


Article

# Asymptotic Theory for Cointegration Analysis When the Cointegration Rank Is Deficient

David H. Bernstein <sup>1</sup> and Bent Nielsen <sup>2,\*</sup> 

<sup>1</sup> Department of Economics, University of Miami, Coral Gables, FL 33146, USA; dbernstein@bus.miami.edu

<sup>2</sup> Department of Economics & Nuffield College & Programme on Economic Modelling, University of Oxford, Oxford OX1 1NF, UK

\* Correspondence: bent.nielsen@nuffield.ox.ac.uk

Received: 3 May 2018; Accepted: 8 January 2019; Published: 18 January 2019

**Abstract:** We consider cointegration tests in the situation where the cointegration rank is deficient. This situation is of interest in finite sample analysis and in relation to recent work on identification robust cointegration inference. We derive asymptotic theory for tests for cointegration rank and for hypotheses on the cointegrating vectors. The limiting distributions are tabulated. An application to US treasury yields series is given.

**Keywords:** cointegration; rank deficiency; weak identification

**JEL Classification:** C32

---

## 1. Introduction

Determination of the cointegration rank is an important part of analyzing the cointegrated vector autoregressive model in the framework of [Johansen \(1988, 1991, 1995\)](#), [Johansen and Juselius \(1990\)](#), and [Juselius \(2006\)](#). We consider the rank deficient case where the cointegration rank of the data generating process is smaller than the rank used in the statistical analysis. In that case, the data generating process has more unit roots than the number of unit roots imposed in the statistical analysis and the usual asymptotic theory fails. We provide asymptotic theory for cointegration rank tests and tests on cointegration vectors along with simulated tables of the asymptotic distributions.

Cointegration analysis is conducted in three steps. First, the specification of the model is checked. Second, the rank is determined using a sequential procedure using Dickey-Fuller type distributions. Third, the cointegrating vectors are estimated and restrictions can be tested using standard inference. Asymptotic theory shows that estimated rank is consistent in the sense that the probability that the estimated rank is not equal to the true rank equals the size of tests, whereas the probability that the estimated rank is too small vanishes, see [Johansen \(1992, 1995\)](#) and [Paruolo \(2001\)](#). Hence, the rank deficiency problem does not arise in the asymptotic analysis.

In practice, rank deficiency matters in two ways. The asymptotic theory often suffers from considerable finite sample distortion. Further, if an investigator wants to focus on the inference on the cointegrating relations then problems can arise if the rank is taken as known when in fact it is deficient. These problems mirror those of instrumental variable estimation with weak instruments, see [Mavroeidis et al. \(2014\)](#).

When conducting inference on the cointegrating vector under near rank deficiency the parameters are weakly identified. At the extreme when testing on the cointegrating vector in the case of a deficient rank the model is mis-specified. This problem arises in cointegration as well as in instrumental variable estimation. In both cases maximum likelihood is conducted using reduced rank regression. The weak identification problem has attracted considerable attention in the instrumental variable literature, see for instance [Mavroeidis et al. \(2014\)](#). [Khalaf and Urga \(2014\)](#) discussed the weak

identification problem for cointegration, that is when testing for a known cointegrating vector in the nearly rank deficient situation. These authors investigate various methods to adjust the asymptotic distribution in the weak identification case. This includes a bounds-based critical value suggested by [Dufour \(1997\)](#). This method requires knowledge of the asymptotic theory for the rank deficient case, which we provide here.

The practical problem of ignoring rank deficiency is illustrated using yield curve data. The expectation hypothesis is often interpreted as follows. Interest rates at different maturities are integrated series, but cointegrate so that spreads are stationary. Spreads are often found to be non-stationary. Thus, it is quite possible that a pair of interest rates do not cointegrate. An investigator may proceed by assuming cointegration when there is none, so that the rank is deficient, and conduct inference on the coefficients on the *alleged* cointegrating vector using standard inference. Our theory shows that the inference is then severely distorted. When the rank is deficient or nearly deficient it is incorrect to use standard inference on the cointegrating vectors. Nonetheless, applying standard inference in the particular example leads to marginal rejection of the hypothesis. Applying the bounds test of [Khalaf and Urga \(2014\)](#) shifts the distribution to the right and there is not much power to reject a hypothesis. If the rank is deficient, which is possible in the example, the alleged cointegrating vector cannot be cointegrating.

Rank deficiency also matters when the rank is determined empirically. Different asymptotic distributions arise in the standard case and when the rank is deficient. The asymptotic distribution tends to give a very good approximation to the finite sample distribution when the rank is far from being deficient, see for instance [Nielsen \(1997, 2004\)](#) When the parameters are in the vicinity of rank deficiency the finite sample distribution tends to be a combination of the two asymptotic distributions. When the parameters are not too close to the rank deficient case a Bartlett correction using a fixed parameter second-order asymptotic expansion works very well, see [Johansen \(2000, 2002\)](#) Bootstrap solutions have been discussed in simulation studies by [Fachin \(2000\)](#); [Gredenhoff and Jacobson \(2001\)](#); [Swensen \(2004\)](#); [Cavaliere et al. \(2012\)](#). When the parameters are closer to rank deficient a local-to-unity asymptotic expansion gives an improvement, see [Nielsen \(2004\)](#) for the cointegration case and [Nielsen \(1999, 2001\)](#) for the corresponding instrumental variable case. A starting point for the finite sample analysis is knowledge of the fixed-parameter first-order asymptotic theory across the parameter space, including rank deficient cases.

We discuss the asymptotic theory for models without and with deterministic terms in Sections 2 and 3, respectively. The implications for finite sample analysis and the weakly identified case are discussed in Section 4 along with an application to US treasury zero coupon yields. Section 5 concludes. Proofs are given in an Appendix A.

## 2. The Model without Deterministic Terms

We consider the Gaussian cointegrated vector autoregressive model in the case with no deterministic terms. The asymptotic theory for tests for reduced cointegration rank and for a known cointegrating vector is derived when the rank is deficient. Finally, we analyze the case of near rank deficiency.

### 2.1. Model and Hypotheses

Consider a  $p$ -dimensional time series  $X_t$  for  $t = 1 - k, \dots, 0, 1, \dots, T$ . The unrestricted vector autoregressive model can be written as

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \quad \text{for } t = 1, \dots, T, \quad (1)$$

where the innovations  $\varepsilon_t$  are independent normal  $N_p(0, \Omega)$ -distributed. The parameters  $\Pi$ ,  $\Gamma_i$ ,  $\Omega$  are freely varying  $p$ -dimensional square matrices so that  $\Omega$  is symmetric, positive definite.

The hypothesis of reduced cointegration rank is formulated as

$$H_z(r) : \quad \text{rank } \Pi \leq r, \quad (2)$$

for some  $0 \leq r \leq p$ . The interpretation of the hypotheses follows from the Granger-Johansen representation presented in Section 2.2 below. The subscript  $z$  indicates that the model has a zero deterministic component. The rank hypotheses are nested so that

$$H_z(0) \subset \cdots \subset H_z(r) \subset \cdots \subset H_z(p). \quad (3)$$

The rank deficiency problem arises when testing the hypothesis  $H_z(r)$  when in fact the sub-hypothesis  $H_z(r-1)$  is satisfied. The rank is determined to be  $r$  if the hypothesis  $H_z(r)$  cannot be rejected while the sub-hypothesis  $H_z(r-1)$  is rejected. As a short-hand we write  $H_z^0(r) = H_z(r) \setminus H_z(r-1)$  for this situation. The rank can be determined along the procedure outlined in Johansen (1992, 1995) [Section 12.1] and Paruolo (2001). In practice, these decisions are often marginal, hence the need to study the asymptotic theory of test statistics in the rank deficient case.

The rank hypothesis can equivalently be written as

$$H_z(r) : \quad \Pi = \alpha\beta', \quad (4)$$

where  $\alpha$  and  $\beta$  are  $p \times r$  matrices. The advantage of this formulation is that  $\alpha$  and  $\beta$  vary in vector spaces. The formulation does, however, allow rank deficiency where the rank of  $\Pi$  is smaller than  $r$ . We follow Johansen (1991, Equation (2.2)) and refer to  $\beta$  as the cointegrating vectors. We find the terminology useful, although it is ambiguous. Indeed, for a particular data generating process where  $\Pi$  has rank less than  $r$  then the identity  $\Pi = \alpha\beta'$  can be satisfied while columns of  $\beta$  may not be row-eigenvectors of  $\Pi$  in which case  $\beta'X_t$  cannot be stationary. Even when  $\Pi$  has rank  $r$  then  $\beta'X_t$  is only (approximately) stationary under the I(1) condition introduced below. However, from a statistical viewpoint, the estimator of  $\Pi$  under the restriction of rank  $r$  will in a finite sample have rank  $r$  with probability one. In practice our only knowledge of the rank arises from inference. Johansen's terminology appears to be focused on the statistical viewpoint which we will follow even when studying the rank deficient cases.

The hypothesis of known cointegration vectors is

$$H_{z,\beta}(r) : \quad \Pi = \alpha b', \quad (5)$$

for some unknown matrix  $\alpha$  and a known matrix  $b$ , both of dimension  $p \times r$ , so that  $b$  has full column rank. The standard analysis is concerned with the situation where  $\alpha$  has full column rank, but in the rank deficient case, it has reduced column rank, so that the hypothesis  $H_z(r-1)$  is satisfied. When referring to  $b$  as the cointegrating vectors, we, once again, follow the terminology of Johansen (1991, Equation (3.1)) even though  $b'X_t$  cannot be stationary under rank deficiency.

## 2.2. Granger-Johansen Representation

The Granger-Johansen representation provides an interpretation of the cointegration model that is useful in the asymptotic analysis. We work with the result stated by Johansen (1995, Theorem 4.2). The theorem requires the following assumption.

**I(1) Condition.** Suppose  $\text{rank } \Pi = s$  where  $s \leq p$ . Consider the characteristic roots satisfying  $0 = \det\{A(z)\}$  where  $A(z) = (1-z)I_p - \Pi z - \sum_{i=1}^{k-1} \Gamma_i z^i (1-z)$ . Suppose there are  $p-s$  unit roots, and that the remaining roots are stationary roots, so satisfying  $|z| > 1$ .

The Granger-Johansen theorem assumes that a process satisfying the model (1) so that  $\text{rank } \Pi = r$  and we can write  $\Pi = \alpha\beta'$  while the I(1) condition holds with  $s = r$ . The process then has the representation

$$X_t = C \sum_{i=1}^t \varepsilon_i + S_t + \tau, \quad (6)$$

where the impact matrix  $C$  for the random walk has rank  $p - r$  and satisfies  $\beta'C = 0$  and  $C\alpha = 0$ , the process  $S_t$  can be given a zero mean stationary initial distribution and  $\tau$  depends on the initial observations in such a way that  $\beta'\tau = 0$ . In other words, the process  $X_t$  behaves like a random walk with cointegrating relations  $\beta'X_t$  that can be given a stationary initial distribution.

### 2.3. Test Statistics

The likelihood ratio test statistic for the reduced rank hypothesis  $H_z(r)$  against the unrestricted model  $H_z(p)$  is found by reduced rank regression, see [Johansen \(1995, Section 6\)](#). It can be described as a two-step procedure. First, the differences  $\Delta X_t$  and the lagged levels  $X_{t-1}$  are regressed on the lagged differences  $\Delta X_{t-i}$ ,  $i = 1, \dots, k - 1$  giving residuals  $R_{0,t}$ ,  $R_{1,t}$ . Secondly, the squared sample correlations,  $1 \geq \hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p \geq 0$  say, of  $R_{0,t}$  and  $R_{1,t}$  are found, by computing product moments  $S_{ij} = T^{-1} \sum_{t=1}^T R_{i,t} R'_{j,t}$  and solving the eigenvalue problem  $0 = \det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01})$ . The log likelihood ratio test statistic for the rank hypothesis is then

$$LR\{H_z(r) | H_z(p)\} = -T \sum_{j=r+1}^p \log(1 - \hat{\lambda}_j). \quad (7)$$

Under the hypothesis of known cointegration vectors, the likelihood is maximised by least squares regression. The log likelihood ratio test statistic against the unrestricted model  $H_z(p)$  is therefore given by

$$LR\{H_{z,\beta}(r) | H_z(p)\} = -T \log \frac{\det\{S_{00} - S_{01} S_{11}^{-1} S_{10}\}}{\det\{S_{00} - S_{01} b(b' S_{11} b)^{-1} b' S_{10}\}}. \quad (8)$$

The log likelihood ratio statistic for the hypothesis of known cointegrating vector against the rank hypothesis is found by combining the statistics in (7) and (8), that is

$$LR\{H_{z,\beta}(r) | H_z(r)\} = LR\{H_{z,\beta}(r) | H_z(p)\} - LR\{H_z(r) | H_z(p)\}. \quad (9)$$

The relationship will be useful in the asymptotic theory. For instance, [Theorems 1 and 2](#) give the asymptotic distributions of  $LR\{H_z(r) | H_z(p)\}$  and  $LR\{H_{z,\beta}(r) | H_z(p)\}$ , respectively. From this we can derive an expression for the distribution of  $LR\{H_{z,\beta}(r) | H_z(r)\}$ . When it comes to tabulation we will need to simulate all three distributions. This would be the case even if the former two statistics were independent.

