June 10, 1998

Testing for the Cointegrating Rank of a VAR Process with an Intercept*

by

Pentti Saikkonen and Helmut Lütkepohl

Department of Statistics
University of Helsinki
P.O. Box 54
SF-00014 University of Helsinki
FINLAND

Tel.: +358-9-1918867
Fax: +358-9-1918872

Institut für Statistik und Ökonometrie
Wirtschaftswissenschaftliche Fakultät
Humboldt–University
Spandauer Str. 1
10178 Berlin
GERMANY
Tel.: +49-30-2093-5718
Fax: +49-30-2093-5712

Abstract

Testing the cointegrating rank of a vector autoregressive process with an intercept is considered. In addition to the likelihood ratio (LR) tests developed by Johansen and Juselius and others we also consider an alternative class of tests which is based on estimating the trend parameters of the deterministic term in a different way. The asymptotic local power of these tests is derived and compared to that of the corresponding LR tests. The small sample properties are investigated by simulations. The new tests are seen to be substantially more powerful than conventional LR tests.

*We are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, for financial support and to Christian Müller and Kirstin Hubrich for performing the computations. Part of this research was done while the first author was visiting the Humboldt University in Berlin and another part was done while the second author was visiting CORE in Louvain-la-Neuve.

1 Introduction

Following the invention of the concept of cointegration by Granger (1981) and Engle & Granger (1987) it has been adopted in many empirical studies. Since inference in and the interpretation of econometric models crucially depends on the existence and the number of cointegration relations within a system of variables, tests for cointegration are now routinely applied at an early stage of an analysis. Whereas some tests are designed for use in single equation models other tests are based on models for the full data generation process (DGP) of a set of variables and enable the analyst to investigate not only the existence of cointegration relations but also their number.

The latter approach usually assumes that the DGP is a finite order vector autoregressive (VAR) process. Notably the likelihood ratio (LR) tests proposed by Johansen (1991, 1995), Johansen & Juselius (1990) and Reinsel & Ahn (1992) are developed in this framework. It is shown that asymptotically valid tests for the number of cointegrating relations can be constructed which do not depend on the short-term dynamics of the DGP. They do depend on the properties of deterministic terms, however. In particular, the specific properties of intercepts and polynomial trend terms in a VAR process have a crucial impact on the asymptotic distribution. In fact, it is shown in Johansen & Juselius (1990) that LR tests for the number of cointegrating relations in a Gaussian VAR process with an intercept term have different limiting distributions under the null hypothesis depending on whether the intercept term generates a linear trend in the variables or it can be absorbed into the cointegrating relations.

Since a VAR process with an intercept term is perhaps the most common model used in applied work we will focus on this case in the following and assume that the intercept cannot be absorbed into the cointegrating relations so that either the system is stationary or a deterministic linear trend is present in at least one of the variables. Under this condition an alternative class of tests for the cointegrating rank is proposed and its limiting null distribution is derived. We will also explore the asymptotic local power of the new tests and of the corresponding LR tests and we will compare it to the local power of other suitable tests for the cointegrating rank when the variables have a deterministic linear trend. It turns out that the new tests have much better local power than the conventional LR tests and other tests that allow for deterministic linear trends. A local power analysis of the LR tests is also performed by Rahbek (1994). In our analysis we will focus on different local alternatives, however. A comparison with Rahbek's approach will be provided. Using Monte Carlo simulations we will also perform a small sample comparison of the new tests and the standard LR tests. It is found that the new tests tend to be more powerful close to the null hypothesis than the standard tests.

The paper is structured as follows. In the next section the basic model is introduced. In Section 3 the test procedures are described and the limiting distributions under the null hypothesis are considered. A local power analysis is performed in Section 4 and Section 5 reports the results of a small sample comparison of the tests. Conclusions are drawn in Section 6. Most proofs are contained in the Appendix.

The following notation is used throughout. The lag and differencing operators are denoted by L and Δ, respectively, that is, $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. The symbol $I(d)$ is used to denote a process which is integrated of order d, that is, it is stationary or asymptotically stationary after differencing d times while it is still nonstationary after differencing just $d-1$ times. The symbol $\stackrel{d}{\rightarrow}$ signifies convergence in distribution or weak convergence. $\lambda_{max}(A)$, $\mathrm{tr}(A)$ and $\mathrm{rk}(A)$ denote the maximal eigenvalue, the trace and the rank of the matrix A, respectively. Moreover, $\|\cdot\|$ denotes the Euclidean norm. If A is an $(n \times m)$ matrix of full column rank $(n > m)$ we denote its orthogonal complement by $A_\perp$. In other words, $A_\perp$ is an $(n \times (n-m))$ matrix of full column rank and such that $A'A_\perp = 0$. The orthogonal complement of a nonsingular square matrix is zero and the orthogonal complement of zero is an identity matrix of suitable dimension. An $(n \times n)$ identity matrix is denoted by I_n. LS and GLS are used to abbreviate least squares and generalized least squares, respectively, RR stands for reduced rank and DGP is short for data generation process. NID means normally independently distributed. A sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

