J. Japan Statist. Soc.
Vol. 37 No. 2 2007 191–205

THE WALD-TYPE TEST OF A NORMALIZATION OF COINTEGRATING VECTORS

Eiji Kurozumi*

This paper proposes a test for the normalization of cointegrating vectors. Our test is constructed using the unrestricted maximum likelihood estimator and then it may be seen as a Wald-type test. The test statistic is shown to be asymptotically bounded above by a chi-square distribution with one degree of freedom (χ_1^2) and then we can conduct a conservative test using critical values of χ_1^2.

Key words and phrases: Cointegration, identification, normalization, vector autoregression.

1. Introduction

Vector autoregressive (VAR) models have often been used in the econometric literature as useful models to describe stationary/non-stationary time series. Additionally cointegrating vectors are of primary interest for researchers who investigate the long-run stable relationship between economic variables. In VAR models it is well known that cointegrating vectors are identifiable only up to their column space. Although Johansen's (1988, 1991) maximum likelihood (ML) method solves the identification problem by imposing the just-identifying restrictions on cointegrating vectors, it is difficult to interpret the identified estimator in an economic sense because Johansen's identifying restrictions are imposed from a statistical point of view. See Johansen and Juselius (1994), Boswijk (1996), Luukkonen *et al.* (1999) and Pesaran and Shin (2002) among others. To obtain an interpretable estimator we may impose some restrictions on cointegrating vectors from an economic point of view, but such restrictions are not necessarily identifiable. We then need to check whether the restrictions imposed on cointegrating vectors are valid or not. General identifying conditions are given by Boswijk (1995), Johansen (1995a), Pesaran and Shin (2002), and Boswijk and Doornik (2003). One of the useful identifying or normalizing conditions is expressed as $c'\beta = I_r$ where β consists of $n \times r$ cointegrating vectors and c is an $n \times r$ known matrix of full column rank, as investigated by Stock (1987), Johansen (1991, 1995b), and Paruolo (1997) among others. Tests for the validity of this normalization are proposed by Boswijk (1996), Luukkonen *et al.* (1999), Saikkonen (1999), and Paruolo (2005). The null hypothesis of the failure of the normalization is considered in Boswijk (1996) and Paruolo (2005), while the other papers propose tests for the null of the validity of the normalization.

In this paper we propose a test for the null hypothesis of the invalid normalization of cointegrating vectors. Since our test is based on the unrestricted ML

Received January 23, 2006. Revised April 18, 2006. Accepted August 3, 2006.
*Department of Economics, Hitotsubashi University, Naka 2-1, Kunitachi, Tokyo 186-8601, Japan.

estimator, it may be seen as a Wald-type test. We show that the proposed test statistic converges in distribution to a chi-square distribution with one degree of freedom (χ_1^2) when the null space of $c'\beta$ is one-dimensional, and it is asymptotically bounded above by χ_1^2 when the dimension of the null space is greater than one; as a result, our test may be conservative in general and hence we can asymptotically control the size of the test. As a by-product, we also develop a test for the rank of a sub-matrix of cointegrating vectors; it may be seen as a generalization of Kurozumi's (2003, 2005) test.

The rest of the paper is organized as follows. We explain the model and assumptions in Section 2, and propose the Wald-type test for the null of the invalid normalization of cointegrating vectors. The asymptotic property of the test is investigated depending on the dimension of the null space of $c'\beta$. The finite sample property is investigated in Section 3. We compare our test with Boswijk's (1996) and Paruolo's (2005) test, which is based on Johansen's likelihood ratio (LR) test. An empirical illustration is given in Section 4, and Section 5 concludes the paper.

2. Test of a normalization

Let us consider the following n-variate vector error-correction model of order p:

$$\Delta y_t = \alpha\beta^{*\prime} y_{t-1}^* + \mu_d d_t + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \tag{2.1}$$

for $t = 1, \ldots, T$, where $\beta^* = [\beta', \mu_c']'$, α and β are $n \times r$ matrices of full column rank, $y_t^* = [y_t', c_t']'$, c_t and d_t are deterministic regressors, $\{\varepsilon_t\} \sim i.i.d.\mathcal{N}(0, \Sigma)$ with Σ being a positive definite variance matrix, and all roots of $|(1-z)I_n - \alpha\beta' z - \sum_{i=1}^{p-1} \Gamma_i z^i (1-z)| = 0$ are outside the unit circle or equal to 1. We consider the two most common cases where $(c_t, d_t) = (1, \emptyset)$ and $(t, 1)$; the former specification corresponds to the case where $\{y_t\}$ is not a trending series with possibly a non-zero mean whereas it is linearly trending for the latter case. The normality assumption on $\{\varepsilon_t\}$ is imposed to obtain the ML estimator, and the asymptotic result in this paper may be obtained under weaker conditions as explained by Pesaran and Shin (2002) and Boswijk and Doornik (2003). We assume that $0 < r < n$ and $|\alpha_\perp' \Gamma \beta_\perp| \neq 0$ where $\Gamma = I_n - \sum_{i=1}^{p-1} \Gamma_i$, so that y_t is cointegrated with cointegrating rank r.

The model (2.1) can be estimated by the ML method and the asymptotic property of the ML estimator is investigated by Johansen (1988, 1991, 1995b). It is known that the cointegrating vectors β can be consistently estimated up to the column space spanned by β, and we then need to normalize the estimator of β. The normalization of the cointegrating vectors is established by assuming that $c'\beta$ is nonsingular for an $n \times r$ known full column rank matrix c. The normalized cointegrating vectors are defined as $\beta_c = \beta(c'\beta)^{-1}$, so that β_c is unique and $c'\beta_c = I_r$. One of the commonly used c is given by $c = [I_r, 0]'$, but we consider a general full column rank matrix c in the following. Similarly, β^* is normalized as $\beta_c^* = \beta^*(c^{*\prime}\beta^*)^{-1}$ where $c^{*\prime} = [c', 0]$. The problem we have here is that the nonsingularity of $c'\beta$ does not necessarily hold. We then consider the following testing problem:

$$H_0 : \mathrm{rk}(c'\beta) < r \qquad \text{v.s.} \qquad H_1 : \mathrm{rk}(c'\beta) = r, \tag{2.2}$$

where $\mathrm{rk}(A)$ signifies the rank of a matrix A. The null hypothesis can also be expressed as $\mathrm{rk}(c^{*\prime}\beta^*) < r$. As a related hypothesis, we also consider

$$H_0^f : \mathrm{rk}(c'\beta) \leq f \qquad \text{v.s.} \qquad H_1^f : \mathrm{rk}(c'\beta) > f, \tag{2.3}$$

for a fixed $f < r$. The testing problem (2.3) has been considered in various situations such as the case when we want to test for the long-run Granger-noncausality as in Yamamoto and Kurozumi (2006). See Kurozumi (2005) and Paruolo (2005) for other useful examples. We first develop a test for (2.3) and next consider the testing problem (2.2).

