T^{-1}) \\
\widehat{Y} - Y &= ((P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}) - (P_{M_r} \otimes P_{N_r}))\widehat{\Omega}((P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}) - (P_{M_r} \otimes P_{N_r})) \\
&\quad + (P_{M_r} \otimes P_{N_r})\widehat{\Omega}((P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}) - (P_{M_r} \otimes P_{N_r})) \\
&\quad + ((P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}) - (P_{M_r} \otimes P_{N_r}))\widehat{\Omega}(P_{M_r} \otimes P_{N_r}) = O_p(T^{-1/2}).
\end{aligned}$$

Next we show that, under either $H_0(r)$ or $H_T(r)$, each of Assumptions A and B imply that $Y^\dagger = O_p(1)$. In particular, we will show that $\text{rank}(Y) = (n-r)(m-r)$ almost surely and $\sigma_{(n-r)(m-r)}(Y) = O_p^{-1}(1)$ under Assumptions A, while $\text{rank}(Y) = (m-r)(m-r+1)/2$ almost surely and $\sigma_{(m-r)(m-r+1)/2}(Y) = O_p^{-1}(1)$ under Assumptions B. Consider Assumptions A first, $\text{rank}((P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r})) = \text{rank}((P_{M_r} \otimes P_{N_r})\widehat{\Omega}^{1/2}) = \text{rank}(P_{M_r} \otimes P_{N_r}) = \text{rank}(P_{N_r})\text{rank}(P_{M_r}) = (n-r)(m-r)$ almost surely. The first two equalities follow from (d) and (b) of Result 0.4.6 of Horn & Johnson (1985) respectively, the third equality follows from Theorem 4.2.15 of Horn & Johnson (1991), and the last from the Spectral Decomposition Theorem along with Exercise 5 of Section 1.1 of Horn & Johnson (1985). Now

$$\begin{aligned}
\sigma_{(n-r)(m-r)}((P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r})) &= \sigma^2_{(n-r)(m-r)}((P_{M_r} \otimes P_{N_r})\widehat{\Omega}^{1/2}) \\
&\geq \sigma^2_{(n-r)(m-r)}(P_{M_r} \otimes P_{N_r})\sigma_1^{-2}(\widehat{\Omega}^{-1/2}) \\
&= \sigma_{nm}(\widehat{\Omega})
\end{aligned}$$

almost surely (Horn & Johnson, 1991, Theorem 3.3.16 (d)), which is bounded away from zero in probability by Assumptions A. On the other hand, Assumptions B imply that $P_{N_r} = P_{M_r}$ and so, by similar arguments to those used earlier, $\text{rank}((P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r})) = \text{rank}((P_{M_r} \otimes P_{M_r})D_m\widehat{\Psi}D_m'(P_{M_r} \otimes P_{M_r})) = \text{rank}((P_{M_r} \otimes P_{M_r})D_m)$ almost surely. Now, for $Z \in \mathbb{S}^m$, $0 = (P_{M_r} \otimes P_{M_r})D_m\text{vech}(Z) = \text{vec}(P_{M_r}ZP_{M_r})$ if and only if $\text{vech}(P_{M_r}ZP_{M_r}) = 0$. But the set of all such Z has dimension $r(r+1)/2 + r(m-r)$ so $\text{rank}((P_{M_r} \otimes P_{M_r})D_m) = m(m+1)/2 - r(r+1)/2 - r(m-r) = (m-r)(m-r+1)/2$. It follows that the rank of $(P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r})$ is $(m-r)(m-r+1)/2$ almost surely. Finally,

$$\begin{aligned}
&\sigma_{(m-r)(m-r+1)/2}((P_{M_r} \otimes P_{N_r})\widehat{\Omega}(P_{M_r} \otimes P_{N_r})) \\
&\qquad = \sigma^2_{(m-r)(m-r+1)/2}((P_{M_r} \otimes P_{M_r})D_m\widehat{\Psi}^{1/2}) \\
&\qquad \geq \sigma^2_{(m-r)(m-r+1)/2}((P_{M_r} \otimes P_{M_r})D_m)\sigma_{m(m+1)/2}(\widehat{\Psi})
\end{aligned}$$

almost surely. The result then follows from the fact that $\sigma_{m(m+1)/2}(\widehat{\Psi}) = O_p^{-1}(1)$ and $\sigma^2_{(m-r)(m-r+1)/2}((P_{M_r} \otimes P_{M_r})D_m)$ is a positive constant as $\text{rank}((P_{M_r} \otimes P_{M_r})D_m) = (m - r)(m - r + 1)/2$.

Since $X = O_p(T^{-1/2})$ under either $H_0(r)$ or $H_T(r)$, Assumption K (i) implies that $L_1(\widehat{X}, X, \widehat{Y}, Y) = O_p(T^{1/2-\theta})$ and $L_2(\widehat{X}, X, \widehat{Y}, Y) = O_p(T^{-\theta})$. Therefore, under $H_0(r)$ or $H_T(r)$, $T^\theta \kappa(X, Y) - T^\theta \kappa(\widehat{X}, \widehat{Y}) = O_p(T^{\theta+1/2-\theta-1}) + O_p(T^{\theta-\theta-1/2}) = O_p(T^{-1/2})$. The weak plug–in principle under $H_0(r)$ and $H_T(r)$ is therefore established.

Consider next the plug–in principle under $H_1(r)$. Since the Euclidean norm can be though of as the Hilbert space norm on the set of matrices $\mathbb{R}^{n\times m}$ associated with inner product $\langle A, B\rangle = \text{tr}(A'B)$, Theorem 4.10 of Rudin (1986) implies that there exists a unique $X^* \in \mathcal{P}$ that minimizes the distance from $\mathcal{P}$ to $P_{N_r}\widehat{B}P_{M_r}$. Since $\inf_{X\in\mathcal{P}} \|P_{N_r}\widehat{B}P_{M_r} - X\| = o_p(1)$, it follows that $X - X^* = o_p(1)$. Since $\widehat{B} - B^* = o_p(1)$ and $P_{N_r}B^*P_{M_r} \neq 0$, it follows that $X^* = O_p^{-1}(1)$. Then Assumption K (ii) implies that $|\kappa(X^*, Y)| = O_p^{-1}(1)$ and Assumption K (i) implies that $|\kappa(X, Y) - \kappa(X^*, Y)| = o_p(1)$. Putting these two together, we have that $|\kappa(X, Y)| \geq |\kappa(X^*, Y)| - |\kappa(X, Y) - \kappa(X^*, Y)| = O_p^{-1}(1) + o_p(1)$. Thus, $|\kappa(X, Y)| = O_p^{-1}(1)$. Next, take $\widehat{X}^* \in \mathcal{P}$ as the closest element of $\mathcal{P}$ to $P_{\widehat{N}_r}\widehat{B}P_{\widehat{M}_r}$ and note that $\|\widehat{X} - \widehat{X}^*\| = \inf_{X\in\mathcal{P}} \|P_{N_r}\widehat{B}P_{M_r} - X\| = o_p(1)$. Lemma 3 (ii) then implies that $\widehat{X} = O_p^{-1}(1)$ so $\widehat{X}^* = O_p^{-1}(1)$ as well. Thus Assumption K (ii) implies that $|\kappa(\widehat{X}^*, \widehat{Y})| = O_p^{-1}(1)$ and Assumption K (i) implies that $|\kappa(\widehat{X}, \widehat{Y}) - \kappa(\widehat{X}^*, \widehat{Y})| = o_p(1)$. Therefore, again, we have that $|\kappa(\widehat{X}, \widehat{Y})| \geq |\kappa(\widehat{X}^*, \widehat{Y})| - |\kappa(\widehat{X}, \widehat{Y}) - \kappa(\widehat{X}^*, \widehat{Y})| = O_p^{-1}(1) + o_p(1)$. Thus, $|\kappa(\widehat{X}, \widehat{Y})| = O_p^{-1}(1)$. This establishes the weak plug–in principle under $H_1(r)$.

