Asymptotic theory for cointegration analysis when the cointegration rank is deficient

David Bernstein
Department of Economics, University of Miami, Coral Gables, FL 33146, U.S.
davebernstein1@gmail.com
Bent Nielsen
Department of Economics, University of Oxford, Oxford OX1 3UQ
Nuffield College, Oxford OX1 1NF, U.K.
Programme on Economic Modelling, INET, University of Oxford
bent.nielsen@nuffield.ox.ac.uk

29 September 2014

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.

1 Introduction

Determination of the cointegration rank is an important part of analyzing the cointegrated vector autoregressive model. 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.

The analysis of the rank deficient case has bearing on two discussions in the literature. First, the results inform finite sample distribution theory for cointegration tests. 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 deficient or it is far from being deficient, see for instance Nielsen (1997). 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). When the parameters are closer to rank deficient a local-to-unity asymptotic expansion gives an improvement, see Nielsen (2004). 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.

Secondly, the results inform the current discussion of inference in cases of weakly identified parameters. Recently, Khalaf and Urga (2014) discussed tests for a known cointegrating vector in the nearly rank deficient situation. They 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.

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

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.

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 is written in equilibrium correction form as

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \qquad \text{for } t = 1, \dots, T, \tag{2.1}$$

where the innovations ϵ_t are independent normal $\mathsf{N}_p(0, \Omega)$-distributed. The parameters Π, Γ_i, Ω are freely varying p-dimensional square matrices so that Ω is symmetric, positive definite.

The hypothesis of reduced cointegration rank is formulated as

$$\mathsf{H}_z(r): \qquad \operatorname{rank} \Pi \leq r, \tag{2.2}$$

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

$$\mathsf{H}_z(0) \subset \cdots \subset \mathsf{H}_z(r) \subset \cdots \subset \mathsf{H}_z(p). \tag{2.3}$$

The rank deficiency problem arises when testing the hypothesis $\mathsf{H}_z(r)$ when in fact the sub-hypothesis $\mathsf{H}_z(r-1)$ is satisfied. The rank is determined to be r if the hypothesis $\mathsf{H}_z(r)$ cannot be rejected while the sub-hypothesis $\mathsf{H}_z(r-1)$ is rejected. As a short-hand we write $\mathsf{H}_z^\circ(r) = \mathsf{H}_z(r) \backslash \mathsf{H}_z(r-1)$ for this situation. The rank can be determined along the procedure outlined in Johansen (1995, §12.1). 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

$$\mathsf{H}_z(r): \qquad \Pi = \alpha\beta', \tag{2.4}$$

where α and β are $p \times r$ matrices. The advantage of this formulation is that α and β vary in vector spaces. The formulation does, however, allow rank deficiency where the rank of Π is smaller than r.

The hypothesis of known cointegration vectors is

$$\mathsf{H}_{z,\beta}(r): \qquad \Pi = \alpha b', \tag{2.5}$$

for some unknown matrix α 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 α has full column rank, but in the rank deficient case, it has reduced column rank, so that the hypothesis $\mathsf{H}_z(r-1)$ is satisfied.

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* $\operatorname{rank}\Pi = s$ *where* $s \leq p$ *such that* $\Pi = \alpha\beta'$ *where* α, β *are* $p \times s$*-matrices with full column rank. 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 satifying* $|z| > 0$.

The Granger-Johansen theorem states that a process satisfying the model (2.1) so that $\operatorname{rank}\Pi = r$ and the I(1) condition holds with $s = r$ has the representation

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + S_t + \tau, \tag{2.6}$$

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 τ 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 $\mathsf{H}_z(r)$ against the unrestricted model $\mathsf{H}_z(p)$ is found by reduced rank regression, see Johansen (1995, §6). It is a two-step procedure. First, the differences ΔX_t and the lagged levels X_{t-1} are regressed on the lagged differences ΔX_{t-i}, $i = 1, \ldots, k-1$ giving residuals $R_{0,t}$, $R_{1,t}$. Secondly the squared sample correlations, $1 \geq \widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_p \geq 0$ 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

$$LR\{\mathsf{H}_z(r) \mid \mathsf{H}_z(p)\} = -T \sum_{j=r+1}^{p} \log(1 - \widehat{\lambda}_j). \tag{2.7}$$

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 $\mathsf{H}_z(p)$ is therefore given by

$$LR\{\mathsf{H}_{z,\beta}(r) \mid \mathsf{H}_z(p)\} = -T \log \frac{\det\{S_{00} - S_{01}S_{11}^{-1}S_{10}\}}{\det\{S_{00} - S_{01}b(b'S_{11}b)^{-1}b'S_{10}\}}. \tag{2.8}$$

The log likelihood ratio statistic for the hypothesis of known cointegrating vector against the rank hypothesis is found by combining the statistics in (2.7), (2.8), that is

$$LR\{\mathsf{H}_{z,\beta}(r) \mid \mathsf{H}_z(r)\} = LR\{\mathsf{H}_{z,\beta}(r) \mid \mathsf{H}_z(p)\} - LR\{\mathsf{H}_z(r) \mid \mathsf{H}_z(p)\}. \tag{2.9}$$

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 2.1. *Consider the rank hypothesis* $\mathsf{H}_z(r)$: $\operatorname{rank}\Pi \leq r$. *Suppose* $\mathsf{H}_z^\circ(s) = \mathsf{H}_z(s)\backslash\mathsf{H}_z(s-1)$ *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* $1 \geq \rho_1 \geq \cdots \geq \rho_{p-s} \geq 0$ *be the eigenvalues of the eigenvalue problem*

$$0 = \det\left\{\rho \int_0^1 F_u F_u' du - \int_0^1 F_u (dB_u)' \int_0^1 (dB_u) F_u'\right\} \tag{2.10}$$

Then, for $T \to \infty$,

$$LR\{\mathsf{H}_z(r) \mid \mathsf{H}_z(p)\} = -T \sum_{j=r+1}^{p} \log(1 - \widehat{\lambda}_j) \overset{\mathsf{D}}{\to} T \sum_{j=r-s+1}^{p-s} \rho_j. \tag{2.11}$$

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 2.1 reports the asymptotic distribution of the rank test reported in Theorem 2.1. The simulation were done using Ox, see Doornik (2007). The simulation design follows that of Johansen (1995, §15). That is, the stochastic integrals in (2.10) were descretized with $T = 1,000$ 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 $r = s$ are analytic values from Nielsen (1997) using the result of Abadir (1995). Bernstein (2014) reports values for higher dimensions.

