# A General Theory of Rank Testing*

Majid M. Al-Sadoon

Universitat Pompeu Fabra & Barcelona GSE

February 7, 2014

### Abstract

This paper develops an approach to rank testing that nests all existing rank tests and simplifies their asymptotics. The approach is based on the fact that implicit in every rank test there are estimators of the null spaces of the matrix in question. The approach yields many new insights about the behavior of rank testing statistics under the null as well as local and global alternatives in both the standard and the cointegration setting. The approach also suggests many new rank tests based on alternative estimates of the null spaces as well as the new fixed–$b$ theory. A brief Monte Carlo study illustrates the results.

JEL Classification: C12, C13, C30.

Keywords: Rank testing, stochastic tests, classical tests, subspace estimation, cointegration.

# 1  Introduction

Rank testing is ubiquitous in empirical data analysis. Examples of applications in econometrics include, the empirical analysis of 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), model reduction (Velu et al., 1986; Camba-Mendez et al., 2003), and Granger causality testing (Al-Sadoon, 2014) (Camba-Mendez & Kapetanios (2009) provide a comprehensive review). Rank tests are also prevalent in statistics (Reinsel & Velu, 1998; Anderson, 2003) as well as systems theory and machine learning (Markovsky, 2012). 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 eigenvalues, eigenvectors, and other products of matrix decompositions, which are quite difficult to handle. There is therefore a great need for an encompassing and simple theory of rank testing.

This paper proposes a class of test statistics called stochastic statistics, which are distinguished from classical statistics in that the restriction matrices are estimated rather than being known a priori. The paper studies conditions under which stochastic and classical statistics are asymptotically equivalent. This is termed the plug–in principle. The paper then demonstrates that all rank test statistics satisfy the plug–in principle, as implicit in every known rank test statistics there are null space estimators, which act as an estimated restriction matrix in a stochastic statistic. In fact, the majority of rank test statistics have very simple structures (e.g. stochastic $F$ statistics or stochastic $t$ statistics – see Table 1). The paper develops a theory of null space estimation based on reduced rank approximations (RRA), which have received a great deal of attention in the numerical analysis literature Hansen (1998). The approach is more general than that proposed by Massmann (2007), which nests some of the likelihood–based tests but not many of the Wald–type tests or the tests for symmetric matrices.

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 null spaces were known. Under the global alternative, we find conditions (conjectured to be generically satisfied) that ensure the stochastic and classical statistics diverge at the same rate. When those conditions are not satisfied, we can still prove consistency by appealing to the properties of null space estimators. We illustrate the simplicity of the approach by proposing new tests based on the QR and Cholesky decompositions as well as the new fixed–$b$ theory proposed by Vogelsang (2001), Kiefer & Vogelsang (2002a), Kiefer & Vogelsang (2002b), and Kiefer & Vogelsang (2005).

Perhaps the greatest advantage of the theory is to the study of cointegration rank testing. 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

Table 1: Summary of the Basic Rank Testing Statistics in the Literature.

| Paper | Statistic | Form | RRA† |
|---|---|---|---|
| Bartlett (1947) | Likelihood Ratio | Asymptotic $F$ | RSD |
| Anderson (1951) | Likelihood Ratio | Asymptotic $F$ | RSD |
| | Trace Statistic | Exact $F$ | RSD |
| Johansen (1991) | Maximum Eigenvalue | $L^2$ Norm | RSD |
| Cragg & Donald (1993) | Wald statistic | Exact $F$ | LU |
| Kleibergen & van Dijk (1994) | Wald statistic | Exact $F$ | LU |
| Cragg & Donald (1997) | Minimum Distance | Exact $F$ | CDA |
| Robin & Smith (1995) | Wald statistic | Exact $F$ | LU |
| Robin & Smith (2000) | Asymptotically Wald | Asymptotic $F$ | RSD |
| Nyblom & Harvey (2000) | Lagrange Multiplier | Exact $t$ | EIG |
| Kleibergen & Paap (2006) | Wald statistic | Exact $F$ | SVD |
| Donald et al. (2007) | Wald statistic | Exact $F$ | LU |
| | Wald statistic | Exact $F$ | SVD |
| | Minimum Distance | Exact $F$ | CDA |
| | Wald statistic | Exact $t$ | EIG |

† CDA: Cragg–Donald approximation. RSD: Robin–Smith decomposition. SVD: Singular value decomposition. LU: LU decomposition. EIG: Eigenvalue decomposition.

regularized rank testing. We illustrate the tractability of the new approach to cointegration rank testing by proposing a new test for common trends as well as deriving the local power of the Nyblom & Harvey (2000) test.

The paper is organized as follows. Section 2 develops the notation of the paper. Section 3 discusses the basic idea of a stochastic test. Section 4 discusses rank testing under standard asymptotics. Section 5 discusses cointegration rank testing. Section 6 discusses alternative statistics to $F$ and $t$. Section 7 provides Monte Carlo evidence. Section 8 concludes and section 9 is an appendix.

# 2 Notation

Let $\mathbb{R}^{n \times m}$ be the set of real $n \times m$ matrices and let $\mathbb{G}^{n \times m}$ be the subset of matrices of full rank. The $ij$–th element of $B$ is denoted by $B_{(i,j)}$. For $i \leq j$ and $k \leq l$, the matrix $B_{(i:j,k:l)}$ will denote the submatrix of $B$ consisting of the rows $i$ to $j$ and columns $k$ to $l$. The diag operator is defined by $\text{diag}(B)_{(i,j)} = B_{(i,j)}\delta_{ij}$, where $\delta_{ij}$ is the Kronecker delta function. For a given Euclidean space, we define $e_i$ to be the $i$–th standard basis vector. We define

$$\text{vec}(B) = \begin{bmatrix} B_{(1:n,1)} \\ B_{(1:n,2)} \\ \vdots \\ B_{(1:n,m)} \end{bmatrix} \qquad \text{vech}(B) = \begin{bmatrix} B_{(1:n,1)} \\ B_{(2:n,2)} \\ \vdots \\ B_{(n,m)} \end{bmatrix}.$$

The duplication matrix $D_n$ is the mapping $\text{vech}(B) \mapsto \text{vec}(B)$ over symmetric matrices $B \in \mathbb{R}^{n \times n}$. 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 $n > m$, the orthogonal complement $B_\perp$ is defined as any matrix in $\mathbb{G}^{n \times n-m}$ satisfying $B'_\perp B = 0$ (the particular choice of $B_\perp$ will not matter for our purposes). The orthogonal projection onto the column space of $B$ is defined as $P_B$. The Euclidean or Frobenius 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 symmetric, positive semi–definite, $\Theta \in \mathbb{R}^{nm \times nm}$. 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 $n = m$ and $B$ is symmetric, we denote the eigenvalues of $B$ as $\lambda_1(B) \geq \lambda_2(B) \geq \cdots \geq \lambda_n(B)$.[1]

Finally, we say that a random sequence of 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 $\delta > 0$ such that $P(\|X_T\| > \delta) > 1 - \varepsilon$ for all $T$. It is easy to show that $\|X_T\|^{-1}$ is bounded away from zero 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 $a_T \to \infty$.

# 3 Stochastic Statistics Under Standard Asymptotics

## 3.1 Basic Stochastic Statistics

We begin with the following general assumption.

(A1) $\beta \in \mathbb{R}^p$. $\Omega \in \mathbb{R}^{p \times p}$ is random, symmetric, and almost surely positive definite. $\widehat{\beta}$ and $\widehat{\Omega}$ are estimators indexed by $T$ such that $(\sqrt{T}(\widehat{\beta} - \beta), \widehat{\Omega}) \xrightarrow{d} (\xi, \Omega)$ as $T \to \infty$, where each $\widehat{\beta} \in \mathbb{R}^p$ is a non–degenerate random vector and each $\widehat{\Omega} \in \mathbb{R}^{p \times p}$ is symmetric and almost surely positive definite.[2]

**Example 1** (Small–$b$ and Fixed–$b$ Asymptotics). Assumption (A1) includes the small–$b$ assumptions of constant $\Omega$ and $\xi \sim N(0, \Phi)$ as a special case. Assumption (A1) also allows for fixed–$b$ asymptotics where $\widehat{\Omega}$ is inconsistent and $\Omega$ is a functional of the Brownian motion that generates $\xi$ (Kiefer et al., 2000; Vogelsang, 2001; Kiefer & Vogelsang, 2002b,a, 2005). □

We would like to test the hypothesis

$$H_0 : R'\beta = 0$$

---

[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 extensively on Dennis Bernstein's magnificent treatise (Bernstein, 2009) although only primary sources are cited in this paper.

[2] A random vector is non–degenerate if it has a probability density function.

for some $R \in \mathbb{G}^{p \times q}$ against either the local alternative

$$H_T : R'\beta^T = \epsilon/\sqrt{T},$$

to be interpreted as a sequence of alternatives where $\beta^T = \beta + T^{-1/2}\delta$,[3] with $\beta$ and $\delta$ fixed, $R'\beta = 0$, and $\epsilon = R'\delta$, as well as the global alternative

$$H_1 : R'\beta \neq 0.$$

Assumption (A1) allows us to use the classical $F$ statistic given by

$$F = T\widehat{\beta}'R\{R'\widehat{\Omega}R\}^{-1}R'\widehat{\beta}. \tag{1}$$

If $q = 1$, we may also form the classical $t$ statistic

$$t = \frac{\sqrt{T}R'\widehat{\beta}}{\sqrt{R'\widehat{\Omega}R}}. \tag{2}$$

We will be interested in testing hypotheses where $R$ is unknown. In this case, it is impossible to formulate the classical statistics (1) and (2). However, if $R$ can be estimated well enough by $\widehat{R}$ then we may try to plug it in for $R$ in order to conduct the above test. Define

$$\widehat{F} = T\widehat{\beta}'\widehat{R}\{\widehat{R}'\widehat{\Omega}\widehat{R}\}^{-1}\widehat{R}'\widehat{\beta} \tag{3}$$

and, for $q = 1$,

$$\widehat{t} = \frac{\sqrt{T}\widehat{R}'\widehat{\beta}}{\sqrt{\widehat{R}'\widehat{\Omega}\widehat{R}}} \tag{4}$$

We will refer to (3) as the stochastic $F$ statistic and to (4) as the stochastic $t$ statistic. For the next result, we will need the following assumptions:

(A2) $\widehat{R} \in \mathbb{G}^{p \times q}$ converges in probability to $R$.

(A3) $\sqrt{T}\widehat{R}'\beta = o_p(1)$.

(A4) $\widehat{R} \in \mathbb{G}^{p \times q}$ is $O_p(1)$ and $\widehat{R}'\widehat{\beta} = O_p^{-1}(1)$.

**Lemma 1** (The Plug–in Principle I). Assume that (A1) holds.

(i) Under either $H_0$ or $H_T$, if (A2) and (A3) hold, then $\widehat{F} - F = o_p(1)$ and, for $q = 1$, $\widehat{t} - t = o_p(1)$.

(ii) Under $H_1$, if (A2) holds, then $T^{-1}(\widehat{F} - F) = o_p(1)$ and, for $q = 1$, $T^{-1/2}(\widehat{t} - t) = o_p(1)$.

(iii) Under $H_1$, if (A4) holds, then $T^{-1}\widehat{F} = O_p^{-1}(1)$ and so $\widehat{F} \overset{p}{\to} \infty$. For $q = 1$, $T^{-1/2}|\widehat{t}| = O_p^{-1}(1)$ and so $|\widehat{t}| \overset{p}{\to} \infty$.

---

[3]To be clear, under $H_T$, $(\sqrt{T}(\widehat{\beta} - \beta^T), \widehat{\Omega}) \overset{d}{\to} (\xi, \Omega)$. See chpater 8 of White (1994) and chapter 12 of Lehmann & Romano (2005) for more on the theory of local power.

Lemma 1 provides conditions under which stochastic $F$ and $t$ tests behave similarly to classical $F$ and $t$ tests. If $\widehat{R}$ estimates $R$ consistently and annihilates $\beta$ at an appropriate rate under $H_0$ and $H_T$, then the stochastic tests have the same asymptotic size and local power as the classical tests. Geometrically, if $\widehat{R}$ is consistent and the angles between $\mathrm{span}(\widehat{R})$ and $\mathrm{span}(\beta_\perp)$ diminish at the rate of $o_p(T^{-1/2})$, then the stochastic and classical statistics are asymptotically equivalent (Figure 1).[4] Now assumption (A3) may seem quite demanding at this stage. We will see in the next section that it is actually very easily achievable.

Figure 1: The Plug–in Principle Under $H_0$ and $H_T$.
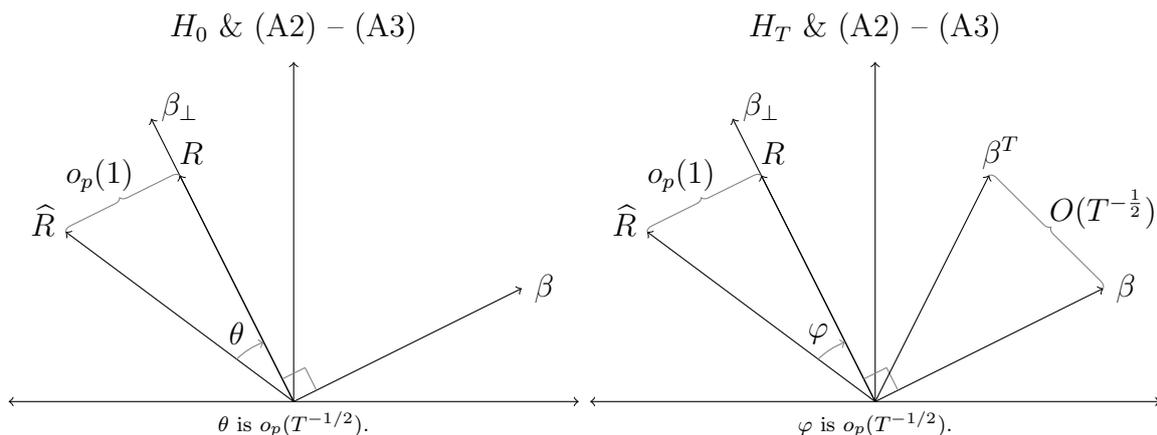


Under $H_1$, if $\widehat{R}$ estimates $R$ consistently, then the stochastic statistic diverges at the same rate as the classical statistic. Alternatively, if $\widehat{R}$ is not necessarily consistent but fails to annihilate $\widehat{\beta}$, then the stochastic statistics diverge to infinity in probability (Figure 2).

Figure 2: The Plug–in Principle Under $H_1$.
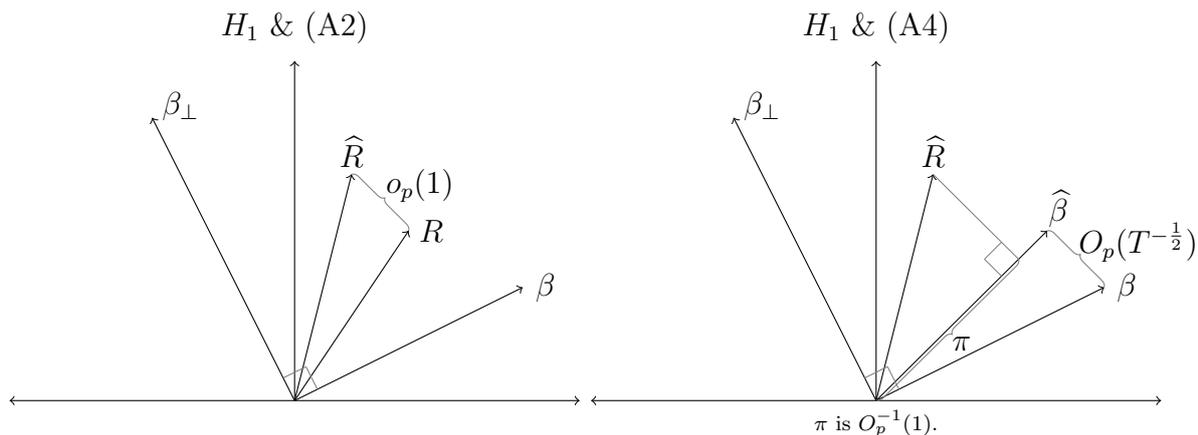


Lemma 1 is the back bone of all rank testing. In fact, the rest of the paper is – simply put – a series of restatements of the lemma under a variety of specializations and generalizations.

---

[4]See Golub & Van Loan (1996) and Gohberg et al. (2006) for more on the angles between subspaces.

It is possible to relax the assumptions of Lemma 1 in various ways. For example, (A1) can be relaxed to $(\sqrt{T}(\widehat{\beta}-\beta), \widehat{\Omega}) = O_p(1)$, provided $\lambda_p(\widehat{\Omega})$ is bounded away from zero in probability. (A1) may also be replaced by the assumption that $\widehat{\beta} = \widehat{\beta}_1 + \widehat{\beta}_2$, with $\widehat{\beta}_1 - \beta = o_p(1)$, $R'\widehat{\beta}_1 = 0$ and $(\sqrt{T}\widehat{\beta}_2, \widehat{\Omega}) \xrightarrow{d} (\xi, \Omega)$ so that only along span$(R)$ is the limiting distribution of $\widehat{\beta}$ known (Donald et al., 2007). This highlights the fact that the only part of the limiting distribution that is relevant for the asymptotics is the component along $R$. We may also relax the condition that $\Omega$ be almost surely invertible and use the generalized inverse à la Moore (1977), provided we make provisions for the conditions found in Andrews (1987). Regularized tests may also be constructed along the lines of Lutkepohl & Burda (1997) and Dufour & Valery (2011). Finally, we may consider complex matrices as in the case of spectral analysis (Camba-Mendez & Kapetanios, 2005). However, such extensions are straightforward and are omitted.

Note that multiplying $\widehat{R}$ on the right by any nonsingular matrix leaves $\widehat{F}$ and $|\widehat{t}|$ invariant. This invariance implies that the particular choice of columns of $\widehat{R}$ is not relevant, only the subspace spanned by these columns is relevant.[5] We therefore relax our assumptions as follows.

(A2)* $\widehat{R} \in \mathbb{G}^{p \times q}$ and $P_{\widehat{R}} - P_R = o_p(1)$.

(A3)* $\sqrt{T} P_{\widehat{R}} P_\beta = o_p(1)$.

(A4)* $\widehat{R} \in \mathbb{G}^{p \times q}$ and $P_{\widehat{R}} P_{\widehat{\beta}} = O_p^{-1}(1)$.

**Lemma 2** (The Plug–in Principle II). Assume that (A1) holds.

    (i) Under either $H_0$ or $H_T$, if (A2)* and (A3)* hold, then $\widehat{F} - F = o_p(1)$ and, for $q = 1$, $|\widehat{t}| - |t| = o_p(1)$.

    (ii) Under $H_1$, if (A2)* holds, then $T^{-1}(\widehat{F}-F) = o_p(1)$ and, for $q = 1$, $T^{-1/2}(|\widehat{t}|-|t|) = o_p(1)$.

    (iii) Under $H_1$, if (A4)* holds, then $T^{-1}\widehat{F} = O_p^{-1}(1)$ and so $\widehat{F} \xrightarrow{p} \infty$. For $q = 1$, $T^{-1/2}|\widehat{t}| = O_p^{-1}(1)$ and so $|\widehat{t}| \xrightarrow{p} \infty$.

Lemma 2 proves that $\widehat{R}$ does not need to be consistent for the plug–in principle to hold under $H_0$ or $H_T$. $\widehat{R}$ only needs to specify a subspace that merges with the subspace spanned by $R$. Additionally, we require span$(\widehat{R})$ to be super–consistent in coalescing with span$(\beta_\perp)$ for the plug–in principle to hold. Under $H_1$, the consistency of span$(\widehat{R})$ is sufficient, although the test continues to have power if the angles between span$(\widehat{R})$ and span$(\widehat{\beta})$ are bounded away from orthogonality.

The conditions of Lemma 2 obviate the need to impose identification restrictions on $\widehat{R}$, as is common in the rank testing literature. In fact, at no point are identification restrictions employed in this paper. Instead, we rely on the one–to–one correspondence between a subspace and the orthogonal projection onto that subspace. Convergence of a subspace estimator is then completely characterized by convergence of the associated projection matrix.[6]

---

[5]Chapter 6 of Lehmann & Romano (2005) provides an exhaustive study of invariance in hypothesis testing.