For the testing problem (2.3), it seems natural to investigate the number of zero-eigenvalues of $c'\beta$. However, since $c'\beta$ is asymmetric in general, some of the eigenvalues possibly take complex values and hence it would be inconvenient to deal with $c'\beta$. Instead, we consider the quadratic form of $c'\beta$ such as $(c'\beta)\Psi(\beta'c)\Phi^{-1}$ for some positive definite and symmetric matrices[1] Ψ and Φ, noting that $\mathrm{rk}(c'\beta) = \mathrm{rk}((c'\beta)\Psi(\beta'c)\Phi^{-1})$. The advantage of considering the quadratic form is that all the eigenvalues, which are given by the solutions of

$$|(c'\beta)\Psi(\beta'c) - \lambda\Phi| = 0, \tag{2.4}$$

take non-negative real values, so that the null hypothesis H_0^f is equivalent to the hypothesis that the $r - f$ smallest eigenvalues equal zero. This strategy is taken by Robin and Smith (2000) in different situations.

In the following, $\beta_\perp$ signifies an orthogonal complement to β such that $\beta'\beta_\perp = 0$. Let $\hat{\alpha}$, $\hat{\beta}^*$, $\hat{\beta}$, $\hat{\beta}_\perp$, and $\hat{\Sigma}$ be the ML estimators of α, β^*, β, $\beta_\perp$, and Σ. Note that $\hat{\beta}$ consists of the first n rows of $\hat{\beta}^*$ while $\hat{\beta}_\perp$ is constructed such that $\hat{\beta}'\hat{\beta}_\perp = 0$. Theoretically, we can choose any positive definite and symmetric matrices Φ and Ψ, but in practice we need to choose them such that the limiting distributions of the $r-f$ smallest eigenvalues become free of nuisance parameters under H_0^f. The sample analogue of (2.4) we consider is given by

$$|(c'\hat{\beta})\hat{\Psi}(\hat{\beta}'c) - \lambda\hat{\Phi}| = 0, \tag{2.5}$$

where $\hat{\Psi} = \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ and

$$\hat{\Phi} = c'\hat{\beta}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'c + c'\bar{\hat{\beta}}_\perp L'(\Upsilon_T' S_{11}^* \Upsilon_T)^{-1} L \bar{\hat{\beta}}_\perp' c,$$

where $\bar{\hat{\beta}}_\perp = \hat{\beta}_\perp(\hat{\beta}_\perp'\hat{\beta}_\perp)^{-1}$, $S_{11}^* = T^{-1}\sum_{t=1}^T R_{1t}^* R_{1t}^{*\prime}$ with R_{1t}^* being a regression residual of y_{t-1}^* on d_t and $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$, $L = [I_{n-r}, 0]'$ is $(n-r+1)\times(n-r)$, and Υ_T is defined by

$$\Upsilon_T = \begin{bmatrix} T^{-1/2}\bar{\hat{\beta}}_\perp & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \Upsilon_T = \begin{bmatrix} T^{-1/2}\bar{\hat{\beta}}_\perp & 0 \\ 0 & T^{-1} \end{bmatrix}$$

---

1 When Ψ and/or Φ are stochastic, they are positive definite almost surely (a.s.).

for $(c_t, d_t) = (1, \emptyset)$ and $(c_t, d_t) = (t, 1)$, respectively.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$ be the ordered eigenvalues of (2.5). We first give the asymptotic behavior of the $r - f$ smallest eigenvalues, $\lambda_{f+1}, \ldots, \lambda_r$, under H_0^f.

Theorem 1. *When the rank of $c'\beta$ equals f $(< r)$, $T^2\lambda_{f+1}, \ldots, T^2\lambda_r$ converge in distribution to the ordered eigenvalues of*

$$|N'N - \rho I_{r-f}| = 0, \tag{2.6}$$

where N is $(r-f) \times (r-f)$ and $\mathrm{vec}(N) \sim \mathcal{N}(0, I_{(r-f)^2})$.

Let us consider the testing problem (2.3). Since the null hypothesis H_0^f is equivalent to the hypothesis that the $r - f$ smallest eigenvalues equal zero as explained above, it is natural to consider

$$W_T^f = T^2 \sum_{j=f+1}^{r} \lambda_j$$

as a test statistic. Note that W_T^f may be seen as a Wald-type test statistic because we use only the unrestricted ML estimator to construct the test statistic. The asymptotic property of this statistic is given by the following corollary.

Corollary 1. (i) *Under H_0^f, W_T^f converges in distribution to $\chi^2_{(r-f)^2}$.* (ii) *Under H_1^f, W_T^f diverges to infinity at a rate of T^2.*

Remark 1. We may also consider $T^2\lambda_{f+1}$ as a test statistic, but we will not investigate it because the finite sample performance of $T^2\lambda_{f+1}$ is shown to be similar to that of W_T^f from preliminary simulations.

Remark 2. When $c = [I_r, 0]$ and $(c_t, d_t) = (1, \emptyset)$, the testing problem (2.3) becomes the same as considered in Theorem 3 of Kurozumi (2005). Then, our testing problem includes Kurozumi's (2005) as a special case.

Next, we develop a test for (2.2). Noting that $H_0 = H_0^0 \cup H_0^1 \cup \cdots \cup H_0^{r-1}$ and $H_0^0 \subset H_0^1 \subset \cdots \subset H_0^{r-1}$, it is sufficient for us to consider only H_0^{r-1}. Then, the test statistic for the testing problem (2.2) is given by the minimum eigenvalue of (2.5) normalized by T^2:

$$W_T^{\min} = T^2\lambda_r.$$

The following corollary gives the asymptotic property of this test statistic.