### 2.4. Asymptotic Theory for the Rank Test

In the asymptotic analysis it is possible to relax the assumption to the innovations. While the likelihood is derived under the assumption of independent, identically Gaussian distributed innovations less is needed for the asymptotic theory. [Johansen \(1995\)](#) assumes the innovations are independent, identically distributed with mean zero and finite variance and uses linear process results from [Phillips and Solo \(1992\)](#). This could be relaxed further to, for instance, a martingale difference assumption. However, for expositional simplicity we follow [Johansen's](#) argument and assumptions.

**Theorem 1.** Consider the rank hypothesis  $H_z(r) : \text{rank } \Pi \leq r$ . Suppose  $H_z^o(s) = H_z(s) \setminus H_z(s - 1)$  holds for some  $s \leq r$  and that the I(1) condition holds for that  $s$ . Let  $F_u = B_u$  be a  $p - s$ -dimensional standard Brownian motion on  $[0, 1]$ . Let  $1 \geq \rho_1 \geq \dots \geq \rho_{p-s} \geq 0$  be the eigenvalues of the eigenvalue problem

$$0 = \det \left\{ \rho \int_0^1 F_u F_u' du - \int_0^1 F_u (dB_u)' \int_0^1 (dB_u) F_u' \right\} \tag{10}$$

Then, for  $T \rightarrow \infty$ ,

$$LR\{H_z(r) | H_z(p)\} = -T \sum_{j=r+1}^p \log(1 - \hat{\lambda}_j) \xrightarrow{D} \sum_{j=r-s+1}^{p-s} \rho_j. \tag{11}$$

In the standard non-deficient situation where  $r = s$  the result reduces to the result of Johansen (1995, Theorem 6.1). The rank deficient case was also discussed by Johansen (1995, p. 158) and Nielsen (2004, Theorem 6.1).

Table 1 reports the asymptotic distribution of the rank test reported in Theorem 1. The simulation were done using Ox (Doornik 2007). The simulation design follows that of Johansen (1995, Section 15). That is, the stochastic integrals in (10) were discretized with  $T = 1000$  and zero initial observations with one million repetitions. The table reports simulated quantiles and moments for  $r - s = 0, 1, 2$  and  $p - r = 1, 2, 3, 4$ . However, the case of  $p - r = 1$  and  $r - s = 0$  are analytic values from Nielsen (1997) and where the quantiles were provided by Karim Abadir using his results in Abadir (1995). Bernstein (2014) reports values for higher dimensions. The 85% quantile has not been computed analytically in this case.

**Table 1.** Quantiles, mean, and variance of  $LR\{H_z(r)|H_z(p)\}$ , where the data generating process has rank  $s = \text{rank } \Pi \leq r$ .

$r - s$	$p - r$	50%	80%	85%	90%	95%	97.5%	99%	Mean	Var
0	1	0.60	1.88	—	2.98	4.13	5.32	6.94	1.14	2.22
	2	5.48	8.48	9.31	10.44	12.30	14.07	16.34	6.09	10.61
	3	14.39	18.94	20.13	21.70	24.22	26.54	29.37	15.02	25.13
	4	27.29	33.35	34.88	36.91	40.04	42.93	46.45	27.93	45.66
1	1	0.36	1.13	1.38	1.74	2.35	2.98	3.81	0.67	0.70
	2	4.27	6.25	6.78	7.50	8.65	9.76	11.14	4.61	4.66
	3	11.92	15.20	16.04	17.14	18.88	20.50	22.48	12.31	13.22
	4	23.47	28.09	29.25	30.76	33.10	35.21	37.83	23.89	26.96
2	1	0.30	0.97	1.18	1.48	1.98	2.47	3.11	0.56	0.48
	2	3.93	5.57	6.01	6.59	7.51	8.38	9.46	4.18	3.24
	3	11.04	13.82	14.53	15.46	16.91	18.24	19.87	11.34	9.63
	4	21.84	25.83	26.82	28.11	30.09	31.91	34.13	22.18	20.21

The first panel of Table 1 reports the distribution for the standard case where  $s = r$ . This corresponds to Table 15.1 of Johansen (1995). The second and third panel of Table 1 report the distribution for the rank deficient case where  $s = r - 1$  so  $r - s = 1$  and where  $s = r - 2$  so  $r - s = 2$ . The first entry in panel 2 for  $s = r - 1$  and  $p - r = 1$ , so  $r - s = 1$ , corresponds to Table 6 of Nielsen (2004). It is seen that as the rank becomes more deficient the distribution shifts to the left. It should be noted that if the rank is non deficient, but the I(1) condition is not satisfied then the distribution would tend to shift to the right, see Nielsen (2004) for a discussion. The simulations reported in Table 8 of that paper indicates that the distribution is between these extremes if the rank is deficient and the I(1) condition fails.

The rank test statistic in (7) has been analyzed analytically for the canonical correlation problem in cross-sectional models in Nielsen (1999, 2001) This test also corresponds to the test for relevance in the instrument variable problem. In that case, analytic expressions are available when  $p = 2, r = 1$

and  $s = 0, 1$ . When  $s = 1$  we have a  $\chi^2$ -distribution with mean 1 and variance 2. When  $s = 0$  the mean is 0.429 and the variance is  $0.575 - (0.429)^2 = 0.391$ , see Nielsen (1999). Thus, the impact of rank deficiency is similar to what is seen in Table 1 for cointegration rank testing.

### 2.5. Asymptotic Theory for the Test on the Cointegrating Vectors

In the analysis of the test for known cointegrating vectors, we focus on the situation where the data generating process has rank  $s = 0$ . In this situation the asymptotic distribution is relatively simple to describe, because it does not depend on the value of the hypothesized cointegrating vectors  $b$ . This is adequate for a discussion of aspects of situations considered in Khalaf and Urga (2014). If the rank is non-zero but deficient so  $0 < s < r$ , then the data generating process will have cointegrating vectors  $\beta_0$  of dimension  $p \times s$  and the asymptotic theory will depend on  $\beta_0$  and  $b$ . In practice, it is rare to test for simple hypotheses when there is more than one hypothesized cointegrating vector, so we do not pursue this complication.

The analysis of the test for known cointegrating vectors is somewhat different from the analysis in Johansen (1995). His analysis is aimed at the situation where different restrictions are imposed on the cointegrating vectors. The argument then involves an intriguing consistency proof for the estimated cointegrating vectors. However, when testing the hypothesis of known cointegrating vectors the likelihood is maximized by the least squares method and the consistency argument is not needed. The asymptotic theory can then be described by the following result.

**Theorem 2.** Consider the hypothesis  $H_{z,\beta}(r) : \Pi = \alpha b'$ , where  $\alpha, b$  have dimension  $p \times r$  and where  $\alpha$  is unknown and  $b$  is known with full column rank. Suppose  $H_z(0)$  is satisfied, so that  $\alpha = 0$  and  $s = 0$ , and that the  $I(1)$  condition is satisfied with  $s = 0$ . Let  $B_u$  be a  $p$ -dimensional standard Brownian motion on  $[0, 1]$  with components  $B_{1,u}$  and  $B_{2,u}$  of dimension  $r$  and  $p - r$ , respectively. Then, for  $T \rightarrow \infty$ ,

$$LR\{H_{z,\beta}(r) \mid H_z(p)\} \xrightarrow{D} \text{tr}\left\{ \int_0^1 dB_u B_u' \left( \int_0^1 B_u B_u' du \right)^{-1} \int_0^1 B_u (dB_u)' - \int_0^1 dB_u B_{1,u}' \left( \int_0^1 B_{1,u} B_{1,u}' du \right)^{-1} \int_0^1 B_{1,u} (dB_u)' \right\}. \quad (12)$$

The convergence of the test statistic  $LR\{H_{z,\beta}(r) \mid H_z(p)\}$  holds jointly with the convergence for the rank test statistic  $LR\{H_z(r) \mid H_z(p)\}$ , for  $s = 0$ , in Theorem 1. Thus, when  $s = 0$  the formula (9) implies that the limit distribution of the test statistic for known  $\beta$  within the model with rank of at most  $r$  can be found as the difference of the two limiting variables.

Table 2 reports the asymptotic distribution of the test for known cointegrating vector in the model where the rank is at most  $r$ . When  $s = r$  the asymptotic distribution is  $\chi^2$  with  $r(p - r)$  degrees of freedom, see Johansen (1995, Theorem 7.2.1). When  $s = 0$  the asymptotic distribution reported in Theorem 2 applies. The simulation design is as before. It is seen that in the rank deficient case the distribution is shifted to the right. This matches the finite sample simulations reported by Johansen (2000, Table 2).

**Table 2.** Quantiles, mean, and variance of  $LR\{H_{z,\beta}(r)|H_z(r)\}$ , where the data generating process has rank  $s = \text{rank } \Pi \leq r$ .

$p$	$r$	$s$	50%	80%	85%	90%	95%	97.5%	99%	Mean	Var
2	1	1	0.45	1.64	2.07	2.71	3.84	5.02	6.63	1	2
		0	2.62	5.44	6.22	7.30	9.05	10.75	12.96	3.31	8.71
3	2	2	1.39	3.22	3.79	4.61	5.99	7.38	9.21	2	4
		0	5.80	9.42	10.40	11.71	13.82	15.77	18.27	6.42	15.53
3	1	1	1.39	3.22	3.79	4.61	5.99	7.38	9.21	2	4
		0	6.79	10.58	11.57	12.89	15.02	17.00	19.49	7.33	17.52

Table 3 reports the simulated asymptotic distribution of the test for known cointegrating vector in the model where the rank is unrestricted. The distribution is shifted to the right in the rank deficient case. Note, that the table reports the distribution of the convolution of the statistics simulated in Tables 1 and 2, see (9). Thus, up to a simulation error the expectations reported in Tables 1 and 2 add up to the expectation reported in Table 3. In the full rank case  $r = s$  the statistics in Tables 1 and 2 are independent, as proved below, so also the variances are additive.

**Theorem 3.** Consider the hypothesis  $H_{z,\beta}^{\circ}(r)$ . Suppose  $H_z^{\circ}(r) = H_z^{\circ}(r)/H_z^{\circ}(r - 1)$  is satisfied and that the  $I(1)$  condition holds with  $s = r$ . Then the rank test statistic  $LR\{H_z^{\circ}(r)|H_z^{\circ}(p)\}$  and the statistic  $LR\{H_{z,\beta}^{\circ}(r)|H_z^{\circ}(r)\}$  for testing a simple hypothesis on the cointegrating vector are asymptotically independent.

The asymptotic distribution of the rank statistic  $LR\{H_z^{\circ}(r)|H_z^{\circ}(p)\}$  is given in Theorem 1, while the statistic for the cointegrating vector  $LR\{H_{z,\beta}^{\circ}(r)|H_z^{\circ}(r)\}$  is asymptotically  $\chi^2\{r(p - r)\}$ .

**Table 3.** Quantiles, mean, and variance of  $LR\{H_{z,\beta}(r)|H_z(p)\}$ , where the data generating process has rank  $s = \text{rank } \Pi \leq r$ .