Finally, if we additionally assume that the underlying RRA is continuous at B^*, Lemma 3 (iii) implies that $\widehat{X} - X = o_p(1)$ and $\widehat{Y} - Y = o_p(1)$ and since $L_1(\widehat{X}, X, \widehat{Y}, Y) = O_p(1)$ and $L_2(\widehat{X}, X, \widehat{Y}, Y) = O_p(1)$, the strong plug–in principle holds. □

Proof of Corollary 1. The proof for general κ follows from Theorem 1. For the LRA, F, and JA statistics, the proof follows from the fact that $\kappa(X, Y) = \varphi(Y^{\dagger/2}\text{vec}(X)) + O(\|Y^{\dagger/2}\text{vec}(X)\|^3)$ as $Y^{\dagger/2}\text{vec}(X) \to 0$, where φ is homogenous of degree 2. □

Proof of Corollary 2. The proof for general κ, LRA, F, and JA statistics follows the same logic as in Corollary 1. For the t statistic, Theorem 1 implies that it has the same limiting distribution as the infeasible statistic $t(\widehat{B}, \widehat{\Psi}, P_{M_r})$. Now the numerator of the infeasible

statistic has the limiting distribution $\mathrm{tr}(M_r \mathrm{mat}(D_{m-r}\xi_r)M_r')) = \mathrm{tr}(\mathrm{mat}(D_{m-r}\xi_r)))$. On the other hand, the square of the denominator can be written as

$$\begin{aligned}
&\mathrm{vec}'(I_m)((M_rM_r' \otimes M_rM_r')D_m\widehat{\Psi}D_m'(M_rM_r' \otimes M_rM_r'))\mathrm{vec}(I_m) \\
&= \mathrm{vec}'(I_m)((M_r \otimes M_r)(M_r \otimes M_r)'D_m\widehat{\Psi}D_m'(M_r \otimes M_r)(M_r \otimes M_r)')\mathrm{vec}(I_m) \\
&= \mathrm{vec}'(I_m)((M_r \otimes M_r)(M_r \otimes M_r)'D_mD_m^{\dagger}D_m\widehat{\Psi}D_m'D_m^{\dagger}{}'D_m'(M_r \otimes M_r)(M_r \otimes M_r)')\mathrm{vec}(I_m) \\
&= \mathrm{vec}'(I_m)((M_r \otimes M_r)D_{m-r}D_{m-r}^{\dagger}(M_r \otimes M_r)'D_m\widehat{\Psi}D_m'(M_r \otimes M_r)D_{m-r}^{\dagger}{}'D_{m-r}'(M_r \otimes M_r)')\mathrm{vec}(I_m),
\end{aligned}$$

where the first equality follows from the properties of Kronecker products, the second from the properties of the generalized inverse, and the third from the properties of the commutator (Magnus & Neudecker, 1999, Theorems 3.9 and 3.12). This then has a limiting distribution of $\mathrm{vec}'(I_m)((M_r \otimes M_r)D_{m-r}\Omega_rD_{m-r}'(M_r \otimes M_r)')\mathrm{vec}(I_m)$. The final step follows from the fact that for any symmetric matrix $A \in \mathbb{R}^{(m-r)\times(m-r)}$, $\mathrm{vec}'(I_m)(M_r \otimes M_r)D_{m-r}\mathrm{vech}(A) = \mathrm{trace}(M_rAM_r') = \mathrm{trace}(A) = \mathrm{vec}'(I_{m-r})\mathrm{vec}(A) = \mathrm{vec}'(I_{m-r})D_{m-r}\mathrm{vech}(A)$, thus $\mathrm{vec}'(I_m)(M_r \otimes M_r)D_{m-r} = \mathrm{vec}'(I_{m-r})D_{m-r}$. The final substitution follows from the fact "these convergences are jointly valid" as Nyblom and Harvey would say (Nyblom & Harvey, 2000, p. 194). □

Proof of Lemma 4. First, note that the singular values of $[\ N_{r\perp} \quad N_{rT}\ ]$ and $[\ M_{r\perp} \quad M_{rT}\ ]$ are bounded away from zero as $\widehat{B}_T - \widehat{B}_T^* \to 0$. To see this, write

$$\begin{aligned}
\sigma_n([\ N_{r\perp} \quad P_{N_{rT}}N_r\ ]) &\le \sigma_n([\ N_{r\perp} \quad N_{rT}\ ])\sigma_1\left(\begin{bmatrix} I_r & 0 \\ 0 & (N_{rT}'N_{rT})^{-1}N_{rT}'N_r \end{bmatrix}\right) \\
&\le \sigma_n([\ N_{r\perp} \quad N_{rT}\ ])(1 + \|(N_{rT}'N_{rT})^{-1}N_{rT}'N_r\|_2) \\
&\le \sigma_n([\ N_{r\perp} \quad N_{rT}\ ])(1 + \|(N_{rT}'N_{rT})^{-1/2}\|_2\|(N_{rT}'N_{rT})^{-1/2}N_{rT}'\|_2\|N_r\|_2) \\
&= \sigma_n([\ N_{r\perp} \quad N_{rT}\ ])\left(1 + \frac{\|N_r\|_2}{\sigma_{n-r}(N_{rT})}\right),
\end{aligned}$$

where the first inequality follows from Theorem 3.3.16 (d) of Horn & Johnson (1991). Now since $[\ N_{r\perp} \quad N_r\ ]$ is of full rank and $P_{N_{rT}}N_r \to N_r$ as $\widehat{B}_T - \widehat{B}_T^* \to 0$, the left hand side is bounded away from zero as $\widehat{B}_T - \widehat{B}_T^* \to 0$. Since $\sigma_{n-r}(N_{rT})$ is bounded away from zero as $\widehat{B}_T - \widehat{B}_T^* \to 0$, it follows that $\sigma_n([\ N_{r\perp} \quad N_{rT}\ ])$ is also bounded away from zero. A similar argument proves that $\sigma_m([\ M_{r\perp} \quad M_{rT}\ ])$ is also bounded away from zero as $\widehat{B}_T - \widehat{B}_T^* \to 0$.

Next, define $\widehat{B}^* = [\ N_{r\perp} \quad N_{rT}\ ]^{-1\prime}\,\widehat{B}_T^*\,[\ M_{r\perp} \quad M_{rT}\ ]^{-1}$ and note that

$$\|\widehat{B} - \widehat{B}^*\| \le \|[\ N_{r\perp} \quad N_{rT}\ ]^{-1}\|\|[\ M_{r\perp} \quad M_{rT}\ ]^{-1}\|\|\widehat{B}_T - \widehat{B}_T^*\|.$$

Similarly

$$\|\widehat{B}^*\| \leq \|[\ N_{r\perp} \quad N_{rT}\ ]^{-1}\|\|[\ M_{r\perp} \quad M_{rT}\ ]^{-1}\|\|\widehat{B}^*_T\|.$$