Corollary 2. (i) *Under H_0,*

$$W_T^{\min} \xrightarrow{d} \begin{cases} \chi^2_1 & \textit{when} \quad \mathrm{rk}(c'\beta) = r - 1 \\ \underline{\rho} \leq \chi^2_1 & \textit{when} \quad \mathrm{rk}(c'\beta) < r - 1, \end{cases}$$

where $\underline{\rho}$, which is bounded above by χ^2_1, is the minimum eigenvalue of (2.6) *when the true rank of $c'\beta$ equals $f < r - 1$.* (ii) *Under H_1, $W_T^{\min}$ diverges to infinity at a rate of T^2.*

From this theorem we can see that if we use critical values of χ_1^2, our test has an asymptotically exact size under the null hypothesis if $\mathrm{rk}(c'\beta) = r-1$, and it is conservative if $\mathrm{rk}(c'\beta) < r-1$. Then, we can control the size of the test at least asymptotically.

3. Finite sample evidence

In this section we investigate the finite sample property of the test proposed in the previous section. For the purpose of comparison, we also investigate the LR test proposed by Boswijk (1996) and Paruolo (2005), according to which the null hypothesis is expressed as $H_0' : \beta^* = [c_\perp^* \phi, \psi]$ where $(\phi, \psi) \in R^{(n-r+1)\times 1} \times R^{(n+1)\times(r-1)}$. The LR test statistic is calculated by the switching algorithm of Johansen and Juselius (1992), and it has a similar asymptotic property to $W_T^{\min}$; its limiting distribution is χ_1^2 when $\mathrm{rk}(c'\beta) = r-1$, while it is bounded above by χ_1^2 when $\mathrm{rk}(c'\beta) < r-1$. The data generating process (DGP) we considered is basically the same as that of Luukkonen *et al.* (1999),

$$y_{1t} = Ay_{2t} + w_{1t}, \qquad \Delta y_{2t} = w_{2t}, \qquad \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} = \begin{bmatrix} b_1 I_k & 0 \\ b_2 I_k & 0 \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix},$$

where $\varepsilon_t = [\varepsilon_{1t}', \varepsilon_{2t}']' \sim i.i.d.\mathcal{N}(0, I_{2k})$. We consider two cases where $k = 2$ (DGP1) and 3 (DGP2). We set $b_1 = 0.4$ and 0.8, while $b_2 > 0$ is determined to satisfy the relation of $R^2(b_1, b_2) = b_2^2/(1-b_1^2-b_2^2) = 0.4$ and 0.8, where $R^2(b_1, b_2)$ is a measure of cross-correlation between w_{1t} and w_{2t}. See Luukkonen *et al.* (1999) for details. The process $y_t = [y_{1t}', y_{2t}']'$ has VAR(2) representation and the cointegrating vectors are expressed as $\beta' = [I_k, -A]$, so that y_t is 4-dimensional for $k = 2$ (DGP1) and 6-dimensional for $k = 3$ (DGP2). The matrices β and c are chosen as

$$\beta' = \begin{bmatrix} 1 & 0 & a & -1 \\ 0 & 1 & 1-a & 1 \end{bmatrix}, \qquad c' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad \text{for DGP1,}$$

$$\beta' = \begin{bmatrix} 1 & 0 & 0 & a & 0 & 1 \\ 0 & 1 & 0 & 0 & a & 1 \\ 0 & 0 & 1 & 1 & 1-a & 0 \end{bmatrix}, \qquad c' = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad \text{for DGP2.}$$

Then, $c'\beta$ becomes

$$\begin{bmatrix} 0 & 1 \\ a & 1-a \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 0 & 0 & 1 \\ a & 0 & 1 \\ 0 & a & 1-a \end{bmatrix},$$

so that the rank of $c'\beta$ is 1 and 2 for $a = 0$ and $a \neq 0$, respectively, for DGP1, while it is 1 and 3 for DGP2. Then, under the null hypothesis, both the test statistics are asymptotically chi-square distributed for DGP1 while they are bounded above by χ_1^2 for DGP2. We set all the initial values to be zero and the first 100

Table 1. Nominal size and size-unadjusted power of the tests.

			DGP1				DGP2			
$(c_t, d_t) = (1, \emptyset)$			$T = 100$		$T = 200$		$T = 100$		$T = 200$	
b_1	R^2	a	$W_T^{\min}$	LR	$W_T^{\min}$	LR	$W_T^{\min}$	LR	$W_T^{\min}$	LR
		0.050	0.648	0.543	0.942	0.918	0.310	0.186	0.760	0.687
		0.020	0.273	0.187	0.512	0.458	0.079	0.034	0.170	0.125
		0.015	0.216	0.145	0.368	0.316	0.060	0.026	0.093	0.063
0.4	0.4	0.010	0.170	0.110	0.225	0.186	0.048	0.020	0.045	0.030
		0.005	0.141	0.087	0.128	0.099	0.039	0.016	0.022	0.014
		0.000	0.133	0.084	0.084	0.064	0.036	0.014	0.016	0.009
		0.050	0.546	0.410	0.884	0.841	0.251	0.109	0.608	0.501
		0.020	0.264	0.156	0.437	0.370	0.105	0.036	0.135	0.083
		0.015	0.222	0.125	0.316	0.257	0.088	0.030	0.082	0.046
0.8	0.4	0.010	0.193	0.103	0.203	0.158	0.083	0.024	0.046	0.024
		0.005	0.172	0.091	0.130	0.093	0.074	0.021	0.024	0.012
		0.000	0.161	0.089	0.097	0.070	0.070	0.021	0.019	0.009
		0.050	0.997	0.992	1.000	1.000	0.979	0.944	1.000	1.000
		0.020	0.851	0.800	0.996	0.992	0.571	0.451	0.961	0.945
		0.015	0.718	0.645	0.973	0.964	0.376	0.270	0.864	0.824
0.4	0.8	0.010	0.507	0.437	0.869	0.846	0.187	0.118	0.591	0.530
		0.005	0.250	0.190	0.501	0.464	0.063	0.031	0.162	0.129
		0.000	0.106	0.072	0.075	0.060	0.024	0.012	0.010	0.007
		0.050	0.994	0.987	1.000	1.000	0.946	0.893	1.000	1.000
		0.020	0.840	0.785	0.995	0.993	0.519	0.385	0.960	0.942
		0.015	0.716	0.643	0.978	0.971	0.352	0.231	0.869	0.831
0.8	0.8	0.010	0.519	0.440	0.893	0.870	0.187	0.112	0.613	0.549
		0.005	0.263	0.197	0.553	0.513	0.079	0.036	0.185	0.142
		0.000	0.119	0.077	0.083	0.065	0.037	0.015	0.013	0.008

observations are discarded. The level of significance is set to be 0.05 and the number of replications is 10,000 in all experiments.

