The Stochastic Wald Test*

Majid M. Al-Sadoon†

November 24, 2010

Abstract

This paper develops a general and very simple framework for tests of rank. We show that most existing tests of rank, including almost all cointegration rank tests, are, either exactly or asymptotically, of the form of a stochastic Wald test where the deterministic constraint matrix in the Wald (1943) statistic is replaced by its estimator. We provide very general conditions under which this plug-in for the constraint matrix is valid, thus nesting a wide variety of tests but also allowing for many new tests of rank including ones based on the new fixed-b inference theory of Kiefer & Vogelsang (2005). We apply the new fixed-b rank test to subspace causality testing in time series, where it is shown that the fixed-b test performs comparably to the bootstrap but is more accurate than the bootstrap in terms of accuracy of rank detection.

JEL Classification: C12, C13, C30.

Keywords: Rank test, Wald test, stochastic Wald test, hypothesis testing, cointegration, HAC tests, fixed-b inference, small-b inference, subspace causality.

*I would like to thank professors Sean Holly, Hashem Pesaran, George Kapetanios, and Richard Smith for helpful comments and suggestions. Thanks also go to seminar participants at the University of Cambridge.

†Contact details are available at: www.econ.cam.ac.uk/phd/mma48/.

1 Introduction

Rank tests are ubiquitous in economics. The substitution matrix in a demand system is required to be have reduced rank, providing a testable implication of utility theory (Lewbel, 1991; Gill & Lewbel, 1992). Identification in instrument variable estimation (Cragg & Donald, 1993) and GMM (Wright, 2003; Arellano et al., 2009) require rank conditions to hold in order to have valid estimation and inference. Rank tests also appear as ingredients to inference schemes in the case of weak instruments (Kleibergen, 2005). Cointegration tests are also inherently rank tests as the number of cointegration relationships in a multivariate time series is related to the rank of the series' long run coefficient matrix (Johansen, 1991). Rank testing is also used in model reduction where it is desired to approximate the coefficients of a VAR by a matrix of lower rank (Velu et al., 1986; Camba-Mendez et al., 2003). Finally, and most importantly for this paper, rank tests arise in time series Granger causality testing (Al-Sadoon, 2009b) where rank is a measure of the strength of association between two series.[1]

Wald tests, on the other hand, have occupied a prominent part of statistical and econometric hypothesis testing since the seminal work of Wald (1943). It is one of the *Holy Trinity* of tests, along with the likelihood ratio test and the lagrange multiplier test. Yet its application to rank testing is relatively recent. The first test of rank was given by Bartlett (1947), who tested the rank of a covariance matrix based on canonical correlation analysis. Subsequently, Anderson (1951) gave the first likelihood ratio test of rank in the context of multivariate regression. It was not until Gill & Lewbel (1992) that the first attempt at a Wald test was made. Unfortunately, this proved to be unsuccessful as shown by Cragg & Donald (1996), who constructed the first Wald test of rank, based on the LU decomposition. This was followed by Ratsimalahelo (2003) and Kleibergen & Paap (2006), whose test is based on the singular value decomposition. Important other approaches include Robin & Smith (2000) who construct their test based on a decomposition inspired by canonical correlations analysis and reduced rank regression and Cragg & Donald (1997) who construct their test based on the asymptotics of a minimum distance estimator. There is also a large literature that generalizes the original Anderson (1951) paper (see Anderson (1999) and section 2.1 of Reinsel & Velu (1998) for a review).

[1]This is just a sample of the many applications of rank tests. See the review article Camba-Mendez & Kapetanios (2009) and the monograph by Reinsel & Velu (1998) for further applications.

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. In contrast, our approach is based on the key insight that all of the rank test statistics above have a common structure either exactly or asymptotically. Each looks like a Wald test, where the constraint matrix has been exchanged for its estimator. If we can justify plugging-in these estimators – and we do – then all of the rank tests above are asymptotically equivalent to a Wald test. We refer to this equivalence as the plug-in principle. We show that the constraint matrices span the left and right null spaces of the population matrix, while the estimators of the constraint matrices are obtainable by reduced rank approximations. It follows then that the different rank tests differ only in the way they estimate the left and right null spaces but are asymptotically equivalent.[2] Application is then almost entirely mechanical: one first solves for the asymptotic distribution of the Wald statistic, then uses estimators of the left and right null spaces obtained by any of the methods we propose, to construct the stochastic Wald statistic. This affords tremendous tractability to the asymptotics of these tests because it is very easy to derive the asymptotics of the Wald test but extremely difficult to derive the asymptotics of the rank test.

Our approach allows us to nest most of the rank testing literature.[3] It also allows us to obtain new tests. In particular, we propose a new rank tests based on the QR decomposition as well as refinements to rank tests based on the LU decomposition, which was used in Cragg & Donald (1996). These new test statistics are far more economical in terms of computation time than the other approaches in the literature. We also derive new tests of rank based on the heteroskedasticity and autocorrelation consistent (HAC) inference methods proposed of Vogelsang (2001), Kiefer & Vogelsang (2002a), Kiefer & Vogelsang (2002b), and Kiefer & Vogelsang (2005). This type of inference, known as fixed-b inference, utilizes an inconsistent estimate of the covariance matrix in the Wald test statistic but its asymptotic distribution is pivotal nevertheless. The size and power properties of fixed-b tests have been extensively studied in the aforementioned literature, particularly relative to small-b inference where the covariance

[2]This also allows us to see how rank test statistics behave under misspecification. Robin & Smith (2000) is the only paper offering any results on misspecification in the rank testing literature.

[3]The only rank testing approaches known to the author that this paper does not nest are Nyblom & Harvey (2000) and Donald et al. (2007), although the general idea of the plug-in principle is still applicable.

is consistently estimated. We present evidence that fixed-b inference is more accurate than small-b inference in detecting rank.

Our approach also allows for a novel and simplified approach to cointegration rank testing. We prove the plug-in principle for cointegration analysis allowing for tremendous simplification of the asymptotics of cointegration ranks tests. Again, the researcher is only required to obtain the asymptotic distribution of the Wald statistic (a usually trivial exercise) and check a few simple conditions (usually checked in the process of cointegration analysis anyway), before concluding that the asymptotic distribution of the rank statistic is equal to that of the Wald statistic. Our approach nests the results of Johansen (1991), Cavaliere et al. (2007), Cavaliere et al. (2009), and – with a few modifications – is capable to nesting Caner (1998) and Johansen (2006).

We then apply the new fixed-b rank test to subspace causality testing. Al-Sadoon (2009b) presented a generalization of Granger causality that is capable of capturing cross-sectional conditional relationships. The researcher is often interested in what portion of the covariation of two variables is due to a third variable (e.g. the covariation of output growth and inflation due to interest rates). On the other hand, sometimes the research is interested in what kinds of covariations of two variables Granger cause a third (e.g. the covariation of interest rates and the money supply that Granger cause output growth). Al-Sadoon (2009b) has shown that these effects are captured by subspace causality, where a rank test is performed instead of the test for zero block restrictions in Dufour & Renault (1998). We follow Dufour et al. (2006) and Al-Sadoon (2009a) in analyzing the predictive effect of monetary policy (consisting of non–borrowed reserves and the federal funds rate) on output growth and inflation at horizons 1–24 months in the Bernanke & Mihov (1998) data set. We find that the fixed-b test has a slight tendency to over-reject in small samples but when correcting for this size distortion, it performs comparably to the bootstrap. The particular empirical findings are: (1) We find that the federal funds rate predicts output growth and inflation only along a single very flat line in output–growth–inflation space for most of the forecast horizons we consider. Therefore the predictive power of the federal funds rate is primarily for output growth rather than inflation. (2) We find that the variations of non–borrowed reserves and the federal funds rate along certain directions have no predictive power for output growth and inflation at most of the horizons considered. The slopes of these directions are consistently negative, indicating a tradeoff between the federal funds rate and non–borrowed reserves *viz–a–viz* output growth and inflation. Thus one can interpret this result as a confirmation of the idea that the central bank may control either the money supply or interest rates but not both at the same time.

The paper is organized as follows. Section 2 develops the notation of the paper. Section 3 gives the basic theory behind the stochastic Wald test. Section 4 specializes the stochastic Wald test to the rank testing setting. Section 5 discusses the variety of available rank revealing decompositions. Section 6 then shows how the stochastic Wald encompasses most of the existing tests of rank as well as presenting a new fixed-b theory of rank testing. Section 7 discusses the extension to cointegration. Section 8 applies the new fixed-b tests to subspace causality testing. Section 9 concludes and section 10 is an appendix.

2 Notation

The set of $n \times m$ matrices is denoted by $\mathbb{R}^{n\times m}$. The reconstitution operator is defined as the inverse of the vectorization operator $\mathrm{mat} = \mathrm{vec}^{-1}$ where $\mathrm{vec} : \mathbb{R}^{n\times m} \to \mathbb{R}^{nm}$ vectorizes a matrix by vertically stacking its columns. The range of mat will be clear from the context. For a matrix $A \in \mathbb{R}^{n\times m}$, we define the Euclidean norm $\|A\| = (\sum_{i=1}^{n}\sum_{j=1}^{m} A_{ij}^2)^{1/2}$, where A_{ij} is the (i, j)-th element of A. We will also utilize the L^2 norm, defined as $\|A\|_2 = \max_{\|x\|=1} \|Ax\|$ and the Mahalanobis norm $\|A\|_{\Xi} = (\mathrm{vec}'(A)\Xi^{-1}\mathrm{vec}(A))^{1/2}$, for symmetric, positive definite, $\Xi \in \mathbb{R}^{nm\times nm}$. When $n > m$, the orthogonal complement of A is denoted by $A_\perp$ and defined as any matrix satisfying $A'_\perp A = 0$ and $\mathrm{rank}([\,A \quad A_\perp\,]) = n$. The particular choice of $A_\perp$ will not matter for our purposes. When $n = m$, we denote the matrix of diagonal elements of A as $\mathrm{diag}(A) = (A_{11}, \ldots, A_{nn})'$. The orthogonal projection onto the column space of A is defined as P_A. Finally, if $B \in \mathbb{R}^{p\times q}$ is another matrix then $A \oplus B = \left[\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right] \in \mathbb{R}^{(n+p)\times(m+q)}$. Finally, we will require the following well known identity.

$$I_n = A^{\frac{1}{2}}B(B'AB)^{-1}B'A^{\frac{1}{2}} + A^{-\frac{1}{2}}B_\perp(B'_\perp A^{-1}B_\perp)^{-1}B'_\perp A^{-\frac{1}{2}}, \tag{2.1}$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric and positive definite and $B \in \mathbb{R}^{n\times m}$ is of rank m.

3 The Basic Stochastic Wald Test

We will work under the following general assumptions and modify or extend as needed for the various applications we consider.

(A1) $X \in \mathbb{R}^p$. $\Omega \in \mathbb{R}^{p\times p}$ is (possibly) random, symmetric, and almost surely positive definite.

(A2) $\widehat{X}$ and $\widehat{\Omega}$ are estimators indexed by T such that $(\sqrt{T}(\widehat{X} - X), \widehat{\Omega}) \xrightarrow{d} (\xi, \Omega)$ as $T \to \infty$, where $\xi \in \mathbb{R}^p$ is a random vector and each $\widehat{\Omega} \in \mathbb{R}^{p\times p}$ is symmetric and almost surely positive definite.

Suppose we would like test the null hypothesis

$$H_0 : C'X = 0 \tag{3.1}$$

for some $C \in \mathbb{R}^{p\times q}$ of rank q. Then, under H_0 and assumptions (A1) and (A2), the *Wald statistic*,

$$\mathscr{W} = T\widehat{X}'C\{C'\widehat{\Omega}C\}^{-1}C'\widehat{X}, \tag{3.2}$$

converges in distribution to $\psi = \xi'C\{C'\Omega C\}^{-1}C'\xi$. We will be interested in testing H_0 when C is unknown.

Note that our assumptions includes the linear version of the original Wald (1943) test as a special case. Wald assumes Ω is constant, $\xi \sim N(0, \Omega)$, and C is known. Our more general assumptions allow us to derive results useful for misspecification, where Ω is not the covariance matrix of ξ. It also allows us to use recent fixed-b inference methods, where Ω is random. It is, of course, possible to relax the assumptions to obtain greater generality. For example, the plug-in principle that we are about to introduce holds if we replace (A2) by $(\sqrt{T}(\widehat{X} - X), \widehat{\Omega}) = O_p(1)$. We may also relax the condition that Ω 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 stochastic Wald tests may also be constructed along the lines of Lutkepohl & Burda (1997) and Dufour & Valery (2009). However, we choose to simplify the exposition by using these more basic assumptions.

Now because C is unknown it is impossible to formulate the Wald statistic. However, suppose we had the additional assumption,

(A3)* There is an estimator $\widehat{C} \xrightarrow{p} C$.

The question now is: can we plug–in $\widehat{C}$ in place of C in (3.2)? Clearly, assumptions (A1), (A2), and (A3)* imply that $T\widehat{X}'C\{\widehat{C}'\widehat{\Omega}\widehat{C}\}^{-1}C'\widehat{X} = \mathscr{W} + o_p(1)$. In order to make the final substitution of $\widehat{C}$ for C then it suffices to have $\sqrt{T}\widehat{C}'\widehat{X} = \sqrt{T}C'\widehat{X} + o_p(1)$. By (A2), $\sqrt{T}(\widehat{X} - X) = O_p(1)$ and so (A3)* implies that $\sqrt{T}(\widehat{C} - C)'(\widehat{X} - X) = o_p(1)$, which implies that $\sqrt{T}(\widehat{C} - C)'\widehat{X} = \sqrt{T}(\widehat{C} - C)'X + o_p(1) = \sqrt{T}\widehat{C}'X + o_p(1)$, under the null. Therefore, $\sqrt{T}\widehat{C}'X = o_p(1)$ is sufficient for

$$\mathscr{W}_s = T\widehat{X}'\widehat{C}\{\widehat{C}'\widehat{\Omega}\widehat{C}\}^{-1}\widehat{C}'\widehat{X}, \tag{3.3}$$

to have an asymptotic ψ distribution. We will refer to this test statistic as the *stochastic Wald statistic* and tests based on it as *stochastic Wald tests*. The additional assumption we need is,

(A4)* $\sqrt{T}\widehat{C}'X = o_p(1)$.