Thus $\widehat{B} - \widehat{B}^* = O(\|\widehat{B}_T - \widehat{B}^*_T\|)$ and $\widehat{B}^* = O(1)$ as $\widehat{B}_T - \widehat{B}^*_T \to 0$. It also follows that $|\sigma_r(\widehat{B}) - \sigma_r(\widehat{B}^*)| = O(\|\widehat{B}_T - \widehat{B}^*_T\|)$ as $\widehat{B}_T - \widehat{B}^*_T \to 0$ (Horn & Johnson, 1991, Theorem 3.3.16 (c)). Thus $\sigma_r(\widehat{B}^*)$ is bounded away from zero as $\widehat{B}_T - \widehat{B}^*_T \to 0$ if and only if $\sigma_r(\widehat{B})$ is. Since $\sigma_r(N'_{r\perp}\widehat{B}M_{r\perp}) \leq \|N_{r\perp}\|_2\|M_{r\perp}\|_2\sigma_r(\widehat{B})$ (Horn & Johnson, 1991, Theorem 3.3.16 (d)), $\sigma_r(\widehat{B})$ is bounded away from zero as $\widehat{B}_T - \widehat{B}^*_T \to 0$ and therefore so is $\sigma_r(\widehat{B}^*)$. Property (ii) of Definition 2 then implies that $\|\widehat{B} - \widehat{B}^{DBA}_i\| = O(\sigma_{r+1}(\widehat{B}))$ as $\widehat{B}_T - \widehat{B}^*_T \to 0$. Now,

$$\sigma_{r+1}(\widehat{B}) \leq \|[\ N_{r\perp} \quad N_{rT}\ ]^{-1}\|_2\|[\ M_{r\perp} \quad M_{rT}\ ]^{-1}\|_2\sigma_{r+1}(\widehat{B}_T)$$

by Theorem 3.3.16 (d) of Horn & Johnson (1991). Thus $\sigma_{r+1}(\widehat{B}) = O(\sigma_{r+1}(\widehat{B}_T))$. Now since $\text{rank}(\widehat{B}^*_T) = r$, it follows from the properties of singular values that $\sigma_{r+1}(\widehat{B}_T) = O(\|\widehat{B}_T - \widehat{B}^*_T\|)$ and therefore $\|\widehat{B} - \widehat{B}^{DBA}_i\| = O(\|\widehat{B}_T - \widehat{B}^*_T\|)$ as $\widehat{B}_T - \widehat{B}^*_T \to 0$.

Next, recall from Lemma 7 (iii) that $\widehat{B}^{DBA}_i = (I_n - \widehat{Q}_i)\widehat{B}$, where $\widehat{Q}_i = O(1)$ as $\widehat{B} - \widehat{B}^* \to 0$. By a similar argument to that used above, $\widehat{Q}_i = O(1)$ as $\widehat{B}_T - \widehat{B}^*_T \to 0$. This then implies that $\widehat{B}^{DBA}_i M_{rT} = O(\|\widehat{B}M_{rT}\|) = O(\|\widehat{B}_T - \widehat{B}^*_T\|)$ as $\widehat{B}_T - \widehat{B}^*_T \to 0$. A similar argument applies for $N'_{rT}\widehat{B}^{DBA}_i$. □

Proof of Lemma 5. First, note that since $[\ N_{r\perp} \quad N_{rT}\ ]$ and $[\ M_{r\perp} \quad M_{rT}\ ]$ have singular values bounded away from zero (see the proof of Lemma 4), so does Z_T (Horn & Johnson, 1991, Theorem 4.2.15).

Next, define $\widehat{B}^{CDA}_{iT} = [\ N_{r\perp} \quad N_{rT}\ ]'\widehat{B}^{CDA}_i[\ M_{r\perp} \quad M_{rT}\ ]$. Then just as in the proof of Lemma 2 (ii),

$$\begin{aligned}
\|\widehat{B}_T - \widehat{B}^{CDA}_{iT}\|^2 &= \|Z'_T\text{vec}(\widehat{B} - \widehat{B}^{CDA}_i)\|^2 \\
&\leq \lambda_1(\Theta_T)\|Z'_T\text{vec}(\widehat{B} - \widehat{B}^{CDA}_i)\|^2_{\Theta_T} \\
&= \lambda_1(\Theta_T)\|\widehat{B} - \widehat{B}^{CDA}_i\|^2_{\Theta}.
\end{aligned}$$

Since $\widehat{B}^* = [\ N_{r\perp} \quad N_{rT}\ ]^{-1\prime}\, \widehat{B}^*_T\, [\ M_{r\perp} \quad M_{rT}\ ]^{-1}$ has a rank r,

$$\begin{aligned}
\|\widehat{B}_T - \widehat{B}^{CDA}_{iT}\|^2 &\leq \lambda_1(\Theta_T)\|\widehat{B} - \widehat{B}^*\|^2_\Theta \\
&= \lambda_1(\Theta_T)\|Z'_T \mathrm{vec}(\widehat{B} - \widehat{B}^*)\|^2_{\Theta_T} \\
&\leq \lambda_1(\Theta_T)\lambda_1(\Theta_T^{-1})\|Z'_T \mathrm{vec}(\widehat{B} - \widehat{B}^*)\|^2 \\
&= \mathrm{cond}(\Theta_T)\|\widehat{B}_T - \widehat{B}^*_T\|^2,
\end{aligned}$$

where the last equality follows from the definition of $\widehat{B}_T$ and $\widehat{B}^*_T$. It follows that $\widehat{B}^{CDA}_i M_{rT}$ and $N'_{rT}\widehat{B}^{CDA}_i$ are $O(\|\widehat{B}_T - \widehat{B}^*_T\|)$. The fact that $\widehat{B} - \widehat{B}^{CDA}_i$ is also $O(\|\widehat{B}_T - \widehat{B}^*_T\|)$ follows from the fact that $\|\widehat{B} - \widehat{B}^{CDA}_i\| \leq \|Z_T^{-1}\|\|\widehat{B}_T - \widehat{B}^{CDA}_{iT}\|$ and the boundedness away from zero of the singular values of Z_T. Finally, the bounds on $(I_n - P_{\widehat{N}_r})P_{\widehat{N}_i}$ and $(I_m - P_{\widehat{M}_r})P_{\widehat{M}_i}$ developed in the proof of Lemma 2 (iii) along with the results above imply that they are both $O(\|\widehat{B}_T - \widehat{B}^*_T\|)$. □

Proof of Lemma 6. (i) By Lemmas 4 and 5, $T^\gamma\|\widehat{B} - \widehat{B}^{RRA}_r\| = O_p(1)$. On the other hand, multiplying and dividing $T^\gamma(\widehat{B} - \widehat{B}^*)$ by $[\ N_{r\perp} \quad N_{rT}\ ]'$ and $[\ M_{r\perp} \quad M_{rT}\ ]$ on the left and right, $T^\gamma(\widehat{B} - \widehat{B}^*) = [\ N_{r\perp} \quad N_{rT}\ ]^{-1\prime} \begin{bmatrix} 0 & T^\gamma N'_{r\perp}\widehat{B}M_{rT} \\ T^\gamma N'_{rT}\widehat{B}M_{r\perp} & T^\gamma N'_{rT}\widehat{B}M_{rT} \end{bmatrix} [\ M_{r\perp} \quad M_{rT}\ ]^{-1} = O_p(1)$. Lemmas 4 and 5 also imply that $T^\gamma N'_{rT}\widehat{B}^{RRA}_r$ and $T^\gamma \widehat{B}^{RRA}_r M_{rT}$ are $O_p(1)$. Therefore, $T^\gamma \widehat{V}'_{\cdot 1} M_{rT} = O_p(1)$, where $\widehat{U}_{\cdot 1}\widehat{S}_{11}\widehat{V}'_{\cdot 1}$ is the SVD of $\widehat{B}^{RRA}_r$, since $\widehat{U}_{\cdot 1} = O_p(1)$ and $\widehat{S}^{-1}_{11} = O_p(1)$. This then implies that $T^\gamma(P_{M_{rT}} - P_{\widehat{M}_r})M_{rT} = T^\gamma(I_m - P_{\widehat{M}_r})M_{rT} = T^\gamma \widehat{V}_{\cdot 1}\widehat{V}'_{\cdot 1}M_{rT} = O_p(1)$. If we now let V_{rT} be a matrix of orthonormal columns that spans M_{rT}, then since the singular values of M_{rT} are $O_p^{-1}(1)$, we have that $T^\gamma \widehat{V}'_{\cdot 1}V_{rT} = O_p(1)$, which implies that $T^\gamma(P_{M_{rT}} - P_{\widehat{M}_r}) = O_p(1)$ by Theorem 2.6.1 of Golub & Van Loan (1996). In particular, $T^\gamma(P_{M_{rT}} - P_{\widehat{M}_r})M_{r\perp} = O_p(1)$. An analogous argument proves the rates for the left null space estimator.