The first panel of Table 2.1 reports the distribution for the standard case where $s = r$. This corresponds to Table 15.1 of Johansen (1995). The second and third panel

$r-s$	$p-r$	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

Table 2.1: Quantiles, mean and variance of $LR\{\mathsf{H}_z(r)|\mathsf{H}_z(p)\}$ where the data generating process has rank $s = \operatorname{rank}\Pi \leq r$.

of Table 2.1 report the distribution for the rank deficient case where $s = r - 1$ and $s = r - 2$. The first entry for $s = r - 1$, $r = 1$ 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. Presumably the distribution would be inbetween these extremes if the rank is deficient and the I(1) condition fails.

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 also suffices to discuss the situation 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 β_0 of dimension $p \times s$ and the asymptotic theory will depend on β_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 2.2. *Consider the hypothesis* $\mathsf{H}_{z,\beta}(r) : \Pi = \alpha b'$ *for unknown* α *and a known, full column rank* b *of dimensions* $p \times r$ *are satisfied. Suppose* $\mathsf{H}_z(0)$ *is satisfied, so that* $\alpha = 0$ *and* $s = 0$*, and that the I(1) condition is satisfied. 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* $T \to \infty$*,*

$$LR\{\mathsf{H}_{z,\beta}(r) \mid \mathsf{H}_z(p)\} \overset{\mathsf{D}}{\to} \mathrm{tr}\{\int_0^1 dB_u B_u' (\int_0^1 B_u B_u' du)^{-1} \int_0^1 B_u (dB_u)' - \int_0^1 dB_u B_{1,u}' (\int_0^1 B_{1,u} B_{1,u}' du)^{-1} \int_0^1 B_{1,u} (dB_u)'\}. \quad (2.12)$$

The convergence of the test statistic $LR\{\mathsf{H}_{z,\beta}(r) \mid \mathsf{H}_z(p)\}$ *holds jointly with the convergence for the rank test statistic* $LR\{\mathsf{H}_z(r) \mid \mathsf{H}_z(p)\}$*, for* $s = 0$*, in Theorem 2.1. Thus, when* $s = 0$ *the formula (2.9) implies that the limit distribution of the test statistic for known* β *within the model with rank of at most* r *can be found as the difference of the two limiting variables.*

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 2.2: Quantiles, mean and variance of $LR\{\mathsf{H}_{z,\beta}(r)|\mathsf{H}_z(r)\}$ where the data generating process has rank $s = \mathrm{rank}\,\Pi \leq r$.

Table 2.2 reports the asymptotic distribution of the test for known cointegrating vector in the model where the rank is at most r. When $r = s$ the asymptotic distribution is χ^2 with $r(p-r)$ degrees of freedom, see Johansen (1995, Theorem 7.2.1). When $s = 0$ the asymptotic distribution reported in Theorem 2.2 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 2.3 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 Table 2.1 and Table 2.2, see (2.9). Thus, up to a simulation error the expectations reported in Tables 2.1, 2.2 add up to the expectation reported in Table 2.3. In the full rank case $r = s$ the statistics in Tables 2.1, 2.2 are independent, as proved below, so also the variances are additive.

Theorem 2.3. *Consider the hypothesis* $\mathsf{H}^\circ_{z,\beta}(r)$*. Suppose* $\mathsf{H}^\circ_z(r) = \mathsf{H}^\circ_z(r)/\mathsf{H}^\circ_z(r-1)$ *is satisfied and that the* $I(1)$ *condition holds with* $s = r$*. Then the rank test statistic* $LR\{\mathsf{H}^\circ_z(r)|\mathsf{H}^\circ_z(p)\}$ *and the statistic* $LR\{\mathsf{H}^\circ_{z,\beta}(r)|\mathsf{H}^\circ_z(r)\}$ *for testing a simple hypothesis on the cointegrating vector are asymptotically independent.*

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

Table 2.3: Quantiles, mean and variance of $LR\{\mathsf{H}_{z,\beta}(r)|\mathsf{H}_z(p)\}$ where the data generating process has rank $s = \operatorname{rank}\Pi \leq r$.

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

$$\Delta X_t = \Pi X_{t-1} + \mu + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \qquad \text{for } t = 1, \ldots, T, \tag{3.1}$$

where the innovations ε_t are independent normal $\mathsf{N}_p(0, \Omega)$-distributed. The parameters are the p-dimensional square matrices Π, Γ_i, Ω and the p-vector μ. They vary freely so that Ω is symmetric, positive definite.

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

$$\mathsf{H}_{c\ell}(r): \qquad \operatorname{rank}\Pi \leq r, \tag{3.2}$$

$$\mathsf{H}_{c}(r): \qquad \operatorname{rank}(\Pi, \mu) \leq r. \tag{3.3}$$

Their interpretations follow from the Granger-Johansen representation which is reviewed in §3.2 below. In short, if there are no rank deficiencies the first hypothesis $\mathsf{H}_{c\ell}$ 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

$$\mathsf{H}_c(0) \subset \mathsf{H}_{c\ell}(0) \subset \cdots \subset \mathsf{H}_{c\ell}(r-1) \subset \mathsf{H}_c(r) \subset \mathsf{H}_{c\ell}(r) \subset \cdots \subset \mathsf{H}_c(p) = \mathsf{H}_{c\ell}(p). \tag{3.4}$$