$p$	$r$	$s$	50%	80%	85%	90%	95%	97.5%	99%	Mean	Var
2	1	1	1.54	3.43	4.01	4.83	6.22	7.62	9.47	2.15	4.23
		0	3.35	6.11	6.89	7.95	9.70	11.38	13.57	3.98	8.82
3	2	2	2.52	4.85	5.53	6.48	8.07	9.60	11.62	3.15	6.26
		0	6.36	9.96	10.92	12.22	14.32	16.29	18.79	6.98	15.35
3	1	1	7.50	11.03	11.98	13.27	15.34	17.30	19.81	8.13	14.73
		0	11.33	15.73	16.88	18.41	20.83	23.09	25.91	11.96	23.31

### 2.6. The Case of Nearly Deficient Rank

With the above results we have two extremes. First, the full rank case where standard results apply, that is Johansen’s Dickey-Fuller type distribution for rank testing and  $\chi^2$  inferences for testing constraints on the cointegrating vectors. Second, the rank deficient case where new Dickey-Fuller type distributions apply both for rank testing and for testing constraints on the cointegrating vectors. In between these extremes we have the nearly rank deficient case corresponding to weak identification in the instrumental variable literature. These nearly deficient cases can be analyzed using local-to-unity parametrization. However, a full theory is notationally complicated as there will be many nuisance parameters. We therefore consider a simple special case inspired by the power analysis of Johansen (1995, Section 14) and distribution analysis of Nielsen (2004).

The main finding is that the appropriate local rate is  $T^{-1}$  as in power analysis for unit tests and cointegration rank tests as opposed to  $T^{-1/2}$  for stationary models as in Andrews and Cheng (2012). Consider a bivariate, first order, local-to-unity vector autoregressive model where

$$\Delta X_t = \frac{1}{T} \begin{pmatrix} b_1 & b_2 \\ 0 & 0 \end{pmatrix} X_{t-1} + \varepsilon_t \quad \text{for } t = 1, \dots, T, \tag{13}$$

where the innovations  $\varepsilon_t$  are independent normal  $N_2(0, I_2)$ -distributed where  $b_1 \neq 0$ .

We now have the following variant of the result for the rank test in Theorem 1.

**Theorem 4** (Nielsen 2004, Theorem 6.2). Consider the data generating process (13). Let  $B_u$  be a bivariate standard Brownian motion on  $[0, 1]$  and let  $J_u$  be the bivariate Ornstein-Uhlenbeck process given by

$$J_u = \begin{pmatrix} b_1 & b_2 \\ 0 & 0 \end{pmatrix} \int_0^u J_s ds + B_u.$$

Let  $1 \leq \rho_1 \leq \rho_2 \leq 0$  be the eigenvalues of the eigenvalue problem

$$0 = \det \left\{ \rho \int_0^1 J_u J_u' du - \int_0^1 J_u (dB_u)' \int_0^1 (dB_u) J_u' \right\}$$

Then, for  $T \rightarrow \infty$ ,

$$LR\{H_z(1) \mid H_z(2)\} \xrightarrow{D} \rho_2.$$

The limit distribution is tabulated in Nielsen (2004, Table 8).

We now consider the test for known cointegrating vector,  $b = (b_1, b_2)'$ . The result in Theorem 2 is modified as follows.

**Theorem 5.** Consider the data generating process (13). Let  $B_u, J_u$  be defined as in Theorem 4 and let  $J_{1,u} = b' J_u$ . Then

$$LR\{H_{z,\beta}(1) \mid H_z(2)\} \xrightarrow{D} \text{tr} \left\{ \int_0^1 dB_u J_u' \left( \int_0^1 J_u J_u' du \right)^{-1} \int_0^1 J_u (dB_u)' \int_0^1 dB_u J_{1,u}' \left( \int_0^1 J_{1,u} J_{1,u}' du \right)^{-1} \int_0^1 J_{1,u} (dB_u)' \right\}.$$

### 3. The Model with a Constant

We now consider the model augmented with a constant. In the cointegrated model the constant is restricted to the cointegrating space. Thus, the cointegrating vectors consist of vectors relating the dynamic variable extended by a further coordinate for the constant. There are now two rank conditions; one related to the dynamic part of these extended cointegrating vectors and one relating to the deterministic part of the cointegrating vectors. The condition to the cointegration rank in the standard theory can therefore fail in two ways.

#### 3.1. Model and Hypotheses

The unrestricted vector autoregressive model is

$$\Delta X_t = \Pi X_{t-1} + \mu + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \quad \text{for } t = 1, \dots, T, \tag{14}$$

where the innovations  $\varepsilon_t$  are independent normal  $N_p(0, \Omega)$ -distributed. The parameters are the  $p$ -dimensional square matrices  $\Pi, \Gamma_i, \Omega$  and the  $p$ -vector  $\mu$ . They vary freely so that  $\Omega$  is symmetric, positive definite.

For the model with a constant there are two types of cointegration rank hypotheses:

$$H_{c\ell}(r) : \quad \text{rank } \Pi \leq r, \tag{15}$$

$$H_c(r) : \quad \text{rank } (\Pi, \mu) \leq r. \tag{16}$$



Their interpretations follow from the Granger-Johansen representation which is reviewed in Section 3.2 below. In short, if there are no rank deficiencies the first hypothesis  $H_{c\ell}$  gives cointegrating relations with a constant level and common trends with a linear trend. The second hypothesis  $H_c$  has a constant level both for the cointegrating relations and the common trends. The hypotheses are nested so that

$$H_c(0) \subset H_{c\ell}(0) \subset \dots \subset H_{c\ell}(r-1) \subset H_c(r) \subset H_{c\ell}(r) \subset \dots \subset H_c(p) = H_{c\ell}(p). \tag{17}$$

This nesting structure is considerably more complicated than the structure (3) for the model without deterministic terms. A practical investigation may start in three different ways. First, the model (14) is taken as the starting point. Both types of hypotheses come into play and the rank is determined as outlined in Johansen (1995, Section 12). Secondly, if visual inspection of the data indicates that linear trends are not present the hypotheses  $H_{c\ell}$  may be ignored. Thirdly, if visual inspection of the data indicates that a linear trend could be present, the model (14) should be augmented with a linear trend term and we move outside the present framework. Nielsen and Rahbek (2000) discuss the latter two possibilities. Here, we are concerned with the first two possibilities.

The rank hypotheses can equivalently be formulated as

$$H_{c\ell}(r) : \quad \Pi = \alpha\beta', \tag{18}$$

$$H_c(r) : \quad (\Pi, \mu) = \alpha(\beta', \beta'_c). \tag{19}$$

The hypotheses of known cointegrating vectors are therefore

$$H_{c\ell,\beta}(r) : \quad \Pi = \alpha b', \tag{20}$$

$$H_{c,\beta}(r) : \quad (\Pi, \mu) = \alpha(b', b'_c). \tag{21}$$

for a known  $(p \times r)$ -matrix  $b$  with full column rank and, in the second case, also a known  $(1 \times r)$ -matrix  $b_c$  so that  $b^* = (b', b'_c)'$  has full column rank.

### 3.2. Granger-Johansen Representation

We give a Granger-Johansen representation for each of the two reduced rank hypotheses. Both results follow from Theorem 4.2 and Exercise 4.5 of Johansen (1995). First, consider the hypothesis  $H_{c\ell}(r)$ . Suppose that the sub-hypothesis  $H_c(r)$  does not hold and that the I(1) condition holds with  $s = r$ . Thus, the  $(p \times r)$ -matrices  $\alpha, \beta$  have full column rank but  $\alpha'_{\perp} \mu \neq 0$ , so that the matrix  $\Pi^* = (\Pi, \mu)$  has rank  $r + 1$ . Then, the Granger-Johansen representation is

$$X_t = C \sum_{i=1}^t \varepsilon_i + S_t + \tau_c + \tau_{\ell} t, \tag{22}$$

where the impact matrix  $C$  has rank  $p - r$  and satisfies  $\beta' C = 0$  and  $C\alpha = 0$  while  $\tau_{\ell} = C\mu \neq 0$ . As a consequence, the process has a linear trend, but the cointegrating relations  $\beta' X_t$  do not have a linear trend, since  $\beta' C = 0$ .

Secondly, consider the hypothesis  $H_c(r)$ . Suppose that the sub-hypothesis  $H_{c\ell}(r - 1)$  does not hold and that the I(1) condition holds with  $s = r$ . Thus, the  $(p \times r)$ -matrices  $\alpha, \beta$  have full column rank, and the  $\{(p + 1) \times r\}$ -matrix  $\beta^* = (\beta, \beta'_c)'$  has full column rank. Then, the Granger-Johansen representation (22) holds with  $\tau_{\ell} = 0$ , while  $\tau_c$  has the property that  $\beta' \tau_c = -\beta'_c$ . In other words, the process  $X_t$  behaves like a random walk where  $\beta' X_t$  has an invariant distribution with a non-zero mean, while  $\beta' X_t + \beta'_c t$  has a zero mean invariant distribution.

### 3.3. Test Statistics

The test statistics are variations of those for the model without deterministic terms. The differences relate to the formation of the residuals  $R_{0,t}$  and  $R_{1,t}$

First, consider the reduced rank hypothesis  $H_{c\ell}(r)$  and the corresponding hypothesis  $H_{c\ell,\beta}(r)$  of known cointegrating vectors. The residuals  $R_{0,t}$  and  $R_{1,t}$  are formed by regressing the differences  $\Delta X_t$  and the lagged levels  $X_{t-1}$  on an intercept and the lagged differences  $\Delta X_{t-i}$ ,  $i = 1, \dots, k - 1$ . In the second step, compute the canonical correlations  $1 \geq \hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p \geq 0$  of  $R_{0,t}$  and  $R_{1,t}$ . The rank test statistic  $LR\{H_{c\ell}(r)|H_{c\ell}(p)\}$  then has the form (7). The test statistic for known cointegrating vectors  $LR\{H_{c\ell,\beta}(r)|H_{c\ell}(p)\}$  has the form (8), using the same residuals  $R_{0,t}$  and  $R_{1,t}$ , and the hypothesized cointegrating vectors  $b$ .

Secondly, consider the reduced rank hypothesis  $H_c(r)$  and the corresponding hypothesis  $H_{c,\beta}(r)$  of known cointegrating vectors. The residuals  $R_{0,t}$  and  $R_{1,t}$  are formed by regressing the differences  $\Delta X_t$  and the vector formed by stacking the lagged levels and an intercept  $X_{t-1}^* = (X'_{t-1}, 1)'$  on the lagged differences  $\Delta X_{t-i}$ ,  $i = 1, \dots, k - 1$ . In the second step, compute the canonical correlation of these  $R_{0,t}$  and  $R_{1,t}$ . The rank test statistic  $LR\{H_c(r)|H_c(p)\}$  then has the form (7). The test statistic for known cointegrating vectors  $LR\{H_{c,\beta}(r)|H_c(p)\}$  has the form (8), using the same residuals  $R_{0,t}$  and  $R_{1,t}$ , and the hypothesized cointegrating vectors  $b^* = (b', b'_c)'$ .

### 3.4. Asymptotic Theory for the Rank Tests

There are now four situations to consider. Indeed, the nesting structure in (17) shows that each of the two rank hypotheses  $H_{c\ell}(r)$  and  $H_c(r)$  can be rank deficient in two ways when either of  $H_{c\ell}^\circ(s) = H_{c\ell}(s)/H_c(s)$  or  $H_c^\circ(s) = H_c(s)/H_{c\ell}(s - 1)$  holds. In three cases the limiting distribution is of the same form as in Theorem 1, albeit with a different limiting random function  $F_u$ . In the fourth case the limiting distribution has nuisance parameters. The nuisance parameter case arises when testing  $H_c(r)$  with a data generating process satisfying  $H_{c\ell}^\circ(s) = H_{c\ell}(s)/H_c(s)$ . This is the case that can often be ruled out through visual inspection of the data as mentioned in Section 3.1.