2 The Framework of Analysis

Our point of departure is the DGP of an n-dimensional multiple time series $y_t = (y_{1t}, \ldots, y_{nt})'$ given by the VAR(p) process

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{2.1}$$

where ν is an unknown, fixed $(n \times 1)$ intercept vector, the A_j are $(n \times n)$ coefficient matrices and ε_t is an $(n \times 1)$ stochastic error term which we assume to be a martingale difference sequence with $E(\varepsilon_t|\varepsilon_s, s < t) = 0$, nonstochastic positive definite conditional covariance matrix $E(\varepsilon_t\varepsilon_t'|\varepsilon_s, s < t) = \Omega$ and bounded fourth moments. The choice of initial values y_t $(t = 1, \ldots, p)$ will be discussed later. Subtracting y_{t-1} on both sides of (2.1) and rearranging terms gives the error correction (EC) form

$$\Delta y_t = \nu + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{2.2}$$

where $\Pi = -(I_n - A_1 - \cdots - A_p)$ and $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ $(j = 1, \ldots, p-1)$ are $(n \times n)$. We assume that the components of the process y_t are at most $I(1)$ and possibly cointegrated so that we can write

$$\Pi = \alpha\beta', \tag{2.3}$$

where α and β are $(n \times r)$ matrices of full column rank and $0 \leq r < n$. Here r is the cointegrating rank. Note that we exclude the possibility that y_t is $I(0)$, that is, the cointegrating rank cannot be equal to n. This assumption is often reasonable because stationarity can often be ruled out on the basis of prior information on the data and variables. Formally the reason for ruling out stationarity is that for a stationary process an intercept term cannot generate a deterministic linear trend and is hence inconsistent with our assumptions. Note that our model excludes a linear trend in the cointegration relations, that is, the variables are assumed not to cointegrate with a deterministic linear trend. This condition is sometimes imposed in the form $\alpha_\perp'\nu \neq 0$. In order to exclude $I(2)$ processes we require that the characteristic equation

$$\det(I_n - A_1 z - \cdots - A_p z^p) = \det\left(\left(I_n - \sum_{j=1}^{p-1} \Gamma_j z^j\right)(1-z) - \alpha\beta' z\right) = 0 \tag{2.4}$$

has exactly $n - r$ roots equal to one and all other roots outside the unit circle.

A process of this type can generate deterministic linear trends in the variables and, as mentioned earlier, we assume that at least some component of y_t has such a trend. For our purposes it will be convenient to separate the deterministic part from the stochastic part of the process. Therefore we write y_t in the form

$$y_t = \mu_0 + \mu_1 t + x_t, \qquad t = 1, 2 \ldots, \tag{2.5}$$

where μ_0 and μ_1 are $(n \times 1)$ vectors with $\mu_1 \neq 0$ reflecting the fact that at least one component has a deterministic linear trend. The process x_t is an unobservable error term which is easily seen to have a VAR(p) representation

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t \tag{2.6}$$

and hence an EC from

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \qquad t = 1, 2, \ldots \tag{2.7}$$

(see, e.g., Lütkepohl & Saikkonen (1997) (henceforth L&S)). For convenience, we impose the initial value condition $x_t = 0$, $t \leq 0$. Our results remain valid if the initial values have some fixed distribution which does not depend on the sample size. The initial values of y_t $(t = 1, \ldots, p)$ are assumed to be the ones implied by these assumptions for the x_t process.

Under our present assumptions, it follows from Johansen's (1991, 1995) formulation of Granger's representation theorem that

$$x_t = C \sum_{i=1}^{t} \varepsilon_i + \xi_t, \qquad t = 1, 2, \ldots, \tag{2.8}$$

where, apart from the specification of initial values, ξ_t is a stationary process and $C = \beta_\perp(\alpha'_\perp \Psi \beta_\perp)^{-1}\alpha'_\perp$ with $\Psi = I_n - \Gamma_1 - \cdots - \Gamma_{p-1} = I_n + \sum_{j=1}^{p-1} j A_{j+1}$. An immediate consequence of (2.8) is that the process x_t obeys the functional central limit theorem

$$T^{-1/2} x_{[Ts]} \xrightarrow{d} C\mathbf{B}(s), \qquad 0 \leq s \leq 1, \tag{2.9}$$

where $\mathbf{B}(s)$ is a Brownian motion with covariance matrix Ω.