(ii) The proof for the DBA is identical to that used in Lemma 3 (ii). As for the CDA, the argument requires a slight modification because $\widehat{P}_i$ may not be bounded in probability. However, $\widehat{P}_{iT} = Z'_T\widehat{P}_i Z_T^{-1\prime} = Z'_T\Theta^{1/2}\Theta^{-1/2}\widehat{P}_i\Theta^{1/2}\Theta^{-1/2}Z_T^{-1\prime} = O_p(\mathrm{cond}^{1/2}(\Theta_T)) = O_p(1)$ as $\Theta^{-1/2}\widehat{P}_i\Theta^{1/2}$ is an orthogonal projection matrix. Note that $\widehat{P}_{iT}$ satisfies (6) with all the matrices substituted for ones with subindex T. This follows from the fact that $(\widehat{M}_{iT} \otimes \widehat{N}_{iT})_\perp =$

$(Z_T^{-1}(\widehat{M}_i \otimes \widehat{N}_i))_\perp = Z_T(\widehat{M}_i \otimes \widehat{N}_i)_\perp$. Now

$$\begin{aligned}
\sigma_r^2(\widehat{B}) &\leq \|\widehat{B} - \widehat{B}_i^{CDA}\|^2 \\
&\leq \lambda_1(\Theta)\|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta^2 \\
&\leq \sigma_1^2(Z_T^{-1})\lambda_1(\Theta_T)\|\widehat{B} - \widehat{B}_i^{CDA}\|_\Theta^2 \\
&= \frac{\lambda_1(\Theta_T)}{\sigma_{nm}^2(Z_T)}\|\widehat{B}_T - \widehat{B}_{iT}^{CDA}\|_{\Theta_T}^2,
\end{aligned}$$

where $\widehat{B}_{iT}^{CDA} = [\ N_{r\perp} \quad N_{rT}\ ]'\widehat{B}_i^{CDA}[\ M_{r\perp} \quad M_{rT}\ ]$. By the properties of quadratic forms, we then have

$$\begin{aligned}
\sigma_r^2(\widehat{B}) &\leq \frac{\mathrm{cond}(\Theta_T)}{\sigma_{nm}^2(Z_T)}\|\widehat{B}_T - \widehat{B}_{iT}^{CDA}\|^2 \\
&= \frac{\mathrm{cond}(\Theta_T)}{\sigma_{nm}^2(Z_T)}\|(I_{nm} - \widehat{P}_{iT})\mathrm{vec}(\widehat{B}_T)\|^2 \\
&= \frac{\mathrm{cond}(\Theta_T)}{\sigma_{nm}^2(Z_T)}\|(I_{nm} - \widehat{P}_{iT})(P_{\widehat{M}_{iT}} \otimes P_{\widehat{N}_{iT}})\mathrm{vec}(\widehat{B}_T)\|^2 \\
&\leq \frac{\mathrm{cond}(\Theta_T)}{\sigma_{nm}^2(Z_T)}\|I_{nm} - \widehat{P}_{iT}\|^2\|P_{\widehat{N}_{iT}}\widehat{B}_T P_{\widehat{M}_{iT}}\|^2.
\end{aligned}$$

Since $\sigma_r(\widehat{B})$ and $\sigma_{nm}(Z_T)$ are $O_p^{-1}(1)$, while $\mathrm{cond}(\Theta_T)$ and $\widehat{P}_{iT}$ are $O_p(1)$, it follows that $P_{\widehat{N}_{iT}}\widehat{B}_T P_{\widehat{M}_{iT}} = O_p^{-1}(1)$. The final boundedness result follows from the fact that $P_{\widehat{N}_{iT}}\widehat{B}_T P_{\widehat{M}_{iT}} = P_{\widehat{N}_{iT}}[\ N_{r\perp} \quad N_{rT}\ ]'P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i}[\ M_{r\perp} \quad M_{rT}\ ]P_{\widehat{M}_{iT}}$.

When $n = m$ and the null spaces are estimated by DBA, $P_{\widehat{N}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$ implies that $\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$ and therefore $\widehat{B}^*P_{\widehat{M}_i} = O_p^{-1}(1)$. Now factoring out the rescaling matrices, $\inf_{X\in\mathbb{P}^m}\|\widehat{B}^* - X\| \leq \|Z_T^{-1}\|\inf_{X\in\mathbb{P}^m}\left\|\begin{bmatrix} M_{r\perp}'\widehat{B}M_{r\perp} & 0 \\ 0 & 0\end{bmatrix} - X\right\| = \|Z_T^{-1}\|\inf_{X\in\mathbb{P}^r}\|M_{r\perp}'\widehat{B}M_{r\perp} - X\| = o_p(1)$. Let X be the closest elements of $\mathbb{P}^m$ to $\widehat{B}^*$ (Rudin, 1986, Theorem 4.10). Then, $XP_{\widehat{M}_i} = O_p^{-1}(1)$ and $X^{1/2}P_{\widehat{M}_i} = O_p^{-1}(1)$, which in turn implies that $P_{\widehat{M}_i}XP_{\widehat{M}_i} = O_p^{-1}(1)$. Since $\widehat{B} - X = o_p(1)$, we have that $P_{\widehat{M}_i}\widehat{B}P_{\widehat{M}_i} = O_p^{-1}(1)$. A similar argument proves that $P_{\widehat{N}_i}\widehat{B}P_{\widehat{N}_i} = O_p^{-1}(1)$ as well as the results for the CDA.

(iii) The proof is identical to that of Lemma 3 (iii). □

Proof of Theorem 2. Suppose κ is invariant with respect to the transformations given in (i). Then, under either $H_0(r)$ or $H_T(r)$, we may write the statistic as $T^{2\gamma\theta}\kappa(\widehat{X}, \widehat{Y})$ with $\widehat{X} = P_{\widehat{N}_{rT}}\widehat{B}_T P_{\widehat{M}_{rT}}$ and $\widehat{Y} = (P_{\widehat{M}_{rT}} \otimes P_{\widehat{N}_{rT}})\widehat{\Omega}_T(P_{\widehat{M}_{rT}} \otimes P_{\widehat{N}_{rT}})$, where $\widehat{\Omega}_T = Z_T'\widehat{\Omega}Z_T$, where $Z_T = [\ M_{r\perp} \quad M_{rT}\ ] \otimes [\ N_{r\perp} \quad N_{rT}\ ]$. The infeasible analogue is then given by $X = P_{\widehat{N}_r^*}\widehat{B}_T P_{\widehat{M}_r^*} = \begin{bmatrix} 0 & 0 \\ 0 & N_{rT}'\widehat{B}M_{rT}\end{bmatrix}$ and $Y = \left(\left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r}\end{smallmatrix}\right] \otimes \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{n-r}\end{smallmatrix}\right]\right)\widehat{\Omega}_T\left(\left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r}\end{smallmatrix}\right] \otimes \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{n-r}\end{smallmatrix}\right]\right)$. The proof of the plug–in principle under $H_0(r)$ and $H_T(r)$ then proceeds by demonstrating that $P_{\widehat{N}_{rT}}$ and $P_{\widehat{M}_{rT}}$ are