We start with the test for the hypothesis  $H_{c\ell}(r)$  in the rank deficient case where  $H_{c\ell}^\circ(s) = H_{c\ell}(s)/H_c(s)$  holds for  $s < r$ . Johansen (1995) discusses the possibility  $H_c^\circ(r)$ . The asymptotic theory is as follows.

**Theorem 6.** Consider the rank hypothesis  $H_{c\ell}(r) : \text{rank } \Pi \leq r$ . Suppose  $H_{c\ell}^\circ(s) = H_{c\ell}(s) \setminus H_c(s)$  holds for some  $s \leq r$ , so that  $\text{rank } \Pi = s$  and  $\text{rank}(\Pi, \mu) = s + 1$  and that the  $I(1)$  condition is satisfied for that  $s$ . Let  $B_u$  be a  $(p - s)$ -dimensional standard Brownian motion on  $[0, 1]$ . Define a  $(p - s)$ -dimensional vector  $F_u$  with coordinates

$$F_{i,u} = \begin{cases} B_{i,u} - \bar{B}_i & \text{for } i = 1, \dots, p - s - 1 \\ u - 1/2 & \text{for } i = p - s \end{cases}$$

Then  $LR\{H_{c\ell}(r) | H_{c\ell}(p)\}$  converges as in (11) using the present  $F$ .

Table 4 reports the simulated asymptotic distribution of the rank test reported in Theorem 6. The first panel gives the standard case where  $s = r$  and corresponds to Johansen (1995, Table 15.3). For  $p - r = 1$  the asymptotic distribution is actually  $\chi^2$  and the numbers are the standard numerically calculated ones rather than simulated ones. The second and the third panel report the distribution for the rank deficient case  $H_{c\ell}^\circ(s)$  where  $H_{c\ell}(s)$  holds, but  $H_c(s)$  fails. The distribution is shifted to the left when  $r - s > 0$  as in Table 1.

**Table 4.** Quantiles, mean, and variance of  $LR\{H_{c\ell}(r)|H_{c\ell}(p)\}$ , where the data generating process satisfies  $H_c^\circ(s) = H_{c\ell}(s) \setminus H_c(s)$  with  $s \leq r$ .

$r - s$	$p - r$	50%	80%	85%	90%	95%	97.5%	99%	Mean	Var
0	1	0.45	1.64	2.07	2.71	3.84	5.02	6.63	1	2
	2	7.61	11.09	12.04	13.30	15.35	17.27	19.74	8.24	14.29
	3	18.66	23.72	25.03	26.76	29.47	31.95	34.99	19.29	31.38
	4	33.52	40.07	41.71	43.86	47.22	50.21	53.94	34.15	53.86
1	1	0.38	1.33	1.66	2.13	2.93	3.72	4.74	0.79	1.08
	2	6.01	8.34	8.96	9.78	11.10	12.34	13.87	6.37	6.53
	3	15.49	19.14	20.08	21.30	23.21	24.99	27.14	15.88	16.73
	4	28.82	33.82	35.07	36.70	39.20	41.50	44.27	29.24	31.96
2	1	0.34	1.19	1.47	1.87	2.55	3.19	4.00	0.69	0.79
	2	5.43	7.34	7.84	8.51	9.57	10.56	11.81	5.70	4.46
	3	14.17	17.26	18.04	19.05	20.64	22.09	23.86	14.48	12.00
	4	26.62	30.92	31.98	33.38	35.52	37.46	39.79	26.95	23.82

The second case is the test for the same hypothesis  $H_{c\ell}(r)$  in the rank deficient case where  $H_c^\circ(s) = H_c(s)/H_{c\ell}(s - 1)$  holds for  $s \leq r$ .

**Theorem 7.** Consider the rank hypothesis  $H_{c\ell}(r) : \text{rank } \Pi \leq r$ . Suppose  $H_c^\circ(s) = H_c(s) \setminus H_{c\ell}(s - 1)$  holds for some  $s \leq r$ , so that  $\text{rank } \Pi = \text{rank } \Pi^* = s$  and that the  $I(1)$  condition is satisfied for that  $s$ . Let  $B_u$  be a  $(p - s)$ -dimensional standard Brownian motion on  $[0, 1]$ . Define a  $(p - s)$ -dimensional vector  $F_u$  as the de-meant Brownian motion

$$F_u = B_u - \bar{B} = B_u - \int_0^1 B_v dv.$$

Then  $LR\{H_{c\ell}(r) | H_{c\ell}(p)\}$  converges as in (11) using the present  $F$ .

Table 5 reports the simulated asymptotic distribution of the rank test reported in Theorem 7. The first panel where  $s = r$  and corresponds to Table A.2 of Johansen and Juselius (1990). It is shifted to the right when compared to the first panel of Table 4. The second and the third panel of Table 5 report the distribution for the rank deficient case  $H_c^\circ(s)$  for  $s < r$ . In those case the distribution is shifted to the left relative to the first panel as in Tables 1 and 4.

**Table 5.** Quantiles, mean, and variance of  $LR\{H_{c\ell}(r)|H_{c\ell}(p)\}$ , where the data generating process satisfies  $H_c^\circ(s) = H_c(s) \setminus H_{c\ell}(s - 1)$  with  $s \leq r$ .

$r - s$	$p - r$	50%	80%	85%	90%	95%	97.5%	99%	Mean	Var
0	1	2.45	4.90	5.60	6.56	8.15	9.72	11.71	3.04	6.95
	2	9.39	13.36	14.41	15.80	18.03	20.14	22.80	10.03	18.66
	3	20.30	25.70	27.09	28.89	31.75	34.37	37.61	20.95	35.73
	4	35.19	42.01	43.71	45.94	49.38	52.52	56.31	35.84	58.26
1	1	1.51	3.12	3.55	4.12	5.04	5.92	7.03	1.87	2.72
	2	7.21	9.95	10.66	11.61	13.09	14.47	16.21	7.60	8.95
	3	16.78	20.75	21.75	23.08	25.13	26.98	29.32	17.20	19.57
	4	30.25	35.49	36.81	38.51	41.15	43.56	46.46	30.69	35.22
2	1	1.16	2.54	2.89	3.36	4.09	4.76	5.62	1.48	1.81
	2	6.38	8.66	9.25	10.03	11.26	12.40	13.80	6.69	6.23
	3	15.27	18.64	19.49	20.61	22.35	23.94	25.88	15.61	14.27
	4	28.00	32.45	33.58	35.05	37.32	39.37	41.85	28.26	26.55

In the third case we consider the test for the hypothesis  $H_c(r)$  in the rank deficient case where  $H_c^\circ(s) = H_c(s)/H_{c\ell}(s - 1)$  holds for  $s < r$ .

**Theorem 8.** Consider the rank hypothesis  $H_c(r) : \text{rank} \Pi \leq r$ . Suppose  $H_c^\circ(s) = H_c(s) \setminus H_{c\ell}(s-1)$  holds for some  $s \leq r$  so that  $\text{rank} \Pi = \text{rank}(\Pi, \mu) = s$  and that the  $I(1)$  condition is satisfied for that  $s$ . Let  $B_u$  be a  $(p-s)$ -dimensional standard Brownian motion on  $[0, 1]$ . Define a  $(p-s+1)$ -dimensional vector  $F_u$  given as

$$F_u = \begin{pmatrix} B_u \\ 1 \end{pmatrix}. \tag{23}$$

Then  $LR\{H_c(r) | H_c(p)\}$  converges as in (11) using the present  $F$ .

Table 6 reports the simulated asymptotic distribution of the rank test reported in Theorem 8. The first panel gives the standard case where  $s = r$  and corresponds to Johansen (1995, Table 15.2). The second and the third panel report the distribution for the rank deficient case  $H_c^\circ(s)$  for  $s < r$ . Once again, the distribution shifts to the left in the rank deficient case.

**Table 6.** Quantiles, mean, and variance of  $LR\{H_c(r)|H_c(p)\}$ , where the data generating process satisfies  $H_c^\circ(s) = H_c(s) \setminus H_{c\ell}(s-1)$  with  $s \leq r$ .

$r-s$	$p-r$	50%	80%	85%	90%	95%	97.5%	99%	Mean	Var
0	1	3.44	5.86	6.56	7.52	9.13	10.69	12.74	4.04	6.89
	2	11.40	15.43	16.49	17.91	20.18	22.33	25.03	12.02	19.50
	3	23.31	28.86	30.28	32.15	35.06	37.74	41.04	23.95	38.13
	4	39.20	46.23	47.99	50.28	53.82	57.05	61.01	39.84	62.48
1	1	2.74	4.27	4.70	5.27	6.21	7.10	8.25	3.05	2.75
	2	9.47	12.30	13.04	14.01	15.54	16.96	18.74	9.84	9.81
	3	20.04	24.19	25.25	26.63	28.76	30.71	33.13	20.45	21.78
	4	34.51	40.03	41.40	43.17	45.93	48.43	51.41	34.95	39.09
2	1	2.62	3.89	4.22	4.68	5.41	6.10	6.96	2.84	1.87
	2	8.86	11.26	11.87	12.67	13.93	15.10	16.54	9.14	7.06
	3	18.77	22.37	23.27	24.43	26.23	27.88	29.91	19.09	16.34
	4	32.40	37.23	38.43	39.98	42.35	44.52	47.08	32.76	30.09

The final case is the test for the hypothesis  $H_c(r)$  in the rank deficient case where  $H_{c\ell}^\circ(s) = H_{c\ell}(s-1)/H_c(s-1)$  for  $s < r$ . In this case the limiting distribution has nuisance parameters. We do not give the result here, since it is complicated to state and it does not seem particularly useful in practice. Indeed in practical work, this type of data generating process can often be ruled through visual data inspection as discussed in Section 3.1. Furthermore, it would be hard to deal with the nuisance parameters in applications.

It is worth noting that the proof in this final case would be somewhat different from the proof of Theorems 1, 6–8. They are all proved by modifying the argument of Johansen (1995, Sections 10 and 11). However, in the final case, a cointegration vector with random coefficients arise. Therefore, the analysis is best carried out in terms of the dual eigenvalue problem  $0 = \det(\lambda S_{00} - S_{01} S_{11}^{-1} S_{10})$  as opposed to the standard eigenvalue problem  $0 = \det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01})$ .

### 3.5. Asymptotic Theory for the Test on the Cointegrating Vectors

We now consider the tests on the cointegrating vectors in the rank deficient case when a constant is present in the model. There is now a wide range of possible limit distributions. Only a few of these will be discussed.

The unrestricted model is  $H_c(r)$  where the constant is restricted to the cointegrating space. Thus, in the full rank case the Granger-Johansen representation (22) has a zero linear slope  $\tau_\ell = 0$  and level satisfying  $\beta' \tau_c = -\beta_c$ .

Consider now the hypothesis of a known cointegrating vector, (21). It is now important whether the hypothesized level for the cointegrating vector,  $b_c$  is zero or not. If  $b_c \neq 0$  then a nuisance parameter

depending on  $b, b_c$  would appear in the limit distributions in the rank deficient case. If  $b_c = 0$  then the limit distributions are simpler. Fortunately, the zero level case is the most natural hypothesis in most applications. The asymptotic theory for the test statistic is described in the following theorems.

**Theorem 9.** Consider the hypothesis  $H_{c,\beta}(r) : (\Pi, \mu) = \alpha b^{*'} where  $b^* = (b', b_c)'$ . Here,  $\alpha, b$  have dimension  $p \times r$  while  $b_c'$  is an  $r$ -vector, where  $\alpha$  is unknown and  $b^*$  is known and  $b$  has full column rank. Suppose  $H_z(0)$  is satisfied so that  $\Pi = 0, \mu = 0$ , and  $s = 0$  and that the  $I(1)$  condition is satisfied. Let  $B$  be a  $p$ -dimensional standard Brownian motion on  $[0, 1]$ , where the first  $r$  components are denoted  $B_1$ . Define the  $(p - s + 1)$ -dimensional process  $F_u = (B_u', 1)$  as in (23). Then it holds, for  $T \rightarrow \infty$ , that$

$$LR\{H_{z,\beta}(r) \mid H_z(p)\} \xrightarrow{D} \text{tr}\left\{ \int_0^1 dB_u F_u' \left( \int_0^1 F_u F_u' du \right)^{-1} \int_0^1 F_u (dB_u)' - \int_0^1 dB_u B_{1,u}' \left( \int_0^1 B_{1,u} B_{1,u}' du \right)^{-1} \int_0^1 B_{1,u} (dB_u)' \right\}. \quad (24)$$

The convergence of the test statistic  $LR\{H_{c,\beta}(r) \mid H_c(p)\}$  holds jointly with the convergence for the rank test statistic  $LR\{H_c(r) \mid H_c(p)\}$ , for  $s = 0$ , in Theorem 8. Thus, when  $s = 0$  a formula of the type (9) implies that the limit distribution of the test statistic for known  $\beta$  within the model with rank of at most  $r$  satisfies can be found as the difference of the two limiting variables.