To summarize what we have found so far:

Lemma 1. Under assumptions (A1), (A2), (A3)*, and (A4)*, $\mathscr{W}_s - \mathscr{W} = o_p(1)$ and so $\mathscr{W}_s \xrightarrow{d} \psi$.

We will say that the *plug-in principle* holds whenever $\mathscr{W}_s - \mathscr{W} = o_p(1)$. Next, note from (3.3) that multiplying $\widehat{C}$ on the right by any nonsingular matrix – even if it is inconsistent in T – leaves the test statistic and its asymptotic distribution invariant. This implies that the particular choice of columns of $\widehat{C}$ is not relevant, only the space spanned by these columns is relevant. We therefore modify assumptions (A3)* and (A4)* as,

(A3) $P_{\widehat{C}} \xrightarrow{p} P_C$.

(A4) $\sqrt{T}P_{\widehat{C}}X = o_p(1)$.

Theorem 1. The results of Lemma 1 continue to hold under assumptions (A1) – (A4).

Theorem 1 proves that $\widehat{C}$ does not need to be element-wise consistent. All we need is that $\widehat{C}$ be consistent for the subspace spanned by the columns of C and that it have a rate of convergence faster than $\sqrt{T}$ along X. We will see in the next section that (A4) is in fact very easily satisfied.

Because $\mathscr{W}_s - \mathscr{W} = o_p(1)$, stochastic Wald tests also inherit the power properties of the Wald test. Under the alternative, $H_1 : C'X = \delta$, for some constant $\delta \neq 0$, and assumptions (A1), (A2), and (A3)*, $T^{-1}\mathscr{W}_s \xrightarrow{d} \delta'(C'\Omega C)^{-1}\delta$, so that $\mathscr{W}_s \xrightarrow{p} \infty$. Thus, the stochastic Wald test is consistent. Under the local alternative, $H_T : C'X = \delta/\sqrt{T}$, if $\sqrt{T}P_{\widehat{C}}X = C(C'C)^{-1}\delta + o_p(1)$

(i.e. if the estimator of C converges fast enough), then $\mathscr{W}_s \xrightarrow{d} (C'\xi + \delta)'(C'\Omega C)^{-1}(C'\xi + \delta)$. Under the Wald (1943) assumptions, this reduces to the familiar noncentral χ^2_q distribution with noncentrality parameter $\delta'(C'\Omega C)^{-1}\delta$.

4 Testing Rank in Theory

We may now specialize the theory above to the context of rank testing. In particular, we will utilize the following assumption,

(A5) $X = \text{vec}(A)$ and $\widehat{X} = \text{vec}(\widehat{A})$ for $A, \widehat{A} \in \mathbb{R}^{n\times m}$.

For $r < l = \min\{n, m\}$ we are interested in testing the hypotheses,

$$H_0(r) : \text{rank}(A) = r. \tag{4.1}$$

Now $H_0(r)$ implies the existence of left and right null spaces as follows,

$$H_0(r) \Rightarrow \begin{cases} \exists\ N \in \mathbb{R}^{n\times(n-r)} \text{ s.t. } \text{rank}(N) = n - r \text{ and } N'A = 0. \\ \exists\ M \in \mathbb{R}^{m\times(m-r)} \text{ s.t. } \text{rank}(M) = m - r \text{ and } AM = 0. \end{cases} \tag{4.2}$$

Thus we shall also assume that the constraint matrix takes a particular form.

(A6) $C = M \otimes N$, where M and N are as in (4.2). There are null space estimators $\widehat{M}$ and $\widehat{N}$ from which we form the estimator $\widehat{C} = \widehat{M} \otimes \widehat{N}$.[4]

Under assumptions (A1) – (A6) and $H_0(r)$, the Wald statistic,

$$\mathscr{W} = T\text{vec}'(\widehat{A})(M \otimes N)\{(M \otimes N)'\widehat{\Omega}(M \otimes N)\}^{-1}(M \otimes N)'\text{vec}(\widehat{A}), \tag{4.3}$$

converges in distribution to ψ. On the other hand, the stochastic Wald statistic,

$$\mathscr{W}_s = T\text{vec}'(\widehat{A})(\widehat{M} \otimes \widehat{N})\{(\widehat{M} \otimes \widehat{N})'\widehat{\Omega}(\widehat{M} \otimes \widehat{N})\}^{-1}(\widehat{M} \otimes \widehat{N})'\text{vec}(\widehat{A}). \tag{4.4}$$

converges to ψ if $\sqrt{T}P_{\widehat{N}}AP_{\widehat{M}} = o_p(1)$, by Theorem 1. Note that we could have also tried formulating stochastic Wald statistic for the restrictions $N'A = 0$ (resp. $AM = 0$) but for to have the correct asymptotic distribution, we would then need $\sqrt{T}P_{\widehat{N}}A = o_p(1)$ (resp. $\sqrt{T}AP_{\widehat{M}} = o_p(1)$), which are not easily satisfied. Indeed, every known test of rank is asymptotically (if not exactly) of the form (4.4) – i.e. it takes estimators of both the left and right null spaces.

[4]The particular choice of N, $\widehat{N}$, M, and $\widehat{M}$ (hence C and $\widehat{C}$) does not matter as we have already discussed.

We now show that the condition $\sqrt{T} P_{\widehat{N}} A P_{\widehat{M}} = o_p(1)$ is in fact very easily satisfied. All we need is a consistent estimator for the null space on one side of A and a $\sqrt{T}$-consistent estimator for the other null space. Suppose then that $\sqrt{T}(P_{\widehat{N}} - P_N) = O_p(1)$ and $P_{\widehat{M}} - P_M = o_p(1)$. Then,

$$\begin{aligned} P_{\widehat{C}} - P_C &= P_{\widehat{M}} \otimes P_{\widehat{N}} - P_M \otimes P_N \\ &= P_{\widehat{M}} \otimes (P_{\widehat{N}} - P_N) + (P_{\widehat{M}} - P_M) \otimes P_N \\ &= o_p(1), \end{aligned} \tag{4.5}$$

and,

$$\begin{aligned} \sqrt{T} P_{\widehat{C}} X &= \sqrt{T}(P_{\widehat{M}} \otimes P_{\widehat{N}}) \text{vec}(A) \\ &= \sqrt{T} \left\{ P_{\widehat{M} \otimes M} \otimes P_{\widehat{N}} - P_M \otimes P_N \right\} \text{vec}(A) \\ &= \sqrt{T} \left\{ P_{\widehat{M}} \otimes (P_{\widehat{N}} - P_N) + (P_{\widehat{M}} - P_M) \otimes P_N \right\} \text{vec}(A) \\ &= \underbrace{(I_m \otimes \sqrt{T}(P_{\widehat{N}} - P_N))}_{O_p(1)} \underbrace{(P_{\widehat{M}} \otimes I_n) \text{vec}(A)}_{o_p(1)} \\ &\qquad + \sqrt{T}((P_{\widehat{M}} - P_M) \otimes I_n) \underbrace{(I_m \otimes P_N) \text{vec}(A)}_{=0} \\ &= o_p(1). \end{aligned} \tag{4.6}$$

A similar argument holds if $P_{\widehat{N}} - P_N = o_p(1)$ and $\sqrt{T}(P_{\widehat{M}} - P_M) = O_p(1)$. Thus we have proven the following result:

Lemma 2. Under Assumptions (A1) – (A6) and $H_0(r)$, let $\widehat{N} \in \mathbb{R}^{n \times (n-r)}$ and $\widehat{M} \in \mathbb{R}^{m \times (m-r)}$ be estimators for the left and right null spaces of A and suppose one of the following conditions hold,

(i) $\sqrt{T}(P_{\widehat{N}} - P_N) = O_p(1)$ and $P_{\widehat{M}} - P_M = o_p(1)$.

(ii) $P_{\widehat{N}} - P_N = o_p(1)$ and $\sqrt{T}(P_{\widehat{M}} - P_M) = O_p(1)$.

Then $\mathscr{W}_s - \mathscr{W} = o_p(1)$.

Lemma 2 is remarkable in that it requires very minimal assumptions on rates of convergence. Only one side needs to be $\sqrt{T}$-consistent, the other only needs to be consistent. The question now is: how do we obtain estimators for the left and right null spaces of A? This is the topic of the next section.

5 Reduced Rank Approximations

It has long been recognized that the algebraic definition of rank is unsuitable for a wide variety of applied work. For example, $A = \left[\begin{smallmatrix} 1 & 1 \\ 0 & \varepsilon \end{smallmatrix}\right]$ has rank 2 for every $\varepsilon \neq 0$, but for small values of ε, its "effective rank" is 1 with "effective left null space," $\operatorname{span}\left(\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]\right)$ and "effective right null space" $\operatorname{span}\left(\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]\right)$. The basic idea here is that the effective rank of A is the rank of the reduced rank matrix that best approximates A and its effective left and right null spaces are then the left and right null spaces of the approximating matrix. In the numerical analysis literature the quality of the approximation is judged by tolerance levels that are of the order magnitude of machine epsilon. For econometric applications, on the other hand, we utilize the stochastic structure of the matrix to determine its effective rank.[5]

In most applications, our estimator $\widehat{A}$ will have full rank l, even though, under $H_0(r)$, $\widehat{A}$ gets closer and closer to a matrix of rank r. Its "effective" rank, on the other hand, clearly converges to r and so can be used to test $H_0(r)$. In fact, if the rate of convergence of the rank-r approximation to $\widehat{A}$ is fast enough, then the stochastic Wald test has the correct asymptotic distribution.

Theorem 2. Under assumptions (A1) – (A6) and $H_0(r)$, if $\widehat{A}_r$ is a rank r approximation to $\widehat{A}$, $\widehat{N} \in \mathbb{R}^{n\times(n-r)}$ and $\widehat{M} \in \mathbb{R}^{m\times(m-r)}$ span the left and right null spaces of $\widehat{A}_r$, and $\sqrt{T}(\widehat{A}_r - A) = O_p(1)$ (or equivalently, $\sqrt{T}(\widehat{A} - \widehat{A}_r) = O_p(1)$), then $\mathscr{W}_s - \mathscr{W} = o_p(1)$.

We discuss a number of reduced rank approximations below. Throughout the following it is assumed that Assumptions (A1) – (A6) and $H_0(r)$ hold.

The Singular Value Decomposition. The most common way of obtaining reduced rank approximations is by so called rank revealing decompositions, the most important of which is the singular value decomposition (SVD). It is also the one on which the Ratsimalahelo-Kleibergen-Paap test static is based (Ratsimalahelo, 2003; Kleibergen & Paap, 2006).[6] The SVD of $\widehat{A}$ is of the form $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}'$, 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 $\sigma_1(\widehat{A}) \geq \sigma_2(\widehat{A}) \geq \cdots \geq \sigma_l(\widehat{A}) \geq 0$. Under $H_0(r)$, we partition

[5] See Golub & Van Loan (1996) or Hansen (1998) for more details.

[6] Interestingly, the singular value decomposition has a long history in applied mathematics as a rank revealing decomposition (Stewart, 1993) and yet it was the last decomposition to be used in a rank test.

the SVD of $\widehat{A}$ as,

$$\widehat{A} = \begin{bmatrix} \widehat{U}_{\cdot 1} & \widehat{U}_{\cdot 2} \end{bmatrix} \begin{bmatrix} \widehat{S}_1 & 0 \\ 0 & \widehat{S}_2 \end{bmatrix} \begin{bmatrix} \widehat{V}_{\cdot 1}' \\ \widehat{V}_{\cdot 2}' \end{bmatrix} = \widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}' + \widehat{U}_{\cdot 2}\widehat{S}_2\widehat{V}_{\cdot 2}', \tag{5.1}$$

where $\widehat{S}_1 \in \mathbb{R}^{r\times r}$. Then, as is well known $\widehat{A}_r^{SVD} = \widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}'$ is the closest matrix to $\widehat{A}$ of rank r in Euclidian distance. In particular, $\sum_{i=r+1}^{l} \sigma_i^2(\widehat{A}) = \|\widehat{A} - \widehat{A}_r^{SVD}\|^2 \leq \|\widehat{A} - A\|^2$ (Horn & Johnson, 1985, example 7.4.1). Therefore,

$$\sqrt{T}\sigma_i(\widehat{A}) = O_p(1), \quad i = r+1, \ldots, l, \tag{5.2}$$

and so $\sqrt{T}\|\widehat{A} - \widehat{A}_r^{SVD}\| = O_p(1)$. By Theorem 2 then, $\mathscr{W}_s$ with $\widehat{N} = \widehat{U}_{\cdot 2}$ and $\widehat{M} = \widehat{V}_{\cdot 2}$, is asymptotically distributed as ψ.

The Robin-Smith Decomposition. The decomposition, which we will abbreviate as RSD, appears explicitly for the first time in Robin & Smith (2000) although it appears implicitly in the context of maximum likelihood estimation under rank restrictions (Anderson, 1951; Johansen, 1991), canonical correlations analysis (Hotelling, 1936; Bartlett, 1947), and reduced rank regression (Izenman, 1975).[7]

The RSD of $\widehat{A}$ takes symmetric positive definite matrices $\widehat{\Sigma} \in \mathbb{R}^{n\times n}$ and $\widehat{\Gamma} \in \mathbb{R}^{m\times m}$ and obtains $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}'$, 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 $\widehat{U}^{-1\prime}$ are generalized eigenvectors of $(\widehat{A}\widehat{\Gamma}^{-1}\widehat{A}', \widehat{\Sigma})$.

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

(iii) $\widehat{S}$ is diagonal with diagonal entries, $\tau_1(\widehat{A}) \geq \tau_2(\widehat{A}) \geq \cdots \geq \tau_l(\widehat{A}) \geq 0$.