T^{γ}–consistent. To that end, write

$$\begin{aligned}\left\| P_{\widehat{M}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r} \end{smallmatrix}\right] \right\|_2 &\leq K \left\| \widehat{M}_{rT} Q - \left[\begin{smallmatrix} 0 \\ I_{m-r} \end{smallmatrix}\right] \right\|_2 \\ &= K \left\| [\, M_{r\perp} \quad M_{rT} \,]^{-1} (\widehat{M}_r Q - M_{rT}) \right\|_2 \\ &= K \left\| [\, M_{r\perp} \quad M_{rT} \,]^{-1} \right\|_2 \left\| \widehat{M}_r Q - M_{rT} \right\|,\end{aligned}$$

where $K > 0$ and depends only on $\left[\begin{smallmatrix} 0 \\ I_{m-r} \end{smallmatrix}\right]$ and $Q \in \mathbb{G}^{(m-r)\times(m-r)}$ is arbitrary (Gohberg et al., 2006, Theorem 13.5.1). We now claim that the infimum of $\|\widehat{M}_r Q - M_{rT}\|$ over $Q \in \mathbb{G}^{(m-r)\times(m-r)}$ is $\|(P_{\widehat{M}_r} - I_m) M_{rT}\|$. To see this, note that theory of least squares approximation implies that the infimum of $\|\widehat{M}_r Q - M_{rT}\|$ over $Q \in \mathbb{R}^{(m-r)\times(m-r)}$ is $\|(P_{\widehat{M}_r} - I_m) M_{rT}\|$ and is therefore less than or equal to the infimum over the subset $\mathbb{G}^{(m-r)\times(m-r)}$. On the other hand, the fact that $\mathbb{G}^{(m-r)\times(m-r)}$ is dense in $\mathbb{R}^{(m-r)\times(m-r)}$ (Horn & Johnson, 1985, Exercise 5.6.8) and the continuity for the norm furnish the opposite inequality. Continuing then,

$$T^{\gamma} \left\| P_{\widehat{M}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r} \end{smallmatrix}\right] \right\|_2 \leq K \left\| [\, M_{r\perp} \quad M_{rT} \,]^{-1} \right\|_2 \left\| T^{\gamma} (P_{\widehat{M}_r} - P_{M_{rT}}) M_{rT} \right\| = O_p(1).$$

A similar expression holds for the left null space estimator. Following the same logic as in Theorem 1 then, $\widehat{X} - X = O_p(T^{-2\gamma})$, $\widehat{Y} - Y = O_p(T^{-\gamma})$, $Y^{\dagger} = O_p(1)$, and $X = O_p(T^{-\gamma})$ under either $H_0(r)$ or $H_T(r)$. Condition (i) of Theorem 1 then implies that $L_1(\widehat{X}, X, \widehat{Y}, Y) = O_p(T^{-\gamma(2\theta-1)})$ and $L_2(\widehat{X}, X, \widehat{Y}, Y) = O_p(T^{-2\gamma\theta})$. Thus, under $H_0(r)$ or $H_T(r)$, $T^{2\gamma\theta}\kappa(\widehat{X}, \widehat{Y}) - T^{2\gamma\theta}\kappa(X, Y) = O_p(T^{2\gamma\theta - 2\gamma\theta + \gamma - 2\gamma}) + O_p(T^{2\gamma\theta - 2\gamma\theta - \gamma}) = O_p(T^{-\gamma})$. The weak plug–in principle under $H_0(r)$ and $H_T(r)$ is therefore established.

Consider next the plug–in principle under $H_1(r)$. For the DBA the result follows as in Theorem 1. For the CDA on the other hand, let $X^* \in \mathcal{P}$ be the unique matrix that minimizes the distance from $\mathcal{P}$ to $P_{\widehat{N}^*_{rT}} \widehat{B}_T P_{\widehat{M}^*_{rT}}$. Since $\inf_{X \in \mathcal{P}} \|P_{\widehat{N}^*_{rT}} \widehat{B}_T P_{\widehat{M}^*_{rT}} - X\| = o_p(1)$, it follows that $X - X^* = o_p(1)$.

Recalling that $\widehat{B}^*_T = \left[\begin{smallmatrix} N'_{r\perp} \widehat{B} M_{r\perp} & 0 \\ 0 & 0 \end{smallmatrix}\right]$ satisfies $\widehat{B}_T - \widehat{B}^*_T = o_p(1)$ and $P_{\widehat{N}^*_{rT}} \widehat{B}^*_T P_{\widehat{M}^*_{rT}} = O_p^{-1}(1)$ (apply Lemma 6 (ii) to $\widehat{B}^*$), it follows that $X^* = O_p^{-1}(1)$. Then Assumption K (ii) implies that $|\kappa(X^*, Y)| = O_p^{-1}(1)$ and Assumption K (i) implies that $|\kappa(X, Y) - \kappa(X^*, Y)| = o_p(1)$. Putting these two together, we have that $|\kappa(X, Y)| \geq |\kappa(X^*, Y)| - |\kappa(X, Y) - \kappa(X^*, Y)| = O_p^{-1}(1) + o_p(1)$. Thus, $|\kappa(X, Y)| = O_p^{-1}(1)$. Next, take $\widehat{X}^* \in \mathcal{P}$ as the closest element of $\mathcal{P}$ to $P_{\widehat{N}_{rT}} \widehat{B}_T P_{\widehat{M}_{rT}}$ and note that $\|\widehat{X} - \widehat{X}^*\| = \inf_{X \in \mathcal{P}} \|P_{\widehat{N}_{rT}} \widehat{B}_T P_{\widehat{M}_{rT}} - X\| = o_p(1)$. Lemma 6 (ii) then implies that $\widehat{X} = O_p^{-1}(1)$ so $\widehat{X}^* = O_p^{-1}(1)$ as well. Thus Assumption K (ii) implies

that $|\kappa(\widehat{X}^*, \widehat{Y})| = O_p^{-1}(1)$ and Assumption K (i) implies that $|\kappa(\widehat{X}, \widehat{Y}) - \kappa(\widehat{X}^*, \widehat{Y})| = o_p(1)$. Therefore, again, we have that $|\kappa(\widehat{X}, \widehat{Y})| \geq |\kappa(\widehat{X}^*, \widehat{Y})| - |\kappa(\widehat{X}, \widehat{Y}) - \kappa(\widehat{X}^*, \widehat{Y})| = O_p^{-1}(1) + o_p(1)$. Thus, $|\kappa(\widehat{X}, \widehat{Y})| = O_p^{-1}(1)$. This establishes the weak plug–in principle under $H_1(r)$. The strong plug–in principle under $H_1(r)$ follows from the same argument used in the proof of Theorem 1.

Now suppose $Z_T = O_p(1)$. Let $\widehat{X}$ and $\widehat{Y}$ be as in Theorem 1 and let $X = P_{N_{rT}} \widehat{B} P_{M_{rT}}$ and $Y = (P_{M_{rT}} \otimes P_{N_{rT}}) \widehat{\Omega} (P_{M_{rT}} \otimes P_{N_{rT}})$. Then by the same logic as before $\widehat{X} - X = O_p(T^{-2\gamma})$, while $\widehat{Y} - Y = O_p(T^{-\gamma})$ under either $H_0(r)$ or $H_T(r)$. In order to prove that $Y^\dagger = O_p(1)$, it suffices to show that its rank is almost surely constant and the smallest non–zero singular value is $O_p^{-1}(1)$. Following the same steps as used in the proof of Theorem 1, we can show that Assumptions C imply that $\text{rank}(Y) = (n-r)(m-r)$ almost surely. However, we cannot use the same bounds on $Y^\dagger$ as Assumptions C do not ensure that $\sigma_{nm}(\widehat{\Omega})$ is $O_p^{-1}(1)$. Instead, we use Theorem 3.3.16 (d) of Horn & Johnson (1991) to write

$$\begin{aligned}
\sigma_{(n-r)(m-r)}((P_{M_{rT}} \otimes P_{N_{rT}}) \widehat{\Omega} (P_{M_{rT}} \otimes P_{N_{rT}})) &\geq \frac{\sigma_{(n-r)(m-r)}((M_{rT} \otimes N_{rT}) \widehat{\Omega} (M_{rT} \otimes N_{rT}))}{\sigma_1((M_{rT}' M_{rT})^{1/2} \otimes (N_{rT}' N_{rT})^{1/2})} \\
&= \frac{\sigma_{(n-r)(m-r)}((M_{rT} \otimes N_{rT}) \widehat{\Omega} (M_{rT} \otimes N_{rT}))}{\sigma_1(M_{rT}) \sigma_1(N_{rT})} \\
&= O_p^{-1}(1).
\end{aligned}$$