Table 1 reports the empirical size and the size-unadjusted power of the tests. For the case where there is no linear trend in the series, $(c_t, d_t) = (1, \emptyset)$, we can see that $W_T^{\min}$ tends to overly reject the null hypothesis $(a = 0)$ for DGP1 when $T = 100$, while the empirical size of the LR test is closer to the nominal one, although it is greater than 0.05. However, the problem of the over-rejection of $W_T^{\min}$ is mitigated when $T = 200$. The parameter b_1 seems to have only a small effect on the finite sample performance of both the tests, while R^2 does affect the property of the tests. When the value of R^2 increases, the empirical size of the tests becomes closer to the nominal one and the power of the tests increases dramatically.

For DGP2, the LR test is too conservative and it suffers from reduced power. The Wald-type test also becomes conservative, but not as much as the LR test, so that $W_T^{\min}$ is more powerful than the LR test. Note that the small empirical size

Table 1. (continued).

			DGP1				DGP2			
$(c_t, d_t) = (t, 1)$			$T = 100$		$T = 200$		$T = 100$		$T = 200$	
b_1	R^2	a	$W_T^{\min}$	LR	$W_T^{\min}$	LR	$W_T^{\min}$	LR	$W_T^{\min}$	LR
		0.050	0.487	0.364	0.855	0.806	0.208	0.096	0.591	0.491
		0.020	0.231	0.139	0.346	0.283	0.071	0.026	0.092	0.057
		0.015	0.205	0.119	0.247	0.195	0.063	0.022	0.052	0.031
0.4	0.4	0.010	0.185	0.103	0.167	0.125	0.055	0.019	0.030	0.016
		0.005	0.165	0.095	0.113	0.081	0.054	0.019	0.018	0.010
		0.000	0.161	0.092	0.093	0.067	0.055	0.019	0.017	0.010
		0.050	0.408	0.258	0.751	0.672	0.192	0.065	0.413	0.294
		0.020	0.245	0.128	0.292	0.224	0.115	0.032	0.075	0.038
		0.015	0.228	0.115	0.219	0.160	0.105	0.030	0.052	0.023
0.8	0.4	0.010	0.214	0.108	0.162	0.115	0.102	0.029	0.035	0.016
		0.005	0.205	0.105	0.126	0.083	0.098	0.029	0.028	0.011
		0.000	0.203	0.104	0.114	0.074	0.099	0.028	0.026	0.011
		0.050	0.988	0.973	1.000	1.000	0.936	0.862	1.000	1.000
		0.020	0.705	0.618	0.980	0.970	0.380	0.263	0.904	0.868
		0.015	0.544	0.452	0.920	0.899	0.235	0.136	0.725	0.660
0.4	0.8	0.010	0.356	0.276	0.726	0.686	0.113	0.062	0.390	0.323
		0.005	0.197	0.134	0.329	0.288	0.052	0.022	0.083	0.059
		0.000	0.129	0.083	0.083	0.063	0.036	0.014	0.014	0.009
		0.050	0.974	0.947	1.000	1.000	0.858	0.740	1.000	0.999
		0.020	0.672	0.580	0.978	0.970	0.332	0.204	0.888	0.843
		0.015	0.524	0.426	0.924	0.903	0.211	0.110	0.716	0.647
0.8	0.8	0.010	0.352	0.263	0.755	0.711	0.117	0.053	0.397	0.326
		0.005	0.200	0.130	0.362	0.317	0.062	0.023	0.089	0.062
		0.000	0.140	0.086	0.088	0.067	0.043	0.015	0.014	0.009

results in power reduction and this deviation from the nominal size is caused by two factors; the small sample bias and the test being asymptotically conservative. To evaluate the degree of power reduction caused by the latter factor, we increase sample size to 1,000 such that the small sample bias is negligible and the empirical size becomes close to the theoretical conservative one for both the tests. The simulation result is summarized in Table 2 for the parameter a ranging from 0 to 0.01 in increments of 0.02. We observe that the empirical size of both the tests is only 0.005 to 0.007. The Wald-type test is slightly more powerful than the LR test in almost all the cases but the difference is not as much addressed as in the small sample cases such as $T = 100$ and 200. This result implies that the degree of power reduction caused by being asymptotically conservative is similar for both the Wald-type and LR tests and that the difference of power between the two tests in small samples is mainly due to the small sample bias.

We also calculate the size-adjusted power of the tests for $T = 100$ and 200 to see the difference of the theoretical performance of the two tests. In Table 3 we

Table 2. Nominal size and size-unadjusted power of the tests.

DGP2, $T = 1,000$			$(c_t, d_t) = (1, \emptyset)$		$(c_t, d_t) = (t, 1)$	
b_1	R^2	a	$W_T^{\min}$	LR	$W_T^{\min}$	LR
		0.010	0.777	0.765	0.609	0.589
		0.008	0.601	0.583	0.398	0.379
		0.006	0.352	0.337	0.195	0.183
0.4	0.4	0.004	0.134	0.126	0.055	0.053
		0.002	0.022	0.022	0.012	0.012
		0.000	0.005	0.005	0.006	0.006
		0.010	0.688	0.668	0.495	0.468
		0.008	0.497	0.478	0.309	0.288
		0.006	0.284	0.268	0.138	0.126
0.8	0.4	0.004	0.099	0.091	0.042	0.039
		0.002	0.020	0.019	0.012	0.011
		0.000	0.006	0.006	0.007	0.007
		0.010	1.000	1.000	1.000	1.000
		0.008	1.000	1.000	1.000	1.000
		0.006	1.000	1.000	0.999	0.998
0.4	0.8	0.004	0.979	0.978	0.943	0.937
		0.002	0.605	0.596	0.408	0.395
		0.000	0.005	0.006	0.006	0.006
		0.010	1.000	1.000	1.000	1.000
		0.008	1.000	1.000	1.000	1.000
		0.006	1.000	1.000	1.000	0.999
0.8	0.8	0.004	0.993	0.992	0.976	0.973
		0.002	0.730	0.719	0.540	0.525
		0.000	0.005	0.007	0.006	0.006

observe that the Wald-type test is more powerful than the LR test in almost all the cases, but the difference is only slight. Since the difference of the two size-adjusted powers is small, we may see that the Wald-type test has the theoretical property similar to the LR test.

For the case where there is a linear trend in the series, $(c_t, d_t) = (t, 1)$, the overall property of the tests seems to be preserved compared with the non-trending case.