[6]Chapter 13 of Gohberg et al. (2006) provides a detailed discussion of the convergence of subspaces. Note that the theory of this paper can be recast in terms of convergence of points on Grassmannian manifolds.

## 3.2 Rank Tests as Stochastic Tests

Suppose that $B \in \mathbb{R}^{n \times m}$. For $0 \leq r < \min\{n, m\}$, we will be interested in testing the hypotheses

$$H_0(r) : \mathrm{rank}(B) = r$$

against the local alternative

$$H_T(r) : \mathrm{rank}(B^T - D/\sqrt{T}) = r,$$

to be interpreted as a sequence of alternatives where $B^T = B + T^{-1/2}D$ with $B$ and $D$ fixed and $\mathrm{rank}(B) = r$, as well as the global alternative

$$H_1(r) : \mathrm{rank}(B) > r.$$

It follows from basic linear algebra that:

$H_0(r)$ is equivalent to the existence of matrices $N_r \in \mathbb{G}^{n \times (n-r)}$ and $M_r \in \mathbb{G}^{m \times (m-r)}$, unique up to right multiplication by invertible matrices, such that $N_r'B = 0$ and $BM_r = 0$. Therefore, $\mathrm{vec}(N_r'BM_r) = (M_r \otimes N_r)'\mathrm{vec}(B) = 0$.

$H_T(r)$ is equivalent to the existence of matrices $N_r \in \mathbb{G}^{n \times (n-r)}$ and $M_r \in \mathbb{G}^{m \times (m-r)}$, unique up to right multiplication by invertible matrices, such that $N_r'B^T = N_r'D/\sqrt{T}$ and $B^T M_r = DM_r/\sqrt{T}$. Therefore, $(M_r \otimes N_r)'\mathrm{vec}(B^T) = (M_r \otimes N_r)'\mathrm{vec}(D)/\sqrt{T}$.

$H_1(r)$ is equivalent to the absence of any matrices $N_r \in \mathbb{G}^{n \times (n-r)}$ and $M_r \in \mathbb{G}^{m \times (m-r)}$ satisfying $N_r'B = 0$ or $BM_r = 0$. $N_r$ and $M_r$ may (and will) be chosen so that $N_r'BM_r \neq 0$. Therefore, $(M_r \otimes N_r)'\mathrm{vec}(B) \neq 0$.

Based on the observations above, $H_0(r)$, $H_T(r)$, and $H_1(r)$ are equivalent to $H_0$, $H_T$, and $H_1$ respectively, with $\beta = \mathrm{vec}(B)$, $\beta^T = \mathrm{vec}(B^T)$, and $R_r = M_r \otimes N_r$. If we have estimators $\widehat{\beta} = \mathrm{vec}(\widehat{B})$ for $\beta$, $\widehat{N}_r$ for $N_r$, and $\widehat{M}_r$ for $M_r$, then we may form the estimated restriction matrix $\widehat{R}_r = \widehat{M}_r \otimes \widehat{N}_r$ and the stochastic $F$ statistic

$$\widehat{F}(r) = T\mathrm{vec}'(\widehat{N}_r'\widehat{B}\widehat{M}_r)\{(\widehat{M}_r \otimes \widehat{N}_r)'\widehat{\Omega}(\widehat{M}_r \otimes \widehat{N}_r)\}^{-1}\mathrm{vec}(\widehat{N}_r'\widehat{B}\widehat{M}_r) \tag{5}$$

and compare it to the classical statistic

$$F(r) = T\mathrm{vec}'(N_r'\widehat{B}M_r)\{(M_r \otimes N_r)'\widehat{\Omega}(M_r \otimes N_r)\}^{-1}\mathrm{vec}(N_r'\widehat{B}M_r). \tag{6}$$

If $B$ and $\widehat{B}$ are restricted to be symmetric $m \times m$ matrices, then one simply works with $\beta = \mathrm{vech}(B)$ and $\widehat{\beta} = \mathrm{vech}(\widehat{B})$. We take $D$ under $H_T(r)$ to be symmetric. It follows from basic linear algebra again that:

$H_0(r)$ is equivalent to the existence of a matrix $M_r \in \mathbb{G}^{m \times (m-r)}$, unique up to right multiplication by an invertible matrix, such that $BM_r = 0$. Therefore, $\mathrm{vech}(M_r'BM_r) = D_{m-r}^{\dagger}(M_r \otimes M_r)'D_m\mathrm{vech}(B) = 0$.

$H_T(r)$ is equivalent to the existence of a matrix $M_r \in \mathbb{G}^{m \times (m-r)}$, unique up to right multiplication by an invertible matrix, such that $B^T M_r = DM_r/\sqrt{T}$. Therefore, $D_{m-r}^\dagger (M_r \otimes M_r)'D_m \text{vech}(B^T) = D_{m-r}^\dagger (M_r \otimes M_r)'D_m \text{vech}(D)/\sqrt{T}$.

$H_1(r)$ is equivalent to the absence of any matrix $M_r \in \mathbb{G}^{m \times (m-r)}$ satisfying $BM_r = 0$. $M_r$ may (and will) be chosen so that $M_r'BM_r \neq 0$. Therefore, $D_{m-r}^\dagger (M_r \otimes M_r)'D_m \text{vech}(B) \neq 0$.

If we now have an estimator $\widehat{M}_r$ for $M_r$, then we may form the estimated restriction matrix $\widehat{R}_r = D_m'(\widehat{M}_r \otimes \widehat{M}_r)D_{m-r}^{\dagger\,\prime}$ and the stochastic $F$ statistic

$$\widehat{F}^s(r) = T\text{vech}'(\widehat{M}_r'\widehat{B}\widehat{M}_r)\{D_{m-r}^\dagger(\widehat{M}_r \otimes \widehat{M}_r)'D_m\widehat{\Omega}D_m'(\widehat{M}_r \otimes \widehat{M}_r)D_{m-r}^{\dagger\,\prime}\}^{-1}\text{vech}(\widehat{M}_r'\widehat{B}\widehat{M}_r) \quad (7)$$

and compare it to the classical statistic

$$F^s(r) = T\text{vech}'(M_r'\widehat{B}M_r)\{D_{m-r}^\dagger(M_r \otimes M_r)'D_m\widehat{\Omega}D_m'(M_r \otimes M_r)D_{m-r}^{\dagger\,\prime}\}^{-1}\text{vech}(M_r'\widehat{B}M_r).$$
$$(8)$$

The reader may wonder why we have formulate the classical and stochastic $F$ statistics with both null spaces rather than using just one (e.g. $N_r'B = 0$ instead of $N_r'BM_r = 0$ or $BM_r = 0$ instead of $M_r'BM_r = 0$). It will be seen in the proof of Theorem 1 that using the two null spaces allows us to satisfy condition (A3) (resp. (A3)$^*$), as two null space estimators annihilate $B$ faster than just one null space estimator.

When testing $H_0(r)$ for $r < \min\{n, m\} - 1$, there are clearly too many degrees of freedom to be able to use a stochastic $t$ test. However, if the parameter space is restricted enough, stochastic $t$ testing may be feasible. This is the case for the class of positive semi–definite matrices. The intuition here is that $t$ testing operates on scalar–valued statistics and positive semi–definite matrices are scalar–like matrices (Horn & Johnson, 1985, chapter 7).[7]

Suppose, as before, that $B$ and $\widehat{B}$ are restricted to be symmetric $m \times m$ matrices and let $\beta = \text{vech}(B)$ and $\widehat{\beta} = \text{vech}(\widehat{B})$. We take $B$ to be positive semi–definite and $D$ under $H_T(r)$ to be positive semi–definite. Then under $H_0(r)$ it is clear that $\lambda_{r+1}(B) + \cdots + \lambda_m(B) = 0$. We can therefore hope to construct a stochastic $t$ test around the smallest $m - r$ eigenvalues of $\widehat{B}$. If we define $\widehat{M}_r$ to consist of the eignevectors associated with the $m - r$ smallest eigenvalues of $\widehat{B}$, then $\sqrt{T}(\lambda_{r+1}(\widehat{B}) + \cdots + \lambda_m(\widehat{B})) = \sqrt{T}\text{tr}(P_{\widehat{M}_r}\widehat{B})$ is a stochastic $t$ statistic with $\widehat{\beta} = \text{vech}(\widehat{B})$ and restriction matrix $\widehat{R}_r = D_m'\text{vec}(P_{\widehat{M}_r})$.

More generally, for a given null space estimator $\widehat{M}_r$ and normalization matrix $\widehat{\Omega}$, we may

---

[7]The applicability of stochastic $t$ tests extends to matrices of the form $B = F^{-1}G$, estimated by $\widehat{B} = \widehat{F}^{-1}\widehat{G}$, where $F$, $\widehat{F}$, and $\widehat{G}$ are positive definite and $G$ is positive semi–definite (Nyblom & Harvey (2000) formulate their stochastic $t$ test in this form). $F^{-1}G$ has a rank of $r$ if and only if $G$ has a rank of $r$ so if $\widehat{M}_r$ is a right null space estimator of $B$, then a left null space estimator is given by $\widehat{N}_r = \widehat{F}\widehat{M}_r$ and $\widehat{N}_r'\widehat{B}\widehat{M}_r = \widehat{M}_r'\widehat{G}\widehat{M}_r$, which is symmetric and therefore amenable to the analysis of this section.

define the stochastic $t$ test

$$\widehat{t}(r) = \frac{\sqrt{T}\operatorname{tr}(P_{\widehat{M}_r}\widehat{B})}{\sqrt{\operatorname{vec}'(P_{\widehat{M}_r})D_m\widehat{\Omega}D_m'\operatorname{vec}(P_{\widehat{M}_r})}} \tag{9}$$

and compare it to the classical $t$ test

$$t(r) = \frac{\sqrt{T}\operatorname{tr}(P_{M_r}\widehat{B})}{\sqrt{\operatorname{vec}'(P_{M_r})D_m\widehat{\Omega}D_m'\operatorname{vec}(P_{M_r})}}, \tag{10}$$

where $M_r$ spans the null space of $B$. In case we wish not to normalize, the implicit $\widehat{\Omega}$ is $\frac{1}{m-r}(D_m'D_m)^{-1}$, which satisfies (A1) automatically.[8]

We will see that most of the rank test statistics proposed in the literature are either exactly or asymptotically of the forms (5), (7) or (9) and differ only by their exponent on $T$, their methods of identifying and estimating the null spaces, and their choices of normalizing matrix $\widehat{\Omega}$. We will, however, discuss alternative rank testing statistics in section 6.

# 4 Rank Testing Under Standard Asymptotics

## 4.1 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 generally assume that $\widehat{B} \in \mathbb{G}^{n \times m}$. This is guaranteed under assumption (A1) because $\operatorname{vec}(\widehat{B})$ is non–degenerate and the set of rank–deficient matrices is of measure zero.[9]

The problem of identifying and 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 2.** Suppose $\{\varepsilon_1, \varepsilon_2\} \subset (0,1)$ and consider the set of matrices

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

---

[8]Since $D_m(D_m'D_m)^{-1}D_m' : \operatorname{vec}(B) \mapsto \operatorname{vec}(B)$ for every symmetric $B \in \mathbb{R}^{m \times m}$ (Magnus & Neudecker, 1999, Theorem 3.12 (b)), $\operatorname{vec}'(P_{\widehat{M}_r})D_m\widehat{\Omega}D_m'\operatorname{vec}(P_{\widehat{M}_r}) = \frac{1}{m-r}\|\operatorname{vec}(P_{\widehat{M}_r})\|^2 = \frac{1}{m-r}\|P_{\widehat{M}_r}\|^2 = 1$.

[9]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 symmetric matrices in $\mathbb{R}^{m \times m}$ of rank $r$ is a submanifold 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 symmetric matrices in $\mathbb{R}^{m \times m}$ is also of measure zero.

When $\varepsilon_2$ is very small relative to $\varepsilon_1$ and 1, we may approximate $\widehat{B}$ by $\widehat{B}_2$ and estimate the null spaces by $\widehat{M}_2 = \widehat{N}_2 = [\,0 \quad 0 \quad 1\,]'$. When both $\varepsilon_1$ and $\varepsilon_2$ are small, we may approximate $\widehat{B}$ by $\widehat{B}_1$ and estimate the null spaces by $\widehat{M}_1 = \widehat{N}_1 = [\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}]'$. 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 $\varepsilon_1$ and $\varepsilon_2$. If $\varepsilon_1$ is very small relative to $\varepsilon_2$ and 1, we estimate the null spaces by $\widehat{M}_2 = \widehat{N}_2 = [\,0 \quad 1 \quad 0\,]'$ instead. 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.

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

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

Thus, to estimate null spaces we think of $\widehat{B}$ as a perturbation of the rank–$r$ matrix $B$. If we can find a reduced rank approximation (RRA) $\widehat{B}_r$ of rank $r$ that approximates $\widehat{B}$ well enough, then $\widehat{B}_r$ will approximate $B$ well and we may estimate the null spaces of $B$ as the null spaces of $\widehat{B}_r$. If on the other hand, we approximate $\widehat{B}$ by a matrix of rank $l > r$, then $\widehat{B}_l$ will be consistent for $B$ and we might expect the null space estimators to merge with the null spaces of $B$, although we cannot, in general, expect the null space estimators to converge. Finally, if $l < r$, then $\widehat{B}_l$ cannot possibly converge to $B$ and we may choose null space estimators, $\widehat{N}_l$ and $\widehat{M}_l$, so that $\widehat{N}_l'\widehat{B}\widehat{M}_l$ is bounded away from zero.

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

**Definition 1** (Norm–based Approximations)**.** Here we approximate $\widehat{B}$ by the rank–$r$ matrix that minimizes the Mahalanobis distance to $\widehat{B}$,

$$\widehat{B}_r^{CDA} \in \operatorname{argmin}\{\|\widehat{B} - A\|_{\widehat{\Theta}} : \operatorname{rank}(A) \leq r\}, \tag{11}$$

where $\widehat{\Theta}$ is symmetric and positive definite. We term this the Cragg and Donald approximation (CDA), after Cragg & Donald (1997), who first proposed it in econometrics.[10] $\qquad\square$

The idea behind the CDA is quite simply to find the closest rank–$r$ matrix according to the Mahalanobis metric. Cragg & Donald (1997) prove the existence of the CDA by standard methods. Uniqueness may not hold (see appendix 9). However, this will not affect our results as any solution to (11) will do. We show in the appendix that $\widehat{B}_r^{CDA}$ is a good approximation to $\widehat{B}$, provided $\widehat{B}$ converges to a matrix of rank $r$ and $\widehat{\Theta}$ does not give disproportionate weight

---

[10]The approximation was proposed much earlier by Gabriel & Zamir (1979).

to any given component in the approximation (i.e. cond($\widehat{\Theta}$) remains bounded). We also show that rank($\widehat{B}_r^{CDA}$) $= r$ so that the null space estimators are well defined. Finally, we note that Cragg & Donald (1997) utilize the statistic $T\|\widehat{B} - \widehat{B}_r^{CDA}\|_{\widehat{\Omega}}^2$ to test rank hypotheses and we show in the appendix that this statistic is exactly of the form of a stochastic $F$ statistic.

The CDA nests a number of other RRAs as special cases. When $\widehat{\Theta}$ is the identity matrix we obtain the RRA used in Ratsimalahelo (2003), Kleibergen & Paap (2006), and Donald et al. (2007), the rank–$r$ singular value decomposition approximation of $\widehat{B}$. For specificity, we will refer to this RRA as the SVD approximation and denote it by $\widehat{B}_r^{SVD}$. When $\widehat{B} \in \mathbb{G}^{m \times m}$ is symmetric, $\widehat{B}_r^{SVD}$ is the RRA obtained by eliminating from $\widehat{B}$ its components associated with the closest $m - r$ eigenvalues to zero.

More generally, when $\widehat{\Theta} = \widehat{\Gamma} \otimes \widehat{\Sigma}$, with $\widehat{\Gamma} \in \mathbb{R}^{m \times m}$ and $\widehat{\Sigma} \in \mathbb{R}^{n \times n}$, we obtain the RRAs implicit in Bartlett (1947), Anderson (1951), and Izenman (1975). It appears explicitly only in Robin & Smith (2000) and so we will refer to it as the Robin and Smith decomposition (RSD) approximation and denote it by $\widehat{B}_r^{RSD}$.

When $\widehat{\Theta}$ is not of Kronecker product form, there are no known analytical solutions. However, we detail a novel iterative scheme for obtaining the CDA in the appendix, which works quite well in numerical experiments.

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

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

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) If rank($B$) $= r$, there is a $K_2 > 0$, not dependent on $\widehat{B}$, such that $\sigma_{r+1}(\widehat{S}_{22}) \leq K_2\sigma_{r+1}(\widehat{B})$ as $\widehat{B} \to B$.

The RRA suggested by this decomposition is then

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

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 our next result that the bound on $\sigma_r(\widehat{S}_{11})$ guarantees power of our rank tests, while the bound on $\sigma_{r+1}(\widehat{S}_{22})$ guarantees size. As with the CDA, existence is guaranteed although uniqueness may not hold and, again, this has no

affect on our results. We show in the appendix that $\widehat{B}_r^{DBA}$ provides a good approximation to $\widehat{B}$ provided it is converging to a matrix of rank at most $r$. We also show that $\text{rank}(\widehat{B}_r^{DBA}) = r$ so that the null space estimators are well defined.

In the LU decomposition, 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. Robin & Smith (1995) and Kleibergen & van Dijk (1994) utilize decompositions similar to the LU decomposition. Donald et al. (2007) used an LU decomposition for symmetric matrices. The spectral decomposition for symmetric matrices used in Donald et al. (2007) and Nyblom & Harvey (2000) has $\widehat{S}$ diagonal and $\widehat{U} = \widehat{V}$ and orthogonal. The SVD and RSD may also be considered DBAs, however, they are more naturally analyzed as norm–based RRAs.

There are also a number of DBAs not previously used in the statistics or the econometrics literatures such as the QR and Cholesky decompositions (Higham, 1990; Hansen, 1998). This is quite surprising as many of the numerical analysis algorithms dominate the algorithms used in statistics and econometrics in terms of computational speed.

We now have all the elements necessary in order to estimate null spaces for rank testing.

**Lemma 3.** Let $\widehat{B}$ be an estimator of $B \in \mathbb{R}^{n \times m}$ such that $\widehat{B} \in \mathbb{G}^{n \times m}$, $\text{rank}(B) = r$, and $\sqrt{T}(\widehat{B} - B) = O_p(1)$. Let the RRAs $\{\widehat{B}_i : 0 \leq i < \min\{n, m\}\}$ be either CDAs or DBAs. In the former case, we assume that $\text{cond}(\widehat{\Theta}) = O_p(1)$. Then the following holds:

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

(ii) If, for $0 \leq l < r$, $\widehat{B}_l - B_l = o_p(1)$, then $P_{\widehat{N}_l} - P_{N_l} = o_p(1)$ and $P_{\widehat{M}_l} - P_{M_l} = o_p(1)$, where $N_l$ and $M_l$ span the left and right null spaces of $B_l$, the rank–$l$ RRA of $B$.

(iii) If $0 \leq l < r$ then $P_{\widehat{N}_l} \widehat{B} P_{\widehat{M}_l} = O_p^{-1}(1)$.

Lemma 3 (i) establishes the rate of convergence of CDAs and DBAs, along with the associated null space estimators. Lemma 3 (ii) establishes the consistency of the null space estimators when the rank of $B$ is underestimated and the underlying RRA is continuous at $B$. As continuity can fail, Lemma 3 (iii) provides conditions that ensure power for our tests. Continuity is known to be generic for the SVD and RSD (Stewart & Sun, 1990; Markovsky, 2012). However, there does not seem to be much one can say about the CDA and one has to think very hard to produce a point of discontinuity for a DBA (e.g. matrices with multiple pivots). One might well conjecture that continuity is generic for all RRAs. However, this is beyond the scope of this paper.