This nesting structure is considerably more complicated than the structure (2.3) for the model without deterministic terms. A practical investigation may start in three different ways. First, the model (3.1) is taken as the starting point. Both types of hypotheses come into play and the rank is determined as outlined in Johansen (1995, §12). Secondly, if visual inspection of the data indicates that linear trends are not present the hypotheses $\mathsf{H}_{c\ell}$ may be ignored. Thirdly, if visual inspection of the data indicates that a linear trend could be present, the model (3.1) 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

$$\mathsf{H}_{c\ell}(r): \qquad \Pi = \alpha\beta', \tag{3.5}$$

$$\mathsf{H}_{c}(r): \qquad (\Pi, \mu) = \alpha(\beta', \beta_c'). \tag{3.6}$$

The hypotheses of known cointegrating vectors are therefore

$$\mathsf{H}_{c\ell,\beta}(r): \qquad \Pi = \alpha b', \tag{3.7}$$

$$\mathsf{H}_{c,\beta}(r): \qquad (\Pi, \mu) = \alpha(b', b_c'). \tag{3.8}$$

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

There is 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 $\mathsf{H}_{c\ell}(r)$. Suppose that the sub-hypothesis $\mathsf{H}_{c}(r)$ does not hold and that the I(1) condition holds with $s = r$. Thus, the $(p \times r)$-matrices α, β 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

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + S_t + \tau_c + \tau_\ell t, \tag{3.9}$$

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 $\mathsf{H}_{c}(r)$. Suppose that the sub-hypothesis $\mathsf{H}_{c\ell}(r-1)$ does not hold and that the I(1) condition holds with $s = r$. Thus, the $(p \times r)$-matrices α, β have full column rank, and the $\{(p + 1) \times r\}$-matrix $\beta^* = (\beta, \beta_c')'$ has full column rank. Then, the Granger-Johansen representation (3.9) holds with $\tau_\ell = 0$, while τ_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 $\beta' X_t + \beta_c'$ has a zero mean invariant distribution.

3.3 Test statistics

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

First, consider the reduced rank hypothesis $\mathsf{H}_{c\ell}(r)$ and the corresponding hypothesis $\mathsf{H}_{c\ell,\beta}(r)$ of known cointegrating vectors. The residuals $R_{0,t}$ and $R_{1,t}$ are formed by regressing the differences ΔX_t and the lagged levels X_{t-1} on an intercept and the lagged differences ΔX_{t-i}, $i = 1, \ldots, k-1$. In the second step, compute the canonical correlations $1 \geq \widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_p \geq 0$ of $R_{0,t}$ and $R_{1,t}$. The rank test statistic $LR\{\mathsf{H}_{c\ell}(r)|\mathsf{H}_{c\ell}(p)\}$ then has the form (2.7). The test statistic for known cointegrating vectors $LR\{\mathsf{H}_{c\ell,\beta}(r)|\mathsf{H}_{c\ell}(p)\}$ has the form (2.8), using the same residuals $R_{0,t}$ and $R_{1,t}$, and the hypothesized cointegrating vectors b.

Secondly, consider the reduced rank hypothesis $\mathsf{H}_c(r)$ and the corresponding hypothesis $\mathsf{H}_{c,\beta}(r)$ of known cointegrating vectors. The residuals $R_{0,t}$ and $R_{1,t}$ are formed by regressing the differences ΔX_t and the vector formed by stacking the lagged levels and an intercept $X^*_{t-1} = (X'_{t-1}, 1)'$ on the lagged differences ΔX_{t-i}, $i = 1, \ldots, k-1$. In the second step, compute the canonical correlation of these $R_{0,t}$ and $R_{1,t}$. The rank test statistic $LR\{\mathsf{H}_c(r)|\mathsf{H}_c(p)\}$ then has the form (2.7). The test statistic for known cointegrating vectors $LR\{\mathsf{H}_{c,\beta}(r)|\mathsf{H}_c(p)\}$ has the form (2.8), 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 (3.4) shows that each of the two rank hypotheses $\mathsf{H}_{c\ell}(r)$ and $\mathsf{H}_c(r)$ can be rank deficient in two ways when either of $\mathsf{H}^\circ_{c\ell}(s) = \mathsf{H}_{c\ell}(s)/\mathsf{H}_c(s)$ or $\mathsf{H}^\circ_c(s) = \mathsf{H}_c(s)/\mathsf{H}_{c\ell}(s-1)$ holds. In three cases the limiting distribution is of the same form as in Theorem 2.1, 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 $\mathsf{H}_c(r)$ with a data generating process satisfying $\mathsf{H}^\circ_{c\ell}(s) = \mathsf{H}_{c\ell}(s)/\mathsf{H}_c(s)$. This is the case that can often be ruled out through visual inspection of the data as mentioned in §3.1.

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

Theorem 3.1. *Consider the rank hypothesis* $\mathsf{H}_{c\ell}(r)$: $\operatorname{rank}\Pi \leq r$. *Suppose* $\mathsf{H}^\circ_{c\ell}(s) = \mathsf{H}_{c\ell}(s)\backslash\mathsf{H}_c(s)$ *holds for some* $s \leq r$, *so that rank*$\Pi = s$ *and 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*

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

Then $LR\{\mathsf{H}_{c\ell}(r) \mid \mathsf{H}_{c\ell}(p)\}$ *converges as in (2.11) using the present* F.

Table 3.1 reports the simulated asymptotic distribution of the rank test reported in Theorem 3.1. The first panel gives the standard case where $s = r$ and corresponds to Table 15.3 of Johansen (1995). For $p - r = 1$ the asymptotic distribution is actually χ^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

$r-s$	$p-r$	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

Table 3.1: Quantiles, mean and variance of $LR\{\mathsf{H}_{c\ell}(r)|\mathsf{H}_{c\ell}(p)\}$ where the data generating process satisfies $\mathsf{H}_{c\ell}^{\circ}(s)=\mathsf{H}_{c\ell}(s)\backslash\mathsf{H}_{c}(s)$ with $s\leq r$.

$\mathsf{H}_{c\ell}^{\circ}(s)$ where $\mathsf{H}_{c\ell}(s)$ holds, but $\mathsf{H}_{c}(s)$ fails. The distribution is shifted to the left when $r-s>0$ as in Table 2.1.