The RSD is very easily derived from the SVD: if $\widehat{U}_0\widehat{S}_0\widehat{V}_0'$ is the SVD of $\widehat{\Sigma}^{-\frac{1}{2}}\widehat{A}\widehat{\Gamma}^{-\frac{1}{2}}$ then $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}'$ 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. Clearly the RSD reduces to the SVD when $\widehat{\Sigma} = I_n$ and $\widehat{\Gamma} = I_m$.

In the typical context in which the RSD arises, $\widehat{A}$ is some measure of covariation between two random vectors, while $\widehat{\Sigma}$ and $\widehat{\Gamma}$ are measures of variance for each of the two vectors. In the canonical correlation analysis of two random vectors y and z if the sample covariance matrix, of $(y', z')'$ is $\left[\begin{smallmatrix} \widehat{\Sigma} & \widehat{A} \\ \widehat{A}' & \widehat{\Gamma} \end{smallmatrix}\right]$, the columns found in (i) and (ii) define the coefficients of the canonical variates, while (iii) lists the canonical correlations. In reduced rank regression, on the other

[7] 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).

hand, we take $\widehat{A}$ to be the OLS estimator of A in the regression equation $y = Az + \varepsilon$, while $\widehat{\Sigma}$ is the OLS estimator of the variance of ε and $\widehat{\Gamma}^{-1}$ is the sample second moment of z. Section 2.4.2 of Reinsel & Velu (1998) discusses the relationship between canonical correlations and reduced rank regression in further detail; see also, Anderson (2003) and Reinsel (2003).

Now just as we did in (5.1), write $\widehat{A} = \widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}' + \widehat{U}_{\cdot 2}\widehat{S}_2\widehat{V}_{\cdot 2}'$, where $\widehat{S}_1 \in \mathbb{R}^{r\times r}$ and set $\widehat{A}_r^{RSD} = \widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}'$. Then,

$$\begin{aligned}\|\widehat{A} - \widehat{A}_r^{RSD}\|^2 &\leq \|\widehat{U}_{\cdot 2}\|^2\|\widehat{V}_{\cdot 2}\|^2 \sum_{i=r+1}^{l} \tau_i^2(\widehat{A}) \\ &\leq (n-r)(m-r)\|\widehat{\Sigma}^{\frac{1}{2}}\|^2\|\widehat{\Gamma}^{\frac{1}{2}}\|^2 \sum_{i=r+1}^{l} \tau_i^2(\widehat{A}) \\ &\leq (n-r)(m-r)\|\widehat{\Sigma}^{\frac{1}{2}}\|^2\|\widehat{\Gamma}^{\frac{1}{2}}\|^2\sigma_1^2(\widehat{\Sigma}^{-\frac{1}{2}})\sigma_1^2(\widehat{\Gamma}^{-\frac{1}{2}}) \sum_{i=r+1}^{l} \sigma_i^2(\widehat{A}). \qquad (5.3)\end{aligned}$$

The last inequality follows from the fact that $\tau_i(\widehat{A}) = \sigma_i(\widehat{\Sigma}^{-\frac{1}{2}}\widehat{A}\widehat{\Gamma}^{-\frac{1}{2}}) \leq \sigma_1(\widehat{\Sigma}^{-\frac{1}{2}})\sigma_i(\widehat{A})\sigma_1(\widehat{\Gamma}^{-\frac{1}{2}})$ (Horn & Johnson, 1991, Theorem 3.3.16). If $\widehat{\Sigma}$, $\widehat{\Gamma}$, and their inverses are bounded in probability, then by (5.2), $\sqrt{T}(\widehat{A} - \widehat{A}_r^{RSD}) = O_p(1)$. By Theorem 2 then we may pick any matrix orthogonal to $\widehat{U}_{\cdot 1}$ of rank $n-r$, call it $\widehat{N}$, and any matrix orthogonal to $\widehat{V}_{\cdot 1}$ of rank $m-r$, call it $\widehat{M}$ and construct the stochastic Wald statistic $\mathscr{W}_s$, which then has an asymptotic ψ distribution. Note, in particular, that we may choose $\widehat{N}$ and $\widehat{M}$ so that $\widehat{N}'\widehat{A}\widehat{M} = \widehat{S}_2$.

The Cragg-Donald Approximation. The Cragg & Donald (1997) inference scheme may be used to construct a reduced rank approximation of $\widehat{A}$, which we abbreviate as CDA. The approximation takes a symmetric positive definite matrix $\widehat{\Xi}$ and obtains,

$$\widehat{A}_r^{CD} \in \operatorname{argmin}\{\|\widehat{A} - A\|_{\widehat{\Xi}} : \operatorname{rank}(A) = r\} \qquad (5.4)$$

It can be proven by standard methods that minimizers in (5.4) exist. It is also easy to show that when $\widehat{\Xi} = \widehat{\Gamma} \otimes \widehat{\Sigma}$ then $\widehat{A}_r^{RSD}$ is a minimizer in (5.4). Thus, the CDA generalizes both the RSD and the SVD reduced rank approximations. Uniqueness, on the other hand, may not hold.[8] Therefore, whenever definiteness is required we will agree to pick the minimizer in (5.4) of minimal lexicographical order. This matrix is unique because the set of matrices of rank r of minimum Mahalanobis distance to $\widehat{A}$ is compact.

It is easy to show that the CDA is an admissible reduced rank approximation:

$$\|\widehat{A} - \widehat{A}_r^{CD}\| \leq \sigma_1(\widehat{\Xi})\|\widehat{A} - \widehat{A}_r^{CD}\|_{\widehat{\Xi}} \leq \sigma_1(\widehat{\Xi})\|\widehat{A} - A\|_{\widehat{\Xi}} \leq \sigma_1(\widehat{\Xi})\sigma_1(\widehat{\Xi}^{-1})\|\widehat{A} - A\|, \qquad (5.5)$$

[8] For example, with $\widehat{A} = I_n$, and $\widehat{\Xi} = I_{n^2}$, $\widehat{A}_{n-1}^{CD}$ may be chosen as I_n, with one diagonal element set equal to zero.

where the first and third inequalities follows from standard norm calculus and the second follows from the definition of the CDA. Admissibility then follows if $\widehat{\Xi}$ and its inverse are bounded in probability.

The CDA is unfortunately not analytically tractable for general forms of $\widehat{\Xi}$. Therefore it must be obtained numerically, a serious drawback to its practical implementation. In fact, Cragg and Donald have no use for their reduced rank approximation in their own work. It emerges as a byproduct of their inference scheme, minimum distance estimation. However, it turns out that their test statistic is still a stochastic Wald statistic.

Proposition 1. The Cragg & Donald (1997) test statistic, $T\|\widehat{A} - \widehat{A}_r^{CD}\|_{\widehat{\Xi}}^2$, is identical to the stochastic Wald statistic based on the CDA.

The LU Decomposition. The LU decomposition, derives from extensions to the Gaussian elimination algorithm and is used in Cragg & Donald (1996) to construct a Wald test. Consider the following algorithm, which can be found in section 3.4.8 of Golub & Van Loan (1996); we apply it first to A rather than $\widehat{A}$.

Algorithm 1 (Gaussian Elimination with Complete Pivoting)**.** Set $A^{(0)} = A$. For $i = 1, \ldots, l-1$,

(i) Let (u^*, v^*) be such that $|A_{u^*v^*}| = \max_{u,v \geq i} |A_{uv}^{(i-1)}|$ (this is called the pivot).

(ii) Let E_i be the $n \times n$ permutation matrix that switches the u^*–th and i–th rows.

(iii) Let F_i be the $m \times m$ permutation matrix that switches the v^*–th and i–th columns.

(iv) Let G_i be the matrix that uses the i–th equation of $E_i A^{(i-1)} F_i$ to eliminate the i–th variable from the rows $i+1, \ldots, n$.

(v) Set $A^{(i)} = G_i E_i A^{(i-1)} F_i$.

The algorithm terminates at (i) if the pivot is zero. If the pivot is not unique, we choose according to lexigoraphical order on the indices $\{(u,v) : u, v \geq i\}$. □

Clearly, the algorithm must terminate after the r–th step and the matrices E_i, F_i, $1 \leq i \leq r$ are well defined. It is also possible to prove that if $P = E_r E_{r-1} \cdots E_1$ and $Q = F_1 F_2 \cdots F_r$, then $PAQ = LU$ where L and U' are lower triangular and L has 1's along its diagonal (Golub & Van Loan, 1996, Theorem 3.4.2). It follows that U has zeros below its r–th row and the upper left $r \times r$ corner of PAQ must be of full rank. Setting, $B = PAQ = \left[\begin{smallmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{smallmatrix}\right]$, with $B_{11} \in \mathbb{R}^{r \times r}$, Cragg & Donald (1996) also show that it must be the case that $B_{22} - B_{21}B_{11}^{-1}B_{12} = 0$. Finally, we have that,

$$B = \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} B_{11}^{-1} \begin{bmatrix} B_{11} & B_{12} \end{bmatrix}, \tag{5.6}$$

which is known as the skeleton decomposition (Goreinov et al., 1997).

Next define $\widehat{P}$ and $\widehat{Q}$ as above, the product of the first r permutations which result from the process of Gaussian elimination on $\widehat{A}$ with complete pivoting. Set $\widehat{B} = \widehat{P}\widehat{A}\widehat{Q}$, partition it conformably with B, and note that eventually $\text{rank}(\widehat{B}_{11}) = r$, whence we define its rank r approximation analogously to (5.6),

$$\widehat{B}_r = \begin{bmatrix} \widehat{B}_{11} \\ \widehat{B}_{21} \end{bmatrix} \widehat{B}_{11}^{-1} \begin{bmatrix} \widehat{B}_{11} & \widehat{B}_{12} \end{bmatrix}, \tag{5.7}$$

with $\widehat{A}_r^{LU} = \widehat{Q}'\widehat{B}_r\widehat{P}'$. Therefore, eventually we must have that,

$$\begin{aligned} \|\widehat{A} - \widehat{A}_r^{LU}\| &= \|\widehat{P}(\widehat{A} - \widehat{A}_r^{LU})\widehat{Q}\| = \|\widehat{B} - \widehat{B}_r\| = \|\widehat{B}_{22} - \widehat{B}_{21}\widehat{B}_{11}^{-1}\widehat{B}_{12}\| \\ &= O(\|\widehat{B}\|_2^2\|\widehat{B}_{11}^{-1}\|_2^2\sigma_{r+1}(\widehat{B})) \end{aligned} \tag{5.8}$$

where the last equality follows from equation (1.5) of Goreinov et al. (1997). Since $\widehat{B}$ and $\widehat{B}_{11}^{-1}$ are bounded in probability and the singular values of $\widehat{B}$ are identical to the singular values of $\widehat{A}$, we have that $\widehat{A} - \widehat{A}_r^{LU} = O_p(\sigma_{r+1}(\widehat{A}))$. By (5.2) then, the LU decomposition is admissible as a rank–revealing decomposition in Theorem 2 and we may choose, as Cragg and Donald do, $\widehat{N}' = \left[\, \widehat{B}_{21}\widehat{B}_{11}^{-1} \;\; -I_{n-r} \,\right]\widehat{Q}$ and $\widehat{M} = \widehat{P}\begin{bmatrix} \widehat{B}_{11}^{-1}\widehat{B}_{12} \\ -I_{m-r} \end{bmatrix}$ to formulate $\mathscr{W}_s$.

Unfortunately, the Gaussian elimination algorithm with complete pivoting may fail to detect effective rank (Peters & Wilkinson, 1970). The LU decomposition is also computationally burdensome and so is rarely included in computer packages.[9] To address these shortcomings, a number of rank–revealing LU decompositions have been proposed in the numerical analysis literature; these include Hwang et al. (1992) and Pan (2000) who propose different pivoting algorithms to obtain $\widehat{P}\widehat{A}\widehat{Q} = \widehat{L}\widehat{U}$, with $\widehat{U} = \begin{bmatrix} \widehat{U}_{11} & \widehat{U}_{12} \\ 0 & \widehat{U}_{22} \end{bmatrix}$, $\widehat{L}$ bounded, and $\|\widehat{U}_{22}\| = O(\sigma_{r+1}(\widehat{A}))$. Setting $\widehat{U}_{22}$ equal to zero then we obtain our rank r approximation to $\widehat{A}$ which, by (5.2), clearly satisfies the conditions of Theorem 2.

[9] Gaussian elimination with partial pivoting runs faster but it too can fail to detect the effective rank of some systems. Following the same line of argument as above, it is easy to check that a reduced rank approximation based on the partial pivoting algorithm continues to satisfy the conditions of Theorem 2.

The QR Decomposition. The QR decomposition arises from extensions to the Gram–Schmidt orthogonalization algorithm and, to the author's knowledge, has never been used in a Wald test of rank. Like the LU decomposition, it is not unique as it depends on the particular choice of pivoting strategy. We will discuss a very commonly used QR algorithm, which can be found in section 5.4.1 of Golub & Van Loan (1996).

Algorithm 2 (QR Algorithm with Pivoting)**.** Set $A^{(0)} = A$. For $i = 1, \ldots, l-1$.

(i) Partition $A^{(i-1)} = \begin{bmatrix} A_{11}^{(i-1)} & A_{12}^{(i-1)} \\ 0 & A_{22}^{(i-1)} \end{bmatrix}$ with $A_{11}^{(i-1)} \in \mathbb{R}^{(i-1)\times(i-1)}$ and denote the columns of $A_{22}^{(i-1)}$ as z_j for $1 \leq j \leq m-i+1$.

(ii) Let v^* be such that $\|z_{v^*}\| = \max\{\|z_1\|, \ldots, \|z_{m-i+1}\|\}$ (this is called the pivot).

(iii) Let $\widetilde{\Pi}_i$ be the $(m-i+1) \times (m-i+1)$ permutation matrix that interchanges the first and v^*-th columns and set $\Pi_i = I_{i-1} \oplus \widetilde{\Pi}_i$.