## 4.2 The Stochastic $F$ Test

Now that we have obtained null space estimators for a given matrix, we may attempt to formulate rank testing statistics. We begin with the stochastic $F$ statistic as it is the most commonly encountered statistic in the literature.

**Theorem 1.** Let assumption (A1) hold with $\beta = \text{vec}(B)$, $\widehat{\beta} = \text{vec}(\widehat{B})$, and $B, \widehat{B} \in \mathbb{R}^{n \times m}$.

(i) Under $H_0(r)$ (resp. $H_T(r)$), 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 $B$. If there are null space estimators $\widehat{N}_r$ and $\widehat{M}_r$, such that either $\sqrt{T}(P_{\widehat{N}_r} - P_{N_r}) = O_p(1)$ and $P_{\widehat{M}_r} - P_{M_r} = o_p(1)$ or $P_{\widehat{N}_r} - P_{N_r} = o_p(1)$ and $\sqrt{T}(P_{\widehat{M}_r} - P_{M_r}) = O_p(1)$, then $\widehat{F}(r) - F(r) = o_p(1)$.

(ii) Under $H_1(r)$, let $N_r \in \mathbb{G}^{n \times (n-r)}$ and $M_r \in \mathbb{G}^{m \times (m-r)}$ satisfy $N_r' B M_r \neq 0$. If $P_{\widehat{N}_r} - P_{N_r} = o_p(1)$ and $P_{\widehat{M}_r} - P_{M_r} = o_p(1)$, then $T^{-1}(\widehat{F}(r) - F(r)) = o_p(1)$.

(iii) Under $H_1(r)$, if $P_{\widehat{N}_r} \widehat{B} P_{\widehat{M}_r}$ is $O_p^{-1}(1)$, then $T^{-1}\widehat{F}(r)$ is $O_p^{-1}(1)$ and so $\widehat{F}(r) \overset{p}{\to} \infty$.

Theorem 1 (i) says that under the null or the local alternative, only one null space estimator needs to be $\sqrt{T}$–consistent, the other only needs to be consistent. This is quite surprising in that the RRA we have been studying furnish us with $\sqrt{T}$–consistent null space estimators. Thus they provide much more than is actually necessary for the plug–in principle to hold. This is all the more surprising considering that assumptions (A3) and (A3)$^*$ seemed so demanding.

Theorem 1 (ii) and (iii) are the counterparts to Lemma 2 (ii) and (iii). Under $H_1(r)$, if the RRA is continuous at $B$, the stochastic and classical $F$ statistics diverge at the same rate and if the RRA is not continuous at $B$, then we only have the divergence result. Example 8 provides a Monte Carlo illustration of this phenomenon.

The typical assumptions in the literature are those given in Example 1. In the small–$b$ case, if $\Omega = \Phi$ then the limiting distribution of $\widehat{F}(r)$ is the usual $\chi^2((n-r)(m-r))$ distribution under $H_0(r)$ and $\chi^2((n-r)(m-r), \text{vec}'(D)R_r(R_r'\Omega R_r)^{-1}R_r'\text{vec}(D))$ under $H_T(r)$. If $\Omega \neq \Phi$, the limiting distributions under $H_0(r)$ (and $H_T(r)$) is the quadratic form in $(n-r)(m-r)$ normal random variables familiar from the misspecification literature (White, 1994, chapter 8). The fixed–$b$ asymptotics are more involved and we refer the reader to the literature cited in Example 1. Fixed–$b$ asymptotics are illustrated in section 7.

It follows from the above that the particular choice of null space estimator is immaterial, at least as far as first order asymptotics is concerned. An immediate corollary, for example, 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 calculate Kleibergen & Paap (2006) statistic.

Stochastic $F$ statistics for symmetric matrices satisfy analogous results to Theorems 1.

**Theorem 2.** Let assumption (A1) hold with $B, \widehat{B} \in \mathbb{R}^{m \times m}$ symmetric, $\beta = \text{vech}(B)$, and $\widehat{\beta} = \text{vech}(\widehat{B})$. Under $H_T(r)$, we assume that $D$ is symmetric.

(i) Under $H_0(r)$ (resp. $H_T(r)$), let $M_r \in \mathbb{G}^{m \times (m-r)}$ span the null space of $B$. If there is a null space estimator $\widehat{M}_r$ such that $\sqrt{T}(P_{\widehat{M}_r} - P_{M_r}) = O_p(1)$, then $\widehat{F}^s(r) - F^s(r) = o_p(1)$.

(ii) Under $H_1(r)$, let $M_r \in \mathbb{G}^{n \times (n-r)}$ satisfy $M_r' B M_r \neq 0$. If $P_{\widehat{M}_r} - P_{M_r} = o_p(1)$, then $T^{-1}(\widehat{F}^s(r) - F^s(r)) = o_p(1)$.

(iii) Under $H_1(r)$. If $P_{\widehat{M_r}} \widehat{B} P_{\widehat{M_r}}$ is $O_p^{-1}(1)$, then $\widehat{F}^s(r) \xrightarrow{p} \infty$.

Identical observations apply to Theorem 2 as do to Theorem 1. In particular, Lemma 3 together with Theorem 2 establish the plug–in principle for stochastic $F$ statistics in symmetric matrices when the null spaces are estimated by CDAs or DBAs. Under small–$b$ asymptotics with $\Omega = \Phi$, the limiting distribution of $\widehat{F}^s(r)$ is $\chi^2(m(m+1)/2)$ under $H_0(r)$ (Donald et al., 2007). Under $H_T(r)$, it converges in distribution to a noncentral $\chi^2$ distribution with $m(m+1)/2$ degrees of freedom and noncentrality parameter $\text{vech}'(M_r'DM_r)\{D_{m-r}^{\dagger}(M_r \otimes M_r)'D_m\Omega D_m'(M_r \otimes M_r)D_{m-r}^{\dagger}{}'\}^{-1}\text{vech}(M_r'DBM_r)$. The case $\Omega \neq \Phi$ can easily be derived by the methods of White (1994). Finally, fixed–$b$ theory is also applicable as we illustrate in section 7.

## 4.3 The Stochastic $t$ Test

As we saw in section 3.2, stochastic $t$ tests apply only to symmetric matrices that converge to positive semi–definite matrices. Here we have the following general result.

**Theorem 3.** Let assumption (A1) hold with $\beta = \text{vech}(B)$, $\widehat{\beta} = \text{vech}(\widehat{B})$, and $B, \widehat{B} \in \mathbb{R}^{m \times m}$ are symmetric with $B$ positive semi–definite.

(i) Under $H_0(r)$ (resp. $H_T(r)$), let $M_r \in \mathbb{G}^{m \times (m-r)}$ span the null space of $B$. If there is a null space estimator $\widehat{M_r}$ such that $\sqrt{T}(P_{\widehat{M_r}} - P_{M_r}) = O_p(1)$, then $\widehat{t}(r) - t(r) = o_p(1)$.

(ii) Under $H_1(r)$, let $M_r \in \mathbb{G}^{m \times (m-r)}$ satisfy $M_r'BM_r \neq 0$. If $P_{\widehat{M_r}} - P_{M_r} = o_p(1)$, then $T^{-1/2}(\widehat{t}(r) - t(r)) = o_p(1)$.

(iii) Under $H_1(r)$. If $P_{\widehat{M_r}} \widehat{B} P_{\widehat{M_r}}$ is $O_p^{-1}(1)$, then $|\widehat{t}(r)| \xrightarrow{p} \infty$.

Identical observations apply to Theorem 3 as do to Theorems 1 and 2. In particular, Lemma 3 together with Theorem 3 establish the plug–in principle for stochastic $t$ statistics when the null spaces are estimated by CDAs or DBAs. Finally, under small–$b$ asymptotics and $\Omega = \Phi$, the limiting distribution of $\widehat{t}$ is $N(0, 1)$ under $H_0(r)$ (Donald et al., 2007). The limiting distribution is $N\left(\frac{\text{tr}(P_{M_r}D)}{\sqrt{\text{vec}'(P_{M_r})D_m\Omega D_m'\text{vec}(P_{M_r})}}, 1\right)$ under $H_T(r)$. When $\Omega \neq \Phi$, the results of White (1994) apply. Finally, fixed–$b$ theory is applicable and illustrated in section 7.

# 5 Rank Testing Under General Asymptotics

Cointegration presents some truly fascinating anomalies for rank testing. To see this, we consider the following examples.

**Example 3** (The Cointegrated VAR Model). Let $\{\varepsilon_t : t \geq 1\}$ be i.i.d. with finite and positive definite covariance matrix $\Sigma \in \mathbb{G}^{m \times m}$, let $y_{-1} = y_0 = 0$, and

$$y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \varepsilon_t, \qquad t = 1, \dots, T.$$

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). If $B = \Pi_1 + \Pi_2 - I_m$, then $\text{rank}(B) < m$ is the number of cointegration relationships. Let $N_r \in \mathbb{G}^{m \times m-r}$ and $M_r \in \mathbb{G}^{m \times m-r}$ span the left and right null spaces of $B$. Let $\widehat{B}$ be the OLS estimator of $B$ and let $\widehat{\Omega}$ be the OLS estimator of the asymptotic variance of $\widehat{B}$.

The classical $F$ statistic for testing $H_0(r)$ is given by

$$F(r) = T\text{vec}'(\widehat{B})(M_r \otimes N_r)\left((M_r \otimes N_r)'\widehat{\Omega}(M_r \otimes N_r)\right)^{-1}(M_r \otimes N_r)'\text{vec}(\widehat{B})$$

and it easy to check that it has the same limiting distribution as Johansen's trace statistic (Johansen, 1991), which is a stochastic $F$ statistic based on the RSD. Thus, the plug–in principle would seem to hold in this case.

Unfortunately, however, assumption (A1) is violated as $\widehat{\Omega}$ converges to a singular matrix. In particular, $T(M_r \otimes I_n)'\widehat{\Omega}(M_r \otimes I_n)$ converges in distribution to an almost surely positive definite random matrix. 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, $T\widehat{B}M_r = O_p(1)$. Thus, elements of the side terms of $F(r)$ above converge to zero at a rate that exactly counterbalances the rate at which the corresponding elements of the sandwiched term are exploding.

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 some linear combinations of the columns of $\widehat{B}$ converges to zero at a rate of $T$, while other linear combinations converge to zero at a rate of $T^2$. Thus, there may be heterogenous rates of accelerated convergence that need to be taken into account. □

The example suggests that some form of regularization is necessary if we are to continue to utilize the theory we have developed in section 4. If $\widehat{\Omega}$ is asymptotically singular, we can try to find a regularizing matrix $Z_T$ that takes into account the potentially heterogenous rates of convergence so that $\widehat{\Omega}_T = Z_T'\widehat{\Omega}Z_T$ and its inverse are bounded in probability and, along with $\widehat{\beta}_T = Z_T'\widehat{\beta}$, satisfies the conditions of section 4. We can then attempt to derive a plug–in principle for $F = T\widehat{\beta}'R(R'\widehat{\Omega}R)^{-1}R'\widehat{\beta} = T\widehat{\beta}_T'R_T(R_T'\widehat{\Omega}_T R_T)^{-1}R_T'\widehat{\beta}_T$ and $t = \frac{\sqrt{T}R'\widehat{\beta}}{\sqrt{R'\widehat{\Omega}R}} = \frac{\sqrt{T}R_T'\widehat{\beta}_T}{\sqrt{R_T'\widehat{\Omega}_T R_T}}$, where $R_T = Z_T^{-1}R$. In fact, all of the cointegration rank statistics can be viewed as regularized statistics as we will soon see.

**Example 4** (Common Stochastic Trends). Let $\{\varepsilon_t : t \geq 1\}$ and $\{u_t : t \geq 1\}$ be mutually independent $m$–dimensional Gaussian i.i.d. processes, let $x_0$ be fixed, and define

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

We assume that $\varepsilon$ has a covariance matrix $\Sigma \in \mathbb{G}^{m \times m}$, while $u$ has a covariance matrix $B \in \mathbb{R}^{m \times m}$, whose rank 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 & Harvey (2000) show that $\widehat{B}$ converges in distribution to a random matrix whose null space is exactly the span of $M_r$. In particular, $[\begin{array}{cc} M_{r\perp} & \sqrt{T} M_r \end{array}]' \widehat{B} [\begin{array}{cc} M_{r\perp} & \sqrt{T} M_r \end{array}]$ converges in distribution to an almost surely positive definite matrix. Experience would then suggest applying the plug–in principle to the statistic $\mathrm{tr}(T P_{M_r} \widehat{B})$. However, this statistic does not achieve the same limiting distribution as Nyblom and Harvey's statistic, which is the stochastic $t$ statistic based on the eigenvalues of $\widehat{B}$. Indeed $\mathrm{tr}(T P_{M_r} \widehat{B})$ does not achieve the same limiting distribution as any other statistic that plugs in a reasonable estimator for $M_r$ (see Example 11). It would seem then that the plug–in principle fails.

Figure 3: Convergence of the Nyblom & Harvey (2000) Statistic.



$\widehat{B}$ converges at a rate of $\sqrt{T}$ along $M_r$ but at a rate of $T$ along the curve.
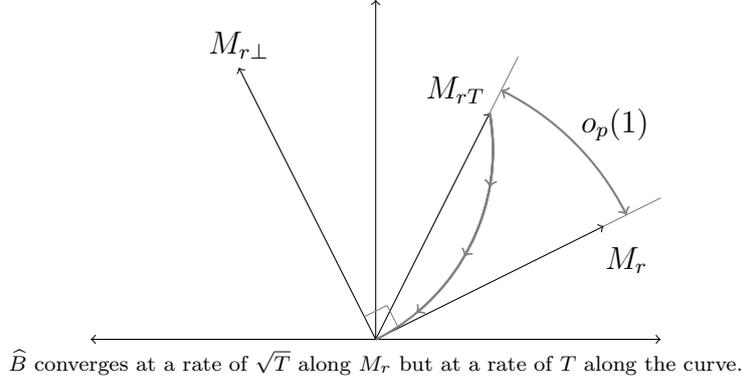
In fact, the plug–in principle still holds but for a different restriction matrix than $M_r$. One can check that $\widehat{B} M_r$ converges at a rate slower than $O_p(T^{-1})$. On the other hand, the Poincaré separation theorem implies that along the eigenvectors associated with the smallest $m - r$ 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 $T \widehat{B} M_{rT} = O_p(1)$. 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. Another, simpler, example of this is $M_{rT} = \sqrt{T} \widehat{\Sigma}^{1/2} (I_m - M_{r\perp}(M_{r\perp}' \widehat{\Gamma} M_{r\perp})^{-1} M_{r\perp}' \widehat{\Gamma}) M_r$, which is bounded in probability and satisfies $T \widehat{B} M_{rT} = O_p(1)$. 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 $[\begin{array}{cc} M_{r\perp} & M_r \end{array}]' \widehat{\Gamma} [\begin{array}{cc} M_{r\perp} & M_r \end{array}]$. In both cases, $\widehat{B}$ converges faster along $M_{rT}$ than it does along $M_r$, even though $P_{M_{rT}}$ converges to $P_{M_r}$ (see Figure 3). 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

$\operatorname{tr}(T P_{M_{rT}} \widehat{B})$ is precisely the limiting distribution of the Nyblom and Harvey statistic.[11]  □

The above example suggests two additional features of cointegration. First, $\widehat{B}$ is not consistent for $B$ in the above example, although its estimated null spaces are consistent for the null spaces of $B$. Thus, we must allow for statistics of varying rates of convergence, including statistics that are inconsistent. Second, the regularizing matrices should be allowed to be random in addition to being time varying.

## 5.1   Estimating the Null Spaces

In order to formulate stochastic statistics for cointegration, we must first understand how RRAs perform under heterogeneous rates of convergence along possibly nonconstant directions. The following is a generalization of Lemma 3.

**Lemma 4.** Let $\widehat{B}$ be an estimator of $B \in \mathbb{R}^{n \times m}$ such that $\widehat{B} \in \mathbb{G}^{n \times m}$, $T^\alpha(\widehat{B} - B) = O_p(1)$ for $\alpha \geq 0$, $\operatorname{rank}(B) = r$, and $\sigma_r(\widehat{B}) = O_p^{-1}(1)$. Suppose there exists sequences of possibly random matrices $N_{rT} \in \mathbb{G}^{n \times (n-r)}$ and $M_{rT} \in \mathbb{G}^{m \times (m-r)}$ such that $[\ N_{r\perp} \quad N_{rT}\ ]$ and $[\ M_{r\perp} \quad M_{rT}\ ]$ have singular values bounded away from zero in probability and, for $\gamma \geq 0$,

$$T^\gamma N'_{rT} \widehat{B} = O_p(1), \qquad T^\gamma \widehat{B} M_{rT} = O_p(1), \qquad T^\gamma N'_{rT} \widehat{B} M_{rT} = O_p(1).$$

Let the RRAs $\{\widehat{B}_i : 0 \leq i < \min\{n, m\}\}$ be either CDAs or DBAs. In the former case, we assume that $\operatorname{cond}(\widehat{\Theta}_T) = O_p(1)$, where $\widehat{\Theta}_T = Z'_T \widehat{\Theta} Z_T$ and $Z_T = [\ M_{r\perp} \quad M_{rT}\ ] \otimes [\ N_{r\perp} \quad N_{rT}\ ]$. Then:

(i)  $T^\gamma(\widehat{B} - \widehat{B}_r) = O_p(1)$, $T^\gamma(P_{\widehat{N}_r} - P_{N_{rT}})N_{rT} = O_p(1)$, and $T^\gamma(P_{\widehat{M}_r} - P_{M_{rT}})M_{rT} = O_p(1)$.

(ii)  If, for $0 \leq l < r$, $\widehat{B}_l - B_l = o_p(1)$, then $P_{\widehat{N}_l} - P_{N_l} = o_p(1)$ and $P_{\widehat{M}_l} - P_{M_l} = o_p(1)$, where $N_l$ and $M_l$ span the left and right null spaces of $B_l$, the rank–$l$ RRA of $B$.

(iii)  Let $0 \leq l < r$. For the DBA, $P_{\widehat{N}_l} \widehat{B} P_{\widehat{M}_l} = O_p^{-1}(1)$. For the CDA, $P_{\widehat{N}_{lT}} \widehat{B}_T P_{\widehat{M}_{lT}} = O_p^{-1}(1)$, where $\widehat{N}_{lT} = [\ N_{r\perp} \quad N_{rT}\ ]^{-1} \widehat{N}_l$, $\widehat{M}_{lT} = [\ M_{r\perp} \quad M_{rT}\ ]^{-1} \widehat{M}_l$, and $\widehat{B}_T = [\ N_{r\perp} \quad N_{rT}\ ]' \widehat{B} [\ M_{r\perp} \quad M_{rT}\ ]$.

Lemma 3 is the special case of Lemma 4 where $\alpha = \gamma = \frac{1}{2}$, $N_{rT} = N_r$, and $M_{rT} = M_r$. In the $I(1)$ case of example 3, $\alpha = \gamma = \frac{1}{2}$, $N_{rT} = N_r$, and $M_{rT} = T^{1/2} M_r$, while in the $I(2)$ case $M_{rT}$ consists of columns of $M_r$ that grow at a rate of $\sqrt{T}$ as well as columns of $M_r$ that grow at a rate of $T^{3/2}$. In example 4, on the other hand, we have $\alpha = 0$, $\gamma = 1$, and $N_{rT} = M_{rT} = \sqrt{T} \widehat{\Sigma}^{1/2}(I_m - M_{r\perp}(M'_{r\perp} \widehat{\Gamma} M_{r\perp})^{-1} M'_{r\perp} \widehat{\Gamma}) M_r$.

The condition that $\sigma_r(\widehat{B})$ be bounded away from zero in probability is clearly redundant when $\alpha > 0$ as $\widehat{B}$ converges to a rank–$r$ matrix. However, when $\alpha = 0$, as in example 4, it allows us to estimate the limiting null spaces of $\widehat{B}$, provided $\widehat{B}$ converges to zero in probability

---

along $M_{rT}$ and $N_{rT}$. The condition that $Z_T$ have singular values bounded away from zero in probability is important for two reasons. First, it requires that $M_{rT}$ be distinguishable from $M_{r\perp}$ in the limit $T \to \infty$ and similarly for $N_{rT}$. Second, it requires that $M_{rT}$ and $N_{rT}$ be almost surely of full rank so that they almost surely define subspaces of the stated dimensions.