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

Theorem 3.2. *Consider the rank hypothesis $\mathsf{H}_{c\ell}(r)$: $\operatorname{rank}\Pi\leq r$. Suppose $\mathsf{H}_{c}^{\circ}(s)=\mathsf{H}_{c}(s)\backslash\mathsf{H}_{c\ell}(s-1)$ holds for some $s\leq r$, so that $\operatorname{rank}\Pi=\operatorname{rank}\Pi^{*}=s$ 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-meaned Brownian motion*

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

Then $LR\{\mathsf{H}_{c\ell}(r)\mid\mathsf{H}_{c\ell}(p)\}$ converges as in (2.11) using the present F.

Table 3.2 reports the simulated asymptotic distribution of the rank test reported in Theorem 3.2. 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 3.1. The second and the third panel of Table 3.2 report the distribution for the rank deficient case $\mathsf{H}_{c}^{\circ}(s)$ for $s<r$. In those case the distribution is shifted to the left relative to the first panel as in Tables 2.1, 3.1.

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

Theorem 3.3. *Consider the rank hypothesis $\mathsf{H}_{c}(r)$: $\operatorname{rank}\Pi\leq r$. Suppose $\mathsf{H}_{c}^{\circ}(s)=\mathsf{H}_{c}(s)\backslash\mathsf{H}_{c\ell}(s-1)$ holds for some $s\leq r$ so that $\operatorname{rank}\Pi=\operatorname{rank}(\Pi,\mu)=s$ 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*

$$F_u=\begin{pmatrix} B_u \\ 1 \end{pmatrix}. \tag{3.10}$$

$r-s$	$p-r$	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

Table 3.2: Quantiles, mean and variance of $LR\{\mathsf{H}_{c\ell}(r)|\mathsf{H}_{c\ell}(p)\}$ where the data generating process satisfies $\mathsf{H}_c^\circ(s) = \mathsf{H}_c(s)\backslash\mathsf{H}_{c\ell}(s-1)$ with $s \leq r$.

Then $LR\{\mathsf{H}_c(r) \mid \mathsf{H}_c(p)\}$ converges as in (2.11) using the present F.

$r-s$	$p-r$	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

Table 3.3: Quantiles, mean and variance of $LR\{\mathsf{H}_c(r)|\mathsf{H}_c(p)\}$ where the data generating process satisfies $\mathsf{H}_c^\circ(s) = \mathsf{H}_c(s)\backslash\mathsf{H}_{c\ell}(s-1)$ with $s \leq r$.

Table 3.3 reports the simulated asymptotic distribution of the rank test reported in Theorem 3.3. The first panel gives the standard case where $s = r$ and corresponds to Table 15.2 of Johansen (1995). The second and the third panel report the distribution for the rank deficient case $\mathsf{H}_c^\circ(s)$ for $s < r$. Once again, the distribution shifts to the left in the rank deficient case.

The final case is the test for the hypothesis $\mathsf{H}_c(r)$ in the rank deficient case where $\mathsf{H}_{c\ell}^\circ(s) = \mathsf{H}_{c\ell}(s-1)/\mathsf{H}_c(s-1)$ 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 §3.1. Futhermore, it would be hard to deal with the nuissance parameters in applications.

It is worth noting that the proof in this final case would be somewhat different from the proof of Theorems 2.1, 3.1, 3.2, 3.3. They are all proved by modifying the argument of Johansen (1995, §10, 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 $\mathsf{H}_c(r)$ where the constant is restricted to the cointegrating space. Thus, in the full rank case the Granger-Johansen representation (3.9) 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, (3.8). 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 3.4. *Consider the hypothesis $\mathsf{H}_{c,\beta}(r) : (\Pi, \mu) = \alpha b^{*\prime}$ where $b^* = (b', b_c')'$ for an unknown α and a known, full column rank b, both of dimension $p \times r$, along with a known r-vector b_c'. Suppose $\mathsf{H}_z(0)$ 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 (3.10). Then it holds, for $T \to \infty$, that*

$$LR\{\mathsf{H}_{z,\beta}(r) \mid \mathsf{H}_z(p)\} \overset{\mathsf{D}}{\to} \mathrm{tr}\{\int_0^1 dB_u F_u'(\int_0^1 F_u F_u' du)^{-1} \int_0^1 F_u (dB_u)' - \int_0^1 dB_u B_{1,u}'(\int_0^1 B_{1,u} B_{1,u}' du)^{-1} \int_0^1 B_{1,u}(dB_u)'\}. \quad (3.11)$$

The convergence of the test statistic $LR\{\mathsf{H}_{c,\beta}(r) \mid \mathsf{H}_c(p)\}$ holds jointly with the convergence for the rank test statistic $LR\{\mathsf{H}_c(r) \mid \mathsf{H}_c(p)\}$, for $s = 0$, in Theorem 3.3. Thus, when $s = 0$ a formula of the type (2.9) implies that the limit distribution of the test statistic for known β within the model with rank of at most r satisfies can be found as the difference of the two limiting variables.

Table 3.4 reports the asymptotic distribution of the test for known cointegrating vector in the model where the rank is at most r. When $r = s$ the asymptotic distribution

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 3.4: Quantiles, mean and variance of $LR\{\mathsf{H}_{c,\beta}(r)|\mathsf{H}_c(r)\}$ where the data generating process satisfies $\mathsf{H}^\circ_{c,\beta}(s)$.

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

Table 3.5: Quantiles, mean and variance of $LR\{\mathsf{H}_{c,\beta}(r)|\mathsf{H}_c(p)\}$ where the data generating process satisfies $\mathsf{H}^\circ_{c,\beta}(s)$.

is χ^2 with $r(p+1-r)$ degrees of freedom, see Johansen and Juselius (1990, p. 193–194), Johansen, Mosconi and Nielsen (2000, Lemma A.5). When $s = 0$ the distribution is simulated according to Theorem 3.4. It is shifted to the right relative to the case $r = s$.

Table 3.5 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 3.3, 3.4 add up to the expectation reported in Table 3.5. In the full rank case $r = s$ the statistics in Tables 3.3, 3.4 are independent, as proved below, so also the variances are additive.