On the other hand, Assumptions D imply that $\text{rank}(Y) = (m-r)(m-r+1)/2$ almost surely and, using similar arguments to those used in the proof of Theorem 1,

$$\begin{aligned}
&\sigma_{(m-r)(m-r+1)/2}((P_{M_{rT}} \otimes P_{M_{rT}}) \widehat{\Omega} (P_{M_{rT}} \otimes P_{M_{rT}})) \\
&\quad \geq \frac{\sigma_{(m-r)(m-r+1)/2}((M_{rT} \otimes M_{rT}) \widehat{\Omega} (M_{rT} \otimes M_{rT}))}{\sigma_1((M_{rT}' M_{rT})^{1/2} \otimes (M_{rT}' M_{rT})^{1/2})} \\
&\quad = \frac{\sigma_{(m-r)(m-r+1)/2}(D_{m-r} D_{m-r}^\dagger (M_{rT} \otimes M_{rT}) D_m \widehat{\Psi} D_m' (M_{rT} \otimes M_{rT})) D_{m-r}^\dagger{}' D_{m-r}')}{\sigma_1^2(M_{rT})} \\
&\quad \geq \sigma_1^{-2}(M_{rT}) \sigma_{(m-r)(m-r+1)/2}(D_{m-r}^\dagger (M_{rT} \otimes M_{rT}) D_m \widehat{\Psi} D_m' (M_{rT} \otimes M_{rT})) D_{m-r}^\dagger{}')) \\
&\quad = O_p^{-1}(1).
\end{aligned}$$

Since $X = O_p(T^{-\gamma})$ under either $H_0(r)$ or $H_T(r)$, condition (i) of Theorem 1 implies that $L_1(\widehat{X}, X, \widehat{Y}, Y) = O_p(T^{-\gamma(2\theta-1)})$ and $L_2(\widehat{X}, X, \widehat{Y}, Y) = O_p(T^{-2\gamma\theta})$. Therefore, under $H_0(r)$ or $H_T(r)$, $T^{2\gamma\theta} \kappa(\widehat{X}, \widehat{Y}) - T^{2\gamma\theta} \kappa(X, Y) = O_p(T^{2\gamma\theta - 2\gamma\theta + \gamma - 2\gamma}) + O_p(T^{2\gamma\theta - 2\gamma\theta - \gamma}) = O_p(T^{-\gamma})$. The weak plug–in principle under $H_0(r)$ and $H_T(r)$ is therefore established. Under $H_1(r)$, the

exact same arguments as in the proof of Theorem 1 are applied to prove the weak and strong plug–in principles. □

Proof of the Asymptotics of Example 14. We will show that Assumptions D are satisfied and derive an expression for ξ_r in Corollary 4. We can simplify the analysis by writing $u_t = \eta_t + \psi_t/T$, where $(\varepsilon_t', \eta_t', \psi_t')' \sim N\left(\left[\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}\Sigma & 0 & 0\\0 & B & 0\\0 & 0 & D\end{smallmatrix}\right]\right)$ are i.i.d. Let $Q_1 = [\ M_{r\perp} \quad M_r\ ]$ be an orthogonal matrix.

The fact that $\mathrm{vec}(\widehat{B})$ is nondegenerate follows from the fact that it is a composite of functions continuous–almost–everywhere of continuous random variables. Now the usual asymptotic arguments imply that for $u \in [0,1]$,

$$\begin{bmatrix} T^{-3/2}\sum_{t=1}^{\lfloor uT\rfloor} M_{r\perp}'(y_t - \overline{y}) \\ T^{-1/2}\sum_{t=1}^{\lfloor uT\rfloor} M_r'(y_t - \overline{y}) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} (M_{r\perp}'\Gamma M_{r\perp})^{1/2}\int_0^u W_1^*(s)ds \\ (M_r'\Sigma M_r)^{1/2}(W_2(u) - uW_2(1)) + (M_r'DM_r)^{1/2}\int_0^u W_3^*(s)ds \end{bmatrix},$$

where W_1 is generated by η, W_2 is generated by ε, and W_3 is generated by ψ (Phillips & Durlauf, 1986; Phillips & Solo, 1992; Billingsley, 1999). Letting $Q_T = [\ M_{r\perp} \quad TM_r\ ]$ and $H = \left[\begin{smallmatrix}(M_{r\perp}'\Gamma M_{r\perp})^{1/2} & 0\\ 0 & (M_r'\Sigma M_r)^{1/2}\end{smallmatrix}\right]$ we have,

$$\begin{bmatrix} H^{-1}Q_T'\widehat{\Gamma}Q_TH^{-1} \\ H^{-1}Q_1'\widehat{\Sigma}Q_TH^{-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} C_{11} & C_{12} \\ C_{12}' & C_{22} \\ \int_0^1 W_1^*(u)W_1^{*\prime}(u)du & \int_0^1 W_1^*(u)dW_2'(u) \\ 0 & I_{m-r} \end{bmatrix}.$$

Thus, $Q_{\sqrt{T}}'\widehat{\Sigma}Q_{\sqrt{T}}$ converges in distribution to a block–diagonal positive definite matrix.

Now consider $Q_1'\widehat{\Sigma}^{1/2}Q_1Q_1'\widehat{\Sigma}^{1/2}Q_1 = Q_1'\widehat{\Sigma}Q_1 = \left[\begin{smallmatrix}O_p(1) & O_p(T^{-1})\\ O_p(T^{-1}) & O_p(T^{-1})\end{smallmatrix}\right]$, where the blocks are conformable with Q_1. The $(2,2)$ block implies that $M_{r\perp}'\widehat{\Sigma}^{1/2}M_rM_r'\widehat{\Sigma}^{1/2}M_{r\perp} + (M_r'\widehat{\Sigma}^{1/2}M_r)^2 = O_p(T^{-1})$. Thus $M_{r\perp}'\widehat{\Sigma}^{1/2}M_r = O_p(T^{-1/2})$ and $M_r'\widehat{\Sigma}^{1/2}M_r = O_p(T^{-1/2})$. Using the first of these in the $(1,1)$ block we have that $(M_{r\perp}'\widehat{\Sigma}^{1/2}M_{r\perp})^2 - M_{r\perp}'\widehat{\Sigma}M_{r\perp} = O_p(T^{-1})$, which implies that $M_{r\perp}'\widehat{\Sigma}^{1/2}M_{r\perp} - (M_{r\perp}'\widehat{\Sigma}M_{r\perp})^{1/2} = O_p(T^{-1/2})$ (Horn & Johnson, 1985, Exercise 7.2.18). Using this in the $(1,2)$ block, we obtain that, in fact, $M_{r\perp}'\widehat{\Sigma}^{1/2}M_r = O_p(T^{-1})$ and not just $O_p(T^{-1/2})$ as found earlier. If we now go back to the $(2,2)$ block we obtain that $T(M_r'\widehat{\Sigma}^{1/2}M_r)^2 - TM_r'\widehat{\Sigma}M_r = O_p(T^{-1})$. Therefore $\sqrt{T}M_r'\widehat{\Sigma}^{1/2}M_r - \sqrt{T}(M_r'\widehat{\Sigma}M_r)^{1/2} = O_p(T^{-1/2})$, a fact that we will need later, and $\widehat{\Sigma}^{1/2}Q_{\sqrt{T}}$ converges in distribution to an almost surely invertible matrix.

Putting it all together, $Q'_{\sqrt{T}}\widehat{B}Q_{\sqrt{T}} = Q'_{\sqrt{T}}\widehat{\Sigma}^{1/2}(Q'_T\widehat{\Sigma})^{-1}Q'_T\widehat{\Gamma}Q_T(\widehat{\Sigma}Q_T)^{-1}\widehat{\Sigma}^{1/2}Q_{\sqrt{T}}$ converges in distribution to an almost surely positive definite matrix and since the submatrix $M'_{r\perp}\widehat{B}M_{r\perp}$ converges in distribution to an almost surely positive definite matrix, $\sigma_r(\widehat{B}) = O_p^{-1}(1)$ (Horn & Johnson, 1991, Corollary 3.1.3).