4. An empirical application

In this section, we apply our test and the LR test by Boswijk (1996) and Paruolo (2005) for normalizing restrictions on the cointegrating vectors to a Finish dataset, which is used in Luukkonen *et al.* (1999). The dataset is comprised of the own-yield of harmonized broad money (IOWN), one- and three-month money market interest rates of money (I1M and I3M), and the five-year bond rate (IBOND), from January 1980 to December 1995. See Luukkonen *et al.* (1999) for details. Of practical interest is how the own-yield of money is in-

Table 3. Size-adjusted power of the tests.

			DGP1				DGP2			
$(c_t, d_t) = (1, \emptyset)$			$T = 100$		$T = 200$		$T = 100$		$T = 200$	
b_1	R^2	a	$W_T^{\min}$	LR	$W_T^{\min}$	LR	$W_T^{\min}$	LR	$W_T^{\min}$	LR
		0.050	0.497	0.458	0.918	0.903	0.346	0.310	0.842	0.824
		0.020	0.141	0.134	0.434	0.418	0.099	0.091	0.291	0.274
0.4	0.4	0.015	0.100	0.094	0.291	0.282	0.076	0.072	0.186	0.178
		0.010	0.072	0.069	0.166	0.158	0.062	0.057	0.107	0.100
		0.005	0.057	0.055	0.083	0.080	0.053	0.050	0.064	0.062
		0.050	0.354	0.327	0.834	0.817	0.207	0.180	0.700	0.670
		0.020	0.103	0.098	0.332	0.326	0.076	0.076	0.210	0.196
0.8	0.4	0.015	0.078	0.075	0.221	0.220	0.067	0.065	0.146	0.138
		0.010	0.063	0.058	0.132	0.132	0.057	0.057	0.092	0.088
		0.005	0.054	0.052	0.070	0.073	0.050	0.053	0.059	0.058
		0.050	0.993	0.990	1.000	1.000	0.985	0.976	1.000	1.000
		0.020	0.782	0.758	0.993	0.992	0.641	0.608	0.982	0.978
0.4	0.8	0.015	0.610	0.589	0.965	0.961	0.456	0.426	0.926	0.919
		0.010	0.389	0.378	0.839	0.832	0.252	0.235	0.736	0.725
		0.005	0.149	0.145	0.441	0.439	0.100	0.096	0.304	0.298
		0.050	0.988	0.981	1.000	1.000	0.953	0.934	1.000	1.000
		0.020	0.762	0.739	0.992	0.991	0.549	0.513	0.978	0.975
0.8	0.8	0.015	0.602	0.587	0.968	0.965	0.385	0.359	0.922	0.913
		0.010	0.389	0.381	0.860	0.854	0.215	0.200	0.738	0.722
		0.005	0.156	0.151	0.479	0.480	0.095	0.092	0.323	0.312
$(c_t, d_t) = (t, 1)$			$W_T^{\min}$	LR	$W_T^{\min}$	LR	$W_T^{\min}$	LR	$W_T^{\min}$	LR
		0.050	0.295	0.262	0.792	0.774	0.199	0.180	0.694	0.675
		0.020	0.090	0.082	0.250	0.246	0.067	0.067	0.175	0.172
0.4	0.4	0.015	0.074	0.066	0.163	0.162	0.058	0.058	0.113	0.113
		0.010	0.061	0.054	0.100	0.099	0.052	0.053	0.075	0.077
		0.005	0.053	0.049	0.061	0.061	0.048	0.050	0.056	0.056
		0.050	0.172	0.154	0.644	0.618	0.116	0.100	0.493	0.462
		0.020	0.066	0.061	0.182	0.175	0.059	0.057	0.124	0.116
0.8	0.4	0.015	0.061	0.054	0.130	0.123	0.055	0.054	0.089	0.083
		0.010	0.056	0.051	0.084	0.082	0.053	0.051	0.069	0.064
		0.005	0.052	0.048	0.059	0.058	0.051	0.049	0.056	0.054
		0.050	0.975	0.961	1.000	1.000	0.946	0.923	1.000	1.000
		0.020	0.562	0.537	0.970	0.966	0.425	0.395	0.949	0.942
0.4	0.8	0.015	0.389	0.370	0.895	0.884	0.273	0.251	0.831	0.817
		0.010	0.216	0.208	0.664	0.656	0.139	0.129	0.551	0.532
		0.005	0.090	0.086	0.261	0.261	0.070	0.068	0.184	0.178
		0.050	0.945	0.926	1.000	1.000	0.867	0.829	1.000	1.000
		0.020	0.514	0.498	0.968	0.962	0.351	0.322	0.934	0.924
0.8	0.8	0.015	0.353	0.345	0.894	0.887	0.229	0.214	0.808	0.796
		0.010	0.194	0.195	0.687	0.675	0.131	0.120	0.537	0.523
		0.005	0.084	0.084	0.284	0.281	0.071	0.068	0.181	0.177

Table 4. Results of an empirical application.

(a) Estimated cointegrating vectors ($r = 2$)

IOWN	I1M	I3M	IBOND	Constant
−0.854	−96.983	41.057	108.559	−6.401
−10.970	−198.224	247.965	−77.691	3.873

(b) p-values of the tests for normalizations with respect to the own-yield of money

c	the Wald test	the LR test	LRS
$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$	0.810	0.810	invalid
$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	0.956	0.960	invalid
$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	0.529	0.562	invalid

(c) p-values of the tests for normalizations with respect to interest rates and bond rate

c	the Wald test	the LR test	LRS
$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	0.000	0.000	valid
$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	0.000	0.000	valid
$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	0.000	0.000	valid

The estimated cointegrating vectors in Panel (a) are normalized according to Johansen (1991). Entries corresponding to the Wald and LR tests are p-values of these tests, while those in the column of "LRS" are the results of Luukkonen *et al.* (1999).

fluenced by the market interest rates and bond rate. This implies that one of the cointegrating vectors, if it exists, should be normalized with respect to the own-yield of money.