Theorem 3.5. *Consider the hypothesis* $\mathsf{H}^\circ_{c,\beta}(r)$. *Suppose* $\mathsf{H}^\circ_c(r) = \mathsf{H}^\circ_c(r)/\mathsf{H}^\circ_{c\ell}(r-1)$ *is satisfied and that the* $I(1)$ *condition holds with* $s = r$. *Then the rank test statistic* $LR\{\mathsf{H}^\circ_c(r)|\mathsf{H}^\circ_c(p)\}$ *and the statistic* $LR\{\mathsf{H}^\circ_{c,\beta}(r)|\mathsf{H}^\circ_c(r)\}$ *for testing a simple hypothesis on the cointegrating vector are asymptotically independent.*

4 Applications of results

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

4.1 Finite sample theory

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 $LR\{\mathsf{H}_{z,\beta}(1)|LR\{\mathsf{H}_z(p)\}$ with the asymptotic χ^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 2.2 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 2.2 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 approaces 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 $\mathsf{H}_{z,\beta}(r)$ and $\mathsf{H}_{c\ell,\beta}(r)$. The test statistics are

$$LR(\mathbf{b}_0) = LR\{\mathsf{H}_{m,\beta}(r)|\mathsf{H}_m(p)\}, \tag{4.1}$$

$$LRC(\mathbf{b}_0) = LR\{\mathsf{H}_{m,\beta}(r)|\mathsf{H}_m(r)\}, \tag{4.2}$$

for $m = z, c\ell$. Moreover they consider the hypothesis $\mathsf{H}_{m,\Pi}(r)$, say, of a known impact matrix Π of rank r. This is tested through the statistic

$$LR_* = LR\{\mathsf{H}_{m,\Pi}(r)|\mathsf{H}_m(p)\}. \tag{4.3}$$

When the rank is not deficient the test statistic $LRC(\mathbf{b}_0)$ is asymptotically $\chi^2_{r(p-r)}$, see Johansen (1995, §7). The test statistic $LR(\mathbf{b}_0)$ has a Dickey-Fuller type distribution as derived in Theorem 2.2 for the case without deterministic terms. Table 2.2 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 2.2.

Khalaf and Urga's Theorem 1 is concerned with bounding the distribution of the statistic $LR(\mathbf{b}_0)$ when the rank is nearly deficient. Suppose the rank is nearly deficient in the sense that $\Pi \approx T^{-1}M$ for some matrix M. 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 2.1, 2.2 imply the following pointwise result.

Theorem 4.1. *Let θ denote the parameters of the model (2.1). Consider the parameter space Θ_z where the hypothesis $\mathsf{H}_{z,\beta}(1) : \Pi = \alpha b'$ holds for unknown α and a known, full column rank b, both of dimension $p \times 1$ so that the data generating process satisfies the $I(1)$ condition with $s = 0$ or $s = 1$. Let $q_{z,s}$ be the $(1-\psi)$ quantile of $LR\{\mathsf{H}_{z,\beta}(1)|\mathsf{H}_z(1)\}$ when the data generating process satisfies $\mathsf{H}^\circ_{z,\beta}(s)$ for $s = 0, 1$. Let $q_{z,*} = \max_{s=0,1} q_{z,s}$. Then it holds for all $\theta \in \Theta_z$ that*

$$\mathsf{P}[LR\{\mathsf{H}_{z,\beta}(1)|\mathsf{H}_z(1)\} \geq q_{z,*}] \leq \psi. \tag{4.4}$$

The local-to-unity motiviation for Theorem 4.1 suggests that a stronger uniform result is true. That is, for all $\epsilon > 0$ there exists a T_0 so that for all $T > T_0$ then

$$\sup_{\theta \in \Theta_z} \mathsf{P}\{LR\{\mathsf{H}_{z,\beta}(1)|\mathsf{H}_z(1)\} \geq q_{z,*}\} \leq \alpha + \epsilon. \tag{4.5}$$

It is, however, beyond the scope of the present paper to prove such a result.

The simulated values in Table 2.2 show that for $\psi = 5\%$ then

$$q_{z,*} = \max(q_{z,0}, q_{z,1}) = \begin{cases} \max(9.05, 3.84) = 9.05 & \text{for } p = 2, \\ \max(13.82, 5.99) = 13.82 & \text{for } p = 3. \end{cases} \tag{4.6}$$

The interpretation is as follows. Suppose the hypothesis $\mathsf{H}_z(1)$ has not been rejected, but it is unclear whether the rank could be nearly deficient. Then the hypothesis of a known β_0 is rejected if the statistic $LR\{\mathsf{H}_{z,\beta}(1)|\mathsf{H}_z(1)\}$ is larger than $q_{z,*}$.

The bound for $q_{z,*}$ seems very extreme. Khalaf and Urga therefore suggest to use the alternative statistic $LR\{\mathsf{H}_{z,\beta}(1)|\mathsf{H}_z(p)\}$. Theorem 4.1 could be modified to cover this statistic. The simulations in Table 2.3 indicate that we would then use bounds

$$\tilde{q}_{z,*} = \max(\tilde{q}_{z,0}, \tilde{q}_{z,1}) = \begin{cases} \max(9.70, 6.22) = 9.70 & \text{for } p = 2, \\ \max(20.83, 15.34) = 20.83 & \text{for } p = 3. \end{cases} \tag{4.7}$$

We can establish a similar result for the constant level model using Theorems 3.3, 3.4. 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 4.2. *Let θ denote the parameters of the model (3.1). Consider the parameter space Θ_c where the hypothesis $\mathsf{H}_{c,\beta}(1) : (\Pi, \mu) = \alpha(b', b_c')$ holds for unknown α and a known, full column rank b, both of dimension $p \times 1$, along with a known scalar b_c so that the data generating process satisfies the $I(1)$ condition with $s = 0$ or $s = 1$. Let $q_{c,s}$ be the $(1-\psi)$ quantile of $LR\{\mathsf{H}_{c,\beta}(1)|\mathsf{H}_c(1)\}$ when the data generating process satisfies $\mathsf{H}^\circ_{c,\beta}(s)$ for $s = 0, 1$. Let $q_{c,*} = \max_{s=0,1} q_{c,s}$. Then it holds for all $\theta \in \Theta_1$ that*

$$\mathsf{P}[LR\{\mathsf{H}_{c,\beta}(1)|\mathsf{H}_c(1)\} \geq q_{c,*}] \leq \psi. \tag{4.8}$$