Table 7 reports the asymptotic distribution of the test for known cointegrating vector in the model where the rank is at most  $r$ . When  $s = r$ , the asymptotic distribution is  $\chi^2$  with  $r(p + 1 - r)$  degrees of freedom, see Johansen and Juselius (1990, p. 193–194), Johansen et al. (2000, Lemma A.5). When  $s = 0$  the distribution is simulated according to Theorem 9. It is shifted to the right relative to the case  $s = r$ .

Table 8 reports the simulated asymptotic distribution of the test for known cointegrating vector in the model where the rank is unrestricted. The distribution is shifted to the right in the rank deficient case. As in the zero level case, the expectations reported in Tables 6 and 7 add up to the expectation reported in Table 8. In the full rank case  $s = r$  the statistics in Tables 6 and 7 are independent, as proved below, so the variances are additive.

**Theorem 10.** Consider the hypothesis  $H_{c,\beta}^\circ(r)$ . Suppose  $H_c^\circ(r) = H_c^\circ(r) / H_{c_\ell}^\circ(r - 1)$  is satisfied and that the  $I(1)$  condition holds with  $s = r$ . Then the rank test statistic  $LR\{H_c^\circ(r) \mid H_c^\circ(p)\}$  and the statistic  $LR\{H_{c,\beta}^\circ(r) \mid H_c^\circ(r)\}$  for testing a simple hypothesis on the cointegrating vector are asymptotically independent. The asymptotic distribution of the rank statistic  $LR\{H_c^\circ(r) \mid H_c^\circ(p)\}$  is given in Theorem 1, while the statistic for the cointegrating vector  $LR\{H_{c,\beta}^\circ(r) \mid H_c^\circ(r)\}$  is asymptotically  $\chi^2\{r(p + 1 - r)\}$ .

**Table 7.** Quantiles, mean, and variance of  $LR\{H_{c,\beta}(r) \mid H_c(r)\}$ , where the data generating process satisfies  $H_{c,\beta}^\circ(s)$ .

$p$	$r$	$s$	50%	80%	85%	90%	95%	97.5%	99%	Mean	Var
2	1	1	1.39	3.22	3.79	4.61	5.99	7.38	9.21	2	4
		0	6.34	9.84	10.78	12.02	14.05	15.96	18.41	6.87	15.09
3	2	2	3.36	5.99	6.75	7.78	9.49	11.14	13.28	4	8
		0	12.45	17.48	18.79	20.53	23.26	25.76	28.91	13.12	30.71
3	1	1	2.37	4.64	5.32	6.25	7.82	9.35	11.35	3	6
		0	10.60	14.82	15.92	17.36	19.66	21.79	24.48	11.07	22.93

**Table 8.** Quantiles, mean, and variance of  $LR\{H_{c,\beta}(r)|H_c(p)\}$ , where the data generating process satisfies  $H_{c,\beta}^0(s)$ .

$p$	$r$	$s$	50%	80%	85%	90%	95%	97.5%	99%	Mean	Var
2	1	1	5.44	8.50	9.34	10.50	12.38	14.17	16.50	6.07	10.98
		0	9.32	13.37	14.44	15.88	18.18	20.31	22.98	9.94	19.72
3	2	2	7.44	11.02	11.99	13.29	15.37	17.37	19.88	8.09	15.09
		0	15.37	20.48	21.80	23.54	26.26	28.78	31.88	15.99	32.22
3	1	1	14.46	19.08	20.28	21.88	24.39	26.74	29.64	15.10	25.77
		0	20.35	25.89	27.31	29.15	32.04	34.72	38.02	20.96	38.07

#### 4. Applications of Results

We discuss how our results apply to the finite sample theory and to identification robust inference. An application to US treasury yields is given.

##### 4.1. Finite Sample Theory

The finite sample distribution of cointegration rank tests have been studied in various ways. When there are no nuisance parameters, the asymptotic distributions generally give good approximations. An example is the test for a unit root in a first order autoregression, where the finite sample distribution and the asymptotic distribution are nearly indistinguishable for  $T = 8$  observations, see [Nielsen \(1997\)](#). A Bartlett correction improves the asymptotic distribution further. Once there are nuisance parameters the situation is different. Under the rank hypothesis the asymptotic distribution differs if there are additional unit roots. This arises either with rank deficiency like here where the distributions tend to be shifted to the left and when there are double roots as in  $I(2)$  systems where the distributions are shifted to the right. [Nielsen \(2004\)](#) analyzed this through simulation and suggested to apply local-to-unity approximation that would average between the different asymptotic distributions. A similar idea was implemented analytically for canonical correlation models in [Nielsen \(1999\)](#). In a follow-up paper, [Nielsen \(2001\)](#) analyzed the effects of plugging parameter estimates into such corrections. [Johansen \(2002\)](#) suggested a Bartlett correction for such models. This works quite well when the nuisance parameters are such that they are far from giving additional unit roots. The issue is that the Bartlett correction asymptotes to infinity when there are additional unit roots. More recently, bootstrap methods have been explored by [Swensen \(2004\)](#) and by [Cavaliere et al. \(2012\)](#).

[Johansen \(2000\)](#) derives a Bartlett-type correction for the tests on the cointegrating relations. In Table 2 he considers the finite sample properties of a test comparing the test statistic  $LR\{H_{z,\beta}(1)|LR\{H_z(1)\}$  with the asymptotic  $\chi^2$ -approximation. Null rejection frequencies are simulated for dimensions  $p = 2, 5$ , a variety of parameter values, and a finite sample size  $T$ . In all the reported simulations the data generating process has rank of unity. The table shows that null rejection frequency can be very much larger for a nominal 5% test when the rank is nearly deficient.

Theorem 2 sheds some light on the behaviour of the test as the rank approaches deficiency. The Theorem shows that the test statistic converges for all deficient ranks. Table 2 indicates that the distribution shifts to the right in the rank deficient case. Thus, we should expect that null rejection frequency increases as the rank approaches deficiency, but it should be bounded away from unity.

##### 4.2. Identification Robust Inference

[Khalaf and Urga \(2014\)](#) were concerned with tests on cointegration vectors in situations where the cointegration rank is nearly deficient. Their results can be developed a little further using the present results.

The notation in Khalaf and Urga (2014) differs slightly from the present notation. The hypothesis of known cointegration vectors is stated as  $\beta_0 = (I_r, \mathbf{b}'_0)'$  for some known  $\mathbf{b}_0$ , corresponding to the present hypotheses  $H_{z,\beta}(r)$  and  $H_{cl,\beta}(r)$ . The test statistics are

$$LR(\mathbf{b}_0) = LR\{H_{m,\beta}(r)|H_m(p)\}, \tag{25}$$

$$LRC(\mathbf{b}_0) = LR\{H_{m,\beta}(r)|H_m(r)\}, \tag{26}$$

for  $m = z, cl$ . Moreover they consider the hypothesis  $H_{m,\Pi}(r)$ , say, of a known impact matrix  $\Pi$  of rank  $r$ . This is tested through the statistic

$$LR_* = LR\{H_{m,\Pi}(r)|H_m(p)\}. \tag{27}$$

When the rank is not deficient the test statistic  $LRC(\mathbf{b}_0)$  is asymptotically  $\chi^2_{r(p-r)}$ , see Johansen (1995, Section 7). The test statistic  $LR(\mathbf{b}_0)$  has a Dickey-Fuller type distribution as derived in Theorem 2 for the case without deterministic terms, contradicting the  $\chi^2$  asymptotics suggested by Khalaf and Urga (2014, Section 4). Table 2 indicates that this distribution is close to, but different from, a  $\chi^2_{p(p-r)}$ -distribution when  $p = 2, 3$  and  $p - r = 1$ . When  $p = 3$  and  $r = 1$ , the limiting distribution is further from a  $\chi^2_{p(p-r)}$ -distribution. Likewise, the statistic  $LR_*$  converges to a Dickey-Fuller-type distribution. This can be proved through a modification of the proof of Theorem 2.

Khalaf and Urga’s Theorem 1 is concerned with bounding the distribution of the likelihood ratio statistic for the hypothesis  $\Pi = ab'$ , where  $a, b$  are known  $p \times r$ -matrices so that  $b$  has rank  $r$ , against the alternative where  $\Pi$  is unrestricted. The idea of their Theorem is to come up with a bound to the critical value when  $a, b$  may have deficient rank  $s \leq r$ . Unfortunately, their theorem evolves around the incorrect  $\chi^2$  distribution although unit root testing is implicitly involved. We therefore reformulate the result in terms of the limiting distributions derived herein.

We consider the test statistic  $LR(\mathbf{b}_0) = LR\{H_{z,\beta}(1)|H_z(1)\}$  when the rank of  $\Pi$  is nearly deficient. Suppose the rank is nearly deficient in the sense that  $\Pi \approx T^{-1}M$  for some matrix  $M$  along the lines of the theory in Section 2.6. Then, intuitively, the limiting distribution will be a combination of those arising when the true rank is 0 and when it is 1. The asymptotic theory developed here gives the relevant bounds. In the case of the zero level model the Theorems 1 and 2 imply the following pointwise result.

**Theorem 11.** *Let  $\theta$  denote the parameters of the model (1). Consider the parameter space  $\Theta_z$  where the hypothesis  $H_{z,\beta}(1) : \Pi = \alpha b'$  holds. Here  $\alpha, b$  are both of dimension  $p \times 1$ . Here  $\alpha$  is unknown, while  $b$  is known and has full column rank. Suppose the data generating process satisfies the  $I(1)$  condition with  $s \leq 1$ . Let  $q_{z,s}$  be the asymptotic  $(1 - \psi)$  quantile of  $LR\{H_{z,\beta}(1)|H_z(1)\}$  when the data generating process satisfies  $H_{z,\beta}^o(s)$  for  $s = 0, 1$ . Let  $q_{z,*} = \max_{s=0,1} q_{z,s}$ . Then it holds for all  $\theta \in \Theta_z$  that*

$$\lim_{T \rightarrow \infty} P[LR\{H_{z,\beta}(1)|H_z(1)\} \geq q_{z,*}] \leq \psi. \tag{28}$$

The simulated values in Table 2 show that for  $\psi = 5\%$  then

$$q_{z,*} = \max(q_{z,0}, q_{z,1}) = \begin{cases} \max(9.05, 3.84) = 9.05 & \text{for } p = 2, \\ \max(13.82, 5.99) = 13.82 & \text{for } p = 3. \end{cases} \tag{29}$$

The interpretation is as follows. Suppose the hypothesis  $H_z(1)$  has not been rejected, but it is unclear whether the rank could be nearly deficient. Then the hypothesis of a known  $\beta_0$  is rejected if the statistic  $LR\{H_{z,\beta}(1)|H_z(1)\}$  is larger than  $q_{z,*}$ .

The bound for  $q_{z,*}$  seems very extreme. Khalaf and Urga therefore suggest to use the alternative statistic  $LR\{H_{z,\beta}(1)|H_z(p)\}$ . Theorem 11 could be modified to cover this statistic. The simulations in Table 3 indicate that we would then use bounds

$$\tilde{q}_{z,*} = \max(\tilde{q}_{z,0}, \tilde{q}_{z,1}) = \begin{cases} \max(9.70, 6.22) = 9.70 & \text{for } p = 2, \\ \max(20.83, 15.34) = 20.83 & \text{for } p = 3. \end{cases} \quad (30)$$

We can establish a similar result for the constant level model using Theorems 8 and 9. However, it is necessary to exclude the possibility of a linear trends in the rank deficient model as this would give a very complicated result.