Lemma 4 (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 right null space estimator in the $I(1)$ case of example 3 converges at a rate of $T$ as $T(P_{\widehat{M}_r} M_r - M_r) = O_p(1)$, while the estimated null spaces converge at a rate of $T$ in example 4 even though $\widehat{B}$ is inconsistent. Second, it allows the subspaces of $\operatorname{span}(\widehat{M}_r)$ to converge at different rates. In the $I(2)$ case of example 3 one subspace of $\operatorname{span}(\widehat{M}_r)$ will converge at the rate of $T$, while its complement in $\operatorname{span}(\widehat{M}_r)$ will converge at a rate of $T^2$. The rates given in Lemma 4 (i) are therefore more parsed descriptions of the rates of convergence of the estimated subspaces.[12]

Lemma 4 (ii) is identical to its counterpart in Lemma 3 (ii). Lemma 4 (iii), on the other hand, is quite different for the CDA. We are only ensured that the regularized $\widehat{B}$ is bounded away from zero in probability along regularized directions. We will see, however, that the CDA is still capable of delivering power in stochastic $F$ tests.

Note finally that nothing in the results of Lemma 4 depends on the overall rate of convergence $\alpha$. As we noted in section 4, the only part of the asymptotics that is relevant is the part along the restriction matrix, in this case $M_{rT} \otimes N_{rT}$, where convergence is at the rate of $T^\gamma$. This is why, we will specify the limiting distribution of $\widehat{B}$ only along $M_{rT} \otimes N_{rT}$ in the remaining sections.

## 5.2 The Stochastic $F$ Test

The following assumptions nest the assumptions of all of the stochastic tests of cointegration rank for nonsymmetric matrices in the literature, including Example 3.

(C1) $B \in \mathbb{R}^{n \times m}$. $\Omega \in \mathbb{R}^{nm \times nm}$ is random, symmetric, and almost surely positive definite. $\widehat{B}$ and $\widehat{\Omega}$ are estimators of $B$ and $\Omega$ indexed by $T$ such that $\operatorname{vec}(\widehat{B})$ is a non–degenerate random vector and each $\widehat{\Omega} \in \mathbb{R}^{nm \times nm}$ is symmetric and almost surely positive definite. If $\operatorname{rank}(B) = r$ and $N_r \in \mathbb{G}^{n \times (n-r)}$ and $M_r \in \mathbb{G}^{m \times (m-r)}$ span the left and right null

---

[12]Formally, set $V_{rT}$ to be a matrix of orthonormal columns that spans $M_{rT}$ and set $\widehat{V}_{\cdot 1}$ to be a matrix of orthonormal columns that spans the row space of $\widehat{B}_r^{CDA}$. Now $T^\gamma(P_{\widehat{M}_r} - P_{M_{rT}})M_{rT} = O_p(1)$ implies that $T^\gamma(P_{\widehat{M}_r} - I_m)M_{rT} = O_p(1)$. Since the singular values of $M_{rT}$ are $O_p^{-1}(1)$ (because those of $[\, M_{r\perp} \quad M_{rT} \,]$ are $O_p^{-1}(1)$ (Horn & Johnson, 1991, Corollary 3.1.3)), 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).

spaces of $B$ respectively, then $\sigma_r(\widehat{B}) = O_p^{-1}(1)$ and there exist sequences of possibly random matrices $N_{rT} \in \mathbb{G}^{n \times (n-r)}$ and $M_{rT} \in \mathbb{G}^{m \times (m-r)}$ such that $[\ N_{r\perp}\ \ N_{rT}\ ]$ and $[\ M_{r\perp}\ \ M_{rT}\ ]$ have singular values bounded away from zero in probability and, for $\alpha, \gamma \geq 0$,

$$T^\alpha(\widehat{B} - B) = O_p(1), \qquad T^\gamma N'_{rT}\widehat{B} = O_p(1), \qquad T^\gamma \widehat{B} M_{rT} = O_p(1),$$

while $(T^\gamma N'_{rT}\widehat{B}M_{rT}, \widehat{\Omega}_T) \xrightarrow{d} (\xi, \Omega)$, where $\widehat{\Omega}_T = Z'_T \widehat{\Omega} Z_T$ and $Z_T = [\ M_{r\perp}\ \ M_{rT}\ ] \otimes [\ N_{r\perp}\ \ N_{rT}\ ]$.

Assumption (C1) specializes the assumptions of Lemma 4 in that $T^\gamma N'_{rT}\widehat{B}M_{cT}$ and $\widehat{\Omega}_T$ converge in distribution rather than just being bounded in probability. It reduces to (A1) in the special case that $\alpha = \gamma = \frac{1}{2}$, $N_{rT}$ and $M_{rT}$ are constant, and all of the given statistics converge jointly in distribution.

The conditions in Assumption (C1) are checked prior to deriving the asymptotics of the cointegration rank test (e.g. chapter 10 of Johansen (1995a)). (C1) is, in fact, much more compact than sets of assumptions usually found in the literature, where one derives the asymptotics of each component of $\widehat{B}$ and $\widehat{\Omega}$ separately, then combines them in the asymptotic analysis.

Under (C1), we will be interested in testing $H_0(r)$ against $H_1(r)$ and $H_T(r)$, which is now defined as

$$H_T(r) : \text{rank}\left(B^T - T^{-\gamma}D^T\right) = r,$$

for any deterministic sequence $D^T$ such that $Z'_T \text{vec}(D^T)$ converges in probability. Note that it reduces to the standard–asymptotics local alternative when $\gamma = \frac{1}{2}$ and $Z_T$ and $D^T$ are time invariant. The cointegration literature typically considers uniformly distant alternatives where $D^T$ is replaced by $T^{-\omega}D$, for a constant $D$ and convergent $T^{-\omega}Z_T$ (Johansen, 1995b, chapter 14). The formulation of $H_T(r)$ above is therefore more general than previously considered in the literature.

It is important to note that $\xi$ and $\Omega$ may be different under $H_T(r)$ than under $H_0(r)$. In the $I(1)$ case of Example 3, for example, $TN'_r\widehat{B}M_r$ and $\widehat{\Omega}$ converge to functionals of a Brownian motion under $H_0(r)$ and to a functionals of an Ornstein-Uhlenbeck processes under $H_T(r)$ (Johansen, 1995a, chapter 14). See the proof of Proposition 1 for another example.

Assumption (C1) allows us to find estimates of the null spaces of $B$ by either the CDA (assuming that $\widehat{\Theta}$ satisfies the necessary condition) or DBA and obtain the rates of convergence $T^\gamma(P_{\widehat{N}_r} - P_{N_{rT}})N_{rT} = O_p(1)$ and $T^\gamma(P_{\widehat{M}_r} - P_{M_{rT}})M_{rT} = O_p(1)$ when $\text{rank}(B) = r$. When the RRA is continuous at $B$, the null space estimators are consistent. Otherwise, when $l < r$ we have that either $P_{\widehat{N}_l}\widehat{B}P_{\widehat{M}_l}$ or $P_{\widehat{N}_{lT}}\widehat{B}_T P_{\widehat{M}_{lT}}$ are $O_p^{-1}(1)$.

Based on the assumed existence of $N_{rT}$ and $M_{rT}$ as well as some null space estimators $\widehat{N}_r$ and $\widehat{M}_r$, we can formulate the following statistics

$$F_c(r) = T^{2\gamma}\mathrm{vec}'(N_{rT}'\widehat{B}M_{rT})\{(M_{rT} \otimes N_{rT})'\widehat{\Omega}(M_{rT} \otimes N_{rT})\}^{-1}\mathrm{vec}(N_{rT}'\widehat{B}M_{rT})$$

$$\widehat{F}_c(r) = T^{2\gamma}\mathrm{vec}'(\widehat{N}_r'\widehat{B}\widehat{M}_r)\{(\widehat{M}_r \otimes \widehat{N}_r)'\widehat{\Omega}(\widehat{M}_r \otimes \widehat{N}_r)\}^{-1}\mathrm{vec}(\widehat{N}_r'\widehat{B}\widehat{M}_r).$$

The following is the cointegration version of Theorem 1.

**Theorem 4.** Let assumptions (C1) hold with $\gamma > 0$.

(i) Under $H_0(r)$ (resp. $H_T(r)$), if there are null space estimators $\widehat{N}_r$ and $\widehat{M}_r$, such that either $T^\gamma(P_{\widehat{N}_r} - P_{N_{rT}})N_{rT} = O_p(1)$ and $(P_{\widehat{M}_r} - P_{M_rT})M_{rT} = o_p(1)$ or $(P_{\widehat{N}_r} - P_{N_{rT}})N_{rT} = o_p(1)$ and $T^\gamma(P_{\widehat{M}_r} - P_{M_rT})M_{rT} = O_p(1)$, then $\widehat{F}_c(r) - F_c(r) = o_p(1)$.

(ii) Under $H_1(r)$, let $N_r \in \mathbb{G}^{n\times(n-r)}$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ satisfy $N_r'BM_r \neq 0$. If $P_{\widehat{N}_r} - P_{N_r} = o_p(1)$ and $P_{\widehat{M}_r} - P_{M_r} = o_p(1)$, then $T^{-2\gamma}(\widehat{F}_c(r) - F_c(r)) = o_p(1)$.

(iii) If $l < r$ and either $P_{\widehat{N}_l}\widehat{B}P_{\widehat{M}_l}$ or $P_{\widehat{N}_{lT}}\widehat{B}_T P_{\widehat{M}_{lT}}$ are $O_p^{-1}(1)$, where $\widehat{N}_{lT}$, $\widehat{M}_{lT}$, and $\widehat{B}_T$ are defined in Lemma 4, then $\widehat{F}_c(l) \xrightarrow{p} \infty$.

Theorem 3 allows the plug–in principle to be used in example 3 to formulate the Johansen (1991), Kleibergen & van Dijk (1994), or Kleibergen & Paap (2006) statistics or any other statistics based on alternative RRAs. It also includes as special cases, the results for conditionally heteroskedastic innovation process (Cavaliere et al., 2010a), non–stationary volatility (Cavaliere et al., 2010b), fractional cointegration (Avarucci & Velasco, 2009), and infinite variance innovations (Caner, 1998).

For cointegration rank testing in symmetric matrices, we modify (C1) as follows.

(CS1) $B \in \mathbb{R}^{m\times m}$. $\Omega \in \mathbb{R}^{m(m+1)/2\times m(m+1)/2}$ is random, symmetric, and almost surely positive definite. $\widehat{B}$ and $\widehat{\Omega}$ are estimators of $B$ and $\Omega$ indexed by $T$ such that $\mathrm{vec}(\widehat{B})$ is a non–degenerate random vector and each $\widehat{\Omega} \in \mathbb{R}^{m(m+1)/2\times m(m+1)/2}$ is symmetric and almost surely positive definite. If $\mathrm{rank}(B) = r$ and $M_r \in \mathbb{G}^{m\times(m-r)}$ spans the null space of $B$, then $\sigma_r(\widehat{B}) = O_p^{-1}(1)$ and there exist sequences of possibly random matrices $M_{rT} \in \mathbb{G}^{m\times(m-r)}$ such that $[\,M_{r\perp}\quad M_{rT}\,]$ has singular values bounded away from zero in probability and, for $\alpha, \gamma \geq 0$,

$$T^\alpha(\widehat{B} - B) = O_p(1), \qquad\qquad T^\gamma \widehat{B}M_{rT} = O_p(1),$$

while $(T^\gamma M_{rT}'\widehat{B}M_{rT}, \widehat{\Omega}_T) \xrightarrow{d} (\xi, \Omega)$, where $\widehat{\Omega}_T = D_m^\dagger Z_T' D_m \widehat{\Omega} D_m' Z_T D_m^{\dagger\prime}$ and $Z_T = [\,M_{r\perp}\quad M_{rT}\,] \otimes [\,M_{r\perp}\quad M_{rT}\,]$.

Just as we did in section 4, under (CS1) and $H_T(r)$, we will require each $D^T$ to be symmetric. We will consider the $F$ statistics for symmetric matrices

$$F_c^s(r) = T^{2\gamma}\mathrm{vech}'(M_{rT}'\widehat{B}M_{rT})\{D_{m-r}^\dagger(M_{rT} \otimes M_{rT})'D_m\widehat{\Omega}D_m'(M_{rT} \otimes M_{rT})D_{m-r}^{\dagger\prime}\}^{-1}\mathrm{vech}(M_{rT}'\widehat{B}M_{rT})$$

$$\widehat{F}_c^s(r) = T^{2\gamma}\mathrm{vech}'(\widehat{M}_r'\widehat{B}\widehat{M}_r)\{D_{m-r}^\dagger(\widehat{M}_r \otimes \widehat{M}_r)'D_m\widehat{\Omega}D_m'(\widehat{M}_r \otimes \widehat{M}_r)D_{m-r}^{\dagger\prime}\}^{-1}\mathrm{vech}(\widehat{M}_r'\widehat{B}\widehat{M}_r)$$

**Theorem 5.** Let assumptions (CS1) hold with $\gamma > 0$.

(i) Under $H_0(r)$ (resp. $H_T(r)$), if there is a null space estimator $\widehat{M}_r$, such that $T^\gamma(P_{\widehat{M}_r} - P_{M_rT})M_{rT} = O_p(1)$, then $\widehat{F}_c^s(r) - F_c^s(r) = o_p(1)$.

(ii) Under $H_1(r)$, let $M_r \in \mathbb{G}^{m \times (m-r)}$ satisfy $M_r'BM_r \neq 0$. If $P_{\widehat{M}_r} - P_{M_r} = o_p(1)$, then $T^{-2\gamma}(\widehat{F}_c^s(r) - F_c^s(r)) = o_p(1)$.

(iii) If $l < r$ and either $P_{\widehat{N}_l}\widehat{B}P_{\widehat{M}_l}$ or $P_{\widehat{N}_{lT}}\widehat{B}_T P_{\widehat{M}_{lT}}$ are $O_p^{-1}(1)$, then $\widehat{F}_c^s(l) \xrightarrow{p} \infty$.

Identical observations apply to Theorem 5 as do to Theorem 4. In particular, Lemma 4 together with Theorem 5 establish the plug–in principle for cointegration stochastic $F$ statistics in symmetric matrices when the null spaces are estimated by CDAs or DBAs. To the author's knowledge, there are no cointegration stochastic $F$ statistics in the literature. Therefore, as an application of the theorem, we propose a new test statistic for the common trends model and illustrate its use in Example 8 .

## 5.3   The Stochastic $t$ Test

Under (CS1), when $\widehat{B}$ is symmetric and converges in distribution to an almost surely positive definite matrix, we may formulate the statistics

$$t_c(r) = \frac{T^\gamma \text{tr}(P_{M_rT}\widehat{B})}{\sqrt{\text{vec}'(P_{M_rT})D_m\widehat{\Omega}D_m'\text{vec}(P_{M_rT})}}$$

$$\widehat{t}_c(r) = \frac{T^\gamma \text{tr}(P_{\widehat{M}_r}\widehat{B})}{\sqrt{\text{vec}'(P_{\widehat{M}_r})D_m\widehat{\Omega}D_m'\text{vec}(P_{\widehat{M}_r})}}.$$

Just as we did in section 4, under (CS1) and $H_T(r)$, we will require each $D^T$ to be symmetric and positive semi–definite.

**Theorem 6.** Let assumption (CS1) hold with $\widehat{B}$ converging in distribution to an almost surely positive definite matrix and $\gamma > 0$.

(i) Under $H_0(r)$ (resp. $H_T(r)$), if there is a null space estimator $\widehat{M}_r$ such that $T^\gamma(P_{\widehat{M}_r} - P_{M_rT})M_{rT} = O_p(1)$, then $\widehat{t}_c(r) - t_c(r) = o_p(1)$.

(ii) Under $H_1(r)$, let $M_r \in \mathbb{G}^{m \times (m-r)}$ satisfy $M_r'BM_r \neq 0$. If $P_{\widehat{M}_r} - P_{M_r} = o_p(1)$, then $T^{-\gamma}(\widehat{t}_c(r) - t_c(r)) = o_p(1)$.

(iii) Under $H_1(r)$. If $P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r}$ is $O_p^{-1}(1)$, then $|\widehat{t}_c(r)| \xrightarrow{p} \infty$.

Identical observations apply to Theorem 6 as do to Theorems 4 and 5. In particular, Lemma 3 together with Theorem 6 establish the plug–in principle for stochastic $t$ statistics when the null spaces are estimated by DBAs. Unfortunately, it is not clear if the CDA is capable of delivering power for cointegration stochastic $t$ statistics.

To the author's knowledge, the local power of the Nyblom & Harvey (2000) has not been studied in the literature. Therefore, as an application of the plug–in principle, we have the following result.

**Proposition 1.** Under $H_T(r)$, with $D^T \to D \in \mathbb{R}^{m \times m}$, the Nyblom & Harvey (2000) statistic in Example 4 converges in distribution to $\text{tr}(C_{22} - C'_{12}C_{11}^{-1}C_{12})$, where

$$C_{11} = \int_0^1 \left[ \int_0^u W_1^*(s)ds \right] \left[ \int_0^u W_1^*(s)ds \right]' du$$

$$C_{12} = \int_0^1 \left[ \int_0^u W_1^*(s)ds \right] K'(u)du$$

$$C_{22} = \int_0^1 K(u)K'(u)du$$

$$W_1^*(u) = W_1(u) - \int_0^1 W_1(s)ds$$

$$K(u) = W_2(u) - uW_2(1) + (M_r'\Sigma M_r)^{-1/2}(M_r'DM_r)^{1/2}\int_0^u W_3^*(s)ds.$$

$(W_1', W_2', W_3')'$ is a standard Brownian motion and the dimensions of its three components are $r$, $m - r$, and $m - r$ respectively.

The limiting distribution above reduces to the one reported by Cappuccio & Lubian (2009) in the univariate case. It depends on $D$ only through the singular values of the local alternative's signal–to–noise ratio along the null spaces of $B$, i.e. $\{\sigma_i((M_r'\Sigma M_r)^{-1/2}(M_r'DM_r)^{1/2})\}$.[13]

# 6 Alternative Statistics

Some rank testing statistics in the literature are neither $\widehat{F}$ nor $\widehat{t}$ statistics. Here we consider how they fit in with the theory above. Let $\widehat{B}$ satisfy the conditions of Theorem 1. The extensions to symmetric matrices and cointegration are straightforward and omitted.