The simulated values in Table 3.4 show that for $\psi = 5\%$ then

$$q_{c,*} = \max(q_{z,0}, q_{z,1}) = \begin{cases} \max(14.05, 5.99) = 14.05 & \text{for } p = 2, \\ \max(19.66, 7.82) = 19.66 & \text{for } p = 3. \end{cases} \tag{4.9}$$

If the alternative is taken as $\mathsf{H}_c(p)$ instead of $\mathsf{H}_c(1)$ the bounds are modified as

$$\tilde{q}_{c,*} = \max(\tilde{q}_{c,0}, \tilde{q}_{c,1}) = \begin{cases} \max(18.18, 12.38) = 18.18 & \text{for } p = 2, \\ \max(32.04, 24.39) = 32.04 & \text{for } p = 3. \end{cases} \tag{4.10}$$

The bounds (4.9), (4.10) for the constant level model appear futher appart than the corresponding bounds (4.6), (4.7) 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).

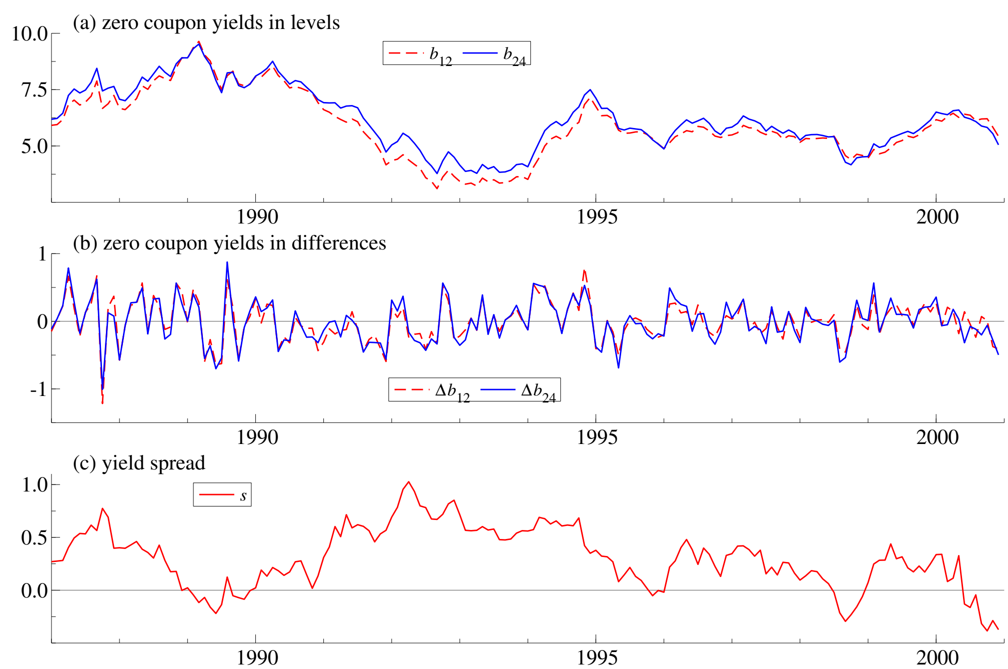

Figure 4.1: Zero coupon yields in (a) levels, (b) differences and (c) spread.

Figure 4.1 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 2 common trends among 5 maturities. The 2 common trends can be interpreted as short-run and long-run forces driving the yield curve. The cointegrating relations match an extended expections 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, Anderson and Granger (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.

A vector autoregression of the form (3.1) with an intercept, $k = 4$ lags as well as a dummy variable for 1987:10 was fitted to the data. Table 4.1 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 autoreregressive and normality tests in for non-stationarity autoregressions see Engler and Nielsen (2009), Kilian and Demiroglu (2000), and Nielsen (2006).

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 4.1: Specification tests for the unrestricted vector autoregression.

The dummy variable matches the policy intervention after the stock market crash on 19 Oct 1987. Empirically, the dummy variable can be justified in two ways. First, the plot of yield differences in Figure 4.1(b) indicate a sharp drop in yields at that point. Secondly, the robustified least squares algorithm analyzed in Johansen and Nielsen (2014) 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, §7.6). Thus we choose a cut-off of 3.02 corresponding to a gauge of 0.25%. When running the autoregression 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.

Table 4.2 reports cointegration rank tests. The fifth column shows conventional p-values based on Table 3.1, 3.3 for $s = r$ corresponding to Johansen (1995, Tables 15.2, 15.3). The sixth column shows p-values based on Table 3.2, 3.3 assuming data have been generating by a model satisfying $\mathsf{H}_c(0) = \mathsf{H}_z(0)$. 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 are higher than the former. Overall this provide overwhelming evidence in favour of a pure random walk model in line with Giese (2008).

If we have a strong belief in the expectation hypothesis we would, perhapse, ignore the rank tests and seek to test the expectations hypothesis directly. If we maintain the model $\mathsf{H}_c(1)$, we could have to contemplate that the cointegration vectors could be

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

Table 4.2: Cointegration rank tests.

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 $\mathsf{H}_c^\circ(0)$ or $\mathsf{H}_c^\circ(1)$, but not by $\mathsf{H}_{c\ell}^\circ(0)$ we can apply the Khalaf-Urga (2014)-type bound test established in Theorem 4.2. The 95% bound in (4.9) is 14.05 so the hypothesis cannot be rejected based on this statistic. 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).

5 Conclusion

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. However, our impression is that the identification robust tests have modest power to reject incorrect restrictions.

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 $t = 1, \ldots, T$ is embedded in $D[0,1]$ through $X_{\mathrm{integer}(Tu)}$ for $0 \leq u \leq 1$. For processes Y_t, Z_t for $t = 1, \ldots, T$ the residuals from regressing Y_t on Z_t are denoted $(Y_t \mid Z_t) = Y_t - \sum_{s=1}^T Y_s Z_s' (\sum_{s=1}^T Z_s Z_s')^{-1} Z_t$.