The results of the tests are summarized in Table 4. We fit a VAR(3) model with cointegrating rank $r = 2$ according to the results in Luukkonen *et al.* (1999). Panel (a) shows the estimated cointegrating vectors that are normalized according to Johansen's (1991) rule, which is difficult to interpret from an economic point of view. We thus consider other normalizations using a pre-specified matrix c. We first consider normalizing one of the cointegrating vectors with respect to the own-yield of money; the matrix c is then chosen such that $c = [e_1, c_2]$ where $e_1 = [1, 0, 0, 0]'$ and c_2 is a 4×1 vector with i-th element equal to 1 ($i = 2$, 3, or 4) and the other elements equal to zero. From Panel (b) we can see that both the Wald-type and LR tests cannot reject the null hypothesis of the invalid normalization even at 10% significance level; p-values of the tests exceed 50%. This is consistent with the results in Luukkonen *et al.* (1999), whose tests reject the null hypothesis of the valid normalization for these c (see the column "LRS" in Table 4). These results imply that the own-yield of money is not included in the cointegrating relationship. On the other hand, the other normalization rules considered in the paper can be seen to be valid from Panel (c). These results imply that we should avoid normalizing the cointegrating vectors by imposing restrictions on the coefficient of the own-yield of money, and should normalize the cointegrating vectors with respect to interest rates and bond rate.

5. Conclusion

In this paper we proposed the Wald-type test of a normalization of cointegrating vectors. The test statistic is constructed using only the unrestricted ML estimator. We showed that our test statistic converges in distribution to a random variable that is bounded above by, or equal to, a chi-square distribution with one degree of freedom, depending on the deficiency of $c'\beta$. As a by-product, we also proposed the test for the null hypothesis of $\mathrm{rk}(c'\beta) = f$ for $f < r$. The finite sample simulations show that we should carefully interpret the result of $W_T^{\min}$ in empirical analysis for small sample sizes such as 100; the LR test is recommended in those cases. Since size distortion of $W_T^{\min}$ is mitigated when $T = 200$ and $W_T^{\min}$ is more powerful than the LR test for DGP2, our test may complement the LR test. We then recommend using both $W_T^{\min}$ and the LR test as well as the tests proposed by Luukkonen *et al.* (1999) in practical analysis to see whether the identifying restrictions imposed on the cointegrating vectors are valid or not.

Appendix

PROOF OF THEOREM 1. (i) We first consider the case where $(c_t, d_t) = (1, \emptyset)$. As did Johansen (1988, 1991, 1995b) we normalize the ML estimators of β^* and α as $\tilde{\beta}^* = \hat{\beta}^*(\bar{\eta}'\hat{\beta}^*)^{-1}$ and $\tilde{\alpha} = \hat{\alpha}\hat{\beta}^{*\prime}\bar{\eta}$, respectively, where $\eta' = [\beta', 0]$. Then, in the same way as Section 13 of Johansen (1995b), we have

$$B_T(\tilde{\beta}^* - \beta^*) \xrightarrow{d} \left(\int_0^1 GG'ds\right)^{-1} \int_0^1 GdV', \qquad \text{where} \quad B_T = \begin{bmatrix} T\beta_\perp' & 0 \\ 0 & T^{1/2} \end{bmatrix},$$

$G(r) = [(\bar{\beta}_\perp CW(r))', 1]'$, $C = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'$, $W(r)$ is an n-dimensional Brownian motion with a variance matrix Σ, $V(r) = (\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}W(r)$, and $G(\cdot)$ and $V(\cdot)$ are independent of each other. Since $\tilde{\beta}$ is the first n rows of $\tilde{\beta}^*$ we have

$$\begin{aligned} T\beta_\perp'(\tilde{\beta} - \beta) &= L'B_T(\tilde{\beta}^* - \beta^*) \\ &\xrightarrow{d} L'\left(\int_0^1 GG'ds\right)^{-1} \int_0^1 GdV'. \end{aligned} \tag{A.1}$$

We can also see that $\tilde{\alpha}$ and $\hat{\Sigma}$ are consistent as shown by Johansen (1995b).

Since $(c'\hat{\beta})\hat{\Psi}(\hat{\beta}'c)\hat{\Phi}^{-1}$ is invariant to the normalizations of $\hat{\alpha}$ and $\hat{\beta}$, we can replace them by $\tilde{\alpha}$ and $\tilde{\beta}$. Then, the eigenvalues are obtained by the following determinant equation:

$$|(c'\tilde{\beta})\tilde{\Psi}(\tilde{\beta}'c) - \lambda\tilde{\Phi}| = |H'||(c'\tilde{\beta})\tilde{\Psi}(\tilde{\beta}'c) - \lambda\tilde{\Phi}||H| = 0, \tag{A.2}$$

where $\tilde{\Psi}$ and $\tilde{\Phi}$ are defined as $\hat{\Psi}$ and $\hat{\Phi}$ using $\tilde{\alpha}$ and $\tilde{\beta}$, and H is a nonsingular matrix. Note that when $\mathrm{rk}(c'\beta) = f$ we can find an $r \times f$ full column rank matrix δ such that $\mathrm{sp}(\delta) = \mathrm{sp}(c'\beta)$. We use $H = [\delta, T\delta_\perp]$ in the following.

Since $\delta_\perp$ is orthogonal to δ, we have $\delta_\perp' c'\beta = 0$, or $c\delta_\perp \in \mathrm{sp}(\beta_\perp)$. Then, we can find an $(n-r) \times (r-f)$ matrix h such that $c\delta_\perp = \beta_\perp h$, or $h = \bar{\beta}_\perp' c\delta_\perp$. Then, (A.1) implies

$$\begin{aligned} T\delta_\perp' c'(\tilde{\beta} - \beta) &= Th'\beta_\perp'(\tilde{\beta} - \beta) \\ &\xrightarrow{d} h'L'\left(\int_0^1 GG'ds\right)^{-1} \int_0^1 GdV' \\ &= \delta_\perp' c'\bar{\beta}_\perp L'\left(\int_0^1 GG'ds\right)^{-1} \int_0^1 GdV' \equiv x', \quad \text{say.} \end{aligned} \tag{A.3}$$