(iv) Let $\widetilde{H}_i$ be the Householder matrix that maps z_{v^*} to $\|z_{v^*}\|e_1$, where e_1 is the first standard basis vector of $\mathbb{R}^{m-i+1}$.[10] Set $H_i = I_{i-1} \oplus \widetilde{H}_i$.

(v) Set $A^{(i)} = H_i A^{(i-1)} \Pi_i$.

The algorithm terminates at (ii) if the pivot is zero. If the pivot is not unique, we choose the smallest of the indices. □

Clearly the algorithm must terminate after the r-th step. If we set $Q' = H_r H_{r-1} \cdots H_1$, $R = A^{(r)}$, and $\Pi = \Pi_1 \Pi_2 \cdots \Pi_r$, then $A\Pi = QR$. Moreover, R has zeros below its r–th row and its upper $r \times r$ corner is of full rank.

Next define $\widehat{\Pi}$, $\widehat{Q}$, and $\widehat{R}$, as above, the resultant of the first r steps of the algorithm. If we partition, $\widehat{R} = \begin{bmatrix} \widehat{R}_{11} & \widehat{R}_{12} \\ 0 & \widehat{R}_{22} \end{bmatrix}$ then it must be the case that $\widehat{R}_{11} \in \mathbb{R}^{r\times r}$ is eventually invertible, while $\widehat{R}_{22}$ should converge to zero. If we approximate $\widehat{A}$ by the rank r matrix which results from setting $\widehat{R}_{22}$ to zero we have,

$$\|\widehat{A} - \widehat{A}_r^{QR}\| = \|\widehat{Q}'(\widehat{A} - \widehat{A}_r^{QR})\widehat{\Pi}\| = \|\widehat{R}_{22}\| \leq \sqrt{n-r}\|\widehat{R}_{11}^{-1}\mathrm{diag}(\widehat{R}_{11})\|_2 \sigma_{r+1}(\widehat{A}), \tag{5.9}$$

where the last inequality is derived in Chandrasekaran & Ipsen (1994). By (5.2) and Theorem 2 then, $\widehat{A}_r^{QR}$ is an admissible rank r approximation.

[10]Householder matrices are orthogonal and symmetric matrices which map a vector to the first standard basis vector (see section 5.1 of Golub & Van Loan (1996) for more details).

The QR algorithm with pivoting has the same drawbacks as the Gaussian elimination algorithm with complete pivoting, although it continues to be sufficient for most purposes. Researchers have therefore considered various modifications to the procedure. Foster (1986) and Chan (1987) provide improvements to the algorithm, which reduce the factor multiplying $\sigma_{r+1}(\widehat{A})$ in (5.9).

Each of the above decompositions (and there are many more in the numerical analysis literature) has its advantages and disadvantages. The SVD is the best detector of rank deficiency but its null space estimators are not interpretable from a statistical point of view, whereas the estimators from the RSD can be seen as coefficients of canonical variates. Both of these, as well as the CDA, are expensive from a numerical point of view, which is why numerical analysts prefer to use rank revealing LU or QR decompositions that can be carried out in a smaller number of operations. This may be an advantage if these decompositions have to be done repeatedly such as in the case of bootstrapping or if the system has a large number of variables. For a more detailed comparison of these, and other, algorithms see Golub & Van Loan (1996) or Hansen (1998).[11]

6 Testing Rank in Practice

We now show how the stochastic Wald test generalizes all preexisting tests of rank. We will require the following assumptions.

(S1) Ω is constant.

(S2) $\xi \sim N(0, \Xi)$, where $\Xi \in \mathbb{R}^{p \times p}$ is constant.

(S3) There is an estimator $\widehat{\Xi} \xrightarrow{p} \Xi$.

(S4) $C'\Xi C \neq 0$.

Under assumptions (A1) – (A6), the plug–in principle holds for any of the null space estimators we have considered in section 5 and so $\mathscr{W}_s \xrightarrow{d} \psi$. Assumptions (S1) – (S4) then establish the distribution of $\psi = \xi' C\{C'\Omega C\}^{-1} C'\xi$, which is simply a quadratic form in the

[11]The reader may wonder why we have not mentioned eigenvalues as a way of detecting effective rank. This is because, in addition to their limitation to square matrices, they do not detect effective rank very well. Horn & Johnson (1985) give an example in problem 21 of section 7.3.

normally distributed vector ξ. Assumption (S4), which appears in Robin & Smith (2000), is required to ensure that the constraints are in some sense "binding." The estimator must have some variation along the constraints otherwise the constraint is meaningless. The following theorem is then an immediate consequence of the well known result on quadratic forms in normally distributed vectors (White, 1994, Lemma 8.2).

Theorem 3. Under assumptions (A1) – (A6), (S1) – (S4), and $H_0(r)$, if the null space estimators satisfy the conditions of Lemma 2, then $\mathscr{W}_s \xrightarrow{d} \sum_{j=1}^{(n-r)(m-r)} \lambda_j Z_j^2$, where $\{\lambda_j\}_{j=1}^{(n-r)(m-r)}$ are the eigenvalues of $(C'\Xi C)^{\frac{1}{2}}(C'\Omega C)^{-1}(C'\Xi C)^{\frac{1}{2}}$ and $\{Z_j\}_{j=1}^{(n-r)(m-r)}$ are independent standard normal random variables.

Theorem 3 generalizes Theorem 3.2 of Robin & Smith (2000), which gives the result only for the case of C estimated using the RSD. It follows that if $\Omega = \Xi$, then $\psi \sim \chi^2_{(n-r)(m-r)}$ so that $\mathscr{W}_s$ is asymptotically pivotal. If, on the other hand, $\Omega \neq \Xi$, then ψ is a weighted sum of $(n-r)(m-r)$ independent χ^2_1 random variables, a familiar outcome of missepcification analysis (White, 1994). In this case, Theorem 4.2 of Robin & Smith (2000) states that we can use the eigenvalues of $(\widehat{C}'\widehat{\Xi}\widehat{C})^{\frac{1}{2}}(\widehat{C}'\widehat{\Omega}\widehat{C})^{-1}(\widehat{C}'\widehat{\Xi}\widehat{C})^{\frac{1}{2}}$ to estimate the distribution of ψ, and get a consistent test with an asymptotically correct size.

Our approach is also general enough to accommodate the new hypothesis testing theory proposed by Kiefer et al. (2000), Vogelsang (2001), Kiefer & Vogelsang (2002a), Kiefer & Vogelsang (2002b), and Kiefer & Vogelsang (2005). Here, we retain assumptions (A1) – (A6) so the plug-in principle continues to hold for any of the null space estimators we have considered in section 5, but $\widehat{\Omega}$ is constructed from fixed–bandwidth long run covariance estimators. In this case, $\widehat{\Omega}$ is an inconsistent estimator of the true asymptotic covariance matrix of $\widehat{A}$. The aforementioned authors show that ψ is still nuisance–parameter–free. In particular, Kiefer & Vogelsang (2005) show that under assumptions that are weaker than those necessary to obtain (S1) – (S4), when the bandwidth is set to bT for $b \in (0, 1]$ we have the following:

(N1) For any constant $D \in \mathbb{R}^{p\times k}$ of rank k, there is a constant, symmetric, positive definite matrix, Ξ, such that $D'\xi = \Xi^{\frac{1}{2}} W_k(1)$, where W_k is a k–dimensional standard Brownian motion.

(N2) For D, Ξ, and W_k as in (N1), $D'\Omega D = \Xi^{\frac{1}{2}} Q_k(b) \Xi^{\frac{1}{2}}$, where $Q_k(b)$ is a known, symmetric, almost surely positive definite random matrix, consisting of non–linear functionals of

W_k, and whose distribution depends on b and the particular choice of kernel used to construct $\widehat{\Omega}$.

Kiefer & Vogelsang (2005) note that $Q_p(b) \xrightarrow{p} I_p$ as $b \to 0$. Thus assumptions (N1) and (N2) reduce to assumptions (S1) – (S4) with $\Xi = \Omega$ when $b \to 0$. The following theorem is then an immediate consequence of the assumptions.

Theorem 4. Under assumptions (A1) – (A6), (N1), (N2), and $H_0(r)$, if the null space estimators satisfy the conditions of Lemma 2, $\mathscr{W}_s \xrightarrow{d} W'_{(n-r)(m-r)}(1) Q^{-1}_{(n-r)(m-r)}(b) W_{(n-r)(m-r)}(1)$.

Under the assumptions of Theorem 4, ψ is free of nuisance parameters and obtainable by standard simulation techniques.

7 Cointegration

Cointegration requires a slightly more delicate approach to rank testing than what we have developed above. Too see this, consider the simplest possible cointegrated VAR(1) model,

$$\Delta y_t = A y_{t-1} + \varepsilon_t, \qquad t = 1, \dots, T, \tag{7.1}$$

where y is n–dimensional, $A = \alpha\beta'$, with $\alpha, \beta \in \mathbb{R}^{n \times r}$, and $r < n$. We assume that $\det(\alpha'_\perp \beta_\perp) \neq 0$ so that y is not $I(2)$. ε is assumed i.i.d. with positive definite covariance matrix Σ. The OLS estimator of A is $\widehat{A} = \sum_{t=1}^T \Delta y_t y_{t-1} \left(\sum_{t=1}^T y_{t-1} y'_{t-1} \right)^{-1}$, while the OLS estimator for Σ is $\widehat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (\Delta y_t - \widehat{A} y_{t-1})(\Delta y_t - \widehat{A} y_{t-1})'$. The OLS estimate of the covariance of $\widehat{A}$ is $\widehat{\Omega} = \left(\frac{1}{T} \sum_{t=1}^T y_{t-1} y'_{t-1} \right)^{-1} \otimes \widehat{\Sigma}$. The Wald test statistic for the rank of A is then given as,

$$\mathscr{W} = T \mathrm{vec}'(\alpha'_\perp \widehat{A} \beta_\perp) \left(\beta'_\perp \left(\frac{1}{T} \sum_{t=1}^T y_{t-1} y'_{t-1} \right)^{-1} \beta_\perp \otimes \alpha'_\perp \widehat{\Sigma} \alpha_\perp \right)^{-1} \mathrm{vec}(\alpha'_\perp \widehat{A} \beta_\perp) \tag{7.2}$$

and standard asymptotic methods can be used to show that $\mathscr{W}$ converges in distribution to the same distribution achieved by the cointegration rank statistic of Johansen (1991). Since Johansen's test statistic is the stochastic Wald statistic based on the RSD, the plug–in principle would seem to hold in this case. Unfortunately, however, assumption (A1) is violated as $\widehat{\Omega}$ converges in probability to a constant matrix that is not positive definite. The problem, quite simply, is that $\widehat{A}$ converges at different rates on each side: $\sqrt{T} \alpha'_\perp \widehat{A} = O_p(1)$ while

$T\widehat{A}\beta_\perp = O_p(1)$. While this suggests that rank testing should be all the more easier, it also implies that $\Omega(\beta_\perp \otimes I_n) = 0$. Therefore, we will need slightly different tools to justify the plug-in principle in cointegration.

Consider the following assumptions.

(C1) $X \in \mathbb{R}^p$. $\Omega \in \mathbb{R}^{p\times p}$ is symmetric and positive semi-definite.

(C2) $\widehat{X}$ and $\widehat{\Omega}$ are estimators indexed by T such that $\sqrt{T}(\widehat{X} - X) = O_p(1)$, $\widehat{\Omega} \xrightarrow{p} \Omega$, and each $\widehat{\Omega} \in \mathbb{R}^{p\times p}$ is almost surely positive definite.

(C3) $X = \text{vec}(A)$ and $\widehat{X} = \text{vec}(\widehat{A})$ for $A, \widehat{A} \in \mathbb{R}^{n\times m}$.

(C4) $C = M \otimes N$, where M and N are as in (4.2). There are null space estimators $\widehat{M}$ and $\widehat{N}$ from which we form the estimator $\widehat{C} = \widehat{M} \otimes \widehat{N}$.

(C5) $\frac{1}{T}(M \otimes I_n)'\widehat{\Omega}^{-1}(M \otimes I_n)$ and its inverse are $O_p(1)$.

(C6) $\widehat{\Omega}^{-1}(M_\perp \otimes I_n) = O_p(1)$.

(C7) $T(M \otimes I_n)'\widehat{X} = O_p(1)$.

Note that our assumption (C3) does not restrict the analysis to square matrices. When the variables of the system are cointegrated with a trend or weakly exogenous variables, A is rectangular (Garrett et al., 2006). We, again, assume that null space estimators are given in (C4). Assumptions (C5) – (C7) are standard in cointegration analysis and are easily checked using population parameters and elementary asymptotics. They hold not only in the cointegration models considered by Johansen (1995) and Lütkepohl (2006) but also in some of its generalizations. For example, assumptions (C1) – (C7) hold in the case where the innovations follow a, possibly non-stationary, conditionally heteroskedastic martingale difference process (Cavaliere et al., 2009) and in the case of non–stationary volatility (Cavaliere et al., 2007). It is, of course, possible to further generalize these assumptions by allowing for more general rates of convergence, which would allow us to cover infinite variance innovations (Caner, 1998) or $I(2)$ processes (Johansen, 2006), but we will stick with the simpler formulation for concreteness.

It is well known in the maximum likelihood estimation of (7.1) that $\widehat{\beta}$ is super-consistent while $\widehat{\alpha}$ is $\sqrt{T}$-consistent, provided these estimators have been suitably identified. If we think of $\widehat{\beta}_\perp$ and $\widehat{\alpha}_\perp$ as estimators of the right and left null space of A then, provided they too have been suitably identified, they will have the same rates of convergence as $\widehat{\beta}$ and $\widehat{\alpha}$ respectively. We now show that in the general case, where the null spaces are not necessarily estimated by maximizing the likelihood, these rates of convergence are necessary for the plug-in principle.

Lemma 3. Under assumptions (C1) – (C7) and $H_0(r)$, if $\sqrt{T}(P_{\widehat{N}} - P_N) = O_p(1)$ and $T(P_{\widehat{M}} - P_M) = O_p(1)$ then $\mathscr{W} - \mathscr{W}_s = o_p(1)$.