**Theorem 12.** Let  $\theta$  denote the parameters of the model (14). Consider the parameter space  $\Theta_c$  where the hypothesis  $H_{c,\beta}(1) : (\Pi, \mu) = \alpha(b', b'_c)$  holds. Here  $\alpha, b$  are both of dimension  $p \times 1$ , while  $b_c$  is a scalar. Further  $b, b_c$  are known and  $b \neq 0$ . Suppose the data generating process satisfies the  $I(1)$  condition with  $s = 0$  or  $s = 1$ . Let  $q_{c,s}$  be the asymptotic  $(1 - \psi)$  quantile of  $LR\{H_{c,\beta}(1)|H_c(1)\}$  when the data generating process satisfies  $H_{c,\beta}^o(s)$  for  $s = 0, 1$ . Let  $q_{c,*} = \max_{s=0,1} q_{c,s}$ . Then it holds for all  $\theta \in \Theta_1$  that

$$\lim_{T \rightarrow \infty} P[LR\{H_{c,\beta}(1)|H_c(1)\} \geq q_{c,*}] \leq \psi. \quad (31)$$

The simulated values in Table 7 show that for  $\psi = 5\%$  then

$$q_{c,*} = \max(q_{z,0}, q_{z,1}) = \begin{cases} \max(14.05, 5.99) = 14.05 & \text{for } p = 2, \\ \max(19.66, 7.82) = 19.66 & \text{for } p = 3. \end{cases} \quad (32)$$

If the alternative is taken as  $H_c(p)$  instead of  $H_c(1)$  the bounds are modified as

$$\tilde{q}_{c,*} = \max(\tilde{q}_{c,0}, \tilde{q}_{c,1}) = \begin{cases} \max(18.18, 12.38) = 18.18 & \text{for } p = 2, \\ \max(32.04, 24.39) = 32.04 & \text{for } p = 3. \end{cases} \quad (33)$$

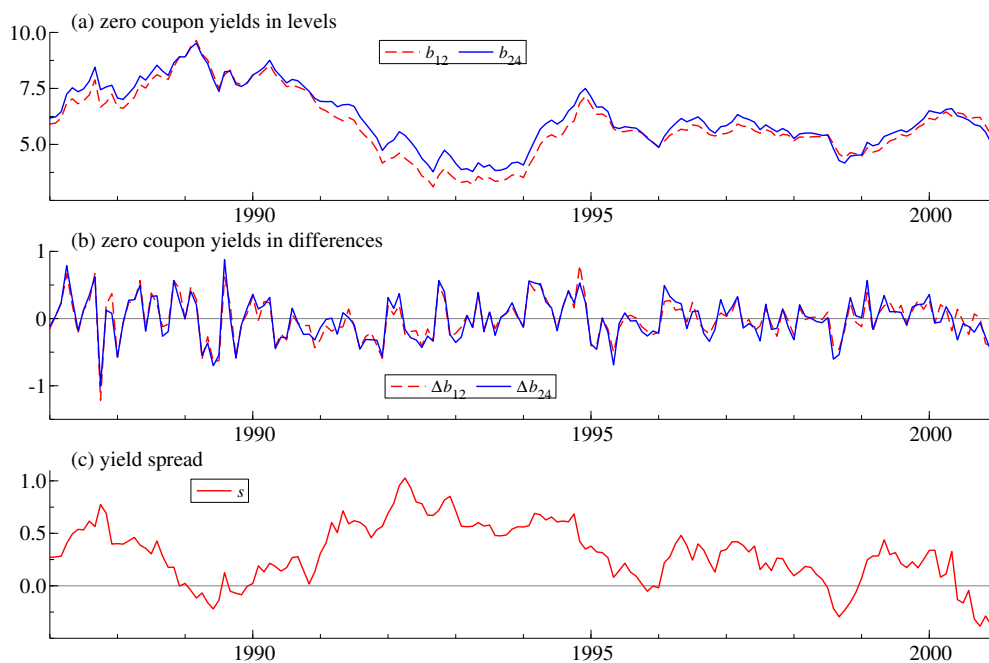
The bounds (32), (33) for the constant level model appear further apart than the corresponding bounds (29), (30) for the zero level model. So in the constant level case there is perhaps less reason to use the test against the unrestricted model.

#### 4.3. Empirical Illustration

The identification robust inference can be illustrated using a series of monthly US treasury zero-coupon yields over the period 1987:8 to 2000:12. The data are taken from Giese (2008) and runs from the start of Alan Greenspan's chairmanship of the Fed and finishes before the burst of the dotcom bubble. Giese considers 5 maturities (1, 3, 18, 48, 120 months), but here we only consider 2 maturities (12, 24 months). The empirical analysis uses OxMetrics, see Doornik and Hendry (2013).

Figure 1 shows the data in levels and differences along with the spread. The spread does not appear to have much of a mean reverting behaviour. It is not crossing the long-run average for periods of up to 4 years. This point towards a random walk behaviour which contradicts the expectations hypothesis in line with Giese's analysis. She finds two common trends among five maturities. The two common trends can be interpreted as short-run and long-run forces driving the yield curve. The cointegrating relations match an extended expectations hypothesis where spreads are not cointegrated but two spreads cointegrate. This is sometimes called butterfly spreads and gives a more flexible match to the yield curve. This is in line with earlier empirical work. Hall et al. (1992), among others, found only one common trend when looking at short-term maturities, while Shea (1992); Zhang (1993) and Carstensen (2003) found more than one common trend when including longer maturities.





**Figure 1.** Zero coupon yields in (a) levels; (b) differences; and (c) spread.

A vector autoregression of the form (14) with an intercept,  $k = 4$  lags as well as a dummy variable for 1987:10 was fitted to the data. This has the form

$$\Delta X_t = \Pi X_{t-1} + \mu + \sum_{i=1}^3 \Gamma_i \Delta X_{t-i} + \Phi 1_{(t=1987:10)} + \varepsilon_t \quad \text{for } t = 1, \dots, T,$$

where  $X_t$  is the bivariate vector of the 12 and 24 month zero-coupon yields and periods  $t = 1$  and  $t = T$  correspond to 1987:8 and 2000:12 giving  $T = 161$ .

Table 9 reports specification test statistics with  $p$ -values in square brackets. The tests do not provide evidence against the initial model. They are the autocorrelation test of Godfrey (1978) the cumulant based normality test, see Doornik and Hansen (2008), and the ARCH test of Engle (1982). For the validity of applying the autoregressive and normality tests for non-stationarity autoregressions, see Engler and Nielsen (2009), Kilian and Demiroglu (2000), and Nielsen (2006).

The dummy variable matches the policy intervention after the stock market crash on 19 October 1987. Empirically, the dummy variable can be justified in two ways. First, the plot of yield differences in Figure 1b indicate a sharp drop in yields at that point. Secondly, the robustified least squares algorithm analyzed in Johansen and Nielsen (2016) could be employed for each of the two equations in the model. The algorithm uses a cut-off for outliers in the residuals that is controlled in terms of the gauge, which is the frequency of falsely detected outliers that can be tolerated. The gauge is chosen small in line with recommendations of Hendry and Doornik (2014, Section 7.6), see also Johansen and Nielsen (2016). Thus, we choose a cut-off of 3.02 corresponding to a gauge of 0.25%. When running the autoregressive distributed lag models without outliers, only 1987:10 has an absolute residual exceeding the cut-off. Next, when re-running the model including a dummy for 1987:10, no further residuals exceed the cut-off. This is a fixed point for the algorithm. The detection of outliers may have some impact on specification tests, estimation, and inference. Johansen and Nielsen (2009, 2016) analyze the impact on estimation when the data generating process has no outliers. They find that outlier detection only gives a modest efficiency loss compared to standard least squares when the cut-off is as large as chosen here. Berenguer-Rico and Nielsen (2017) find a considerable impact on the normality test employed above. At present, there is no theory for these algorithms for data generating processes with

outliers, albeit some results are available for cointegration analysis with known break date, including the broken trend analysis of Johansen et al. (2000) and the structural change model of Hansen (2003).

Table 9. Specification tests for the unrestricted vector autoregression.

Test	$b_{12,t}$	$b_{24,t}$	Test	System
$\chi^2_{normality} (2)$	3.8 [0.15]	4.1 [0.13]	$\chi^2_{normality} (4)$	4.3 [0.36]
$F_{ar,1-7} (7, 144)$	1.7 [0.11]	1.0 [0.45]	$F_{ar,1-7} (28, 272)$	1.2 [0.24]
$F_{arch,1-7} (7, 147)$	1.8 [0.09]	1.0 [0.41]		

Table 10 reports cointegration rank tests. The fifth column shows conventional  $p$ -values based on Tables 4 and 6 for  $s = r$  corresponding to Johansen (1995, Tables 15.2, 15.3). The sixth column shows  $p$ -values based on Tables 5 and 6 assuming data have been generating by a model satisfying  $H_c(0) = H_z(0)$ . In both cases the  $p$ -values are approximated by fitting a Gamma distribution to the reported mean and variance, see Nielsen (1997); Doornik (1998) for details. As expected, the latter  $p$ -values tend to be higher than the former. Overall this provide overwhelming evidence in favour of a pure random walk model in line with Giese (2008).

Table 10. Cointegration rank tests.

Hypothesis	$r$	Likelihood	LR	$p$ -Value	
				$s = r$	$H_c(0)$
$H_{c\ell}(2) = H_c(2)$	2	134.63			
$H_{c\ell}(1)$	1	133.71	1.8	0.18	0.39
$H_c(1)$	1	133.71	1.8	0.80	0.75
$H_{c\ell}(0)$	0	129.70	9.8	0.30	0.46
$H_c(0)$	0	129.21	10.8	0.57	0.57

If we have a strong belief in the expectation hypothesis we would, perhaps, ignore the rank tests and seek to test the expectations hypothesis directly. If we maintain the model  $H_c(1)$ , we could have to contemplate that the cointegration vectors could be nearly unidentified. A mild form of the expectation hypothesis is that the spread is zero mean stationary. Thus, we test the restriction  $b^* = (1, -1, 0)$ . The likelihood ratio statistic is 4.0. Assuming the data generating process satisfies either  $H_c^o(0)$  or  $H_c^o(1)$ , but not by  $H_{c\ell}^o(0)$ , we can apply the Khalaf-Urga (2014)-type bound test established in Theorem 12. The 95% bound in (32) is 14.05 so the hypothesis cannot be rejected based on this test. This contrasts with the above rank tests which gave strong evidence against the expectations hypothesis. The results reconcile if the bounds test does not have much power in the weakly identified case. Indeed, this seems to be the case when looking at Table 3,  $\rho = 0.99$ -panels in Khalaf and Urga (2014), corresponding to near rank deficiency or weak identification. Thus, assuming the rank is one when in fact the data generating process appears to be nearly rank deficient seems to reduce power for tests on the cointegrating vector. That is, when the alleged cointegrating vector is not cointegrating it would be useful to be able to falsify the economic hypothesis. The above mentioned simulations indicate that this is not the case.

### 5. Conclusions

We have derived asymptotic theory for cointegration rank tests and tests on cointegrating vectors in the rank deficient case. The asymptotic distributions have been simulated and tabulated. The results shed some light on the finite sample theory for cointegration analysis. They can be used to improve the theory on identification robust inference developed by Khalaf and Urga (2014). This was applied to two US treasury yield series.

It appears that large distortions arise when applying standard cointegration inference in the situation where the rank is deficient or nearly deficient. The rank hypothesis gives an inequality for the rank, that is  $\text{rank } \Pi \leq r$ . This includes cases where the rank is  $r$  and where it is less than  $r$ . Thus, the parameter space for the model where  $\text{rank } \Pi \leq r$  therefore has a lower dimensional subset where the rank is deficient. Inferential procedures for rank determination are consistent but do leave a positive probability of deciding for a deficient rank in finite samples. In practice, it is therefore possible to end up in a situation of rank deficiency or near deficiency. When proceeding to testing restrictions on the cointegrating vectors, the model is therefore mis-specified or nearly mis-specified.