Proof of Theorem 2.1. This follows the outline of the proof in Johansen (1995, §10, 11). Let $\Pi = \alpha_0 \beta_0'$ for $p \times s$-matrices α_0, β_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 (2.6) 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 $D[0,1]$,

$$T^{-1/2}B_T' X_{\text{integer}(Tu)} = B_T' C T^{-1/2} \sum_{t=1}^{\text{integer}(Tu)} \varepsilon_t + \mathsf{o}_{\mathsf{P}}(1). \tag{A.1}$$

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

$$T^{-1/2}\tilde{M}'(\alpha_{0\perp}'\Omega\alpha_{0\perp})^{-1/2}\alpha_{0\perp}'\Gamma\beta_{0\perp}(\beta_{0\perp}'\beta_{0\perp})^{-1}\beta_{0\perp}'X_{[Tu]} \overset{\mathsf{D}}{\to} B_u \tag{A.2}$$

on $D[0,1]$. □

Proof of Theorem 2.2. 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

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

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

$$LR\{\mathsf{H}_{z,\beta}(r) \mid \mathsf{H}_z(p)\} = \text{tr}\{\Omega^{-1}T(\widehat{\Omega}_R - \widehat{\Omega}_U)\} + \mathsf{o}_{\mathsf{P}}(1), \tag{A.3}$$

following Johansen (1995, p. 224). The consistency of the unrestricted variance estimator $\widehat{\Omega}_U$ follows from Lemma 10.3 of Johansen (1995) 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 α_0, β_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 (A.2) then reduces to

$$T^{-1/2}\tilde{M}'\Omega^{-1/2}\Gamma X_{\text{integer}(Tu)} = T^{-1/2}\tilde{M}'\Omega^{-1/2}\sum_{t=1}^{\text{integer}(Tu)}\varepsilon_t + \mathsf{o}_{\mathsf{P}}(1) \overset{\mathsf{D}}{\to} B_u, \tag{A.4}$$

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

$$\tilde{M} = \left[\begin{array}{c} \{b'\Gamma^{-1}\Omega(\Gamma')^{-1}b\}^{-1/2}b'\Gamma^{-1}\Omega^{1/2} \\ (b_\perp'\Gamma'\Omega^{-1}\Gamma b_\perp)^{-1/2}b_\perp'\Gamma'\Omega^{-1/2} \end{array}\right]. \tag{A.5}$$

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

$$T(\widehat{\Omega}_R - \widehat{\Omega}_U) = T\{S_{\epsilon 1}M(M'S_{11}M)^{-1}M'S_{1\epsilon} - S_{\epsilon 1}b(b'S_{11}b)^{-1}b'S_{1\epsilon}\}, \tag{A.6}$$

for any invertible matrix M and in particular for $M' = \tilde{M}'\Omega^{-1/2}\Gamma$. In light of the identity $\tilde{M}'\tilde{M} = I_p$, the random walk convergence in (A.4) and the rules for the trace write

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

Let $B_{1,u}$, $B_{2,u}$ be the first r and the last $p-r$ coordinates of B_u, respectively. Then the product moment convergence results in Johansen (1995, Lemma 10.3) imply

$$\operatorname{tr}\{\Omega^{-1}T(\widehat{\Omega}_R - \widehat{\Omega}_U)\} \overset{\mathsf{D}}{\to} \operatorname{tr}\{\int_0^1 dB_u B_u'(\int_0^1 B_u B_u' du)^{-1}\int_0^1 B_u(dB_u)' \\ - \int_0^1 dB_u B_{1,u}'(\int_0^1 B_{1,u}B_{1,u}'du)^{-1}\int_0^1 B_{1,u}(dB_u)'\}.$$

This is also the limit of the likelihood ratio test statistic due to (A.3). The convergence holds jointly with the convergence of the likelihood ratio test statistic for rank in Theorem 2.1 since the orthogonal matrix $\tilde{M}$ in (A.2) can be chosen freely. □

Proof of Theorem 2.3. We need a number of results from Johansen (1995). Let B, V be independent standard Brownian motions. His Theorem 11.1 shows

$$LR\{\mathsf{H}_z(r)|\mathsf{H}_z(p)\} \overset{\mathsf{D}}{\to} \operatorname{tr}\{\int_0^1 dB_u B_u'(\int_0^1 B_u B_u' du)^{-1}\int_0^1 B_u dB_u'\}, \tag{A.7}$$

while his Lemma 13.8 shows

$$LR\{\mathsf{H}_{z,\beta}(r)|\mathsf{H}_z(r)\} \overset{\mathsf{D}}{\to} \operatorname{tr}\{\int_0^1 dV_u B_u'(\int_0^1 B_u B_u' du)^{-1}\int_0^1 B_u dV_u'\}. \tag{A.8}$$

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 (A.8) is mixed Gaussian since B, V are independent. Thus, by conditioning on B we see that $LR\{\mathsf{H}_{z,\beta}(r)|\mathsf{H}_z(r)\}$ is asymptotically χ^2 and hence independent of B. In turn the two test statistics are asymptotically independent. □

Proof of Theorem 3.1. Similar to the proof of Theorem 2.1. The relevant Granger-Johansen representation is (3.9) 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 $\gamma \in \operatorname{span}(\beta_{0\perp})$ so that $\gamma'\tau_\ell = 0$ and the expansion (A.1) is replaced by

$$T^{-1/2}B_T' X_{\text{integer}(Tu)} = \left\{ \begin{array}{c} (\gamma'\gamma)^{-1}\gamma' C T^{-1/2}\sum_{t=1}^{\text{integer}(Tu)}\varepsilon_t \\ u \end{array} \right\} + \mathsf{o_P}(1) \tag{A.9}$$

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

Proof of Theorem 3.2. Similar to the proof of Theorem 2.1. Use the Granger-Johansen representation (3.9) 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 (A.1). □