**Example 5** (Anderson (1951)). The likelihood ratio statistic for $H_0(r)$ is

$$A = -T \sum_{i=r+1}^{\min\{n,m\}} \log(1 - \mu_i^2(\widehat{B})),$$

where $\mu_i(\widehat{B}) = \sigma_i(\widehat{\Sigma}^{-\frac{1}{2}}\widehat{B}\widehat{\Gamma}^{\frac{1}{2}})$, where $\widehat{\Sigma}$ and $\widehat{\Gamma}$ converge in probability to positive definite matrices. $A$ is a nonlinear function of $\widehat{N}_r'\widehat{\Sigma}^{-\frac{1}{2}}\widehat{B}\widehat{\Gamma}^{\frac{1}{2}}\widehat{M}_r$, where $\widehat{N}_r$ and $\widehat{M}_r$ are the SVD null space estimators of $\widehat{\Sigma}^{-\frac{1}{2}}\widehat{B}\widehat{\Gamma}^{\frac{1}{2}}$ (see the appendix for more details). Under $H_0(r)$ or $H_T(r)$, $A = T\sum_{i=r+1}^{\min\{n,m\}} \mu_i^2(\widehat{B}) + o_p(1) = T\|\widehat{N}_r'\widehat{\Sigma}^{-\frac{1}{2}}\widehat{B}\widehat{\Gamma}^{\frac{1}{2}}\widehat{M}_r\|^2 + o_p(1)$. Thus, the Anderson (1951)

---

[13]This follows from the fact that the distribution is invariant to the mapping $(K, W_3) \mapsto (U'K, VW_3)$, where $U$ and $V$ are orthogonal. Thus, the distribution depends on $U'(M_r'\Sigma M_r)^{-1/2}(M_r'DM_r)^{1/2}V$ for arbitrary $U$ and $V$, which may therefore be chosen to equal the singular value decomposition left and right rotations respectively.

statistic is asymptotically equivalent a stochastic $F$ statistic under either $H_0(r)$ or $H_T(r)$. Since $T^{-1}A$ is nonzero for a nonzero argument and $\widehat{N}'_r \widehat{\Sigma}^{-\frac{1}{2}} \widehat{B} \widehat{\Gamma}^{\frac{1}{2}} \widehat{M}_r$ converges to a nonzero matrix under $H_1(r)$, $A$ diverges under $H_1(r)$. $\qquad \square$

This example points to the possibility of constructing statistics that are asymptotically of the stochastic $F$ (resp. stochastic $t$) form. For example, we may construct tests based on statistic of the form $Th(\widehat{R}'_r\widehat{\beta}, \widehat{R}'_r\widehat{\Omega}\widehat{R}_r)$, where $h : \mathbb{R}^q \times \mathbb{G}^{q \times q} \to \mathbb{R}$ is continuous and satisfies $h(x, Y) = x'Y^{-1}x + O(\|x\|^3)$ as $x \to 0$. Then the plug–in principle clearly holds under (A1) – (A3) and either $H_0(r)$ or $H_T(r)$. To guarantee power under (A4) and $H_1(r)$, we may require that for each $K_1 > 0$ and $K_2 > 0$, there exists a $K > 0$ such that $\|x\| > K_1$ and $\|Y\| \le K_2$ implies that $h(x, Y) > K$.[14] However, as far as the first–order asymptotics are concerned, there is no advantage to doing so and, when it comes to simplicity, the stochastic $F$ and $t$ statistics are clearly more advantageous.

Another strand of rank testing drops the condition that the statistics be asymptotically $F$ or $t$, as in the following example.

**Example 6** (Johansen (1991))**.** The maximum eigenvalue statistic for $H_0(r)$ is given by

$$J = T\mu_{r+1}^2(\widehat{B}).$$

From the properties of singular values, $J = \sigma_1^2(\sqrt{T}\widehat{N}'_r \widehat{\Sigma}^{-\frac{1}{2}} \widehat{B} \widehat{\Gamma}^{\frac{1}{2}} \widehat{M}_r)$, where $\widehat{N}_r$, and $\widehat{M}_r$ are as in example 5. It then follows that $J = \|\sqrt{T}P_{\widehat{N}_r} \widehat{\Sigma}^{-\frac{1}{2}} \widehat{B} \widehat{\Gamma}^{\frac{1}{2}} P_{\widehat{M}_r}\|_2^2$. The plug–in principle is seen to hold just as before from the $\sqrt{T}$–consistency of $P_{\widehat{N}_r}$ and $P_{\widehat{M}_r}$. Thus, under either $H_0(r)$ or $H_T(r)$, $J = \|\sqrt{T}P_{N_r} \widehat{\Sigma}^{-\frac{1}{2}} \widehat{B} \widehat{\Gamma}^{\frac{1}{2}} P_{M_r}\|_2^2 + o_p(1)$ and converges in distribution to the square of the largest singular value of a random matrix. Divergence under $H_1(r)$ follows the same argument used in Example 5. $\qquad \square$

This suggests statistics of the form $Th(\widehat{R}'_r\widehat{\beta}, \widehat{R}'_r\widehat{\Omega}\widehat{R}_r)$, where $h : \mathbb{R}^q \times \mathbb{G}^{q \times q} \to \mathbb{R}$ is continuous, homogenous of degree 2 in the first argument, and satisfies $h(Qx, Q'YQ) = h(x, Y)$ for all $Q$ within a suitable subgroup of $\mathbb{G}^{q \times q}$. We then have that the statistics depend on $\widehat{R}$ only through some maximal invariant (Lehmann & Romano, 2005, section 6.2). For stochastic $F$ statistics as well as example 6, a maximal invariant for the restriction matrix is $P_{\widehat{M}_r} \otimes P_{\widehat{N}_r}$. This implies that the limiting distribution of the statistic under $H_0(r)$ and $H_T(r)$ can be obtained under suitable restrictions on the rate of convergence of the maximal invariant. Power is guaranteed under the same conditions to the ones given above. Of course, the drawback to this approach is that the limiting distribution is usually nonstandard.

Clearly the aforementioned statistics are rather the exception than the rule in rank testing (see Table 1). In addition to the drawbacks listed above, they also lack the analytic and geometric appeal of stochastic $F$ and $t$ statistics. Formulating the theory to perfectly encompass Examples 5 and 6 (i.e. in terms of the $h$ functions above) would have certainly obscured an already quite abstract theory.

---

[14]This construction is more general than that of Robin & Smith (2000), who allow no role for $\widehat{\Omega}$.

# 7 Monte Carlo

This section illustrates the theory presented in previous sections through a series of PP plots, which plot significance level against nominal size. That is, $p$ on the vertical axis is plotted against $\alpha(p)$, the observed rejection rate at significance $p$. When the two are close under the null, the test has the correct size. When $\alpha(p) > p$, the test has a tendency to over–reject at test significance level $p$.[15]

## 7.1 Standard Asymptotics

**Example 7** (Stochastic $F$ Statistics)**.** Consider the following setting, $\{(x'_t, u'_t)'\}$ is a set of independent vectors in $\mathbb{R}^8$ with distribution $N(0, I_8)$. Let $\{\varepsilon_t\}$ be a stationary process satisfying $\varepsilon_t = 0.5\varepsilon_{t-1} + u_t$. Let $B = \begin{bmatrix} 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.75 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ and $y_t = Bx_t + \varepsilon_t$. Our sample consists of $\{(y'_t, x'_t)' : t = 1, \ldots, 50\}$. We estimate $B$ by OLS and the variance of $\widehat{B}$ nonparametrically using a Bartlett kernel with bandwidth $\lfloor 4(50/100)^{1/4} \rfloor = 3$ for the small–$b$ case and bandwidth 50 for the fixed–$b$ case. The number of replications is set to 2000.

Figure 4: PP Plots for the $\widehat{F}$–tests of Example 7
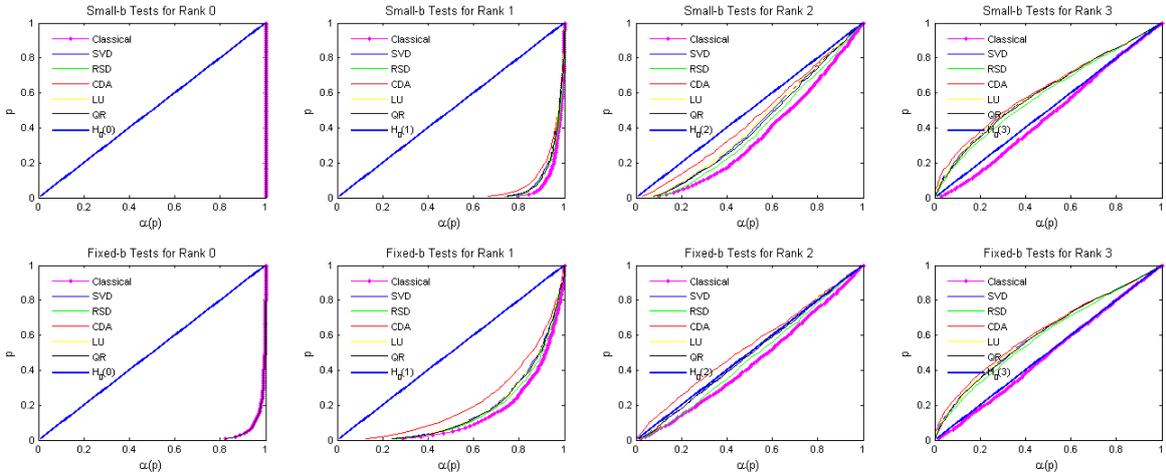


The third column of Figure 7 demonstrates that the rank–2 fixed–$b$ tests outperform the small–$b$ tests in terms of size. The first and second columns demonstrate the power of the test. The fourth column cannot be given an interpretation because we have no guidance in the theory of this paper about the behavior of our test statistics when the rank is overestimated. However, it is easy to prove that $\widehat{F}(2)$ stochastically dominates $\widehat{F}(3)$.[16]

---

[15]The Matlab code for generating these plots is available at `http://www.econ.upf.edu/~alsadoon/`.

[16]For more on the stochastic dominance relationships between rank tests, see Cragg & Donald (1993), Cragg & Donald (1997), and Donald et al. (2007).

**Example 8** (Discontinuity of RRAs). Suppose that instead, we have the population matrix
$B = \begin{bmatrix} 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. 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 stochastic and classical statistics for rank–1 will not necessarily diverge at the same rate. In fact, the stochastic statistics seem to diverge at heterogenous rates. This is illustrated in the second column of Figure 8.

Figure 5: PP Plots for the $\widehat{F}$–tests of Example 8
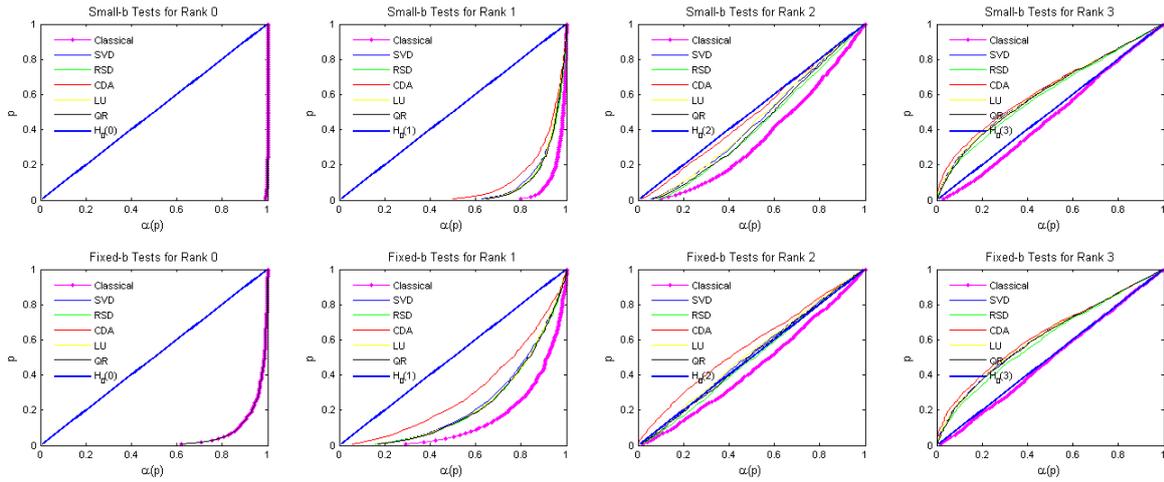


**Example 9** (Symmetric Matrices). Suppose now that we have the same setting above, with
$B = \begin{bmatrix} 0.75 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ and we estimate it by OLS subject to the restriction of symmetry. The $\widehat{F}^s$ tests behave similarly to $\widehat{F}$ (see Figure 9).

Figure 6: PP Plots for the $\widehat{F}^s$–tests of Example 9
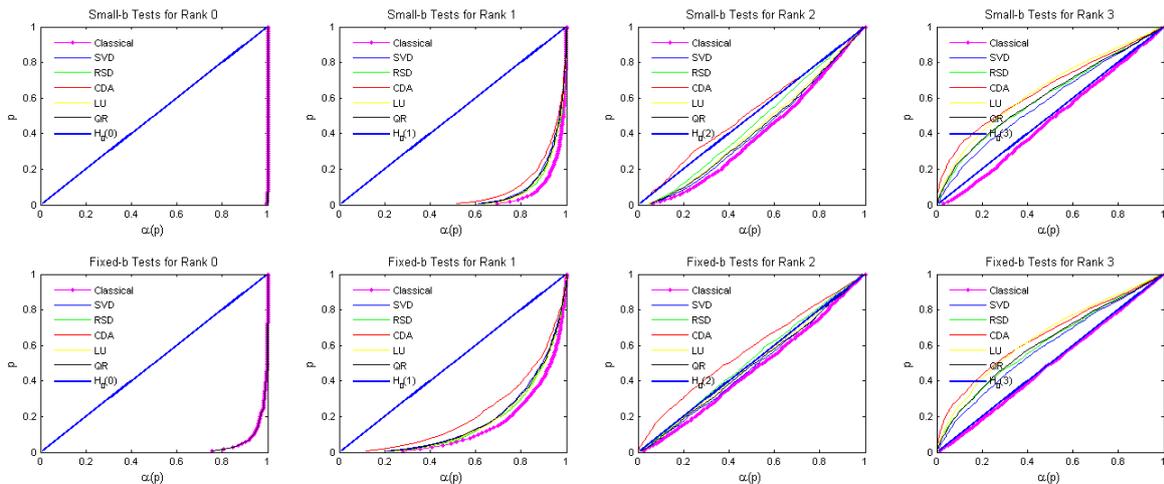


In this setting, we may also use the $\widehat{t}$ tests. The results are plotted in Figure 9.

Figure 7: PP Plots for the $\widehat{t}$–tests of Example 9
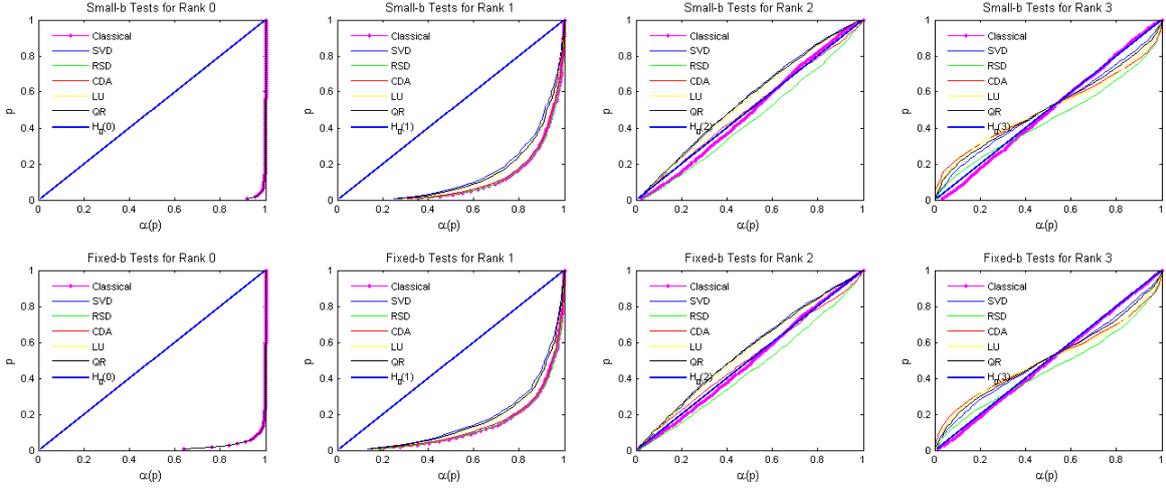


## 7.2 General Asymptotics

**Example 10** (The Cointegrated VAR Model)**.** Consider a submodel of the one given in Example 3 with $\Pi_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, $\Pi_2 = 0$, $y_{-1} = y_0 = 0$, and $T = 100$. This model generates $I(1)$ data. If we proceed as in Example 3, estimating the model by OLS, imposing the restriction $\Pi_2 = 0$, then $\widehat{F}(2)$ has the generalized Dickey Fuller distribution with 2 degrees of freedom, while $\widehat{F}(0)$ and $\widehat{F}(1)$ diverge to infinity (Johansen, 1991). Figure 8 confirms these results for any given stochastic test statistic.

Figure 8: PP Plots for the $\widehat{F}_c$–tests of Example 10
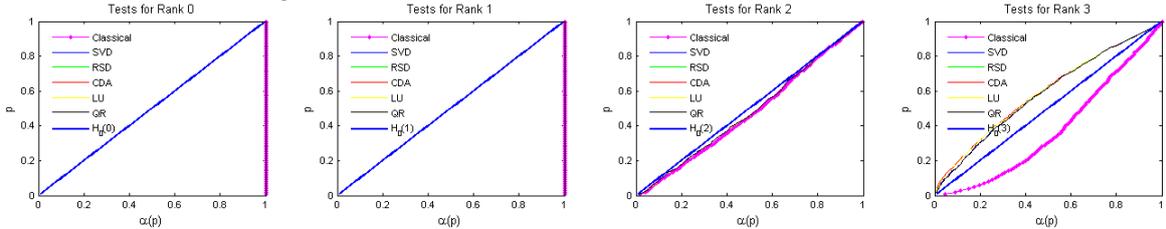


**Example 11** (Common Stochastic Trends)**.** Consider the model given in Example 4 with $\Sigma = I_4$, $\Gamma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, $x_0 = 0$, and $T = 100$. Since $\widehat{B}$ converges in distribution to a positive definite random matrix whose distribution is nuisance–parameter–free and $Z_T$ is bounded, we may formulate $\widehat{F}_c^s$ without normalization (i.e. $\widehat{\Omega} = I_{m(m+1)/2}$). The limiting distribution under $H_0(r)$ (or $H_T(r)$ with $D^T \to D$) can easily be found to be $\mathrm{tr}((C_{22} - C_{12}' C_{11}^{-1} C_{12})^2)$, with the same notation as in Proposition 1. Figure 11 plots the results for $\widehat{F}_c^s$. The classical statistics use $M_{rT}$ in the restriction matrix, while the naive statistics use $M_r$. Clearly, the naive statistic perform quite differently from all other statistics. The figure also displays

results for the Cholesky RRA, which can only be used because $\widehat{B}$ is positive definite. Figure 11 displays the results for $\widehat{t}_c$, to which similar observations apply.

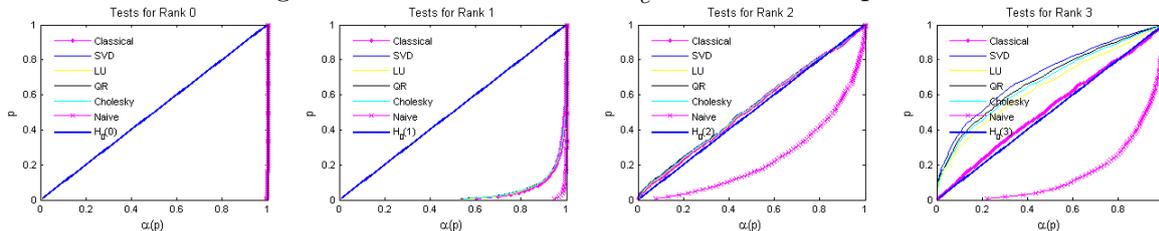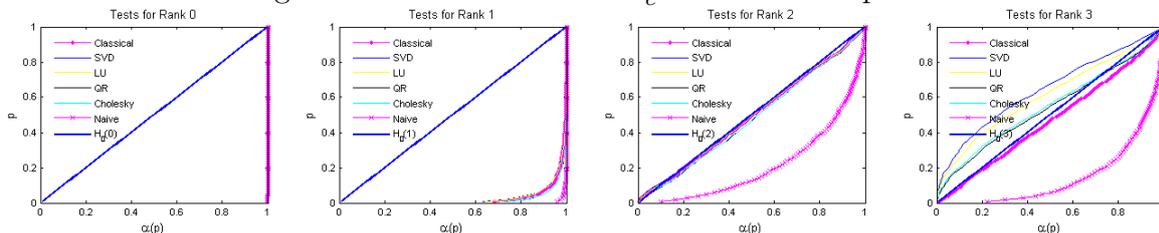Figure 9: PP Plots for the $\widehat{F}_c^s$–tests of Example 11



Figure 10: PP Plots for the $\widehat{t}_c^s$–tests of Example 11



# 8    Conclusion

As discussed in the introduction, rank testing has traditionally been a very difficult subject. This paper has aimed at developing a theory of the basic structure of these tests and to simplify their asymptotics. The main contributions have been discussed in the introduction. Here we briefly discuss a couple of possible extensions.