By defining $\Psi \equiv \alpha'\Sigma^{-1}\alpha$, we obtain

$$H'(c'\tilde{\beta})\tilde{\Psi}(\tilde{\beta}'c)H \xrightarrow{d} \begin{bmatrix} \delta'c'\beta\Psi\beta'c\delta & \delta'c'\beta\Psi x \\ x'\Psi\beta'c\delta & x'\Psi x \end{bmatrix}, \tag{A.4}$$

because $\tilde{\Psi} \xrightarrow{p} \alpha'\Sigma^{-1}\alpha = \Psi$, $\delta'(c\tilde{\beta}) \xrightarrow{p} \delta'c\beta$, and (A.3). Let $T \to \infty$ and $\lambda \to 0$ such that $\rho = T^2\lambda$ is fixed as in Johansen (1995b, p. 159). Then,

$$\lambda H'\tilde{\Phi}H = \frac{\rho}{T^2}H'\tilde{\Phi}H \xrightarrow{d} \begin{bmatrix} 0 & 0 \\ 0 & \rho\delta_\perp' c'\bar{\beta}_\perp L'(\int_0^1 GG'ds)^{-1}L\bar{\beta}_\perp' c\delta_\perp \end{bmatrix}, \tag{A.5}$$

because $\delta'(c\tilde{\beta}_\perp) \xrightarrow{p} \delta'c'\beta_\perp$, $T\delta_\perp'(c'\tilde{\beta}_\perp) = T\delta_\perp' c'\beta_\perp + o_p(T)$, and $\Upsilon_T' S_{11}^* \Upsilon_T \xrightarrow{d} \int_0^1 GG'ds$, which is obtained because $\Upsilon_T R_{1[Tr]}^* \xrightarrow{d} G(r)$. Then, (A.4) and (A.5) imply that the $r-f$ smallest eigenvalues of (A.2) normalized by T^2 converge in distribution to those of the equation,

$$\begin{aligned} 0 &= \begin{vmatrix} \delta'c'\beta\Psi\beta'c\delta & \delta'c'\beta\Psi x \\ x'\Psi\beta'c\delta & x'\Psi x - \rho\delta_\perp' c'\bar{\beta}_\perp L'(\int_0^1 GG'ds)^{-1}L\bar{\beta}_\perp' c\delta_\perp \end{vmatrix} \\ &= |\delta'c'\beta\Psi\beta'c\delta| \\ &\quad \times \left| x'\{\Psi - \Psi\beta'c\delta(\delta'c'\beta\Psi\beta'c\delta)^{-1}\delta'c'\beta\Psi\}x \right. \\ &\qquad \left. - \rho\delta_\perp' c'\bar{\beta}_\perp L'\left(\int_0^1 GG'ds\right)^{-1} L\bar{\beta}_\perp' c\delta_\perp \right| \\ &\propto \left|(x'J)(J'x) - \rho\delta_\perp' c'\bar{\beta}_\perp L'\left(\int_0^1 GG'ds\right)^{-1} L\bar{\beta}_\perp' c\delta_\perp\right| \end{aligned} \tag{A.6}$$

because $|\delta'c'\beta\Psi\beta'c\delta| \neq 0$ under H_0^f, where J is an $r \times (r-f)$ matrix such that $JJ' = \Psi - \Psi\beta'c\delta(\delta'c'\beta\Psi\beta'c\delta)^{-1}\delta'c'\beta\Psi$ and $J'\Psi^{-1}J = I_{r-f}$. See Johansen (1988, p. 246) and Kurozumi (2005). Note that $\mathrm{vec}(x'J)$ is an $(r-f)^2$ random vector and its variance conditioned on $G(\cdot)$ is given by $V_{x'J} \equiv \delta'_{\perp}c'\bar{\beta}_{\perp}L'(\int_0^1 GG'ds)^{-1}L\bar{\beta}'_{\perp}c\delta_{\perp}\otimes I_{r-f}$ from (A.3), using the relation of $J'\Psi^{-1}J = I_{r-f}$. This implies that $V_{x'J}^{-1/2}(x'J) \mid G(\cdot)$ consists of $(r-f)^2$ independent standard normal random variables. Note that the conditional distribution of $N \equiv V_{x'J}^{-1/2}(x'J)$ is free of the conditioning variable $G(\cdot)$, and hence the unconditional distribution of each element is also standard normal. Then, multiplying (A.6) by $V_{x'J}^{-1/2}$ from both sides, we can see that

$$\begin{aligned} 0 &= |(c'\tilde{\beta})\tilde{\Psi}(\tilde{\beta}'c) - \lambda\tilde{\Phi}| \\ &= |V_{x'J}^{-1/2}||H'||(c'\tilde{\beta})\tilde{\Psi}(\tilde{\beta}'c) - \lambda\tilde{\Phi}||H||V_{x'J}^{-1/2}|/|\delta'c'\beta\Psi\beta'c\delta| \\ &\xrightarrow{d} |N'N - \rho I_{r-f}|. \end{aligned}$$

This implies that $\rho = T^2\lambda$ converges in distribution to the roots of (2.6).

The theorem is proved in the same way when $(c_t, d_t) = (t, 1)$, in which we only have to modify the limiting result (A.1) such that

$$T\beta'_{\perp}(\tilde{\beta} - \beta) \xrightarrow{d} L'\left(\int_0^1 \underline{G}\,\underline{G}'ds\right)^{-1}\int_0^1 \underline{G}dV',$$

where $\underline{G}(r) = G_0(r) - \int_0^1 G_0 ds$ with $G_0(r) = [(\bar{\beta}'_{\perp}W(r))', r]'$.

Proof of Corollary 1. Part (i) is immediately obtained because the limiting distribution of W_T^f is a trace of $N'N$. To prove part (ii), let us suppose that the true rank of $c'\beta$ equals, say, $g > f$, under the alternative. Let $H^* = [\delta, \delta_{\perp}]$ where δ and $\delta_{\perp}$ are $r\times g$ and $r\times(r-g)$ matrices such that $\mathrm{sp}(\delta) = \mathrm{sp}(c'\beta)$ and $\mathrm{sp}(\delta_{\perp}) = \mathrm{sp}(c'\beta)^{\perp}$ as in the proof of Theorem 1. Then, in the same way as (A.4) and (A.5), we obtain

$$H^{*\prime}(c'\tilde{\beta})\tilde{\Psi}(\tilde{\beta}'c)H^* \xrightarrow{p} \begin{bmatrix} \delta'c'\beta\Psi\beta'c\delta & 0 \\ 0 & 0 \end{bmatrix} \tag{A.7}$$

and

$$H^{*\prime}\tilde{\Phi}H^* \xrightarrow{d} H^{*\prime}\Phi H^*, \tag{A.8}$$

where $\Phi = c'\beta(\beta'\beta)^{-1}\beta'c + c'\bar{\beta}_{\perp}L'(\int_0^1 GG'ds)^{-1}L\bar{\beta}'_{\perp}c$. Note that $H^{*\prime}\Phi H^*$ is positive definite (a.s.) because $|H^*| \neq 0$ and Φ is positive definite (a.s.). Using (A.7) and (A.8), we have

$$\begin{aligned} &|H^{*\prime}||(c'\tilde{\beta})\tilde{\Psi}(\tilde{\beta}'c) - \lambda\tilde{\Phi}||H^*| \\ &\quad\xrightarrow{d} \begin{vmatrix} \delta'c'\beta\Psi\beta'c\delta - \lambda\delta'\Phi\delta & -\lambda\delta'\Phi\delta_{\perp} \\ -\lambda\delta'_{\perp}\Phi\delta & -\lambda\delta'_{\perp}\Phi\delta_{\perp} \end{vmatrix} \\ &\quad= |-\lambda\delta'_{\perp}\Phi\delta_{\perp}| \\ &\qquad\times|\delta'c'\beta\Psi\beta'c\delta - \lambda(\delta'\Phi\delta - \delta'\Phi\delta_{\perp}(\delta'_{\perp}\Phi\delta_{\perp})^{-1}\delta'_{\perp}\Phi\delta)|. \end{aligned}$$