Since the null space estimators are obtained from reduced rank approximations we need to find out what conditions the reduced rank approximation must satisfy in order to give us estimators that satisfy the plug-in principle.

Theorem 5. Under assumptions (C1) – (C7) and $H_0(r)$, if $\widehat{A}_r$ is a rank r approximation to $\widehat{A}$, $\widehat{N} \in \mathbb{R}^{n\times(n-r)}$ and $\widehat{M} \in \mathbb{R}^{m\times(m-r)}$ span the left and right null spaces of $\widehat{A}_r$, $\sqrt{T}(\widehat{A} - \widehat{A}_r) = O_p(1)$, and $T\widehat{A}_r M = O_p(1)$, then $\widehat{N}$ and $\widehat{M}$ satisfy the conditions of Lemma 3.

The only thing that remains now is to show that if the null spaces are estimated according to any of the methods of section 5, then the left null space estimator is $\sqrt{T}$-consistent while the right null space estimator is super-consistent.

Theorem 6. Under assumptions (C1) – (C7) and $H_0(r)$, if $\widehat{A}_r$ is one of the rank r approximations to $\widehat{A}$ the conditions of Theorem 5 are satisfied.

We note in closing that, to the author's knowledge, there are only two papers in the literature that have used a different reduced rank approximation than that which arises naturally from maximum likelihood estimation (i.e. the RSD). Kleibergen & van Dijk (1994) test for cointegration using the LU decomposition, while Kleibergen & Paap (2006) use the SVD.

8 Application to Granger Causality Testing

Al-Sadoon (2009b) has extended Granger (1969) causality in multivariate time series by showing that the appropriate test in that context is to test rank rather than the block zero restrictions of Dufour & Renault (1998). Al-Sadoon (2009a) developed the inference theory and presented a bootstrapping procedure that achieves the correct size and has power against causal alternatives. We now compare the fixed-b theory to the bootstrap's performance.

We will work with the n-dimensional, stationary VAR(p) processes, W. Dufour et al. (2006) show that every such VAR is representable as

$$W(t+h) = \mu_h + \sum_{j=1}^{p} \pi_j^{(h)} W(t+1-j) + \sum_{j=0}^{h-1} \psi_j a(t+h-j), \quad t = p-1, \ldots, T-h, \tag{8.1}$$

where $h > 0$ and a is a martingale difference sequence with respect to the information set generated by W, with $\mathbb{E}(a(t)a'(t)) = \Omega > 0$ for all $t > 0$. The first p observations of W are assumed given. We will partition W as $W(t) = (X'(t), Y'(t), Z'(t))'$, $t = 1, \ldots, T$, where the dimensions of the components X, Y, and Z are n_X, n_Y, and n_Z respectively. The coefficient matrices are then partitioned conformably with W as

$$\pi_j^{(h)} = \begin{bmatrix} \pi_{XXj}^{(h)} & \pi_{XYj}^{(h)} & \pi_{XZj}^{(h)} \\ \pi_{YXj}^{(h)} & \pi_{YYj}^{(h)} & \pi_{YZj}^{(h)} \\ \pi_{ZXj}^{(h)} & \pi_{ZYj}^{(h)} & \pi_{ZZj}^{(h)} \end{bmatrix}, \qquad j, h \geq 1.$$

Dufour & Renault (1998) define h–step non–causality as follows: Y fails to cause X at a given horizon h if at every time t the forecast of $X(t+h)$ does not depend on current or past Y. We will denote this by $Y \nrightarrow_h X$.

Result 1 (Theorem 3.1 of Dufour & Renault (1998))**.** $Y \nrightarrow_h X$ if and only if, $\pi_{XYj}^{(h)} = 0$ for all $1 \leq j \leq p$.

Al-Sadoon (2009b) has shown that the Dufour & Renault (1998) framework does not capture the full linear structure of dependence in multivariate time series. In particular, if we fail to reject non–causality, it may still be the case that the causal linkages are weak and occur only along certain subspaces of the variations in X and Y. We say that Y along subspace $\mathcal{V} \subseteq \mathbb{R}^{n_Y}$ fails to cause X along subspace $\mathcal{U} \subseteq \mathbb{R}^{n_X}$ at horizon $h \geq 1$ if eliminating the history of variations of Y along $\mathcal{V}$ from the information set does not change the h-step forecast of X in the direction of $\mathcal{U}$. We denote this by $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$ and note that it is equivalent to $P_{\mathcal{V}}Y \nrightarrow_h P_{\mathcal{U}}X$, where $P_{\mathcal{U}}$ and $P_{\mathcal{V}}$ are the orthogonal projection matrices onto $\mathcal{U}$ and $\mathcal{V}$ respectively. The requisite restrictions for this sort of non–causality are as follows.

Result 2 (Theorem 4.1 of Al-Sadoon (2009b))**.** $Y|_{\mathcal{V}} \nrightarrow_h X|_{\mathcal{U}}$ if and only if, $P_{\mathcal{U}} \pi_{XYj}^{(h)} P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p$.

Now if $\mathcal{U}$ and $\mathcal{V}$ are known then testing for subspace non–causality is easily done by employing a Wald test as in Dufour et al. (2006). Typically, however, we will not know *a priori* along which subspaces non–causality occurs. Thus restrictions of the form $P_{\mathcal{U}}\pi^{(h)}_{XYj}P_{\mathcal{V}} = 0$ for all $1 \leq j \leq p$ are actually rank restrictions, which brings us back to the original insight of Anderson (1951), who first proposed these tests: the appropriate extension of zero restrictions in univariate regressions is not zero block restrictions but reduced rank restrictions.

Now define the maximal subspace $\mathcal{U}$ such that $Y \nrightarrow_h X|_{\mathcal{U}}$ to be $\mathcal{U}^{XY}_h$ and let U^{XY}_h be a matrix of orthonormal columns which span $\mathcal{U}^{XY}_h$. Likewise, the maximal subspace $\mathcal{V}$ such that $Y|_{\mathcal{V}} \nrightarrow_h X$ is denoted by $\mathcal{V}^{XY}_h$ and V^{XY}_h is the matrix of orthonormal columns which span $\mathcal{V}^{XY}_h$.[12] Then we can find U^{XY}_h by finding the appropriate U that satisfies

$$U' \begin{bmatrix} \pi^{(h)}_{XY1} & \cdots & \pi^{(h)}_{XYp} \end{bmatrix} = 0. \tag{8.2}$$

Similarly, V^{XY}_h is the appropriate V that satisfies

$$\begin{bmatrix} \pi^{(h)}_{XY1} \\ \vdots \\ \pi^{(h)}_{XYp} \end{bmatrix} V = 0. \tag{8.3}$$

As demonstrated by Dufour et al. (2006), the relevant parameters for these tests can be obtained by OLS in (8.1). However, because the error term is an MA$(h-1)$ process, care must be taken in testing hypotheses on the model. One option is to use a consistent estimate of the long run covariance of the regressors and the residuals, in which case assumptions (A1) – (A6) and (S1) – (S4) hold. The second option is to use a HAC robust test in which the bandwidth is a fixed proportion of T, in which case assumptions (A1) – (A6) and (N1) – (N2) hold. For simplicity, we will use a (Newey & West, 1987) non-parametric estimate of the long run covariance matrix of the estimated parameters. In case of consistent estimation, we fixed the bandwidth at $h-1$. This is perhaps too small, but doubling it produced no significant difference to the results and so we will continue to follow Dufour et al. (2006) who proposed this value. In the case of HAC robust inference, we will consider the case where the bandwidth is set to T, the full sample. As for the null space estimators, we have chosen to work with the SVD as it is the most convenient to compute.[13]

Now consider the Bernanke & Mihov (1998) data set which consists of US monthly data on real GDP growth, GDP, inflation, P, the growth of non–borrowed reserves, NBR, and

[12] Lemma 3.3 of Al-Sadoon (2009b) proves that these subspaces are unique. Therefore the matrices are defined up to a rotation of the columns.

[13] See Al-Sadoon (2009a) for details of the estimation and inference.

the percentage change in the federal funds rate, r for the period January 1965 to December 1996. This is the data set used by Dufour et al. (2006) who chose $p = 16$ as it minimizes the Akaike information criterion.

To compare size and power we followed the procedure of Dufour et al. (2006). First, we estimated the model at horizon $h = 12$ to obtain a matrix of coefficients $\widetilde{B}_{12} = [\,\pi_1^{(12)} \cdots \pi_p^{(12)}\,]$. Then we reestimated the model under the restriction $r \nrightarrow_{12} (GDP, P)|_{\mathcal{U}}$, where $\mathcal{U}$ is the predicted subspace of non-causality. This restriction is imposed as, $D_{12}^1 \text{vec}(\widetilde{B}_{12}^1) = 0$, where $\widetilde{B}_{12}^1$ consists of the parameters of the restricted model and D_{12}^1 is a restriction matrix. Next, we perturbed the restricted model to $\widetilde{B}_{12}^{1,\theta}$ so that, $D_{12}^1 \text{vec}(\widetilde{B}_{12}^{1,\theta}) = \theta e_1$, where e_1 is the first standard basis vector and θ is a small number. We will refer to this alternative as $H_0^\theta(1)$. We then simulated 100 data sets using $\widetilde{B}_{12}^{1,\theta}$, for each $\theta = 0, 0.001, \ldots, 0.01$ (larger values of θ led to explosive models) and tabulated the rejection rates of each test.

Table 8.1 gives the rejection rates of $H_0(0)$ and $H_0(1)$ for the first case of consistent covariance estimators. The test achieves the correct size of 0.05 but its power to reject $H_0(0)$ is not terribly strong as 22% of the $H_0(0)$ tests were accepted. Thus the testing procedure may underestimate the correct subspaces causality rank. As we have shown above, this problem disappears asymptotically but (as discussed at length in Camba-Mendez et al. (2003)) may occur in small samples. The testing producer does, however, exhibit power against the $\theta \neq 0$ cases, particulary in the $H_0(0)$ tests.

Table 8.1: Rejection Frequencies (Consistent Covariance Inference).

Significance 5%	$\theta =$	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.010
Reject $H_0(0)$		0.78	0.81	0.85	0.85	0.87	0.87	0.89	0.9	0.92	0.95	0.93
Reject $H_0(1)$		0.05	0.08	0.08	0.13	0.16	0.18	0.31	0.37	0.45	0.49	0.55
Reject $H_0(0)$ and accept $H_0(1)$		0.74	0.73	0.77	0.72	0.71	0.69	0.58	0.53	0.48	0.47	0.4
Reject $H_0(0)$ and reject $H_0(1)$		0.04	0.08	0.08	0.13	0.16	0.18	0.31	0.37	0.44	0.48	0.53

Table 8.2: Rejection Frequencies (HAC Robust Inference).

Significance 5%	$\theta =$	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.010
Reject $H_0(0)$		0.92	0.9	0.93	0.93	0.94	0.95	0.95	0.97	0.97	0.99	0.98
Reject $H_0(1)$		0.11	0.11	0.15	0.17	0.21	0.26	0.33	0.42	0.52	0.64	0.71
Reject $H_0(0)$ and accept $H_0(1)$		0.82	0.8	0.78	0.76	0.74	0.69	0.63	0.55	0.46	0.35	0.27
Significance 1%	$\theta =$	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.010
Reject $H_0(0)$		0.82	0.78	0.82	0.81	0.83	0.85	0.86	0.9	0.9	0.94	0.96
Reject $H_0(1)$		0.04	0.03	0.05	0.07	0.08	0.09	0.15	0.21	0.29	0.39	0.49
Reject $H_0(0)$ and accept $H_0(1)$		0.78	0.75	0.77	0.74	0.75	0.77	0.72	0.69	0.61	0.56	0.48

Now consider the same exercise with the HAC robust inference instead. At significance

5%, the HAC robust tests are oversized, making it difficult to compare the size and power. So we report the frequency of rejections at 1% significance, which achieves a $H_0(1)$ size slightly less than 4% and may be thought of as a size correction (we are really aiming for a simple rule of thumb). The test at 1% outperforms the bootstrap test in terms of power against $H_0(0)$ and in terms of correct detection of rank but its power against $H_0^\theta(1)$ is lower than the bootstrap.

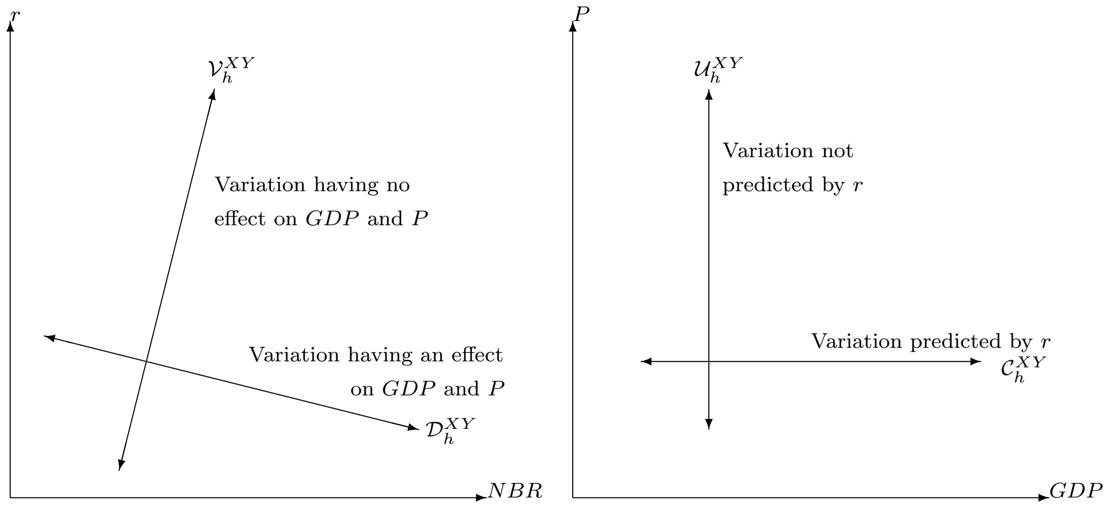
We conclude that the HAC robust inference is a viable alternative to the bootstrap, provided we adjust for size distortions. We then proceed to apply the procedures to the full data set. The results are summarized in Tables 10.1–10.4. Clearly the general qualitative picture of the causal structure is the same under both tests.