First, the natural next step is to utilize the theory to develop rank estimation methods. Preliminary results suggest that estimated null spaces are either exactly or asymptotically nested, imparting stochastic dominance relationships between stochastic statistics, which generalized those found in Cragg & Donald (1993, 1997) and Donald et al. (2007). These may then be used to construct information criteria useful for rank estimation.

Second, rank reduction in high–dimensional data has received a great deal of attention from researchers and practitioners (Bai & Ng, 2008; Onatski, 2009). Although this paper has discussed only the case of fixed dimensions, all of the asymptotics continue to hold if the dimensions grow slowly enough with $T$. However, it would be far more useful to develop tests that place no conditions on the relative rates of growth of $T$, $n$, and $m$. In particular, it would be interesting to see how subspace estimation can be conducted in high dimensions and how that can be translated into practical tests of rank. Alas, this must be left for future research.

# 9 Appendix

## 9.1 Reduced Rank Approximations

In this section, we develop the properties of some of the most popular RRAs in the literature. We assume throughout that $\widehat{B} \in \mathbb{G}^{n \times m}$, $B \in \mathbb{R}^{n \times m}$, and $\widehat{\Theta} \in \mathbb{G}^{nm \times nm}$.

*The Singular Value Decomposition.* The most important reduced rank approximation is the singular value decomposition (SVD) approximation.[17] 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

$$\widehat{B} = \begin{bmatrix} \widehat{U}_{.1} & \widehat{U}_{.2} \end{bmatrix} \begin{bmatrix} \widehat{S}_1 & 0 \\ 0 & \widehat{S}_2 \end{bmatrix} \begin{bmatrix} \widehat{V}'_{.1} \\ \widehat{V}'_{.2} \end{bmatrix} = \widehat{U}_{.1}\widehat{S}_1\widehat{V}'_{.1} + \widehat{U}_{.2}\widehat{S}_2\widehat{V}'_{.2}, \tag{12}$$

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

$$\sum_{i=r+1}^{\min\{n,m\}} \sigma_i^2(\widehat{B}) = \|\widehat{B} - \widehat{B}_r^{SVD}\|^2 \leq \|\widehat{B} - B\|^2 \tag{13}$$

whenever $\text{rank}(B) = r$ (Horn & Johnson, 1985, example 7.4.1). $\widehat{B}_r^{SVD}$ is unique if and only if $\sigma_r(\widehat{B}) \neq \sigma_{r+1}(\widehat{B})$ (see theorem 2.23 of Markovsky (2012)). Finally, the above suggests the null space estimators $\widehat{N}_r = \widehat{U}_{.2}$ and $\widehat{M}_r = \widehat{V}_{.2}$. This implies that $\widehat{N}'_r\widehat{B}\widehat{M}_r = \widehat{S}_2$.

*The Robin–Smith Decomposition.* The RSD of $\widehat{B}$ takes symmetric positive definite matrices $\widehat{\Sigma} \in \mathbb{G}^{n \times n}$ and $\widehat{\Gamma} \in \mathbb{G}^{m \times m}$ and obtains $\widehat{B} = \widehat{U}\widehat{S}\widehat{V}'$, where $\widehat{U} \in \mathbb{R}^{n \times n}$, $\widehat{V} \in \mathbb{R}^{m \times m}$, and $\widehat{S}$ satisfy:

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

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

(iii) $\widehat{S}$ is diagonal with diagonal entries, $\mu_1(\widehat{B}) \geq \mu_2(\widehat{B}) \geq \cdots \geq \mu_{\min\{n,m\}}(\widehat{B}) > 0$.[18]

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

---

[17]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).

[18]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).

Now just as we did in (12), write $\widehat{B} = \widehat{U}_{.1}\widehat{S}_1\widehat{V}_{.1}' + \widehat{U}_{.2}\widehat{S}_2\widehat{V}_{.2}'$, where $\widehat{S}_1 \in \mathbb{R}^{r \times r}$ and set $\widehat{B}_r^{RSD} = \widehat{U}_{.1}\widehat{S}_1\widehat{V}_{.1}'$. We may also write $\widehat{B}_r^{RSD} = \widehat{\Sigma}^{1/2}(\widehat{\Sigma}^{-1/2}\widehat{B}\widehat{\Gamma}^{-1/2})_r^{SVD}\widehat{\Gamma}^{1/2}$, which minimizes $\|\widehat{B} - A\|_{\widehat{\Gamma} \otimes \widehat{\Sigma}} = \|\widehat{\Sigma}^{-1/2}(\widehat{B} - A)\widehat{\Gamma}^{-1/2}\|$ 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 $\widehat{N}_r'\widehat{B}\widehat{M}_r = \widehat{S}_2$.

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 $\widehat{\Sigma}$ and $\widehat{\Gamma}$, 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 of $B$ in the regression equation $y = Bx + \varepsilon$, while $\widehat{\Sigma}$ is the OLS estimator of the variance of $\varepsilon$ and $\widehat{\Gamma}^{-1}$ is the sample second moment of $x$ (Reinsel & Velu, 1998; Anderson, 2003; Reinsel, 2003).

*The Cragg–Donald Approximation.* The RRA implicit in Cragg & Donald (1997) takes a positive definite $\widehat{\Theta}$ and obtains

$$\widehat{B}_r^{CDA} \in \operatorname{argmin}\{\|\widehat{B} - A\|_{\widehat{\Theta}} : \operatorname{rank}(A) \leq r\}.$$

Existence of $\widehat{B}_r^{CDA}$ is proven in Cragg & Donald (1997). There are no known closed–form solutions when $\widehat{\Theta}$ is not of Kronecker form. We have already shown how the SVD and RSD approximations are special cases of the CDA. Next, we provide general results that apply to all CDAs.

**Lemma 5.** Let $\widehat{B} \in \mathbb{G}^{n \times m}$, $\widehat{\Theta} \in \mathbb{R}^{nm \times nm}$ be symmetric and positive definite, 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 $i$, $\operatorname{rank}(\widehat{B}_i^{CDA}) = i$.

(ii) For all $i$,

$$\|\widehat{B} - \widehat{B}_i^{CDA}\|_{\widehat{\Theta}}^2 = \operatorname{vec}'(\widehat{B})(\widehat{M}_i \otimes \widehat{N}_i)\{(\widehat{M}_i \otimes \widehat{N}_i)'\widehat{\Theta}(\widehat{M}_i \otimes \widehat{N}_i)\}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)'\operatorname{vec}(\widehat{B}). \quad (14)$$

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

*Proof of Lemma 5.* (i) Suppose $\operatorname{rank}(\widehat{B}_i^{CDA}) < i$. Then for arbitrary $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and $h \in \mathbb{R}$, $\operatorname{rank}(\widehat{B}_i^{CDA} + hxy') \leq i$ (Horn & Johnson, 1985, section 0.4.5). By the definition of $\widehat{B}_i^{CDA}$,

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

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

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

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

(ii) For $h \in \mathbb{R}^{r^2}$ the matrix, $\widehat{B}_i^{CDA} + \mathrm{vec}^{-1}((\widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp})h)$ has a rank of at most $i$. From the definition of the CDA, we know that

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

and it follows just as in (i) that

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

By the same logic we can show that

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

Since $(\widehat{M}_i \otimes \widehat{N}_i)_\perp = \left[\ \widehat{M}_{i\perp} \otimes \widehat{N}_{i\perp} \quad \widehat{M}_i \otimes \widehat{N}_{i\perp} \quad \widehat{M}_{i\perp} \otimes \widehat{N}_i\ \right]$, we can combine the three equations above to arrive at,

$$
\mathrm{vec}'(\widehat{B} - \widehat{B}_i^{CDA})\widehat{\Theta}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp = 0. \tag{15}
$$

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

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

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 $\mathrm{vec}(\widehat{B}_i) = \widehat{P}_i\mathrm{vec}(\widehat{B})$, where

$$
\widehat{P}_i = (\widehat{M}_i \otimes \widehat{N}_i)_\perp\{(\widehat{M}_i \otimes \widehat{N}_i)'_\perp\widehat{\Theta}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp\}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)'_\perp\widehat{\Theta}^{-1}. \tag{16}
$$

Now using the well known identity

$$
I_n = G^{\frac{1}{2}}H(H'GH)^{-1}H'G^{\frac{1}{2}} + G^{-\frac{1}{2}}H_\perp(H'_\perp G^{-1}H_\perp)^{-1}H'_\perp G^{-\frac{1}{2}}, \tag{17}
$$

for symmetric and positive definite $G \in \mathbb{G}^{n\times n}$ and $H \in \mathbb{G}^{n\times m}$, we have that

$$
\begin{aligned}
\widehat{\Theta}^{-1} &= \widehat{\Theta}^{-\frac{1}{2}}I_{nm}\widehat{\Theta}^{-\frac{1}{2}} \\
&= (\widehat{M}_i \otimes \widehat{N}_i)\{(\widehat{M}_i \otimes \widehat{N}_i)'\widehat{\Theta}(\widehat{M}_i \otimes \widehat{N}_i)\}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)' \\
&\quad + \widehat{\Theta}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp\{(\widehat{M}_i \otimes \widehat{N}_i)'_\perp\widehat{\Theta}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)_\perp\}^{-1}(\widehat{M}_i \otimes \widehat{N}_i)'_\perp\widehat{\Theta}^{-1}. \tag{18}
\end{aligned}
$$

Substituting (15) and (18) into the Cragg and Donald statistic proves (14).

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

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

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

Lemma 5 (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 $\widehat{B}_i^{CDA}$. Lemma 5 (ii) tells us that the Cragg & Donald (1997) statistic given as the left hand side of (14) is a stochastic $F$ statistic. Markovsky & Van Huffel (2007) find a similar representation. Finally, Lemma 5 (iii) proves that the CDA approximates $\widehat{B}$, and therefore $B$, quite well, so long as the rank is overestimated.

In cointegration analysis, $\widehat{B}$ converges towards a rank–$r$ matrix at different rates along different linear combinations of the rows and/or columns and these linear combinations are potentially nonconstant (see Examples 3 and 4). If $\widehat{\Theta}$ matches these speeds of convergence, then the CDA continues to produce good estimates of the null spaces.

**Lemma 6.** Let $\widehat{B}$, $\widehat{\Theta}$, $\widehat{B}_i^{CDA}$, $\widehat{N}_i$, and $\widehat{M}_i$ be as in Lemma 5. Let the sequences of matrices $N_{rT} \in \mathbb{G}^{n \times (n-r)}$ and $M_{rT} \in \mathbb{G}^{m \times (m-r)}$ be such that $[\ N_{r\perp}\ \ N_{rT}\ ]$ and $[\ M_{r\perp}\ \ M_{rT}\ ]$ have singular values bounded away from zero. Let $\widehat{B}_T = [\ N_{r\perp}\ \ N_{rT}\ ]'\widehat{B}[\ M_{r\perp}\ \ M_{rT}\ ]$ and $\widehat{B}_{iT}^{CDA} = [\ N_{r\perp}\ \ N_{rT}\ ]'\widehat{B}_i^{CDA}[\ M_{r\perp}\ \ M_{rT}\ ]$ and denote its null space estimators by $\widehat{N}_{iT} = [\ N_{i\perp}\ \ N_{iT}\ ]^{-1}\widehat{N}_i$ and $\widehat{M}_{iT} = [\ M_{i\perp}\ \ M_{iT}\ ]^{-1}\widehat{M}_i$. Let $\widehat{\Theta}_T = Z_T'\widehat{\Theta}Z_T$, where $Z_T = [\ M_{r\perp}\ \ M_{rT}\ ]\otimes[\ N_{r\perp}\ \ N_{rT}\ ]$. Finally, define $\widehat{B}^*$ and $\widehat{B}_T^*$ by $[\ N_{r\perp}\ \ N_{rT}\ ]'\widehat{B}^*[\ M_{r\perp}\ \ M_{rT}\ ] = \begin{bmatrix} N_{r\perp}'\widehat{B}M_{r\perp} & 0 \\ 0 & 0 \end{bmatrix} = \widehat{B}_T^*$.

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

*Proof of Lemma 6.* Just as in Lemma 5 (iii),

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

Since $\widehat{B}^*$ has a rank of at most $r$,

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

where the last equality follows from the definition of $\widehat{B}_T$ and $\widehat{B}_T^*$. □

An often cited criticism of the CDA in the econometrics literature is that it is difficult to compute for general forms of $\widehat{\Theta}$. 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 5 (ii) (see equation (16)) we find that

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

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 $\widehat{B}_r^{CDA}$ respectively. We may therefore iterate this equation as outlined in the follow 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

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

(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. $\qquad\square$

*The LU Decomposition with Complete Pivoting.* This decomposition arises from Gaussian elimination in linear system and is used in Cragg & Donald (1996) to construct a stochastic $F$ test. 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 $\widehat{S}_{11} \in \mathbb{G}^{r \times r}$.[19] It remains to show that $\sigma_r(\widehat{S}_{11}) \geq K_1 \sigma_r(\widehat{B})$ and $\sigma_1(\widehat{S}_{22}) \leq K_2 \sigma_{r+1}(\widehat{B})$ as $\widehat{B} \to B$ and $\text{rank}(B) = r$.[20]

Let $\widehat{S}^{(r)}$ be the result of the $r$–th permutation and Gaussian elimination of the algorithm. 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, $|\widehat{S}_{(2,2)}| = |\widehat{S}^{(2)}_{(2,2)}| = \max_{i>1}|\widehat{S}^{(1)}_{(i,j)}| \geq \sigma_1(\widehat{S}^{(1)}_{(2:n,1:m)})/\sqrt{(n-1)(m-1)} \geq \sigma_1(\widehat{S}^{(1)}_{(2:n,1:m)})/\sqrt{nm} \geq$

---

[19]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).

[20]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 derived here because they cannot be found elsewhere.

$\sigma_2(\widehat{S}^{(1)})/\sqrt{nm}$, where the last inequality follows from Corollary 3.1.3 of Horn & Johnson (1991). Since the smallest singular value of the $r$–th step Gaussian elimination matrix is bounded below by $(1+n)^{-1}$,[21] 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 a slightly different LU algorithm designed for symmetric matrices. Robin & Smith (1995) and Kleibergen & van Dijk (1994) utilize algorithms based on Gaussian elimination and can therefore be thought of as LU–based algorithms. Other LU–based RRAs can be found in Hansen (1998). Each of these algorithms satisfies the conditions for a DBA using similar arguments to those used above. We mention finally that the LU algorithm with no pivoting and the LU algorithm with partial pivoting are not rank–revealing. For example, if $\widehat{B} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, 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 and to leave the bottom block empty. That is, they fail to satisfy the bounds that define a DBA.

*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 stochastic 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 $\begin{bmatrix} \widehat{S}_{11} & \widehat{S}_{12} \\ 0 & \widehat{S}_{22} \end{bmatrix}$, with $\widehat{S}_{11} \in \mathbb{R}^{r \times r}$,

---

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

then $\sigma_r(\widehat{S}_{11}) \geq (r\max\{n,m\})^{-1/2}2^{-r}\sigma_r(\widehat{B})$.[22] The second inequality can be derived exactly as 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.[23]

*Eigenvalue Decomposition.* If a square matrix is rank deficient, then it has an eigenvalue at zero. By the continuity of the eigenvalues (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 eigenvalues. An $n \times n$ matrix can have all its eigenvalues at zero but have rank–$(n-1)$ (e.g. the Jordan canonical nilpotent matrix). Eigenvalues can also be "too sensitive" to the parameters of the matrix (Horn & Johnson, 1985, problem 21 of section 7.3) limiting their appeal as detectors rank deficiency. Finally, eigenvalues can be complex–valued, necessitating a complex–valued asymptotic theory that may be too technical to operationalize. Therefore, eigenvalues are not recommended for rank detection in general.

The exception to this is symmetric matrices, which have real eigenvalues and eigenvectors (Horn & Johnson, 1985, Theorem 2.5.6) and are well–conditioned with respect to eigenvalue computations 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 $|\lambda_1(\widehat{B})|,\dots,|\lambda_m(\widehat{B})|$ (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.

If $\widehat{B}$ converges to a positive semi–definite matrix, we may utilize the spectral decomposition instead. Here $\widehat{B} = \widehat{V}\widehat{S}\widehat{V}'$ has $\lambda_1(\widehat{B}),\dots,\lambda_m(\widehat{B})$ along the diagonal of $\widehat{S}$. We may then estimate the null space by collecting the eigenvectors associated with $\lambda_{m-r+1}(\widehat{B}),\dots,\lambda_m(\widehat{B})$ (this is the approach taken in Donald et al. (2007)).

*Cholesky Decomposition.* If $\widehat{B}$ is symmetric and positive definite, 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 $\begin{bmatrix} \widetilde{S}_{11} & \widetilde{S}_{12} \\ 0 & \widetilde{S}_{22} \end{bmatrix}$, 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 Cholesky RRA of $\widehat{B}$ is then given by $\widehat{S} = \mathrm{diag}^2(\widetilde{S})$ and $\widehat{U} = \widehat{V}' = \mathrm{diag}^{-1}(\widetilde{S})\widetilde{S}\widetilde{V}'$. The

---

[22]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 $\begin{bmatrix} \widehat{B} & 0 \end{bmatrix}$ when $n > m$ and $\begin{bmatrix} \widehat{B} \\ 0 \end{bmatrix}$ 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 $\widehat{B}$.

[23]The caveat is that this QR algorithm, along with the LU algorithm with complete pivoting, may fail to detect ill conditioning when the dimension of $\widehat{B}$ is unbounded. However, this is not relevant to our exposition as $n$ and $m$ are always assumed fixed. Hwang et al. (1992), Pan (2000), and Chandrasekaran & Ipsen (1994) discuss modifications to the pivoting strategy that overcome this problem.

fact that this decomposition satisfies the conditions for an RRA follows from the results of Higham (1987).

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

**Lemma 7.** Let $\widehat{B} \in \mathbb{G}^{n \times m}$ and let $\widehat{N}_i$ and $\widehat{M}_i$ span the left and right null spaces of $\widehat{B}_i^{DBA}$ respectively.

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

(ii) $\|\widehat{B} - \widehat{B}_i^{DBA}\| = O(\|\widehat{B} - B\|)$ as $\widehat{B} - B \to 0$ for all $i \geq \mathrm{rank}(B)$.

*Proof of Lemma 7.* (i) $\mathrm{rank}(\widehat{B}_i^{DBA}) > i$ 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 $\mathrm{rank}(\widehat{B}_i^{DBA}) < i$, then $\mathrm{rank}([\begin{array}{cc} \widehat{S}_{11} & \widehat{S}_{12} \end{array}]) < i$, which is impossible since $\widehat{S}_{11}$ has full rank by construction.

(ii) It follows from Definition 2 that $\|\widehat{B} - \widehat{B}_i^{DBA}\| \leq \|\widehat{U}\|\|\widehat{V}\|\|\widehat{S}_{22}\| = O(\sigma_{r+1}(\widehat{B}))$ as $\widehat{B} \to B$ and $\mathrm{rank}(B) = r$. The result follows from the fact that $\sigma_{r+1}(\widehat{B}) = O(\|\widehat{B} - B\|)$ as $\widehat{B} \to B$. $\square$

As before, Lemma 7 (i) implies that the null space estimators obtained from DBAs are well defined, while (ii) implies that DBAs are good approximations for rank deficient matrices when the approximating rank is high enough.

Finally, the extension for cointegration is given in the following lemma.

**Lemma 8.** Let $\widehat{B}$, $\widehat{B}_i^{DBA}$, $\widehat{N}_i$, and $\widehat{M}_i$ be as in Lemma 7. Let $N_{rT}$, $M_{rT}$, $\widehat{B}_T$, $\widehat{B}_T^*$, $\widehat{N}_{iT}$, $\widehat{M}_{iT}$, and $Z_T$ be as in Lemma 6 and define $\widehat{B}_{iT}^{DBA}$ analogously to $\widehat{B}_{iT}^{CDA}$.

Then $\|\widehat{B} - \widehat{B}_i^{DBA}\| = O(\|\widehat{B}_T - \widehat{B}_T^*\|)$ as $\widehat{B}_T - \widehat{B}_T^* \to 0$ for all $i \geq \mathrm{rank}(B)$.