This implies that the g largest eigenvalues converge in distribution to the roots of

$$|\delta' c' \beta \Psi \beta' c \delta - \lambda(\delta' \Phi \delta - \delta' \Phi \delta_\perp (\delta_\perp' \Phi \delta_\perp)^{-1} \delta_\perp' \Phi \delta)| = 0. \tag{A.9}$$

Note that both the first and second terms in (A.9) are positive definite (a.s.) from the definition of δ and the fact that $H^{*\prime} \Phi H^*$ is positive definite (a.s.). Then, we can see that all the roots of (A.9) take positive values (a.s.). Therefore, the test statistic diverges to infinity at a rate of T^2.

Proof of Corollary 2. The convergence, $W_T^{\min} \xrightarrow{d} \chi_1^2$ for $\mathrm{rk}(c'\beta) = r-1$, is proved in Corollary 1.

When $\mathrm{rk}(c'\beta) = g < r-1$, the matrices δ and $\delta_\perp$ in the proof of Theorem 1 are $r \times g$ and $r \times (r-g)$, and we obtain (A.6) similarly. In this case, J and $x'J$ are $r \times (r-g)$ and $(r-g) \times (r-g)$ random matrices. Then, we can see that $\rho_r = T^2 \lambda_r$, the smallest eigenvalue multiplied by T^2, converges in distribution to $\rho_{\min}$, the smallest eigenvalue of $|N'N - \rho I_{r-g}| = 0$, where N is $(r-g) \times (r-g)$ and $\mathrm{vec}(N) \sim \mathcal{N}(0, I_{(r-g)^2})$. Let us decompose N into $N = [N_1, N_2]$ where N_1 is $(r-g) \times 1$ and N_2 is $(r-g) \times (r-g-1)$. Note that we can choose a vector $z^*(N_2)$ for given N_2 such that $z^*(N_2)' N_2 = 0$ and $z^*(N_2)' z^*(N_2) = 1$. Then, we can see that

$$\rho_{\min} = \min_z (z'N'Nz)/(z'z) \le z^*(N_2) N N' z^*(N_2) = (z^*(N_2)' N_1)^2,$$

where the first equality holds by equation (6) of Magnus and Neudecker (1988, p. 204). Because $z^*(N_2)$ is uniquely defined by a given N_2, and N_1 is independent of N_2, we can see that

$$z^*(N_2)' N_1 \mid N_2 \sim \mathcal{N}(0, z^*(N_2)' z^*(N_2)) = \mathcal{N}(0, 1),$$

and then $(z^*(N_2)' N_1)^2 \mid N_2 \sim \chi_1^2$, which implies that the unconditional distribution of $(z^*(N_2)' N_1)^2$ is also a chi-square distribution with one degree of freedom. Thus, the statement of the theorem holds for the case where $\mathrm{rk}(c'\beta) < r-1$.

Part (ii) is proved in the same way as Corollary 1.

Acknowledgements

The earlier version of this paper was circulated as "Alternative Tests of a Valid Normalization of Cointegrating Vectors". I am grateful to Professors Ritva Luukkonen, Antti Ripatti, and Pentti Saikkonen for making available their dataset. I am also grateful to two anonymous referees for valuable comments. All errors are my responsibility. This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology under Grants-in-Aid Nos. 17203016 and 18730142.

References

Boswijk, H. P. (1995). Identifiability of cointegrated systems, *Tinbergen Institute Discussion Paper TI 95-78*, Tinbergen Institute.

Boswijk, H. P. (1996). Testing identifiability of cointegrating vectors, *Journal of Business and Economic Statistics*, **14**, 153–160.

Boswijk, H. P. and Doornik, J. A. (2003). Identifying, estimating and testing restricted cointegrated systems: An overview, Manuscript, Universiteit van Amsterdam.

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

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

Johansen, S. (1995a). Identifying restrictions of linear equations with applications to simultaneous equations and cointegration, *Journal of Econometrics*, **69**, 111–132.

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

Johansen, S. and Juselius, K. (1992). Testing structural hypotheses in a multivariate cointegration analysis of the PPP and UIP for UK, *Journal of Econometrics*, **53**, 211–244.

Johansen, S. and Juselius, K. (1994). Identification of the long-run and the short-run structure an application to the ISLM Model, *Journal of Econometrics*, **63**, 7–36.

Kurozumi, E. (2003). Testing the rank of a sub-matrix of cointegration with a deterministic trend, *Discussion Paper #2003-14*, Graduate School of Economics, Hitotsubashi University.

Kurozumi, E. (2005). The rank of a sub-matrix of cointegration, *Econometric Theory*, **21**, 299–325.

Luukkonen, R., Ripatti, A. and Saikkonen, P. (1999). Testing for a valid normalization of cointegrating vectors in vector autoregressive processes, *Journal of Business and Economic Statistics*, **17**, 195–204.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York.

Paruolo, P. (1997). Asymptotic inference on the moving average impact matrix in cointegrated I(1) VAR systems, *Econometric Theory*, **13**, 79–118.

Paruolo, P. (2005). A LR test for a cointegration sub-matrix, Manuscript, University of Insubria.

Pesaran, M. H. and Shin, Y. (2002). Long-run structural modelling, *Econometric Reviews*, **21**, 49–87.

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

Saikkonen, P. (1999). Testing normalization and overidentification of cointegrating vectors in vector autoregressive processes, *Econometric Reviews*, **18**, 235–257.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica*, **55**, 1035–1056.

Yamamoto, T. and Kurozumi, E. (2006). Tests for long-run Granger non-causality in cointegrated systems, *Journal of Time Series Analysis*, **27**, 703–723.