The first four lines are almost identical to the analogous ones in Dufour et al. (2006). The small variations from their results are due to simulation error. The notable effects in the univariate tests above are that r has predictive power for GDP over the range from 7 months ahead to 18 months ahead while NBR is predictive for P for up to three months ahead and has no predictive power beyond that.

From the tests $(NBR, r) \nrightarrow_h (GDP, P)$ we can see that monetary policy has predictive power for output and inflation in the forecast horizons 1–4 and 7–9. From the tests $(NBR, r) \nrightarrow_h (GDP, P)|_{\mathcal{U}}$, we conclude that there is some evidence of subspace non–causality but it is not terribly strong as the subspace non–causality hypothesis is barely accepted. From the tests $(NBR, r)|_{\mathcal{V}} \nrightarrow_h (GDP, P)$, on the other hand, we can see substantial evidence of subspace non–causality. There appears to be a statistically significant relationship between NBR and r along which policy has no predictive power for the other macroeconomic variables. From the associated V_h^{XY} we can see that increases in NBR that are offset by increases r have no effect on GDP and P over all the horizons where monetary policy has effect. Thus the line along $(V_h^{XY})_\perp$ can be interpreted as measuring the statistical tradeoff between policy instruments with respect to output and inflation. This is illustrated in the first graph of Figure 8.1.

Next we look at the predictive power of the individual policy variables for output and inflation. The tests, $NBR \nrightarrow_h (GDP, P)$, show that NBR leads GDP and P for horizons 1–3 months. There is evidence of subspace non–causality as can be seen from the $NBR \nrightarrow_h (GDP, P)|_{\mathcal{U}}$ tests but it is not very strong. As for the effect of r on GDP and P we can see

Figure 8.1: Subspace Causality in Monetary Policy.



that variations of the interest rate have tended to precede variations of output and inflation along an ever so slightly positively sloped line. While this may be consistent with there being a statistical Phillips curve (Stock & Watson, 1999), the slope is too flat to be conclusive and is actually negative at horizon 7. Therefore we conclude that most of the predictive power for r is for GDP rather than P.[14] See Figure 8.1.

Finally, we consider the effect of the policy variables together on output and inflation individually. Again we find substantial scope for subspace non–causality. In particular, the policy tradeoff observed above emerges much more clearly in these tests (see Figure 8.1). Al-Sadoon (2009a) also check for robustness to the assumption of stationarity and finds no significant difference to the results.

9 Conclusion

This paper has developed a simple and unified theory of rank testing, which includes most previous rank tests as special cases. It has clarified the relationships between rank tests and the matrix decompositions on which they are based. It has also proposed new HAC rank tests based on fixed-b asymptotics, which performs comparably to the bootstrap. Here we will discuss venues for future research.

First, we have shown (indirectly) that estimators of zero canonical correlations and singu-

[14] Indeed, we ran the tests on quarterly data and found that the slopes were even less consistently positive.

lar values have simple distributions. This begs the question of whether estimators of non–zero canonical correlations and singular values are just as easily analyzed. It would also be interesting to see whether this work can be extended to eigenvalues in general. Such a theory would be very valuable, not just for multivariate statistical analysis, but also for time series, cointegration, factor analysis, etc.

Second, a recent area of active research has been large dimensional factor analysis (Bai & Ng, 2008). However, to the author's knowledge, only Onatski (2009) formulates a test of the number of factors, which is closely related to rank testing. Extending the stochastic Wald test to this context has the potential of greatly simplifying the analysis of these models as well as relaxing the assumptions necessary for Onatski's result.

Third, rank tests have been used as ingredients of identification-robust hypothesis test statistics (Kleibergen, 2005). It would be interesting to see how the new HAC robust test proposed in this paper can be used in the weak identification literature and whether it can offer improvements over previous tests.

10 Appendix

10.1 Proofs

Proof of Theorem 1. We prove this by constructing a sequence of matrices $\widehat{U}$ which satisfies, $\widehat{U} \xrightarrow{p} U$, $\sqrt{T}\widehat{U}'X = o_p(1)$, $\widehat{U} = \widehat{C}\widehat{\phi}$, and $U = C\phi$, for invertible matrices $\widehat{\phi}$ and ϕ so that the stochastic Wald statistics based on $\widehat{U}$ and $\widehat{C}$ are equivalent and the one based on $\widehat{U}$ satisfies the conditions of Lemma 1.

First, recall that $P_C = UU'$ for some $U \in \mathbb{R}^{p\times q}$ of rank q by the spectral theorem (Theorem 2.5.4 of Horn & Johnson (1985)). Now consider the sequence of matrices $\{P_{\widehat{C}}U\}$. We claim that the number of matrices in this sequence of rank less than q is almost surely finite. Otherwise, there exist an infinite subsequence $\{T_i\}$ and non–zero random vectors $\widehat{x}_i \in \mathbb{R}^q$ such that $P_{\widehat{C}}U\widehat{x}_i = 0$ along T_i. If we normalize these random vectors so that $\|\widehat{x}_i\| = 1$ then the sequence is uniformly bounded and by Prohorov's theorem (van der Vaart, 1998, Theorem 2.4), there exists a further subsequence i_k and a random vector $x \in \mathbb{R}^q$ such that $\widehat{x}_{i_k} \xrightarrow{d} x$ as $k \to \infty$. Then for any given $\varepsilon > 0$, $P(\|x\| > \varepsilon) = P\left(\bigcup_{n=1}^{\infty}\{\|x\| \geq \frac{1}{n}\}\right) \geq P(\|x\| \geq \frac{1}{2}) \geq \limsup_k P(\|\widehat{x}_{i_k}\| \geq \frac{1}{2}) = 1$, where the last inequality follows from Portmanteau's lemma (van der Vaart, 1998, Lemma 2.2). Thus x is almost surely non-zero. Finally, along T_{i_k}, $0 = P_{\widehat{C}} U \widehat{x}_{i_k} \xrightarrow{d} P_C U x = Ux$, a contradiction since U is of full rank. Thus, $N_0 = \inf\{T : \operatorname{rank}(P_{\widehat{C}} U) < q\} < \infty$ almost surely. Now for $T \leq N_0$ choose $\widehat{U}$ so that $P_{\widehat{C}} = \widehat{U}\widehat{U}'$ is a spectral decomposition and for $T > N_0$ set $\widehat{U} = P_{\widehat{C}} U$. Then we have,

$$
\begin{aligned}
P(\|\widehat{U} - U\| > \varepsilon) &= P(\|\widehat{U} - U\| > \varepsilon, T \leq N_0) + P(\|\widehat{U} - U\| > \varepsilon, T > N_0) \\
&\leq P(T \leq N_0) + P(\|P_{\widehat{C}} U - U\| > \varepsilon) \underset{T\uparrow\infty}{\to} 0,
\end{aligned}
$$

because N_0 is almost surely finite and $P_{\widehat{C}} U \xrightarrow{p} U$. Clearly, $P_{\widehat{U}} = P_{\widehat{C}}$ and $P_U = P_C$ so that $\widehat{\phi} = (\widehat{C}'\widehat{C})^{-1}\widehat{C}'\widehat{U}$ and $\phi = (C'C)^{-1}C'U$. Finally, direct substitution into, $P_{\widehat{C}} X = o_p(1)$ yields, $\sqrt{T}\widehat{U}(\widehat{U}'\widehat{U})^{-1}\widehat{X}$ we now multiply by $\widehat{U}'$ we arrive at the desired result. $\square$

Proof of Theorem 2. According to Theorem 13.5.1 of Gohberg et al. (2006), $\|P_{\widehat{M}} - P_M\|_2 \leq K\|\widehat{A}_r - A\|_2$, where $K > 0$ depends only on A. Applying the inequality with $\widehat{A}_r$ and A transposed, we obtain a similar bound on $\|P_{\widehat{N}} - P_N\|_2$. The result then following from Lemma 2 on noting that the right hand sides are $O_p(T^{-\frac{1}{2}})$. $\square$

Proof of Proposition 1. Our approach mirrors the method employed by (Folland, 1999, Theorem 5.24) in proving the projection theorem in Hilbert space. Let $\widehat{N} \in \mathbb{R}^{n\times(n-r)}$ and $\widehat{M} \in \mathbb{R}^{m\times(m-r)}$ be full ranked matrices, orthogonal to the columns and rows of $\widehat{A}_r^{CD}$ respectively. Then for $h \in \mathbb{R}^{r^2}$ the matrix, $\widehat{A}_r^{CD} + \operatorname{mat}((\widehat{M}_\perp \otimes \widehat{N}_\perp)h)$ has a rank of at most r and for small enough $\|h\|$ it has a rank of exactly r. It then follows from the definition of the CDA that for small enough $\|h\|$, $\|\widehat{A} - \widehat{A}_r^{CD} - \operatorname{mat}((\widehat{M}_\perp \otimes \widehat{N}_\perp)h)\|^2_{\widehat{\Omega}} \geq \|\widehat{A} - \widehat{A}_r^{CD}\|^2_{\widehat{\Omega}}$. The left hand side is a quadratic form in h and has a local minimum at $h = 0$. Setting its derivative at $h = 0$ to zero we arrive at,

$$
\operatorname{vec}'(\widehat{A} - \widehat{A}_r^{CD})\widehat{\Omega}^{-1}(\widehat{M}_\perp \otimes \widehat{N}_\perp) = 0.
$$

By the same logic we can show that,

$$
\begin{aligned}
\operatorname{vec}'(\widehat{A} - \widehat{A}_r^{CD})\widehat{\Omega}^{-1}(\widehat{M} \otimes \widehat{N}_\perp) &= 0 \\
\operatorname{vec}'(\widehat{A} - \widehat{A}_r^{CD})\widehat{\Omega}^{-1}(\widehat{M}_\perp \otimes \widehat{N}) &= 0.
\end{aligned}
$$

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

$$
\operatorname{vec}'(\widehat{A} - \widehat{A}_r^{CD})\widehat{\Omega}^{-1}(\widehat{M} \otimes \widehat{N})_\perp = 0. \tag{10.1}
$$

Using (2.1) we get,

$$
\begin{aligned}
\widehat{\Omega}^{-1} &= \widehat{\Omega}^{-\frac{1}{2}} I_{nm} \widehat{\Omega}^{-\frac{1}{2}} \\
&= (\widehat{M} \otimes \widehat{N})\{(\widehat{M} \otimes \widehat{N})'\widehat{\Omega}(\widehat{M} \otimes \widehat{N})\}^{-1}(\widehat{M} \otimes \widehat{N})' \\
&\quad + \widehat{\Omega}^{-1}(\widehat{M} \otimes \widehat{N})_{\perp}\{(\widehat{M} \otimes \widehat{N})'_{\perp}\widehat{\Omega}^{-1}(\widehat{M} \otimes \widehat{N})_{\perp}\}^{-1}(\widehat{M} \otimes \widehat{N})'_{\perp}\widehat{\Omega}^{-1}. \qquad (10.2)
\end{aligned}
$$

Substituting (10.1) and (10.2) into the Cragg and Donald statistic proves the proposition. □

Proof of Lemma 3. Following the same line of argument as we did in section 3, we find that $\mathscr{W}_s - \mathscr{W} = o_p(1)$ if

$$TC'\widehat{\Omega}C - T\widehat{C}'\widehat{\Omega}\widehat{C} = o_p(1) \qquad (10.3)$$

$$T(\widehat{C} - C)'\widehat{X} = o_p(1). \qquad (10.4)$$

By the same construction as that used in the proof of Theorem 1, we can assume that our estimators satisfy $\sqrt{T}(\widehat{N} - N) = O_p(1)$ and $T(\widehat{M} - M) = O_p(1)$. Before we begin proving (10.3) and (10.4), we note the following useful fact

$$
\begin{aligned}
T(\widehat{C} - C)'(P_{M_\perp} \otimes I_n) &= T\widehat{C}'(P_{M_\perp} \otimes I_n) \\
&= T(\widehat{M} \otimes \widehat{N})'(P_{M_\perp} \otimes I_n) \\
&= \underbrace{((T(\widehat{M} - M)'P_{M_\perp}) \otimes I_{n-r})}_{O_p(1)}(I_m \otimes \widehat{N})' = O_p(1). \qquad (10.5)
\end{aligned}
$$

Note that multiplying on the right by $\widehat{X}$ we obtain

$$T(\widehat{C} - C)'(P_{M_\perp} \otimes I_n)\widehat{X} = ((T(\widehat{M} - M)'P_{M_\perp}) \otimes I_{n-r})\underbrace{(I_m \otimes \widehat{N})'\widehat{X}}_{o_p(1)} = o_p(1), \qquad (10.6)$$

because $\widehat{N}'\widehat{A} \xrightarrow{p} N'A = 0$.

Now we decompose the left hand side of (10.3) as follows

$$TC'\widehat{\Omega}(C - \widehat{C}) - T(\widehat{C} - C)'\widehat{\Omega}\widehat{C} = C'T(P_M \otimes I_n)\widehat{\Omega}(P_M \otimes I_n)(C - \widehat{C}) - T(\widehat{C} - C)'\widehat{\Omega}\widehat{C}, \qquad (10.7)$$

Consider the first term in (10.7). Since $\widehat{C} - C = o_p(1)$ it suffices to prove that $T(P_M \otimes I_n)\widehat{\Omega}(P_M \otimes I_n) = O_p(1)$. Now it is easy to show, using equation (2.1) that,

$$
\begin{aligned}
\{T(M \otimes I_n)'\widehat{\Omega}(M \otimes I_n)\}^{-1} = \frac{1}{T}\Big\{&(M \otimes I_n)'\widehat{\Omega}^{-1}(M \otimes I_n) \cdots \\
&- (M \otimes I_n)'\widehat{\Omega}^{-1}(M_\perp \otimes I_n)\{(M_\perp \otimes I_n)'\widehat{\Omega}^{-1}(M_\perp \otimes I_n)\}^{-1}(M_\perp \otimes I_n)'\widehat{\Omega}^{-1}(M \otimes I_n)\Big\},
\end{aligned}
$$

which is $O_p(1)$ by assumptions (C5) and (C6), a fact that we will need in the proof of Theorem 6. It follows that $T(M \otimes I_n)'\widehat{\Omega}(M \otimes I_n) = \frac{1}{T}\{(M \otimes I_n)'\widehat{\Omega}^{-1}(M \otimes I_n)\}^{-1} + o_p(1) = O_p(1)$ by assumption (C5). Thus the first term in (10.7) is $o_p(1)$.