*Proof of Lemma 8.* By construction, $\|\widehat{B} - \widehat{B}_r^{DBA}\| = O(\sigma_{r+1}(\widehat{B}))$ as $\widehat{B} - \widehat{B}^* \to 0$. Since $\|\widehat{B} - \widehat{B}^*\| \leq \|Z_T^{-1}\|\|\widehat{B}_T - \widehat{B}_T^*\|$ it follows that $\|\widehat{B} - \widehat{B}_r^{DBA}\| = O(\sigma_{r+1}(\widehat{B}))$ as $\widehat{B}_T - \widehat{B}_T^* \to 0$.

$$\sigma_{r+1}(\widehat{B}) = \sigma_{r+1}([\begin{array}{cc} N_{r\perp} & N_{rT} \end{array}]^{-1} \widehat{B}_T [\begin{array}{cc} M_{r\perp} & M_{rT} \end{array}]^{-1})$$
$$\leq \|[\begin{array}{cc} N_{r\perp} & N_{rT} \end{array}]^{-1}\|_2 \|[\begin{array}{cc} M_{r\perp} & M_{rT} \end{array}]^{-1}\|_2 \sigma_{r+1}(\widehat{B}_T),$$

which is $O(\|\widehat{B}_T - \widehat{B}_T^*\|)$ since $\sigma_{r+1}(\widehat{B}_T^*) = 0$. $\square$

## 9.2 Proofs

*Proof of Lemma 1.* (i) Under either $H_0$ or $H_T$, (A1) and (A2) imply that

$$F = T\widehat{\beta}' R\{\widehat{R}'\widehat{\Omega}\widehat{R}\}^{-1} R'\widehat{\beta} + o_p(1) \qquad \text{and} \qquad t = \frac{\sqrt{T}R'\widehat{\beta}}{\sqrt{\widehat{R}'\widehat{\Omega}\widehat{R}}} + o_p(1),$$

because $\sqrt{T}R'\widehat{\beta} = O_p(1)$ and $(\widehat{R}'\widehat{\Omega}\widehat{R})^{-1} - (R'\widehat{\Omega}R)^{-1} = o_p(1)$. The second equality is proven as follows. From Wedin's Theorem (Stewart & Sun, 1990, theorem III.3.9),

$$\|(\widehat{R}'\widehat{\Omega}\widehat{R})^{-1} - (R'\widehat{\Omega}R)^{-1}\|_2 \leq \|(\widehat{R}'\widehat{\Omega}\widehat{R})^{-1}\|_2\|(R'\widehat{\Omega}R)^{-1}\|_2\|\widehat{R}'\widehat{\Omega}\widehat{R} - R'\widehat{\Omega}R\|_2.$$

Using the fact that $\|(\widehat{R}'\widehat{\Omega}\widehat{R})^{-1}\|_2 \leq \|(R'\widehat{\Omega}R)^{-1}\|_2 + \|(\widehat{R}'\widehat{\Omega}\widehat{R})^{-1} - (R'\widehat{\Omega}R)^{-1}\|_2$ and rearranging,

$$\|(\widehat{R}'\widehat{\Omega}\widehat{R})^{-1} - (R'\widehat{\Omega}R)^{-1}\|_2 \leq \frac{\|(R'\widehat{\Omega}R)^{-1}\|_2^2 \|\widehat{R}'\widehat{\Omega}\widehat{R} - R'\widehat{\Omega}R\|_2}{1 - \|(R'\widehat{\Omega}R)^{-1}\|_2 \|\widehat{R}'\widehat{\Omega}\widehat{R} - R'\widehat{\Omega}R\|_2}.$$

Since $\widehat{\Omega}$ converges in distribution to an almost surely positive definite matrix, $\|(R'\widehat{\Omega}R)^{-1}\|_2 = \sigma_q^{-1}(R'\widehat{\Omega}R) = O_p(1)$ and the result follows from (A2).

The final substitution of $\widehat{R}$ for $R$ follows from the fact that

$$\sqrt{T}(\widehat{R} - R)'\widehat{\beta} = \sqrt{T}(\widehat{R} - R)'\beta + \sqrt{T}(\widehat{R} - R)'(\widehat{\beta} - \beta) = \sqrt{T}\widehat{R}'\beta + (\widehat{R} - R)'\sqrt{T}(\widehat{\beta} - \beta).$$

(A3) implies the first term is $o_p(1)$. (A1) and (A2) imply the second term is $o_p(1)$.

(ii) The proof is exactly analogous to the proof of (i) and is omitted.

(iii) Theorem 4.2.2 of Horn & Johnson (1985) implies that $T^{-1}\widehat{F} \geq \lambda_q\left((\widehat{R}\widehat{\Omega}\widehat{R})^{-1}\right)\|\widehat{R}'\widehat{\beta}\|^2 = \frac{\|\widehat{R}'\widehat{\beta}\|^2}{\lambda_1(\widehat{R}'\widehat{\Omega}\widehat{R})}$. Assumption (A4) implies that the numerator is $O_p^{-1}(1)$, while (A1) and (A4) imply that the denominator is $O_p(1)$. Therefore $T^{-1}\widehat{F} = O_p^{-1}(1)$ and $\widehat{F} \xrightarrow{p} \infty$. Similarly, for $q = 1$, the numerator of $T^{-1/2}|\widehat{t}|$ is $O_p^{-1}(1)$, while its denominator is $O_p(1)$ so that $T^{-1/2}|\widehat{t}|$ is $O_p^{-1}(1)$ and $\widehat{t}$ is unbounded. $\qquad\square$

*Proof of Lemma 2.* The key to proving this lemma is to note that $F$ and $\widehat{F}$ can be written as $T\widehat{\beta}'P_R(P_R\widehat{\Omega}P_R)^\dagger P_R\widehat{\beta}$ and $T\widehat{\beta}'P_{\widehat{R}}(P_{\widehat{R}}\widehat{\Omega}P_{\widehat{R}})^\dagger P_{\widehat{R}}\widehat{\beta}$ respectively, while $|t|$ and $|\widehat{t}|$ can be written as $\left(T\widehat{\beta}'P_R(P_R\widehat{\Omega}P_R)^\dagger P_R\widehat{\beta}\right)^{1/2}$ and $\left(T\widehat{\beta}'P_{\widehat{R}}(P_{\widehat{R}}\widehat{\Omega}P_{\widehat{R}})^\dagger P_{\widehat{R}}\widehat{\beta}\right)^{1/2}$ respectively. This follows from the fact that for any $H \in \mathbb{G}^{p \times q}$ and any symmetric positive definite $G \in \mathbb{G}^{p \times p}$, $H(H'GH)^{-1}H$ satisfies the four Penrose conditions that characterize the generalized inverse of $P_H G P_H$ (Stewart & Sun, 1990, p. 102).

(i) Since $\widehat{\Omega}$ is almost surely of full rank, $\text{rank}(P_{\widehat{R}}\widehat{\Omega}P_{\widehat{R}}) = \text{rank}(P_R\widehat{\Omega}P_R) = q$. It follows from Wedin's Theorem again (Stewart & Sun, 1990, theorem III.3.9) and the same manipulations used in the proof of Lemma 1 (i) that

$$\|(P_{\widehat{R}}\widehat{\Omega}P_{\widehat{R}})^\dagger - (P_R\widehat{\Omega}P_R)^\dagger\|_2 \leq \frac{\|(P_R\widehat{\Omega}P_R)^\dagger\|_2^2 \|P_{\widehat{R}}\widehat{\Omega}P_{\widehat{R}} - P_R\widehat{\Omega}P_R\|_2}{1 - \|(P_R\widehat{\Omega}P_R)^\dagger\|_2 \|P_{\widehat{R}}\widehat{\Omega}P_{\widehat{R}} - P_R\widehat{\Omega}P_R\|_2}.$$

Now since $\text{rank}(P_R\widehat{\Omega}P_R) = q$, $\|(P_R\widehat{\Omega}P_R)^\dagger\|_2 = \sigma_q^{-1}(P_R\widehat{\Omega}P_R)$, which is bounded in probability as $\widehat{\Omega}$ converges in distribution to an almost surely positive definite matrix. Therefore, $(P_{\widehat{R}}\widehat{\Omega}P_{\widehat{R}})^\dagger - (P_R\widehat{\Omega}P_R)^\dagger = o_p(1)$ by (A2)*. On the other hand, $\sqrt{T}(P_{\widehat{R}} - P_R)\widehat{\beta} = \sqrt{T}P_{\widehat{R}}\beta + \sqrt{T}(P_{\widehat{R}} - P_R)(\widehat{\beta} - \beta) = o_p(1)$ by (A1), (A2)*, and (A3)*.

(ii) The proof is exactly analogous to the proof of (i) and is omitted.

(iii) The invariance allows us to normalize $\widehat{R}$ to have orthonormal columns. Then $\|\widehat{R}'\widehat{\beta}\| = \|P_{\widehat{R}}\widehat{\beta}\| = \|\widehat{\beta}\|\|P_{\widehat{R}}P_{\widehat{\beta}}\|$. By the same argument used in Lemma 1 (iii), $T^{-1}\widehat{F} \geq \frac{\|\widehat{R}'\widehat{\beta}\|^2}{\lambda_1(\widehat{R}'\widehat{\Omega}\widehat{R})} \geq \frac{\|\widehat{\beta}\|^2\|P_{\widehat{R}}P_{\widehat{\beta}}\|^2}{\lambda_1(\widehat{\Omega})} = O_p^{-1}(1)$. A similar argument holds for $|\widehat{t}|$. $\qquad\square$

*Proof of Lemma 3.* (i) The rate of convergence of $\widehat{B}_r$ is proven in Lemma 5 (iii) and Lemma 7 (ii). The rate of convergence of the subspace estimators follows from the inequality

$$\max\{\|P_{\widehat{N}_r} - P_{N_r}\|, \|P_{\widehat{M}_r} - P_{M_r}\|\} \leq K\|\widehat{B}_r - B\|, \tag{19}$$

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

(ii) The result follows from the inequality above applied to $B_l$.

(iii) We will need the following lemma.

**Lemma 9.** Let $\widehat{B}$ be an estimator of $B \in \mathbb{R}^{n \times m}$ such that $\widehat{B} \in \mathbb{G}^{n \times m}$, rank$(B) = r$, and $\sqrt{T}(\widehat{B} - B) = O_p(1)$. Let the RRAs $\{\widehat{B}_i : 0 \leq i < \min\{n, m\}\}$ be either CDAs or DBAs. In the former case, we assume that cond$(\widehat{\Theta}) = O_p(1)$. Then there exists an oblique projection matrix $\widehat{P}_i$ such that:

(i) $\text{vec}(\widehat{B}_i) = \widehat{P}_i \text{vec}(\widehat{B})$.

(ii) $(I_{nm} - \widehat{P}_i)P_{\widehat{M}_i \otimes \widehat{N}_i} = I_{nm} - \widehat{P}_i$

(iii) For $i \leq r$, $\widehat{P}_i = O_p(1)$.

*Proof of Lemma 9.* First, we consider the CDA. (i) is derived in the course of proving (ii) of Lemma 5 (see (16)). From (i) and (17), we find that

$$I_{nm} - \widehat{P}_i = \widehat{\Theta}(\widehat{M}_i \otimes \widehat{N}_i)((\widehat{M}_i \otimes \widehat{N}_i)'\widehat{\Theta}(\widehat{M}_i \otimes \widehat{N}_i))^{-1}(\widehat{M}_i \otimes \widehat{N}_i)', \tag{20}$$

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

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

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

Next, we consider the DBA. Let $\widehat{U} = [\begin{array}{cc} \widehat{U}_{\cdot 1} & \widehat{U}_{\cdot 2} \end{array}]$ and $\widehat{S} = \left[\begin{smallmatrix} \widehat{S}_{1\cdot} \\ \widehat{S}_{2\cdot} \end{smallmatrix}\right]$ 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 the third identity 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 $\widehat{S}_{1\cdot\perp}' = \left[\begin{smallmatrix} \widehat{S}_{11}^{-1}\widehat{S}_{12} \\ -I_{m-i} \end{smallmatrix}\right]$ so that $\widehat{W}_i = \widehat{V}'^{-1}\left[\begin{smallmatrix} 0 & \widehat{S}_{11}^{-1}\widehat{S}_{12}\widehat{S}_{22}^\dagger\widehat{S}_{22} \\ 0 & -\widehat{S}_{22}^\dagger\widehat{S}_{22} \end{smallmatrix}\right]\widehat{V}'$. $\widehat{V}$ 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 in probability. $\qquad\square$

38

Lemma 9 states that $\text{vec}(\widehat{B}_i)$ is obtained from $\text{vec}(\widehat{B})$ by an oblique projection and the two can only be different if they differ along $\widehat{M}_i \otimes \widehat{N}_i$. That is, the RRA is obtained by removing from $\text{vec}(\widehat{B})$ components along the estimated null spaces. The oblique projection is bounded in probability if the rank is underestimated and the condition number of $\widehat{\Theta}$ is bounded in probability.

Now note that $\|\widehat{B}_l - \widehat{B}\|^2 \geq \sigma_{l+1}^2(\widehat{B}) + \cdots + \sigma_{\min\{n,m\}}^2(\widehat{B})$ by (Horn & Johnson, 1985, example 7.4.1). If $l < r$, the right hand side converges in probability to $\sigma_{l+1}^2(B) + \cdots + \sigma_r^2(B) > 0$ and so $\|\widehat{B}_l - \widehat{B}\|^2$ is $O_p^{-1}(1)$. Next, $\|\widehat{B}_l - \widehat{B}\|^2 = \|\text{vec}(\widehat{B}_l - \widehat{B})\|^2 = \|(I_{nm} - \widehat{P}_l)\text{vec}\widehat{B}\|^2$. By the properties of $\widehat{P}_l$ and Cauchy–Schwartz inequality, $\|(I_{nm} - \widehat{P}_l)\text{vec}(\widehat{B})\| = \|(I_{nm} - \widehat{P}_l)(P_{\widehat{M}_l} \otimes P_{\widehat{N}_l})\text{vec}(\widehat{B})\| \leq \|I_{nm} - \widehat{P}_l\| \|P_{\widehat{N}_l} \widehat{B} P_{\widehat{M}_l}\|$. Since $\widehat{P}_l = O_p(1)$, it follows that $P_{\widehat{N}_l} \widehat{B} P_{\widehat{M}_l}$ is $O_p^{-1}(1)$. $\qquad\square$

*Proof of Theorem 1.* (i) The result follows from Lemma 2 (i) if we can show that assumptions $(A2)^*$ and $(A3)^*$ hold for $\widehat{R}_r = \widehat{M}_r \otimes \widehat{N}_r$ and $R_r = \widehat{M}_r \otimes \widehat{N}_r$. $(A2)^*$ follows from the fact that

$$
\begin{aligned}
P_{\widehat{R}_r} - P_{R_r} &= P_{\widehat{M}_r} \otimes P_{\widehat{N}_r} - P_{M_r} \otimes P_{N_r} \\
&= P_{\widehat{M}_r} \otimes (P_{\widehat{N}_r} - P_{N_r}) + (P_{\widehat{M}_r} - P_{M_r}) \otimes P_{N_r} \\
&= o_p(1).
\end{aligned}
$$

On the other hand, if $\sqrt{T}(P_{\widehat{N}_r} - P_{N_r}) = O_p(1)$ and $P_{\widehat{M}_r} - P_{M_r} = o_p(1)$, then

$$
\sqrt{T}P_{\widehat{R}_r}\beta = \sqrt{T}\left((P_{\widehat{M}_r} - P_{M_r}) \otimes (P_{\widehat{N}_r} - P_{N_r})\right)\text{vec}(B) = o_p(1).
$$

Thus $(A3)^*$ holds. An identical argument shows that $(A3)^*$ also hold if $P_{\widehat{N}_r} - P_{N_r} = o_p(1)$ and $\sqrt{T}(P_{\widehat{M}_r} - P_{M_r}) = O_p(1)$.

(ii) The proof is exactly analogous to the proof of (i) and is omitted.

To prove (iii), note that the boundedness away from zero in probability condition in $(A4)^*$ is already satisfied and we only need to show that $\widehat{R}_r$ is of full rank. But this follows from the fact that $\widehat{N}_r \in \mathbb{G}^{n \times (n-r)}$ and $\widehat{M}_r \in \mathbb{G}^{m \times (m-r)}$, which implies that $\widehat{R}_r = \widehat{M}_r \otimes \widehat{N}_r \in \mathbb{G}^{nm \times (n-r)(m-r)}$ (Horn & Johnson, 1991, Theorem 4.2.15). $\qquad\square$

*Proof of Theorem 2.* Define $\widetilde{\beta} = (D_m'D_m)^{1/2}\widehat{\beta}$ and $\widetilde{\Omega} = (D_m'D_m)^{1/2}\widehat{\Omega}(D_m'D_m)^{1/2}$. Also define the restriction matrices

$$
\begin{aligned}
\breve{R}_r &= (D_m'D_m)^{-1/2}D_m'(M_r \otimes M_r)D_{m-r}(D_{m-r}'D_{m-r})^{1/2} \\
\widetilde{R}_r &= (D_m'D_m)^{-1/2}D_m'(\widehat{M}_r \otimes \widehat{M}_r)D_{m-r}(D_{m-r}'D_{m-r})^{1/2}.
\end{aligned}
$$

Then $\widehat{F}^s(r) = T\widetilde{\beta}'\widetilde{R}_r(\widetilde{R}_r\widetilde{\Omega}\widetilde{R}_r)^{-1}\widetilde{R}_r'\widetilde{\beta}$ and $F^s(r) = T\widetilde{\beta}'\breve{R}_r(\breve{R}_r\widetilde{\Omega}\breve{R}_r)^{-1}\breve{R}_r'\widetilde{\beta}$. The proof of the theorem then proceeds by plugging in $P_{\widetilde{R}_r}$ for $P_{\breve{R}_r}$ along the same lines as in the proof of Theorem 1 once we show that

$$
\begin{aligned}
P_{\breve{R}_r} &= (D_m'D_m)^{-1/2}D_m'(P_{M_r} \otimes P_{M_r})D_m(D_m'D_m)^{1/2} \\
P_{\widetilde{R}_r} &= (D_m'D_m)^{-1/2}D_m'(P_{\widehat{M}_r} \otimes P_{\widehat{M}_r})D_m(D_m'D_m)^{1/2}.
\end{aligned}
$$

We will prove the first equation. The second follows an identical argument.

$$P_{\check{R}_r} = (D'_m D_m)^{-1/2} D'_m (M_r \otimes M_r) D_{m-r} (D'_{m-r} D_{m-r})^{1/2} \times$$
$$\times \{(D'_{m-r} D_{m-r})^{1/2} D'_{m-r} (M_r \otimes M_r)' D_m (D'_m D_m)^{-1} D'_m (M_r \otimes M_r) D_{m-r} (D'_{m-r} D_{m-r})^{1/2}\}^{-1} \times$$
$$\times (D'_{m-r} D_{m-r})^{1/2} D'_{m-r} (M_r \otimes M_r)' D_m (D'_m D_m)^{-1/2}.$$

The matrix $D_m (D'_m D_m)^{-1} D'_m$, which appears in the inverted term, is known to possess the commutativity property

$$D_m (D'_m D_m)^{-1} D'_m (M_r \otimes M_r) = (M_r \otimes M_r) D_{m-r} (D'_{m-r} D_{m-r})^{-1} D'_{m-r}.$$

See Theorem 3.9 (a) and Theorem 3.12 (b) of Magnus & Neudecker (1999) for the details. Applying the last equation and canceling out the $(D'_{m-r} D_{m-r})^{1/2}$ terms we obtain

$$P_{\check{R}_r} = (D'_m D_m)^{-1/2} D'_m (M_r \otimes M_r) D_{m-r} \{D'_{m-r} (M'_r M_r \otimes M'_r M_r) D_{m-r}\}^{-1} \times$$
$$\times D'_{m-r} (M_r \otimes M_r)' D_m (D'_m D_m)^{-1/2}.$$

Theorem 3.13 (d) of Magnus & Neudecker (1999) allows us to compute the inverted term and obtain

$$P_{\check{R}_r} = (D'_m D_m)^{-1/2} D'_m (M_r \otimes M_r) D_{m-r} D^{\dagger}_{m-r} \{(M'_r M_r)^{-1} \otimes (M'_r M_r)^{-1}\} D^{\dagger}_{m-r}{}'$$
$$\times D'_{m-r} (M_r \otimes M_r)' D_m (D'_m D_m)^{-1/2}.$$

Applying the commutativity property of $D_m D^{\dagger}_m$ again, we obtain the desired result. $\qquad\square$

*Proof of Theorem 3.* (i) Let $R_r = D'_m \mathrm{vec}(P_{M_r})$ and $\widehat{R}_r = D'_m \mathrm{vec}(P_{\widehat{M}_r})$. Clearly $\widehat{R}_r \xrightarrow{p} R_r$ so (A2) is satisfied. On the other hand, $\sqrt{T} \widehat{R}'_r \beta = \sqrt{T} \mathrm{tr}(P_{\widehat{M}_r} B P_{\widehat{M}_r}) = \mathrm{tr}(\sqrt{T}(P_{\widehat{M}_r} - P_{M_r}) B P_{\widehat{M}_r}) = o_p(1)$. Thus (A3) also holds and the result then follows from Lemma 1 (i).