The asymptotic analysis of the test distributions gives the following results. When testing for cointegration rank, the distribution shifts to the left when the rank is deficient. When testing for restrictions on the cointegrating vector, the distribution shifts to the right when the rank is deficient. When the rank is nearly deficient the distribution will tend to shift in similar directions. As a consequence, a test for cointegration restrictions using conventional critical has a size control problem previously observed by Johansen (2000). One can instead apply identification robust tests as suggested by Khalaf and Urga (2014), but our impression is that while these tests are better behaved in terms of size, they have modest power to reject incorrect restrictions.

Our recommendation is to test for rank before testing restrictions on cointegrating vectors in line with Johansen's framework. If the conclusion from the rank determination is ambiguous it is best to proceed with caution and possibly explore different choices for rank. This is a common theme in the applied work of Juselius.

**Author Contributions:** The authors made equal contributions.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proofs

Processes are considered on the space of right continuous processes with left limits,  $D[0, 1]$ . A discrete time process  $X_t$  for  $t = 1, \dots, T$  is embedded in  $D[0, 1]$  through  $X_{\text{integer}(Tu)}$  for  $0 \leq u \leq 1$ . For processes  $Y_t, Z_t$  for  $t = 1, \dots, T$  the residuals from regressing  $Y_t$  on  $Z_t$  are denoted  $(Y_t | Z_t) = Y_t - \sum_{s=1}^T Y_s Z_s' (\sum_{s=1}^T Z_s Z_s')^{-1} Z_t$ .

**Proof of Theorem 1.** This follows the outline of the proof in Johansen (1995, §10, 11). Let  $\Pi = \alpha_0 \beta_0'$  for  $p \times s$ -matrices  $\alpha_0, \beta_0$  with full column rank. Let  $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$ . Under the I(1) condition the Granger-Johansen representation (6) holds with rank  $s$  and Johansen's Lemma 10.1 stands with  $r$  replaced by  $s$ . His Lemmas 10.2, 10.3 hold with  $B_T = \beta_{0\perp} (\beta_{0\perp}' \beta_{0\perp})^{-1}$  so that, on  $D[0, 1]$ ,

$$T^{-1/2} B_T' X_{\text{integer}(Tu)} = B_T' C T^{-1/2} \sum_{t=1}^{\text{integer}(Tu)} \varepsilon_t + o_P(1). \quad (\text{A1})$$

For later use we will note that the Brownian motion  $B$  can be chosen as follows. For any orthogonal square matrix  $\tilde{M}$  so  $\tilde{M}' \tilde{M} = I_{p-s}$  choose the  $(p-s)$ -dimensional standard Brownian motion  $B$  so that

$$T^{-1/2} \tilde{M}' (\alpha_{0\perp}' \Omega \alpha_{0\perp})^{-1/2} \alpha_{0\perp}' \Gamma \beta_{0\perp} (\beta_{0\perp}' \beta_{0\perp})^{-1} \beta_{0\perp}' X_{[Tu]} \xrightarrow{D} B_u \quad (\text{A2})$$

on  $D[0, 1]$ .  $\square$

**Proof of Theorem 2.** Introduce the notation  $\widehat{\Omega}_U = S_{00} - S_{01}S_{11}^{-1}S_{10}$  for the unrestricted variance estimator and  $\widehat{\Omega}_R = S_{00} - S_{01}b(b'S_{11}b)^{-1}b'S_{10}$  for the restricted variance estimator. Then the likelihood ratio test statistic satisfies

$$LR\{H_{z,\beta}(r) \mid H_z(p)\} = -T \log \frac{\det(\widehat{\Omega}_U)}{\det(\widehat{\Omega}_R)} = T \log \det\{I_p + \widehat{\Omega}_U^{-1}(\widehat{\Omega}_R - \widehat{\Omega}_U)\}.$$

If it is shown that  $\widehat{\Omega}_U$  is consistent and  $T(\widehat{\Omega}_R - \widehat{\Omega}_U)$  converges in distribution then

$$LR\{H_{z,\beta}(r) \mid H_z(p)\} = \text{tr}\{\Omega^{-1}T(\widehat{\Omega}_R - \widehat{\Omega}_U)\} + o_p(1), \tag{A3}$$

following Johansen (1995, p. 224). The consistency of the unrestricted variance estimator  $\widehat{\Omega}_U$  follows from Johansen (1995, Lemma 10.3) used with  $r = s = 0$  and  $B_T = I_p$ .

Consider  $T(\widehat{\Omega}_R - \widehat{\Omega}_U)$ . Note first that the data generating process has cointegration rank  $s = 0$ . Thus  $\alpha_0, \beta_0$  are empty matrices so that their complements can be chosen as the identity matrix. The I(1) condition then implies that  $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$  is invertible. The asymptotic convergence in (A2) then reduces to

$$T^{-1/2}\tilde{M}'\Omega^{-1/2}\Gamma X_{\text{integer}(Tu)} = T^{-1/2}\tilde{M}'\Omega^{-1/2} \sum_{t=1}^{\text{integer}(Tu)} \varepsilon_t + o_p(1) \xrightarrow{D} B_u, \tag{A4}$$

where  $B$  is a standard Brownian motion of dimension  $p$  and for any orthonormal  $\tilde{M}$  so that  $\tilde{M}'\tilde{M} = I_p$ . In particular, we will choose  $\tilde{M}$  so

$$\tilde{M} = \begin{bmatrix} \{b'\Gamma^{-1}\Omega(\Gamma')^{-1}b\}^{-1/2}b'\Gamma^{-1}\Omega^{1/2} \\ (b'_\perp\Gamma'\Omega^{-1}\Gamma b_\perp)^{-1/2}b'_\perp\Gamma'\Omega^{-1/2} \end{bmatrix}. \tag{A5}$$

Let  $B_{1,u}, B_{2,u}$  be the first  $r$  and the last  $p - r$  coordinates of  $B_u$ , respectively. Then we get

$$\{b'\Gamma^{-1}\Omega(\Gamma')^{-1}b\}^{-1/2}b'X_{\text{integer}(Tu)} \xrightarrow{D} B_{1,u}.$$

The variance estimators are  $\widehat{\Omega}_R = S_{\varepsilon\varepsilon} - S_{\varepsilon 1}b(b'S_{11}b)^{-1}b'S_{1\varepsilon}$  and  $\widehat{\Omega}_U = S_{\varepsilon\varepsilon} - S_{\varepsilon 1}S_{11}^{-1}S_{1\varepsilon}$ . In particular, the difference of the variance estimators is

$$T(\widehat{\Omega}_R - \widehat{\Omega}_U) = T\{S_{\varepsilon 1}M(M'S_{11}M)^{-1}M'S_{1\varepsilon} - S_{\varepsilon 1}bm(m'b'S_{11}bm)^{-1}mb'S_{1\varepsilon}\}, \tag{A6}$$

for any invertible matrices  $M, m$  and in particular for  $M' = \tilde{M}'\Omega^{-1/2}\Gamma$  and  $m = \{b'\Gamma^{-1}\Omega(\Gamma')^{-1}b\}^{-1/2}$ . In light of the identity  $\tilde{M}'\tilde{M} = I_p$ , the random walk convergence in (A4), the rules for the trace and the notation  $v = mb$  write

$$\begin{aligned} \text{tr}\{\Omega^{-1}T(\widehat{\Omega}_R - \widehat{\Omega}_U)\} &= \text{tr}\{\tilde{M}'\Omega^{-1/2}T(\widehat{\Omega}_R - \widehat{\Omega}_U)\Omega^{-1/2}\tilde{M}\} \\ &= \text{tr}[\tilde{M}'\Omega^{-1/2}T\{S_{\varepsilon 1}M(M'S_{11}M)^{-1}M'S_{1\varepsilon} - S_{\varepsilon 1}v(v'S_{11}v)^{-1}v'S_{1\varepsilon}\}\Omega^{-1/2}\tilde{M}]. \end{aligned}$$

Then the product moment convergence results in Johansen (1995, Lemma 10.3) imply

$$\begin{aligned} \text{tr}\{\Omega^{-1}T(\widehat{\Omega}_R - \widehat{\Omega}_U)\} &\xrightarrow{D} \text{tr}\left\{\int_0^1 dB_u B_u' \left(\int_0^1 B_u B_u' du\right)^{-1} \int_0^1 B_u (dB_u)'\right. \\ &\quad \left. - \int_0^1 dB_u B_{1,u}' \left(\int_0^1 B_{1,u} B_{1,u}' du\right)^{-1} \int_0^1 B_{1,u} (dB_u)'\right\}. \end{aligned}$$

This is also the limit of the likelihood ratio test statistic due to (A3). The convergence holds jointly with the convergence of the likelihood ratio test statistic for rank in Theorem 1 since the orthogonal matrix  $\tilde{M}$  in (A2) can be chosen freely.  $\square$

**Proof of Theorem 3.** We need a number of results from Johansen (1995). Let  $B, V$  be independent standard Brownian motions. His Theorem 11.1 shows

$$LR\{H_z(r)|H_z(p)\} \xrightarrow{D} \text{tr} \left\{ \int_0^1 dB_u B_u' \left( \int_0^1 B_u B_u' du \right)^{-1} \int_0^1 B_u dB_u' \right\}, \tag{A7}$$

while his Lemma 13.8 shows

$$LR\{H_{z,\beta}(r)|H_z(r)\} \xrightarrow{D} \text{tr} \left\{ \int_0^1 dV_u B_u' \left( \int_0^1 B_u B_u' du \right)^{-1} \int_0^1 B_u dV_u' \right\}. \tag{A8}$$

Johansen does not explicitly argue that the convergence results hold jointly. This can be done by going into the proofs of the results, find the asymptotic expansions of the test statistic, and express them in terms of random walks that converge to the processes  $B, V$  when normalized by  $T^{1/2}$ . The asymptotic distribution in (A8) is mixed Gaussian since  $B, V$  are independent. Thus, by conditioning on  $B$  we see that  $LR\{H_{z,\beta}(r)|H_z(r)\}$  is asymptotically  $\chi^2$  and hence independent of  $B$ . In turn the two test statistics are asymptotically independent.  $\square$

**Proof of Theorem 5.** We follow Stockmarr and Jacobsen (1994) or Johansen (1995, Theorem 14.1, Lemma 14.3) and find that  $T^{-1/2}X_{\text{integer}(Tu)}$  converges to  $J_u$  as a process on  $D[0,1]$  while  $(S_{00}, S_{1e}, S_{11}/T)$  converges in distribution to  $(I_2, \int_0^1 J_u dB_u', \int_0^1 J_u J_u' du)$ .