Just as before, let $M_{rT} = \sqrt{T}\widehat{\Sigma}^{1/2}(I_m - M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma})M_r$ and compute

$$\begin{aligned}
\widehat{B}M_{rT} &= (Q'_{\sqrt{T}}\widehat{\Sigma}^{1/2})^{-1}\sqrt{T}Q'_{\sqrt{T}}\widehat{\Gamma}(I_m - M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma})M_r \\
&= (Q'_{\sqrt{T}}\widehat{\Sigma}^{1/2})^{-1}\begin{bmatrix} \sqrt{T}M'_{r\perp}\widehat{\Gamma}(I_m - M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma})M_r \\ TM'_r\widehat{\Gamma}(I_m - M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma})M_r \end{bmatrix} \\
&= O_p(1)\begin{bmatrix} 0 \\ O_p(T^{-1}) \end{bmatrix}.
\end{aligned}$$

Thus, $T^\gamma\widehat{B}M_{rT} = O_p(1)$ with $\gamma = 1$. On the other hand, using the asymptotics of $\widehat{\Sigma}^{1/2}$ that we obtained earlier, we have

$$\begin{aligned}
M_{rT} &= \sqrt{T}\widehat{\Sigma}^{1/2}(I_m - M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma})M_r \\
&= \sqrt{T}\widehat{\Sigma}^{1/2}M_r - \sqrt{T}\widehat{\Sigma}^{1/2}M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma}M_r \\
&= \sqrt{T}\widehat{\Sigma}^{1/2}M_r + O_p(T^{-1/2}) \\
&= \sqrt{T}P_{M_r}\widehat{\Sigma}^{1/2}M_r + \sqrt{T}P_{M_{r\perp}}\widehat{\Sigma}^{1/2}M_r + O_p(T^{-1/2}) \\
&= \sqrt{T}P_{M_r}\widehat{\Sigma}^{1/2}M_r + O_p(T^{-1/2}) \\
&= M_r(M'_r\Sigma M_r)^{1/2} + o_p(1).
\end{aligned}$$

It follows that $[\ M_{r\perp} \quad M_{rT}\ ]$ has singular values bounded away from zero in probability and $M_{rT}(M'_r\widehat{\Sigma}M_r)^{-1/2}$ as well as $M_{rT}(M'_{rT}M_{rT})^{-1/2}$ converge to M_r in probability.

Finally, since there is no normalization, $\widehat{\Psi} = \Psi = \frac{1}{m-r}(D'_mD_m)^{-1}$, which satisfies the conditions in Assumptions D since Z_T and its inverse are bounded in probability. On the other hand, $TM'_{rT}\widehat{B}M_{rT} = T^2M'_r\widehat{\Gamma}M_r - T^2M'_r\widehat{\Gamma}M_{r\perp}(M'_{r\perp}\widehat{\Gamma}M_{r\perp})^{-1}M'_{r\perp}\widehat{\Gamma}M_r \xrightarrow{d} (M'_r\Sigma M_r)^{1/2}(C_{22} - C_{21}C_{11}^{-1}C_{12})(M'_r\Sigma M_r)^{1/2}$. Thus Assumptions D are satisfied. By Theorem 2, the limiting distribution of the Nyblom & Harvey (2000) statistic under $H_T(r)$ is the limiting distribution of $T\mathrm{tr}(P_{M_{rT}}\widehat{B}) = T\mathrm{tr}((M'_{rT}M_{rT})^{-1/2}M'_{rT}\widehat{B}M_{rT}(M'_{rT}M_{rT})^{-1/2})$, which is as stated. □

References

Al-Sadoon, M. M. (2010). *Causality Along Subspaces*. PhD thesis, University of Cambridge.

Al-Sadoon, M. M. (2014). Geometric and long run aspects of granger causality. *Journal of Econometrics*, *178, Part 3*(0), 558 – 568.

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

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

Arellano, M., Hansen, L. P., & Sentana, E. (2012). Underidentification? *Journal of Econometrics*, *170*(2), 256–280.

Avarucci, M. & Velasco, C. (2009). A wald test for the cointegration rank in nonstationary fractional systems. *Journal of Econometrics*, *151*(2), 178–189.

Aznar, A. & Salvador, M. (2002). Selecting the rank of the cointegration space and the form of the intercept using an information criterion. *Econometric Theory*, *18*(4), pp. 926–947.

Bartlett, M. S. (1947). Multivariate analysis. *Supplement to the Journal of the Royal Statistical Society*, *9*(2), 176–197.

Bernstein, D. S. (2009). *Matrix Mathematics: Theory, Facts, and Formulas* (2 ed.). Princeton, NJ: Princeton University Press.

Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics*, *77*(2), 379–404.

Billingsley, P. (1999). *Convergence of Probability Measures* (2 ed.). Wiley Series in Probability and Statistics. New York, USA: John Wiley & Sons.

Boswijk, H. & Lucas, A. (2002). Semi-nonparametric cointegration testing. *Journal of Econometrics*, *108*(2), 253 – 280.

Boswijk, H. P., Jansson, M., & rregaard Nielsen, M. (2015). Improved likelihood ratio tests for cointegration rank in the var model. *Journal of Econometrics*, *184*(1), 97 – 110.

Camba-Mendez, G. & Kapetanios, G. (2005). Estimating the rank of the spectral density matrix. *Journal of Time Series Analysis*, *26*(1), 37–48.

Camba-Mendez, G. & Kapetanios, G. (2009). Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling. *Econometric Reviews*, *28*(6), 581–611.

Camba-Mendez, G., Kapetanios, G., Smith, R. J., & Weale, M. R. (2003). Tests of rank in reduced rank regression models. *Journal of Business & Economic Statistics*, *21*(1), 145–155.

Caner, M. (1998). Tests for cointegration with infinite variance errors. *Journal of Econometrics*, *86*(1), 155 – 175.

Cappuccio, N. & Lubian, D. (2009). The fragility of the kpss stationarity test. Working Paper Series 67, Department of Economics, University of Verona.

Cavaliere, G., De Angelis, L., Rahbek, A., & Taylor, A. R. (2014). Determining the co-integration rank in heteroskedastic var models of unknown order. Mimeo.

Cavaliere, G., Rahbek, A., & Taylor, A. R. (2010a). Cointegration rank testing under conditional heteroskedasticity. *Econometric Theory*, *26*(06), 1719–1760.

Cavaliere, G., Rahbek, A., & Taylor, A. R. (2010b). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics*, *158*(1), 7–24.

Chandrasekaran, S. & Ipsen, I. C. F. (1994). On rank-revealing factorisations. *SIAM Journal on Matrix Analysis and Applications*, *15*(2), 592–622.

Cragg, J. G. & Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, *9*(02), 222–240.

Cragg, J. G. & 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*(435), 1301–1309.

Cragg, J. G. & Donald, S. G. (1997). Inferring the rank of a matrix. *Journal of Econometrics*, *76*(1-2), 223–250.

Cushing, M. J. & McGravey, M. G. (1999). Covariance matrix estimation. In L. Maátyás (Ed.), *Generalized Method of Moments Estimation*, volume 1 (pp. 63–95). Cambridge University Press.

den Haan, W. J. & Levin, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In G. S. Maddala & C. R. Rao (Eds.), *Handbook of Statistics*, volume 15 (pp. 299–342). Elsevier Science B. V.

Dolado, J. J., Gonzalo, J., & Marmol, F. (2001). Cointegration. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, Blackwell Companions to Contemporary Economics chapter 30, (pp. 634–654). Blackwell Publishing.

Donald, S. G., Fortuna, N., & Pipiras, V. (2007). On rank estimation in symmetric matrices: The case of indefinite matrix estimators. *Econometric Theory*, *23*(06), 1217–1232.

Dufour, J.-M. & Valery, P. (2011). Hypothesis tests when rank conditions fail: a smooth regularization approach. mimeo, HEC Montréal. Available at http://neumann.hec.ca/pages/pascale.valery/pascalevalery.htm.

Duplinskiy, A. (2014). Is regularization necessary? a wald-type test under non-regular conditions. Research Memorandum 025, Maastricht University, Graduate School of Business and Economics (GSBE).

Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Probability and Mathematical Statistics. A Series of Monographs and Textbooks. New York, USA.: Academic Press, Inc.

Gabriel, K. R. & Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, *21*, 489–498.

Gill, L. & 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*(419), 766–776.

Gohberg, I., Lancaster, P., & Rodman, L. (2006). *Invariance Subspaces and Matrices with Applications.* Classics in Applied Mathematics. Philadelphia, USA: SIAM.

Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations* (3 ed.). Classics in Applied Mathematics. Baltimore, USA: Johns Hopkins University Press.

Gonzalo, J. & Pitarakis, J.-Y. (1999). Dimensionality effect in cointegration analysis. In R. Engle & H. White (Eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger* (pp. 212?229). Oxford University Press.

Guillemin, V. & Pollack, A. (1974). *Differential Topology.* New Jersey, USA: Prentice Hall, Inc.

Hallin, M., van den Akker, R., & Werker, B. (2012). Rank-Based Tests of the Cointegrating Rank in Semiparametric Error Correction Models. Working Papers ECARES ECARES 2012-042, ULB – Universite Libre de Bruxelles.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hansen, P. C. (1998). *Rank-Deficient and Discrete Ill-Posed Problems.* Monographs on Mathematical Modelling and Computation. Philadelphia, USA: Society for Industrial and Applied Mathematics.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning.* Springer Series in Statistics. New York, NY, USA: Springer New York Inc.

Higham, N. J. (1987). A survey of condition number estimation for triangular matrices. *SIAM Review*, *29*, 575–596.

Higham, N. J. (1990). Analysis of the cholesky decomposition of a semidefnite matrix. In M. G. Cox & S. J. Hammarling (Eds.), *Reliable Numerical Computation* (pp. 161–185). Oxford University Press.

Horn, R. A. & Johnson, C. R. (1985). *Matrix Analysis.* Cambridge, United Kingdom: Cambridge University Press.

Horn, R. A. & Johnson, C. R. (1991). *Topics in Matrix Analysis.* Cambridge, United Kingdom: Cambridge University Press.

Hubrich, K., Lütkepohl, H., & Saikkonen, P. (2001). A Review Of Systems Cointegration Tests. *Econometric Reviews*, *20*(3), 247–318.

Izenman, A. J. (1975). Reduced-rank regression for multivariate linear models. *Journal of Multivariate Analysis*, *5*, 248–264.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2?3), 231 – 254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, *59*(6), 1551–1580.

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

Johansen, S. (1995b). A stastistical analysis of cointegration for i(2) variables. *Econometric Theory*, *11*(01), 25–59.

Kiefer, N. M. & Vogelsang, T. J. (2002a). Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica*, *70*(5), 2093–2095.

Kiefer, N. M. & Vogelsang, T. J. (2002b). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, *18*(06), 1350–1366.

Kiefer, N. M. & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, *21*(06), 1130–1164.

Kiefer, N. M., Vogelsang, T. J., & Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, *68*(3), 695–714.

Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica*, *73*(4), 1103–1123.

Kleibergen, F. & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126.

Kleibergen, F. & van Dijk, H. K. (1994). Direct cointegration testing in error correction models. *Journal of Econometrics*, *63*(1), 61–103.

Lehmann, E. L. & Romano, J. P. (2005). *Testing Statistical Hypotheses* (3 ed.). Springer Texts in Statistics. New York, USA: Springer.

Lewbel, A. (1991). The rank of demand systems: Theory and nonparametric estimation. *Econometrica*, *59*(3), 711–30.

Lucas, A. (1996). *Outlier Robust Unit Root Analysis.* PhD thesis, Tinbergen Institute, Amsterdam, Holland.

Lucas, A. (1997). Cointegration Testing Using Pseudolikelihood Ratio Tests. *Econometric Theory*, *13*(02), 149–169.

Lucas, A. (1998). Inference on cointegrating ranks using lr and lm tests based on pseudolikelihoods. *Econometric Reviews*, *17*(2), 185–214.

Lütkepohl, H. & Burda, M. M. (1997). Modified wald tests under nonregular conditions. *Journal of Econometrics*, *78*(2), 315–332.

Lütkepohl, H. & Saikkonen, P. (1999). A lag augmentation test for the cointegrating rank of a var process. *Economics Letters*, *63*(1), 23 – 27.

Magnus, J. R. & Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley Series in Probability and Statistics. New York, USA: John Wiley & Sons.

Markovsky, I. (2012). *Low Rank Approximation: Algorithms, Implementation, Applications.* Communications and Control Engineering. New York, USA: Springer.

Markovsky, I. & Van Huffel, S. (2007). Left vs right representations for solving weighted low-rank approximation problems. *Linear Algebra and its Applications*, *422*(23), 540 – 552.

Massmann, M. (2007). Reduced rank regressions and the poincaré separation theorem. Mimeo, Institute of Econometrics. University of Bonn.

Moore, D. S. (1977). Generalized inverses, wald's method, and the construction of chi-squared tests of fit. *Journal of the American Statistical Association*, *72*(357), 131–137.

Nyblom, J. & Harvey, A. (2000). Tests of common stochastic trends. *Econometric Theory*, *16*(02), 176–199.

Phillips, P. C. B. & Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, *53*(4), 473–495.

Phillips, P. C. B. & Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, *20*(2), 971–1001.

Quintos, C. E. (1998). Fully modified vector autoregressive inference in partially nonstationary models. *Journal of the American Statistical Association*, *93*(442), pp. 783–795.

Ratsimalahelo, Z. (2003). Rank test based on matrix perturbation theory. Econometrics 0306008, EconWPA.

Reinsel, G. C. & Velu, R. P. (1998). *Multivariate Reduced-Rank Regression.* Lecture Notes in Statistics. New York, USA: Springer.

Robin, J. & Smith, R. (1995). Tests of rank. Cambridge Working Papers in Economics 9521, Faculty of Economics, University of Cambridge.

Robin, J.-M. & Smith, R. J. (2000). Tests of rank. *Econometric Theory*, *16*(02), 151–175.

Rudin, W. (1986). *Real and Complex Analysis* (3 ed.). New York, USA.: McGraw–Hill Book Company.

Saikkonen, P. & Lütkepohl, H. (1999). Local power of likelihood ratio tests for the cointegrating rank of a var process. *Econometric Theory*, *15*(1), pp. 50–78.

Stewart, G. W. (1993). On the early history of the singular value decomposition. *SIAM Review*, *35*(4), 551–566.

Stewart, G. W. & Sun, J. (1990). *Matrix Perturbation Theory.* New York, USA: Academic Press, Inc.

Stock, J. H. & Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, *83*(404), pp. 1097–1107.

Van Loan, C. & Pitsianis, N. (1993). Approximation with kronecker products. In M. Moonen, G. Golub, & B. De Moor (Eds.), *Linear Algebra for Large Scale and Real-Time Applications*, volume 232 of *NATO ASI Series* (pp. 293–314). Springer Netherlands.

Van Loan, C. F. (1976). Generalizing the singular value decomposition. *SIAM Journal on Numerical Analysis*, *13*(1), 76–83.

Velu, R. P., Reinsel, G. C., & Wichern, D. W. (1986). Reduced rank models for multiple time series. *Biometrika*, *73*(1), 101–118.

Vogelsang, T. J. (2001). Testing in gmm models without truncation. Working Papers 01-12, Cornell University, Center for Analytic Economics.

White, H. (1994). *Estimation, Inference, and Specification Analysis*. Econometric Society Monographs. Cambridge, UK: Cambridge University Press.

Wilkinson, J. H. (1961). Error analysis of direct methods of matrix inversion. *Journal of the Association for Computing Machinery*, *8*(3), 281–330.

Wright, J. H. (2003). Detecting lack of identification in gmm. *Econometric Theory*, *19*(02), 322–330.

Yang, M. & Bewley, R. (1996). On cointegration tests for var models with drift. *Economics Letters*, *51*(1), 45 – 50.