Proof of Theorem 3.3. Similar to the proof of Theorem 2.1. Use the Granger-Johansen representation (3.9) with rank s, and τ_ℓ. Use Johansen's Lemmas 10.2, 10.3

with X_t, B_T and the expansion (A.1) replaced by, respectively, $X_t^* = (X_t', 1)'$, the block diagonal matrix $B_T^* = \mathrm{diag}\,(B_T, T^{1/2})$ where $B_T = \beta_{0\perp}(\beta_{0\perp}'\beta_{0\perp})^{-1}$, and

$$T^{-1/2} B_T^{*\prime} X^*_{\mathrm{integer}(Tu)} = \begin{pmatrix} B_T' C T^{-1/2} \sum_{t=1}^{\mathrm{integer}(Tu)} \varepsilon_t \\ 1 \end{pmatrix} + \mathsf{o_P}(1) \qquad \text{(A.10)}$$

on $D[0, 1]$. □

Proof of Theorem 3.4. The proof of Theorem 2.2 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 (A.5). Replace (A.4) by

$$\begin{pmatrix} T^{-1/2}\tilde{M}'\Omega^{-1/2}\Gamma & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_{\mathrm{integer}(Tu)} \\ 1 \end{pmatrix} \overset{\mathsf{D}}{\to} F_u. \qquad \text{(A.11)}$$

The difference of variance estimators in (A.6) is now

$$T(\widehat{\Omega}_R - \widehat{\Omega}_U) = T\{S_{\epsilon 1} M (M' S_{11} M)^{-1} M' S_{1\epsilon} - S_{\epsilon 1} b^* (b^{*\prime} S_{11} b^*)^{-1} b^{*\prime} S_{1\epsilon}\}, \qquad \text{(A.12)}$$

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

$$M = \begin{Bmatrix} b'\Gamma^{-1}\Omega(\Gamma')^{-1}b & 0 & 0 \\ 0 & b_\perp'\Gamma'\Omega^{-1}\Gamma b_\perp & 0 \\ 0 & 0 & 1 \end{Bmatrix}^{-1/2} \begin{pmatrix} b' & b_c' \\ b_\perp'\Gamma'\Omega^{-1}\Gamma & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{(A.13)}$$

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

$$\begin{pmatrix} T^{-1/2} I_p & 0 \\ 0 & 1 \end{pmatrix} M \begin{pmatrix} X_{\mathrm{integer}(Tu)} \\ 1 \end{pmatrix} \overset{\mathsf{D}}{\to} F_u. \qquad \text{(A.14)}$$

Moreover, the first r coordinates of $MR_{1,t}$ are proportional to $b^{*\prime} R_{1,t}$. Thus the argument can be completed as in the proof of Theorem 2.2. □

Proof of Theorem 3.5. The proof of Theorem 2.3 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, Mosconi and Nielsen (2000, Theorem 3.1, Lemma A.5). □

References

Abadir, K.M. (1995) The limiting distribution of the t ratio under a unit root. *Econometric Theory* 11, 775–793.

Bernstein, D. (2014) Asymptotic theory for unidentified cointegration estimators. M.Phil. thesis, University of Oxford.

Billingsley, P. (1968) *Convergence of Probability Measures* New York: Wiley.

Carstensen, K. (2003) Nonstationary term premia and cointegration of the term structure. *Economics Letters* 80, 409–413.

Doornik, J.A. (1998) Approximations to the Asymptotic Distribution of Cointegration Tests. *Journal of Economic Surveys* 12, 573–593.

Doornik, J.A. (2007) *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com.

Doornik, J.A. and Hansen, H. (2008) An Omnibus Test for Univariate and Multivariate Normality. *Oxford Bulletin of Economics and Statistics* 70, 927–939

Doornik, J.A. and Hendry, D.F. (2013) *Empirical Econometric Modelling - PcGive 14, volume 1.* London: Timberlake Consultants.

Dufour, J.-M. (1997) Some impossibility theorems in econometrics with applications to structural and dynamic methods. *Econometrica* 65, 1365–87.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1108.

Engler, E. and Nielsen, B. (2009) The empirical process of autoregressive residuals. *Econometrics Journal* 12, 367–381.

Giese, J. (2008) Level, slope, curvature: Characterising the yield curve in a cointegrated VAR model. *Economics* 2, article 28.

Godfrey, L.G. (1978) Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica* 46, 1293–1302.

Hall, A.D., Anderson, H.M. and Granger, C.W. (1992) A cointegration analysis of treasury bill yields. *Review of Economics and Statistics* 74, 116–126.

Hendry, D.F. and Doornik, J.A. (2014) *Empirical Model Discovery and Theory Evaluation*. Cambridge MA: MIT Press.

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

Johansen, S. (2000) A Bartlett Correction Factor for Tests on the Cointegrating Relations. *Econometric Theory* 16, 740–777.

Johansen, S. (2002) A Small Sample Correction of the Test for Cointegrating Rank in the Vector Autoregressive Model. *Econometrica* 70, 1929–1961.

Johansen, S. and Juselius, K. (1990) Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52, 169–210.

Johansen, S., Mosconi, R. and Nielsen, B. (2000) Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal* 3, 216–249.

Johansen, S. and Nielsen, B. (2014) Outlier detection algorithms for least squares time series. Nuffield Discussion Paper 2014-W04.

Kilian, L. and Demiroglu, U. (2000) Residual based tests for normality in autoregressions: asymptotic theory and simulations. *Journal of Economic Business and Control* 18, 40–50.

Khalaf, L. and Urga, G. (2014) Identification robust inference in cointegrating regressions. *Journal of Econometrics* 182, 385–396.

Nielsen, B. (1997) Bartlett Correction of the Unit Root Test in Autoregressive Models. *Biometrika* 84, 500–04.

Nielsen, B. (2004) On the Distribution of Likelihood Ratio Test Statistics for Cointegration Rank. *Econometric Reviews* 23, 1–23.

Nielsen, B. (2006) Order determination in general vector autoregressions. In Ho, H.-C., Ing, C.-K., and Lai, T.L. (eds): *Time Series and Related Topics: In Memory of Ching-Zong Wei*. IMS Lecture Notes and Monograph Series 52, 93-112.

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

Shea, G.S. (1992) Benchmarking the expectations hypothesis of the interest-rate term structure: An analysis of cointegration vectors. *Journal of Business and Economic Statistics* 10, 347–366.

Zhang, H. (1993) Treasury yield curves and cointegration. *Applied Economics* 25, 361–367.