Now, proceed as in the proof of Theorem 2. It has to be argued that  $\hat{\Omega}_U$  converges in probability to  $I_2$  and that  $T(\hat{\Omega}_R - \hat{\Omega}_U)$  has the limit distribution postulated in the Theorem. The convergence of the  $\hat{\Omega}_U$  follows from the listed properties of the product moment matrices. For  $T(\hat{\Omega}_R - \hat{\Omega}_U)$  we have as in Equation (A6) that

$$T(\hat{\Omega}_R - \hat{\Omega}_U) = T\{S_{e1}(S_{11})^{-1}S_{1e} - S_{e1}b(b'S_{11}b)^{-1}b'S_{1e}\}.$$

Again, we can apply the listed properties of the product moment matrices.  $\square$

**Proof of Theorem 6.** Similar to the proof of Theorem 1, the relevant Granger-Johansen representation is (22) with rank  $s$ . Use Johansen's Lemmas 10.2, 10.3 with  $B_T = \{\gamma(\gamma'\gamma)^{-1}, T^{-1/2}\tau_\ell(\tau_\ell'\tau_\ell)^{-1}\}$ , where  $\tau_\ell = C\mu$ , while  $\gamma \in \text{span}(\beta_{0\perp})$  so that  $\gamma'\tau_\ell = 0$  and the expansion (A1) is replaced by

$$T^{-1/2}B_T'X_{\text{integer}(Tu)} = \left\{ \begin{matrix} (\gamma'\gamma)^{-1}\gamma'CT^{-1/2} \sum_{t=1}^{\text{integer}(Tu)} \varepsilon_t \\ u \end{matrix} \right\} + o_P(1) \tag{A9}$$

on  $D[0,1]$ . Thus,  $\Delta X_t$  has a non-zero level, but this is eliminated by regression on the intercept.  $\square$

**Proof of Theorem 7.** Similar to the proof of Theorem 1. Use the Granger-Johansen representation (22) with rank  $s$  and  $\tau_\ell = C\mu = 0$ , and Johansen's Lemmas 10.2, 10.3 with  $B_T = \beta_{0\perp}(\beta_{0\perp}'\beta_{0\perp})^{-1}$  so that  $T^{-1/2}B_T'X_{\text{integer}(Tu)}$  has expansion (A1).  $\square$

**Proof of Theorem 8.** Similar to the proof of Theorem 1. Use the Granger-Johansen representation (22) with rank  $s$ , and  $\tau_\ell$ . Use Johansen's Lemmas 10.2, 10.3 with  $X_t, B_T$  and the expansion (A1) replaced by, respectively,  $X_t^* = (X_t', 1)'$ , the block diagonal matrix  $B_T^* = \text{diag}(B_T, T^{1/2})$  where  $B_T = \beta_{0\perp}(\beta_{0\perp}'\beta_{0\perp})^{-1}$ , and

$$T^{-1/2}B_T^*X_{\text{integer}(Tu)}^* = \left( \begin{matrix} B_T'CT^{-1/2} \sum_{t=1}^{\text{integer}(Tu)} \varepsilon_t \\ 1 \end{matrix} \right) + o_P(1) \tag{A10}$$

on  $D[0,1]$ .  $\square$

**Proof of Theorem 9.** The proof of Theorem 2 is modified noting that  $R_{1,t}$  is the  $(p + 1)$ -vector  $(X_{t-1}, 1)'$  corrected for lagged differences instead of  $X_{t-1}$  corrected for lagged differences. Choose  $\tilde{M}$  as in (A5). Replace (A4) by

$$\begin{pmatrix} T^{-1/2}\tilde{M}'\Omega^{-1/2}\Gamma & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_{\text{integer}(Tu)} \\ 1 \end{pmatrix} \xrightarrow{D} F_u. \tag{A11}$$

The difference of variance estimators in (A6) is now

$$T(\widehat{\Omega}_R - \widehat{\Omega}_U) = T\{S_{\varepsilon 1}M(M'S_{11}M)^{-1}M'S_{1\varepsilon} - S_{\varepsilon 1}b^*(b^{*'}S_{11}b^*)^{-1}b^{*'}S_{1\varepsilon}\}, \tag{A12}$$

where the invertible  $(p + 1)$ -dimensional matrix  $M$  now is chosen as

$$M = \begin{pmatrix} b'\Gamma^{-1}\Omega(\Gamma')^{-1}b & 0 & 0 \\ 0 & b'_\perp\Gamma'\Omega^{-1}\Gamma b_\perp & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1/2} \begin{pmatrix} b' & b'_c \\ b'_\perp\Gamma'\Omega^{-1}\Gamma & 0 \\ 0 & 1 \end{pmatrix} \tag{A13}$$

Viewed as a  $(3 \times 2)$ -block matrix, the two upper left equals the previous  $M$ . Since the random walk dominates a constant it holds that

$$\begin{pmatrix} T^{-1/2}I_p & 0 \\ 0 & 1 \end{pmatrix} M \begin{pmatrix} X_{\text{integer}(Tu)} \\ 1 \end{pmatrix} \xrightarrow{D} F_u. \tag{A14}$$

Moreover, the first  $r$  coordinates of  $MR_{1,t}$  are proportional to  $b^{*'}R_{1,t}$ . Thus the argument can be completed as in the proof of Theorem 2.  $\square$

**Proof of Theorem 10.** The proof of Theorem 3 has to be modified to allow for a constant term in the cointegrating vector. The arguments leading to asymptotic results for the test statistics are sketched in Johansen and Juselius (1990) and, with more details, in Johansen et al. (2000, Theorem 3.1, Lemma A.5).  $\square$

**Proof of Theorem 11.** Write

$$LR\{H_{z,\beta}(1)|H_z(1)\} = LR\{H_z(1)|H_z(p)\} - LR\{H_{z,\beta}(1)|H_z(p)\}. \tag{A15}$$

When  $s = 0$  Theorems 1 and 2 give expansions for the right hand expressions of (A15) and in turn for the desired test statistic on the left hand of (A15). This implies an asymptotic distribution with asymptotic  $(1 - \psi)$  quantile  $q_{z,0}$ , say. When  $s = 1$  Theorem 3 in a similar way gives an asymptotic  $(1 - \psi)$  quantile  $q_{z,1}$ . Thus, with  $q_{z,*} = \max_{s=0,1} q_{z,s}$  we get  $\lim_{T \rightarrow \infty} P[LR\{H_{z,\beta}(1)|H_z(1)\} \geq q_{z,*}] \leq \psi$ , both with  $s = 0$  and when  $s = 1$ .  $\square$

**Proof of Theorem 12.** Similar to the proof of Theorem 11, applying Theorems 8–10 instead Theorems 1–3.  $\square$

**References**

Abadir, Karim M. 1995. The limiting distribution of the t ratio under a unit root. *Econometric Theory* 11: 775–93. [CrossRef]

Andrews, Donald W. K., and Xu Cheng. 2012. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80: 2153–11. [CrossRef]

Berenguer-Rico, Vanessa, and Bent Nielsen. 2017. *Marked and Weighted Empirical Processes of Residuals With Applications to Robust Regressions*. Discussion Paper 841. Oxford: Department of Economics, University of Oxford.

Bernstein, David. 2014. *Asymptotic Theory for Unidentified Cointegration Estimators*. M.Phil. thesis, University of Oxford, Oxford, UK.

- Carstensen, Kai. 2003. Nonstationary term premia and cointegration of the term structure. *Economics Letters* 80: 409–13. [\[CrossRef\]](#)
- Cavaliere, Giuseppe, Anders Rahbek, and A. M. Robert Taylor. 2012. Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica* 80: 1721–40.
- Doornik, Jurgen A. 1998. Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* 12: 573–93. [\[CrossRef\]](#)
- Doornik, Jurgen A. 2007. *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake.
- Doornik, Jurgen A., and Henrik Hansen. 2008. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* 70: 927–39. [\[CrossRef\]](#)
- Doornik, Jurgen A. and David F. Hendry. 2013. *PcGive* 14. London: Timberlake, vol. 1.
- Dufour, Jean-Marie. 1997. Some impossibility theorems in econometrics with applications to structural and dynamic methods. *Econometrica* 65: 1365–87. [\[CrossRef\]](#)
- Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50: 987–1108. [\[CrossRef\]](#)
- Engler, Eric, and Bent Nielsen. 2009. The empirical process of autoregressive residuals. *Econometrics Journal* 12: 367–81. [\[CrossRef\]](#)
- Fachin, Stefano. 2000. Bootstrap and asymptotic tests of long-run relationships in cointegrated systems. *Oxford Bulletin of Economics and Statistics* 62: 543–51. [\[CrossRef\]](#)
- Giese, Julia. 2008. Level, slope, curvature: Characterising the yield curve in a cointegrated VAR model. *Economics* 2: 28.
- Godfrey, L. G. 1978. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica* 46: 1293–301. [\[CrossRef\]](#)
- Gredenhoff, Mikael, and Tor Jacobson. 2001. Bootstrap testing linear restrictions on cointegrating vectors. *Journal of Business & Economic Statistics* 19: 63–72.
- Hall, Anthony D., Heather M. Anderson, and Clive W. J. Granger. 1992. A cointegration analysis of treasury bill yields. *Review of Economics and Statistics* 74: 116–26. [\[CrossRef\]](#)
- Hansen, Peter Reinhard. 2003. Structural changes in the cointegrated vector autoregressive model. *Journal of Econometrics* 114: 261–95. [\[CrossRef\]](#)
- Hendry, David F., and Jurgen A. Doornik. 2014. *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*. London: MIT Press.
- Johansen, Søren. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12: 231–54. [\[CrossRef\]](#)
- Johansen, Søren. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59: 1551–580. [\[CrossRef\]](#)
- Johansen, Søren. 1992. Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics* 54: 383–97. [\[CrossRef\]](#)
- Johansen, Søren. 1995. *Likelihood Based Inference on Cointegration in the Vector Autoregressive Model*. Oxford: Oxford University Press.
- Johansen, Søren. 2000. A Bartlett correction factor for tests on the cointegrating relations. *Econometric Theory* 16: 740–77. [\[CrossRef\]](#)
- Johansen, Søren. 2002. A small sample correction of the test for cointegrating rank in the vector autoregressive model. *Econometrica* 70: 1929–61. [\[CrossRef\]](#)
- Johansen, Søren, and Katarina Juselius. 1990. Maximum likelihood estimation and inference on cointegration—With applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52: 169–210. [\[CrossRef\]](#)
- Johansen, Søren, Rocco Mosconi, and Bent Nielsen. 2000. Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal* 3: 216–49. [\[CrossRef\]](#)
- Johansen, Søren, and Bent Nielsen. 2009. Saturation by indicators in regression models. In *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry*. Edited by Jennifer L. Castle and Neil Shephard. Oxford: Oxford University Press, pp. 1–36.
- Johansen, Søren, and Bent Nielsen. 2016. Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scandinavian Journal of Statistics* 43: 321–81. [\[CrossRef\]](#)
- Juselius, Katarina. 2006. *The Cointegrated VAR Model*. Oxford: Oxford University Press.
- Khalaf, Lynda, and Giovanni Urga. 2014. Identification robust inference in cointegrating regressions. *Journal of Econometrics* 182: 385–96. [\[CrossRef\]](#)

- Kilian, Lutz, and Ufuk Demiroglu. 2000. Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence. *Journal of Business & Economic Statistics* 18: 40–50.
- Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock. 2014. Empirical evidence on inflation expectations in the new Keynesian Phillips curve. *Journal of Economic Literature* 52: 124–88. [[CrossRef](#)]
- Nielsen, Bent. 1997. Bartlett correction of the unit root test in autoregressive models. *Biometrika* 84: 500–504. [[CrossRef](#)]
- Nielsen, Bent. 1999. The likelihood ratio test for rank in bivariate canonical correlation analysis. *Biometrika* 86: 279–88. [[CrossRef](#)]
- Nielsen, Bent. 2001. Conditional test for rank in bivariate canonical correlation analysis. *Biometrika* 88: 874–80. [[CrossRef](#)]
- Nielsen, Bent. 2004. On the distribution of likelihood ratio test statistics for cointegration rank. *Econometric Reviews* 23: 1–23. [[CrossRef](#)]
- Nielsen, Bent. 2006. Order determination in general vector autoregressions. In *Time Series And Related Topics: In Memory of Ching-Zong Wei*. Edited by Hwai-Chung Ho, Ching-Kang Ing and Tze Leung Lai. Lecture Notes–Monograph Series. Beachwood: Institute of Mathematical Statistics, vol. 52, pp. 93–112.
- Nielsen, Bent, and Anders Rahbek. 2000. Similarity issues in cointegration models. *Oxford Bulletin of Economics and Statistics* 62: 5–22. [[CrossRef](#)]
- Paruolo, Paolo. 2001. The power of lambda max. *Oxford Bulletin of Economics and Statistics* 63: 395–403. [[CrossRef](#)]
- Phillips, Peter C. B., and Victor Solo. 1992. Asymptotics for linear processes. *Annals of Statistics* 20: 971–1001. [[CrossRef](#)]
- Shea, Gary S. 1992. Benchmarking the expectations hypothesis of the interest-rate term structure: An analysis of cointegration vectors. *Journal of Business & Economic Statistics* 10: 347–366.
- Stockmarr, Anders, and Martin Jacobsen. 1994. Gaussian diffusion and autoregressive processes: Weak convergence and statistical inference. *Scandinavian Journal of Statistics* 21: 403–19.
- Swensen, Anders Rygh 2004. Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* 74: 1699–714. Corrigendum in volume 77: 1703–704. [[CrossRef](#)]
- Zhang, Hua. 1993. Treasury yield curves and cointegration. *Applied Economics* 25: 361–67. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).