Now consider the second term of (10.7). This can be decomposed as follows

$$\underbrace{(\widehat{C}-C)'}_{o_p(1)}\underbrace{T(P_M \otimes I_n)\widehat{\Omega}(P_M \otimes I_n)}_{O_p(1)}\widehat{C} + \underbrace{(\widehat{C}-C)'}_{o_p(1)}(P_M \otimes I_n)\widehat{\Omega}\underbrace{T(P_{M_\perp} \otimes I_n)\widehat{C}}_{O_p(1)}$$
$$+\underbrace{T(\widehat{C}-C)'(P_{M_\perp} \otimes I_n)}_{O_p(1)}\underbrace{\widehat{\Omega}(P_M \otimes I_n)}_{o_p(1)}\widehat{C} + \underbrace{T(\widehat{C}-C)'(P_{M_\perp} \otimes I_n)}_{O_p(1)}\widehat{\Omega}\underbrace{(P_{M_\perp} \otimes I_n)\widehat{C}}_{o_p(1)} = o_p(1).$$

Equation (10.5) is used in all but the first term. As for (10.4),

$$T(\widehat{C}-C)'\widehat{X} = \underbrace{(\widehat{C}-C)'}_{o_p(1)}\underbrace{T(P_M \otimes I_n)\widehat{X}}_{O_p(1)} + \underbrace{T(\widehat{C}-C)'(P_{M_\perp} \otimes I_n)\widehat{X}}_{o_p(1)}, \tag{10.8}$$

by assumption (C7) and equation (10.6). □

Proof of Theorem 5. Because $\sqrt{T}(\widehat{A}-A) = O_p(1)$, it follows, just as in the proof of Theorem 2, that $\sqrt{T}(P_{\widehat{N}} - P_N) = O_p(1)$. As for the second condition, let $\widehat{U}_{\cdot 1}\widehat{S}_1\widehat{V}_{\cdot 1}' = \widehat{A}_r$ be the rank-r SVD approximation to $\widehat{A}_r$. Since $\widehat{U}_{\cdot 1} = O_p(1)$ and $\widehat{S}_1^{-1} = O_p(1)$ under $H_0(r)$, then $T\widehat{A}_r M = O_p(1)$ implies that $T\widehat{V}_{\cdot 1}'M = O_p(1)$. If M is chosen to have orthonormal columns, then by theorem 2.6.1 of Golub & Van Loan (1996), $\|P_{\widehat{M}} - P_M\|_2 = \|\widehat{V}_{\cdot 1}'M\|_2$, which concludes the proof. □

Proof of Theorem 6. First, we show that for any of the given rank r approximation, $\widehat{A}_r$, $\sqrt{T}(\widehat{A}-\widehat{A}_r) = O_p(1)$. Choose a version of $[\, M_\perp \; M \,]$ that is unitary. Then,

$$\sigma_{r+1}(\widehat{A}) = \sigma_{r+1}\left(\widehat{A}\begin{bmatrix} M_\perp & M \end{bmatrix}\right) \leq \sigma_1(\widehat{A}M),$$

where the inequality follows from corollary 3.1.3 of Horn & Johnson (1991). It follows that $T\sigma_{r+1}(\widehat{A}) = O_p(1)$. From equations (5.8) and (5.9), we get $T(\widehat{A}_r^{LU} - \widehat{A}) = O_p(1)$ and $T(\widehat{A}_r^{QR} - \widehat{A}) = O_p(1)$, which is even stronger than what we set out to prove but we will need this later. As for the CDA (and by inclusion, the RSD and SVD), we have that $T\|\widehat{A}-\widehat{A}_r^{CD}\|^2 \leq \sigma_1^2(\widehat{\Omega})\mathscr{W}_s$. Since $\widehat{\Omega}$ converges, it suffices to show that $\mathscr{W}_s = O_p(1)$. From the definition of the CDA,

$$\mathscr{W}_s \leq T\|\widehat{A}-A\|_{\widehat{\Omega}}^2 = \Big\{T\mathrm{vec}'(\widehat{A})(M \otimes I_n)\{T(M \otimes I_n)'\widehat{\Omega}(M \otimes I_n)\}^{-1}T(M \otimes I_n)'\mathrm{vec}(\widehat{A})\cdots$$
$$+\sqrt{T}\mathrm{vec}'(\widehat{A}-A)\widehat{\Omega}^{-1}(M_\perp \otimes I_n)\{(M_\perp \otimes I_n)'\widehat{\Omega}^{-1}(M_\perp \otimes I_n)\}^{-1}(M_\perp \otimes I_n)'\widehat{\Omega}^{-1}\sqrt{T}\mathrm{vec}(\widehat{A}-A)\Big\}.$$

In the first term, $T(M \otimes I_n)'\text{vec}(\widehat{A}) = O_p(1)$ by assumption (C7), while the fact that $\{T(M \otimes I_n)'\widehat{\Omega}(M \otimes I_n)\}^{-1} = O_p(1)$ was shown in the proof of Lemma 3. The second term is $O_p(1)$ by assumptions (C2), (C5) and (C6).

Next, we show that for any of the given rank r approximation, $\widehat{A}_r$, $T\widehat{A}_r M = O_p(1)$. Since $T(\widehat{A}_r^{LU} - \widehat{A}) = O_p(1)$ and $T(\widehat{A}_r^{QR} - \widehat{A}) = O_p(1)$, this implies that $T\widehat{A}_r^{LU} M = O_p(1)$ and $T\widehat{A}_r^{QR} M = O_p(1)$ by assumption (C7). As for the CDA,

$$\begin{aligned} T^2\|(\widehat{A} - \widehat{A}_r^{CD})M\|^2 &= T^2\|(M \otimes I_n)'\text{vec}(\widehat{A} - \widehat{A}_r^{CD})\|^2 \\ &\leq \sigma_1^2(T(M \otimes I_n)'\widehat{\Omega}(M \otimes I_n))T^2\|(M \otimes I_n)'\text{vec}(\widehat{A} - \widehat{A}_r^{CD})\|^2_{T(M\otimes I_n)'\widehat{\Omega}(M\otimes I_n)} \\ &\leq \sigma_1^2(T(M \otimes I_n)'\widehat{\Omega}(M \otimes I_n))\mathscr{W}_s, \end{aligned}$$

where the last inequality follows from equation (2.1) so that $\mathscr{W}_s$ is the stochastic Wald statistic based on the CDA. $\square$

10.2 Empirical Results

Table 10.1: Causality Test Simulated p-values and Non–causal Directions for the Horizons 1-12. (Consistent Covariance Inference).

h	1	2	3	4	5	6	7	8	9	10	11	12
$NBR \nrightarrow_h GDP$	0.199	0.273	0.282	0.394	0.833	0.736	0.815	0.710	0.357	0.277	0.375	0.333
$NBR \nrightarrow_h P$	**0.003**	**0.015**	**0.005**	0.136	0.168	0.168	0.118	0.154	0.219	0.200	0.504	0.582
$r \nrightarrow_h GDP$	0.238	0.172	0.054	0.073	0.080	0.077	**0.004**	**0.004**	**0.001**	**0.001**	**0.003**	**0.001**
$r \nrightarrow_h P$	0.111	0.121	0.139	0.360	0.414	0.398	0.327	0.358	0.202	0.142	0.426	0.757
$(NBR, r) \nrightarrow_h (GDP, P)$	**0.001**	**0.001**	**0.001**	**0.007**	0.059	0.090	**0.016**	**0.019**	**0.025**	0.197	0.784	0.509
$(NBR, r) \nrightarrow_h (GDP, P)\vert_{\mathcal{U}}$	**0.008**	**0.013**	**0.017**	0.051			0.070	0.067	0.255			
U_h^{XY}				$\begin{bmatrix}0.0952\\0.9955\end{bmatrix}$			$\begin{bmatrix}-0.0642\\0.9979\end{bmatrix}$	$\begin{bmatrix}-0.0979\\0.9952\end{bmatrix}$	$\begin{bmatrix}-0.1550\\0.9879\end{bmatrix}$			
$(NBR, r)\vert_{\mathcal{V}} \nrightarrow_h (GDP, P)$	0.274	0.258	0.137	0.192			0.445	0.463	0.212			
V_h^{XY}	$\begin{bmatrix}0.0576\\0.9983\end{bmatrix}$	$\begin{bmatrix}0.0604\\0.9982\end{bmatrix}$	$\begin{bmatrix}0.0412\\0.9992\end{bmatrix}$	$\begin{bmatrix}0.0327\\0.9995\end{bmatrix}$			$\begin{bmatrix}0.1532\\0.9882\end{bmatrix}$	$\begin{bmatrix}0.1631\\0.9866\end{bmatrix}$	$\begin{bmatrix}0.1681\\0.9858\end{bmatrix}$			
$NBR \nrightarrow_h (GDP, P)$	**0.008**	**0.035**	**0.040**	0.148	0.304	0.332	0.144	0.166	0.268	0.362	0.569	0.675
$NBR \nrightarrow_h (GDP, P)\vert_{\mathcal{U}}$	0.056	0.080	0.180									
U_h^{XY}	$\begin{bmatrix}0.1489\\0.9888\end{bmatrix}$	$\begin{bmatrix}0.1206\\0.9927\end{bmatrix}$	$\begin{bmatrix}0.1282\\0.9918\end{bmatrix}$									
$r \nrightarrow_h (GDP, P)$	0.107	0.113	0.076	0.141	0.079	0.204	**0.022**	**0.023**	**0.007**	**0.003**	**0.010**	**0.029**
$r \nrightarrow_h (GDP, P)\vert_{\mathcal{U}}$							0.196	0.309	0.213	0.086	0.353	0.734
U_h^{XY}							$\begin{bmatrix}0.0147\\0.9999\end{bmatrix}$	$\begin{bmatrix}-0.0278\\0.9996\end{bmatrix}$	$\begin{bmatrix}-0.0427\\0.9991\end{bmatrix}$	$\begin{bmatrix}-0.0231\\0.9997\end{bmatrix}$	$\begin{bmatrix}-0.0408\\0.9992\end{bmatrix}$	$\begin{bmatrix}0.0154\\0.9999\end{bmatrix}$
$(NBR, r) \nrightarrow_h GDP$	**0.021**	**0.030**	**0.002**	**0.019**	0.066	0.089	**0.014**	**0.004**	**0.003**	**0.002**	**0.014**	**0.023**
$(NBR, r)\vert_{\mathcal{V}} \nrightarrow_h GDP$	0.497	0.455	0.248	0.113			0.650	0.565	0.511	0.422	0.308	0.075
V_h^{XY}	$\begin{bmatrix}0.0676\\0.9977\end{bmatrix}$	$\begin{bmatrix}0.0689\\0.9976\end{bmatrix}$	$\begin{bmatrix}0.0459\\0.9989\end{bmatrix}$	$\begin{bmatrix}0.0325\\0.9995\end{bmatrix}$			$\begin{bmatrix}0.2054\\0.9787\end{bmatrix}$	$\begin{bmatrix}0.2204\\0.9754\end{bmatrix}$	$\begin{bmatrix}0.2155\\0.9765\end{bmatrix}$	$\begin{bmatrix}0.2087\\0.9780\end{bmatrix}$	$\begin{bmatrix}0.1888\\0.9820\end{bmatrix}$	$\begin{bmatrix}0.1316\\0.9913\end{bmatrix}$
$(NBR, r) \nrightarrow_h P$	**0.001**	**0.001**	**0.001**	**0.043**	0.069	0.066	0.073	**0.037**	0.103	0.288	0.735	0.914
$(NBR, r)\vert_{\mathcal{V}} \nrightarrow_h P$	0.103	0.108	0.137	0.405				0.254				
V_h^{XY}	$\begin{bmatrix}-0.0015\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0070\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0086\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0336\\0.9994\end{bmatrix}$				$\begin{bmatrix}-0.0165\\0.9999\end{bmatrix}$				

Table 10.2: Causality Test Simulated p-values and Non–causal Directions for the Horizons 13-24. (Consistent Covariance Inference).

h	13	14	15	16	17	18	19	20	21	22	23	24
$NBR \nrightarrow_h GDP$	0.504	0.717	0.572	0.773	0.941	0.907	0.753	0.759	0.578	0.513	0.341	0.237
$NBR \nrightarrow_h P$	0.304	0.346	0.280	0.444	0.814	0.875	0.464	0.144	0.262	0.196	0.309	0.568
$r \nrightarrow_h GDP$	**0.002**	**0.016**	**0.014**	**0.025**	**0.047**	**0.033**	0.056	0.057	0.201	0.107	0.210	0.263
$r \nrightarrow_h P$	0.616	0.361	0.365	0.430	0.698	0.492	0.428	0.264	0.185	0.257	0.138	0.283
$(NBR, r) \nrightarrow_h (GDP, P)$	0.090	0.108	0.308	0.417	0.470	0.702	0.661	0.313	0.555	0.681	0.886	0.863
$(NBR, r) \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$												
U_h^{XY}												
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h (GDP, P)$												
V_h^{XY}												
$NBR \nrightarrow_h (GDP, P)$	0.558	0.783	0.541	0.559	0.893	0.724	0.560	0.464	0.561	0.770	0.793	0.689
$NBR \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$												
U_h^{XY}												
$r \nrightarrow_h (GDP, P)$	**0.004**	**0.009**	**0.025**	**0.024**	0.142	0.094	0.265	0.076	0.119	0.233	0.261	0.206
$r \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	0.463	0.248	0.414	0.723								
U_h^{XY}	$\begin{bmatrix} -0.0115 \\ 0.9999 \end{bmatrix}$	$\begin{bmatrix} -0.0320 \\ 0.9995 \end{bmatrix}$	$\begin{bmatrix} -0.0365 \\ 0.9993 \end{bmatrix}$	$\begin{bmatrix} -0.0703 \\ 0.9975 \end{bmatrix}$								
$(NBR, r) \nrightarrow_h GDP$	**0.003**	**0.046**	**0.031**	0.085	0.192	0.247	**0.039**	**0.046**	0.070	**0.032**	0.175	0.127
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h GDP$	0.056	0.070	0.079				**0.035**	**0.038**		0.111		
V_h^{XY}	$\begin{bmatrix} 0.1639 \\ 0.9865 \end{bmatrix}$	$\begin{bmatrix} 0.1305 \\ 0.9914 \end{bmatrix}$	$\begin{bmatrix} 0.1507 \\ 0.9886 \end{bmatrix}$							$\begin{bmatrix} 0.0431 \\ 0.9991 \end{bmatrix}$		
$(NBR, r) \nrightarrow_h P$	0.627	0.602	0.732	0.872	0.932	0.953	0.789	0.233	0.328	0.818	0.680	0.897
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h P$												
V_h^{XY}												