(ii) $P_{\widehat{M}_r} - P_{M_r} = o_p(1)$ implies (A2) and so the result follows from Lemma 1 (ii).

(iii) The columns of $D_m$ are orthogonal and each has either one 1 or two 1's, with all the rest of the elements zero. If we renormalize the columns with two 1's, we obtain the singular value decomposition so that $D_m$ has singular values of either 1's or $\sqrt{2}$'s. Writing $\widehat{R}_r = D'_m D_m \mathrm{vech}(P_{\widehat{M}_r})$ we then have

$$0 \neq \frac{m-r}{\sqrt{m}} = m^{-1/2} \mathrm{tr}(P_{\widehat{M}_r}) \leq \|\mathrm{vech}(P_{\widehat{M}_r})\| \leq \|\widehat{R}_r\| \leq \sqrt{2}\|\mathrm{vec}(P_{\widehat{M}_r})\| = \sqrt{2(m-r)}.$$

The first inequality is follows from the Cauchy–Schwartz inequality applied to the diagonal elements of $P_{\widehat{M}_r}$ and the second and third from knowledge of the singular values of $D_m$. Since $\widehat{B}$ converges to a positive semi–definite matrix, the negative eigenvalues of $P_{\widehat{M}_r} \widehat{B} P_{\widehat{M}_r}$ converge to zero in probability and its largest eigenvalue is $O_p^{-1}(1)$. Therefore the trace of $P_{\widehat{M}_r} \widehat{B} P_{\widehat{M}_r}$ is also $O_p^{-1}(1)$. But its trace is $\widehat{R}'_r \widehat{\beta} = \mathrm{vec}'(P_{\widehat{M}_r}) D_m \widehat{\beta}$. Therefore, (A4) holds and the result follows from Lemma 1 (iii). $\qquad\square$

*Proof of Lemma 4.* (i) Consider the CDA first. By Lemma 6, $\widehat{B}_T - \widehat{B}_{rT}^{CDA} = O_p(\|\widehat{B}_T - \widehat{B}_T^*\|)$. Since $\|\widehat{B}_T - \widehat{B}_T^*\|^2 = \|N_{r\perp}'\widehat{B}M_{rT}\|^2 + \|N_{rT}'\widehat{B}M_{r\perp}\|^2 + \|N_{rT}'\widehat{B}M_{rT}\|^2$, we have $T^\gamma(\widehat{B}_T - \widehat{B}_{rT}^{CDA}) = O_p(1)$. Thus, $\|T^\gamma\widehat{B}_r^{CDA}M_{rT}\| \leq \|[\ N_{r\perp}\quad N_{rT}\ ]^{-1}\|\|T^\gamma[\ N_{r\perp}\quad N_{rT}\ ]'\widehat{B}_r^{CDA}M_{rT}\| = O_p(1)$. Therefore, $T^\gamma\widehat{V}_{\cdot 1}'M_{rT} = O_p(1)$, where $\widehat{B}_r^{CDA} = \widehat{U}_{\cdot 1}\widehat{S}_{11}\widehat{V}_{\cdot 1}'$ is the SVD of $\widehat{B}_r$, since $\widehat{U}_{\cdot 1} = O_p(1)$ and $\widehat{S}_{11}^{-1} = O_p(1)$. This then implies that $T^\gamma(P_{M_{rT}} - P_{\widehat{M}_r})M_{rT} = T^\gamma(I_m - P_{\widehat{M}_r})M_{rT} = T^\gamma\widehat{V}_{\cdot 1}\widehat{V}_{\cdot 1}'M_{rT} = O_p(1)$. An analogous argument proves the rate for the left null space estimator.

Next, consider the DBA. By Lemma 8, $T^\gamma\|\widehat{B} - \widehat{B}_r^{DBA}\| = O_p(1)$. The rate of convergence of the right null space is again proven by showing that $T^\gamma\widehat{B}_r^{DBA}M_{rT} = O_p(1)$. But this follows from the fact that $\widehat{B}_r^{DBA} = (I_n - \widehat{Q}_r)\widehat{B}$, where $\widehat{Q}_r = O_p(1)$ is defined in the proof of Lemma 9 (iii). An analogous argument proves the rate for the left null space estimator.

(ii) The proof is identical to the proof of Lemma 3 (ii).

(iii) The proof for the DBA is identical to that used in Lemma 3 (iii). As for the CDA, the argument requires a slight modification to the one used in Lemma 3 (iii) because $\widehat{P}_l$ may not be bounded in probability. However, $\widehat{P}_{lT} = Z_T'\widehat{P}_l Z_T^{-1\prime} = Z_T'\widehat{\Theta}^{1/2}\widehat{\Theta}^{-1/2}\widehat{P}_l\widehat{\Theta}^{1/2}\widehat{\Theta}^{-1/2}Z_T^{-1\prime} = O_p(\mathrm{cond}^{1/2}(\widehat{\Theta}_T)) = O_p(1)$ as $\widehat{\Theta}^{-1/2}\widehat{P}_l\widehat{\Theta}^{1/2}$ is an orthogonal projection. Note that $\widehat{P}_{lT}$ satisfies (16) with all the matrices substituted for ones with subindex $T$. This follows from the fact that $(\widehat{M}_{lT} \otimes \widehat{N}_{lT})_\perp = (Z_T^{-1}(\widehat{M}_l \otimes \widehat{N}_l))_\perp = Z_T(\widehat{M}_l \otimes \widehat{N}_l)_\perp$. Now

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

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

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

Since $\sigma_r(\widehat{B})$ and $\sigma_{nm}(Z_T)$ are $O_p^{-1}(1)$, while $\mathrm{cond}(\widehat{\Theta}_T)$ and $\widehat{P}_{lT}$ are $O_p(1)$, it follows that $P_{\widehat{N}_{lT}}\widehat{B}_T P_{\widehat{M}_{lT}} = O_p^{-1}(1)$. $\qquad\square$

*Proof of Theorem 4.* (i) First, write $\widehat{F}_c(r)$ as $T^{2\gamma}\widehat{\beta}_T'\widehat{R}_{rT}(\widehat{R}_{rT}'\widehat{\Omega}_T\widehat{R}_{rT})^{-1}\widehat{R}_{rT}'\widehat{\beta}_T$, where $\widehat{\beta}_T = Z_T'\mathrm{vec}(\widehat{B})$ and $\widehat{R}_{rT} = Z_T^{-1}\widehat{R}_r$. Just as in Lemma 2 (i), we may write the statistic as

$T^{2\gamma}\widehat{\beta}'_T P_{\widehat{R}_{rT}}(P_{\widehat{R}_{rT}}\widehat{\Omega}_T P_{\widehat{R}_{rT}})^{\dagger}P_{\widehat{R}_{rT}}\widehat{\beta}_T$ and we will have proven the result if we can show that $(P_{\widehat{R}_{rT}}\widehat{\Omega}_T P_{\widehat{R}_{rT}})^{\dagger} - (P_{R_{rT}}\widehat{\Omega}_T P_{R_{rT}})^{\dagger} = o_p(1)$ and $T^{\gamma}(P_{\widehat{R}_{rT}} - P_{R_{rT}})\widehat{\beta}_T = o_p(1)$, where $R_{rT} = Z_T^{-1}(M_{rT} \otimes N_{rT}) = \left[\begin{smallmatrix} 0 \\ I_{m-r}\end{smallmatrix}\right] \otimes \left[\begin{smallmatrix} 0 \\ I_{n-r}\end{smallmatrix}\right]$. To that end, consider

$$
\begin{aligned}
\|P_{\widehat{R}_{rT}} - P_{R_{rT}}\|_2 &= \left\|P_{\widehat{M}_{rT}} \otimes P_{\widehat{N}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r}\end{smallmatrix}\right] \otimes \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{n-r}\end{smallmatrix}\right]\right\|_2 \\
&\leq \left\|\left(P_{\widehat{M}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r}\end{smallmatrix}\right]\right) \otimes P_{\widehat{N}_{rT}}\right\|_2 + \left\|\left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r}\end{smallmatrix}\right] \otimes \left(P_{\widehat{N}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{n-r}\end{smallmatrix}\right]\right)\right\|_2 \\
&\leq \left\|P_{\widehat{M}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r}\end{smallmatrix}\right]\right\|_2 + \left\|P_{\widehat{N}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{n-r}\end{smallmatrix}\right]\right\|_2.
\end{aligned}
$$

The last inequality follows from the fact that the set of singular values of a Kronecker product is the set of products of the singular values (Horn & Johnson, 1991, Theorem 4.2.15). Now applying (Gohberg et al., 2006, Theorem 13.5.1) again, there exists a $K > 0$ depending only on $\left[\begin{smallmatrix} 0 \\ I_{m-r}\end{smallmatrix}\right]$ such that

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

where $Q \in \mathbb{G}^{(m-r)\times(m-r)}$ is arbitrary. The basic theory of least squares approximation implies that $\inf_Q \|\widehat{M}_r Q - M_{rT}\| = \|(P_{\widehat{M}_r} - I_m)M_{rT}\|$,[24] which is equal to $\|(P_{\widehat{M}_r} - P_{M_{rT}})M_{rT}\|$. Therefore,

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

A similar expression holds for the left null space estimator. Since $\mathrm{rank}(P_{R_{rT}}\widehat{\Omega}_T P_{R_{rT}}) = \mathrm{rank}(P_{\widehat{R}_{rT}}\widehat{\Omega}_T P_{\widehat{R}_{rT}}) = \mathrm{rank}(P_{R_{rT}}\Omega_T P_{R_{rT}}) = (m-r)(n-r)$, the same argument used in Lemma 2 (i) can be used to show that $(P_{\widehat{R}_{rT}}\widehat{\Omega}_T P_{\widehat{R}_{rT}})^{\dagger} - (P_{R_{rT}}\widehat{\Omega}_T P_{R_{rT}})^{\dagger} = o_p(1)$. Finally, set $\widehat{\beta}_T^* = \mathrm{vec}\left(\left[\begin{smallmatrix} N'_{r\perp}\widehat{B}M_{r\perp} & 0 \\ 0 & 0\end{smallmatrix}\right]\right)$ and note that $P_{R_{rT}}\widehat{\beta}_T^* = 0$, $\widehat{\beta}_T^* = O_p(1)$, and $T^{\gamma}(\widehat{\beta}_T - \widehat{\beta}_T^*) = O_p(1)$ so that[25]

$$
\begin{aligned}
T^{\gamma}(P_{\widehat{R}_{rT}} - P_{R_{rT}})\widehat{\beta}_T &= T^{\gamma}(P_{\widehat{R}_{rT}} - P_{R_{rT}})\widehat{\beta}_T^* + T^{\gamma}(P_{\widehat{R}_{rT}} - P_{R_{rT}})(\widehat{\beta}_T - \widehat{\beta}_T^*) \\
&= T^{\gamma}\left\{\left(P_{\widehat{M}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{m-r}\end{smallmatrix}\right]\right) \otimes \left(P_{\widehat{N}_{rT}} - \left[\begin{smallmatrix} 0 & 0 \\ 0 & I_{n-r}\end{smallmatrix}\right]\right)\right\}\widehat{\beta}_T^* + O_p(T^{-\gamma}) \\
&= o_p(1).
\end{aligned}
$$

(ii) From the first inequality in (i), $\|P_{\widehat{R}_{rT}} - P_{R_{rT}}\|_2 = o_p(1)$ and the result follows as in Lemma 2 (ii).

(iii) By the same arguments used in Lemma 2 (iii), $T^{-2\gamma}\widehat{F}_c(r)$ is bounded below by $\frac{\|P_{\widehat{N}_l}\widehat{B}P_{\widehat{M}_l}\|^2}{\lambda_1(\widehat{\Omega})}$ and $\frac{\|P_{\widehat{N}_{lT}}\widehat{B}_T P_{\widehat{M}_{lT}}\|^2}{\lambda_1(\widehat{\Omega}_T)}$. $\square$

---

[24]The minimum of $\|\widehat{M}_r Q - M_{rT}\|$ over $\mathbb{R}^{(m-r)\times(m-r)}$ is $\|(P_{\widehat{M}_r} - I_m)M_{rT}\|$ and is 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.

[25]Note that $\widehat{\beta}_T^* = \mathrm{vec}(\widehat{B}_T^*)$ and $\widehat{B}_T^*$ appears in the proof of Lemma 4 (i).

*Proof of Theorem 5.* Similar to the proofs of Theorems 2 and 4 and is omitted. $\qquad\square$

*Proof of Theorem 6.* Using the Kronecker product properties employed in the proof of Theorem 2, we may write

$$t_c(r) = \frac{T^\gamma R'_{rT}\widehat{\beta}_T}{\sqrt{R'_{rT}\widehat{\Omega}_T R_{rT}}}$$

$$\widehat{t}_c(r) = \frac{T^\gamma \widehat{R}'_{rT}\widehat{\beta}_T}{\sqrt{\widehat{R}'_{rT}\widehat{\Omega}_T \widehat{R}_{rT}}},$$

where $\widehat{\beta}_T = D_m^\dagger Z'_T D_m \text{vech}(\widehat{B}) = D_m^\dagger Z'_T \text{vec}(\widehat{B})$, $R_{rT} = D'_m Z_T^{-1}\text{vec}(P_{M_{rT}})$, and $\widehat{R}_{rT} = D'_m Z_T^{-1}\text{vec}(P_{\widehat{M}_r})$.

(i) We follow the same argument used in Lemma 1 (i). Consider the numerator first,

$$\begin{aligned}
T^\gamma \text{tr}(P_{\widehat{M}_r}\widehat{B}) - T^\gamma \text{tr}(P_{M_{rT}}\widehat{B}) &= T^\gamma \text{tr}(P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r}) - T^\gamma \text{tr}(P_{M_{rT}}\widehat{B}P_{M_{rT}}) \\
&= \text{tr}(T^\gamma(P_{\widehat{M}_r} - P_{M_{rT}})\widehat{B}(P_{\widehat{M}_r} - P_{M_{rT}})) \\
&\quad + \text{tr}(T^\gamma P_{M_{rT}}\widehat{B}(P_{\widehat{M}_r} - P_{M_{rT}})) \\
&\quad + \text{tr}(T^\gamma(P_{\widehat{M}_r} - P_{M_{rT}})\widehat{B}P_{M_{rT}}) \\
&= O_p(T^{-\gamma}).
\end{aligned}$$

Next consider the denominator. Since $\widehat{\Omega}_T$ and its inverse are both bounded in probability, the result follows from the fact that

$$\|\widehat{R}_{rT} - R_{rT}\| \le \|D_m\|\|Z_T^{-1}\|\|P_{\widehat{M}_r} - P_{M_{rT}}\| = O_p(T^{-\gamma}).$$

(ii) Follows from the expressions above just as in Lemma 1 (ii).

(iii) By the same arguments used in Lemma 1 (iii) and using the estimate of $\|P_{\widehat{M}_r}\|$ derived in Theorem 3 (iii), we have that $T^{-\gamma}|\widehat{t}_c(r)|$ is bounded below by $\frac{|\text{tr}(P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r})|}{2(m-r)\lambda_1(\widehat{\Omega})}$. Now since $\widehat{B}$ converges in distribution to a positive semi–definite matrix, the negative eigenvalues of $P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r}$ converge in probability to zero while its largest eigenvalue is $O_p^{-1}(1)$ as $P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r} = O_p^{-1}(1)$. Therefore $\text{tr}(P_{\widehat{M}_r}\widehat{B}P_{\widehat{M}_r}) = O_p^{-1}(1)$ and the result follows. $\qquad\square$

*Proof of Proposition 1.* We will show that (CS1) is satisfied and derive an expression for $\xi$. We can simplify the analysis by writing $\eta_t = \zeta_t + \psi_t/T$, where $\{\zeta_t\}$, $\{\psi_t\}$, and $\{\varepsilon_t\}$ are independent of each other and $\zeta_t \sim N(0,\Gamma)$ and $\psi_t \sim N(0,D)$. Let $M_r$ span the null space of $B$ and assume without loss of generality that $[\,M_{r\perp}\quad M_r\,]$ is an orthogonal matrix.

The fact that $\text{vec}(\widehat{B})$ 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 $u \in [0,1]$,

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

where $W_1$ is generated by $\zeta$, $W_2$ is generated by $\varepsilon$, and $W_3$ is generated by $\psi$ (Phillips & Durlauf, 1986; Phillips & Solo, 1992; Billingsley, 1999). Letting $Q_T = [\; M_{r\perp} \quad TM_r \;]$ and $H = \begin{bmatrix} (M'_{r\perp}\Gamma M_{r\perp})^{1/2} & 0 \\ 0 & (M'_r \Sigma M_r)^{1/2} \end{bmatrix}$ we have,

$$\begin{bmatrix} H^{-1}Q'_T \widehat{\Gamma} Q_T H^{-1} \\ H^{-1}Q'_1 \widehat{\Sigma} Q_T H^{-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} C_{11} & C_{12} \\ C'_{12} & C_{22} \\ \int_0^1 W_1^*(u)W_1^{*\prime}(u)du & \int_0^1 W_1^*(u)dW_2'(u) \\ 0 & I_{m-r} \end{bmatrix}.$$

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

Now consider $Q'_1 \widehat{\Sigma}^{1/2} Q_1 Q'_1 \widehat{\Sigma}^{1/2} Q_1 = Q'_1 \widehat{\Sigma} Q_1 = \begin{bmatrix} O_p(1) & O_p(T^{-1}) \\ O_p(T^{-1}) & O_p(T^{-1}) \end{bmatrix}$, where the blocks are conformable with $Q_1$. The $(2,2)$ block implies that $M'_{r\perp}\widehat{\Sigma}^{1/2} M_r M'_r \widehat{\Sigma}^{1/2} M_{r\perp} + (M'_r \widehat{\Sigma}^{1/2} M_r)^2 = O_p(T^{-1})$. 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 so $\alpha = 0$ 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, $\widehat{\Omega} = \Omega = \frac{1}{m-r}(D_m'D_m)^{-1}$, which satisfies the conditions of assumption (CS1) 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 (CS1) is satisfied. By Theorem 6, 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. $\qquad\square$

# References

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.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3 ed.). Hoboken, NJ: John Wiley and Sons Inc.

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.

Bai, J. & Ng, S. (2008). *Large Dimenstional Factor Analysis*, volume 3 of *Foundations and Trends in Econometrics*. Delft, the Netherlands: now Publishers Inc.

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.

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

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., 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.

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.

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

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.

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

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.

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.

Hwang, T.-M., Lin, W.-W., & Yang, E. K. (1992). Rank revealing lu factorizations. *Linear Algebra and its Applications, 175*, 115 – 141.

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

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.

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

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.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, *77*(5), 1447–1479.

Pan, C. T. (2000). On the existence and computation of rank-revealing lu factorizations. *Linear Algebra and its Applications*, *316*(1-3), 199 – 222.

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.

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

Reinsel, G. C. (2003). *Elements of Multivariate Time Series Analysis* (2 ed.). Springer Series in Statistics. New York, USA: Springer.

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.

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.

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.