Table 10.3: Causality Test p-values and Non–causal Directions for the Horizons 1-12. (HAC Robust Inference).

h	1	2	3	4	5	6	7	8	9	10	11	12
$NBR \nrightarrow_h GDP$	**0.0083**	0.2125	0.0853	0.1012	0.5253	0.3672	0.4726	0.2133	0.0645	0.0598	0.1466	0.1159
$NBR \nrightarrow_h P$	**0.0028**	**0.0013**	**0.0005**	**0.0072**	0.0312	0.0602	0.0318	0.0602	0.1729	0.0689	0.1898	0.2970
$r \nrightarrow_h GDP$	0.2151	0.0732	0.0433	0.0637	0.0809	0.0693	0.0474	**0.0045**	**0.0003**	**0.0000**	**0.0001**	**0.0000**
$r \nrightarrow_h P$	0.2983	0.1044	0.1486	0.5907	0.2325	0.5050	0.4533	0.5579	0.3276	0.4471	0.6331	0.7763
$(NBR, r) \nrightarrow_h (GDP, P)$	**0.0000**	**0.0004**	**0.0000**	**0.0006**	**0.0011**	**0.0041**	**0.0005**	**0.0015**	**0.0013**	0.0274	0.2915	0.0994
$(NBR, r) \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	**0.0085**	**0.0053**	**0.0007**	**0.0055**	**0.0021**	**0.0019**	0.0148	0.0109	0.0288			
U_h^{XY}							$\begin{bmatrix}-0.0642\\0.9979\end{bmatrix}$	$\begin{bmatrix}-0.0979\\0.9952\end{bmatrix}$	$\begin{bmatrix}-0.1550\\0.9879\end{bmatrix}$			
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h (GDP, P)$	0.5004	0.2168	0.1368	0.3142	0.0170	0.3507	0.1030	0.0742	0.0176			
V_h^{XY}	$\begin{bmatrix}0.0576\\0.9983\end{bmatrix}$	$\begin{bmatrix}0.0604\\0.9982\end{bmatrix}$	$\begin{bmatrix}0.0412\\0.9992\end{bmatrix}$	$\begin{bmatrix}0.0327\\0.9995\end{bmatrix}$	$\begin{bmatrix}0.0615\\0.9981\end{bmatrix}$	$\begin{bmatrix}0.1510\\0.9885\end{bmatrix}$	$\begin{bmatrix}0.1532\\0.9882\end{bmatrix}$	$\begin{bmatrix}0.1631\\0.9866\end{bmatrix}$	$\begin{bmatrix}0.1681\\0.9858\end{bmatrix}$			
$NBR \nrightarrow_h (GDP, P)$	**0.0039**	**0.0028**	**0.0026**	0.0173	0.0467	0.0655	0.0617	0.0396	0.0501	0.1067	0.2456	0.3022
$NBR \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$	0.0653	0.0259	0.0221									
U_h^{XY}	$\begin{bmatrix}0.1489\\0.9888\end{bmatrix}$	$\begin{bmatrix}0.1206\\0.9927\end{bmatrix}$	$\begin{bmatrix}0.1282\\0.9918\end{bmatrix}$									
$r \nrightarrow_h (GDP, P)$	0.3734	0.1160	0.1407	0.3514	0.0196	0.2871	**0.0067**	0.0297	**0.0007**	**0.0003**	**0.0011**	**0.0004**
$r \nrightarrow_h (GDP, P)\|_{\mathcal{U}}$							0.3330		0.3784	0.5214	0.6452	0.7339
U_h^{XY}							$\begin{bmatrix}0.0147\\0.9999\end{bmatrix}$		$\begin{bmatrix}-0.0427\\0.9991\end{bmatrix}$	$\begin{bmatrix}-0.0231\\0.9997\end{bmatrix}$	$\begin{bmatrix}-0.0408\\0.9992\end{bmatrix}$	$\begin{bmatrix}0.0154\\0.9999\end{bmatrix}$
$(NBR, r) \nrightarrow_h GDP$	**0.0026**	**0.0033**	**0.0001**	**0.0029**	0.0372	0.0739	**0.0091**	**0.0008**	**0.0000**	**0.0001**	**0.0030**	**0.0003**
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h GDP$	0.5400	0.4053	0.2543	0.2341			0.5295	0.4729	0.1946	0.2104	0.1419	**0.0091**
V_h^{XY}	$\begin{bmatrix}0.0676\\0.9977\end{bmatrix}$	$\begin{bmatrix}0.0689\\0.9976\end{bmatrix}$	$\begin{bmatrix}0.0459\\0.9989\end{bmatrix}$	$\begin{bmatrix}0.0325\\0.9995\end{bmatrix}$			$\begin{bmatrix}0.2054\\0.9787\end{bmatrix}$	$\begin{bmatrix}0.2204\\0.9754\end{bmatrix}$	$\begin{bmatrix}0.2155\\0.9765\end{bmatrix}$	$\begin{bmatrix}0.2087\\0.9780\end{bmatrix}$	$\begin{bmatrix}0.1888\\0.9820\end{bmatrix}$	
$(NBR, r) \nrightarrow_h P$	**0.0001**	**0.0002**	**0.0000**	**0.0033**	**0.0028**	**0.0032**	0.0285	0.0182	0.1287	0.2056	0.4983	0.7924
$(NBR, r)\|_{\mathcal{V}} \nrightarrow_h P$	0.2183	0.1958	0.1970	0.4895	0.1594	0.4409						
V_h^{XY}	$\begin{bmatrix}-0.0015\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0070\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0086\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0336\\0.9994\end{bmatrix}$	$\begin{bmatrix}0.0067\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0043\\1.0000\end{bmatrix}$						

Table 10.4: Causality Test p-values and Non–causal Directions for the Horizons 13-24. (HAC Robust Inference).

h	13	14	15	16	17	18	19	20	21	22	23	24
$NBR \nrightarrow_h GDP$	0.2204	0.4914	0.4556	0.6637	0.7852	0.8011	0.5844	0.5678	0.4197	0.3456	0.2126	0.1708
$NBR \nrightarrow_h P$	0.1816	0.2803	0.2654	0.2252	0.7441	0.7730	0.2453	0.1237	0.2744	0.1750	0.3465	0.4138
$r \nrightarrow_h GDP$	**0.0000**	**0.0016**	**0.0014**	**0.0039**	0.0457	0.0107	0.0234	0.0271	0.1065	0.1071	0.0755	0.1225
$r \nrightarrow_h P$	0.5379	0.2808	0.3722	0.2637	0.5767	0.4395	0.5928	0.3090	0.1245	0.2542	0.1019	0.1829
$(NBR, r) \nrightarrow_h (GDP, P)$	**0.0086**	**0.0059**	0.0891	0.2911	0.2712	0.4034	0.2864	0.0245	0.0649	0.3216	0.6343	0.6409
$(NBR, r) \nrightarrow_h (GDP, P)\vert_{\mathcal{U}}$	0.9672	0.9006										
U_h^{XY}	$\begin{bmatrix}-0.1783\\0.9840\end{bmatrix}$	$\begin{bmatrix}-0.1882\\0.9821\end{bmatrix}$										
$(NBR, r)\vert_{\mathcal{V}} \nrightarrow_h (GDP, P)$	**0.0047**	0.0318										
V_h^{XY}		$\begin{bmatrix}0.1074\\0.9942\end{bmatrix}$										
$NBR \nrightarrow_h (GDP, P)$	0.3149	0.6781	0.4696	0.4210	0.8281	0.6607	0.3494	0.3730	0.4769	0.6412	0.7440	0.7465
$NBR \nrightarrow_h (GDP, P)\vert_{\mathcal{U}}$												
U_h^{XY}												
$r \nrightarrow_h (GDP, P)$	**0.0001**	**0.0032**	**0.0040**	0.0105	0.1285	0.0525	0.1349	0.0160	0.0140	0.0801	0.1048	0.1251
$r \nrightarrow_h (GDP, P)\vert_{\mathcal{U}}$	0.4268	0.2379	0.4034									
U_h^{XY}	$\begin{bmatrix}-0.0115\\0.9999\end{bmatrix}$	$\begin{bmatrix}-0.0320\\0.9995\end{bmatrix}$	$\begin{bmatrix}-0.0365\\0.9993\end{bmatrix}$									
$(NBR, r) \nrightarrow_h GDP$	**0.0000**	**0.0010**	**0.0004**	**0.0078**	0.0571	0.1039	**0.0069**	**0.0052**	**0.0091**	**0.0044**	0.0354	0.0397
$(NBR, r)\vert_{\mathcal{V}} \nrightarrow_h GDP$	**0.0018**	**0.0071**	**0.0095**	**0.0090**			0.0143	0.0206	0.0594	0.0952		
V_h^{XY}							$\begin{bmatrix}0.0014\\1.0000\end{bmatrix}$	$\begin{bmatrix}0.0153\\0.9999\end{bmatrix}$	$\begin{bmatrix}-0.0444\\0.9990\end{bmatrix}$	$\begin{bmatrix}0.0431\\0.9991\end{bmatrix}$		
$(NBR, r) \nrightarrow_h P$	0.5302	0.5803	0.6497	0.7435	0.8676	0.9379	0.6635	0.2183	0.3329	0.7577	0.5984	0.8038
$(NBR, r)\vert_{\mathcal{V}} \nrightarrow_h P$												
V_h^{XY}												

References

Al-Sadoon, M. (2009a). Causality along subspaces: Inference. Technical report.

Al-Sadoon, M. (2009b). Causality along subspaces: Theory. Cambridge Working Papers in Economics 0919, Faculty of Economics, University of Cambridge.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, *22*, 327351.

Anderson, T. W. (1999). Asymptotic distribution of the reduced rank regression estimator under general conditions. *The Annals of Statistics*, *27*(4), 1141–1154.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd Edition.* 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. (2009). Underidentification? Working Papers wp2009-0905, CEMFI.

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.

Bernanke, B. S. & Mihov, I. (1998). Measuring monetary policy. *The Quarterly Journal of Economics*, *113*(3), 869–902.

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.

Cavaliere, G., Rahbek, A., & Taylor, A. M. R. (2007). Testing for co-integration in vector autoregressions with non-stationary volatility. Discussion Papers 07/02, University of Nottingham, Granger Centre for Time Series Econometrics.

Cavaliere, G., Rahbek, A., & Taylor, A. M. R. (2009). Co-integration rank tests under conditional heteroskedasticity. Discussion Papers 09/02, University of Nottingham, Granger Centre for Time Series Econometrics.

Chan, T. F. (1987). Rank revealing qr factorizations. *Linear Algebra and its Applications*, *88-89*, 67 – 82.

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.

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., Pelletier, D., & Renault, E. (2006). Short run and long run causality in time series: inference. *Journal of Econometrics*, *127*(2), 337–362.

Dufour, J.-M. & Renault, E. (1998). Short run and long run causality in time series: Theory. *Econometrica*, *66*(5), 1099–1125.

Dufour, J.-M. & Valery, P. (2009). Hypothesis tests when rank conditions fail: a smooth regularization approach. mimeo. Available at http://neumann.hec.ca/pages/pascale.valery/Dufour_Valery_RegInfer_2009_09_10.pdf.

Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications* (2 ed.). Pure and Applied Mathematics. New York, USA: Wiley-Interscience.

Foster, L. V. (1986). Rank and null space calculations using matrix decomposition without column interchanges. *Linear Algebra and its Applications*, *74*, 47 – 71.

Garrett, T., Lee, K., Pesaran, H., & Shin, Y. (2006). *Global and National Macroeconometric Modelling: A Long Run Structural Approach*. Oxford, UK: Oxford University Press.

Gill, L. & Lewbel, A. (1992). Testing the rank and definiteness of estimated matrices with applications to factor, state-space and arma models. *Journal of the American Statistical Association*, *87*(419), 766–776.

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

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

Goreinov, S. A., Tyrtyshnikov, E. E., & Zamarashkin, N. L. (1997). A theory of pseudoskeleton approximations. *Linear Algebra and its Applications*, *261*(1-3), 1 – 21.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 428–38.

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

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.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *28*(3/4), 321–377.

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

Johansen, S. (2006). Statistical analysis of hypotheses on the cointegrating relations in the i(2) model. *Journal of Econometrics*, *132*(1), 81–115.

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.

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

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis.* Springer.

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

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.

Newey, W. K. & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*(3), 777–87.

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.

Peters, G. & Wilkinson, J. (1970). The least-squares problem and pseudo-inverses. *The Computer Journal*, *13*, 309–316.

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.-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.

Stock, J. H. & Watson, M. W. (1999). Business cycle fluctuations in us macroeconomic time series. In J. B. Taylor & M. Woodford (Eds.), *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics* chapter 1, (pp. 4–64). Elsevier.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press.

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.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, *54*(3), 426–482.

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

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