Without any restrictions for the trend parameters μ_0 and μ_1 a process of the form (2.5) has a VAR(p) representation

$$y_t = \nu + \nu_1 t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = p+1, p+2, \ldots,$$

where $\nu = -\Pi\mu_0 + (\Psi + \Pi)\mu_1$ and $\nu_1 = -\Pi\mu_1$ (see L&S). Hence, since in our DGP (2.1) the term $\nu_1 t$ does not appear, we have

$$\nu_1 = -\Pi\mu_1 = 0 \tag{2.10}$$

which implies $\beta'\mu_1 = 0$. In other words, it is assumed that $\mu_1 \in \text{span}(\beta_\perp)$.

In this framework we are interested in testing

$$H(r_0) : \mathrm{rk}(\Pi) = r_0 \qquad \text{vs.} \qquad \bar{H}(r_0) : \mathrm{rk}(\Pi) > r_0, \tag{2.11}$$

that is, the cointegrating rank being r_0 is tested against a rank greater than r_0 for $0 \leq r_0 < n-1$. Note that since $\mu_1 \neq 0$ is assumed, $\beta_\perp$ must at least consist of one column. Hence, β cannot span the full n-dimensional Euclidean space. Thus, the alternative $\mathrm{rk}(\Pi) = n$ (stationarity) is excluded by assumption and therefore r_0 can be at most $n-2$. This is quite plausible because a stationary process with an intercept cannot generate a linear trend, as mentioned earlier. If a trend-stationary system is regarded as a possibility then (2.2) is not a suitable model.

One possible test for the pair of hypotheses in (2.11) is based on trend-adjusting y_t first. For this purpose we will need estimators of the trend parameters μ_0 and μ_1. Here any estimators may be used which satisfy the following properties:

$$\beta'(\tilde{\mu}_0 - \mu_0) = O_p(T^{-1/2}), \tag{2.12}$$

$$\beta_\perp'(\tilde{\mu}_0 - \mu_0) = O_p(1), \tag{2.13}$$

$$\beta'(\tilde{\mu}_1 - \mu_1) = O_p(T^{-3/2}), \tag{2.14}$$

$$T^{1/2}\beta_\perp'(\tilde{\mu}_1 - \mu_1) \xrightarrow{d} \beta_\perp' C\mathbf{B}(1). \tag{2.15}$$

In (2.15) we have used the same Brownian motion as in (2.9) to indicate that (2.15) holds jointly with other relevant weak convergencies which appear later. The estimators of the trend parameters discussed in L&S and Saikkonen & Lütkepohl (1997) satisfy the above requirements. Since we now assume the a priori restriction (2.10) the estimator in L&S which explicitly takes this restriction into account appears convenient here. For our present purposes the precise form of the estimators is not important. We will therefore not elaborate on them here.

We also need appropriate estimators for the parameters α, β and Ω. Again, we can use any estimators with the properties

$$D_T'(\tilde{\beta} - \beta) = O_p(1), \tag{2.16}$$

$$T^{1/2}(\tilde{\alpha} - \alpha) = O_p(1) \tag{2.17}$$

and

$$T^{1/2}(\tilde{\Omega} - \Omega) = O_p(1), \tag{2.18}$$

where $D_T = [T^{3/2}\mu_1 : T\gamma]$ and γ is any $(n \times (n-r-1))$ matrix orthogonal to β and μ_1 such that $[\beta : \mu_1 : \gamma]$ is of full rank. Note that here we implicitly assume that the parameter matrix β and the estimator $\tilde{\beta}$ have been made unique by a suitable normalization (cf. Johansen (1995, Chapter 13.2) and Paruolo (1997)). Such a normalization also implies a normalization of α and its estimator which are also assumed here. These normalizations have no effect on the new tests presented in the next section because the test statistics are invariant to normalizations of this kind. It is well-known that the usual RR estimators based on (2.2) satisfy (2.16) – (2.18) (see Johansen (1995, Lemma 13.2) and Paruolo (1997, Lemma 5.1)). Alternative possibilities will be discussed in the next section.

3 Tests for the Cointegrating Rank

In this section we present tests for the pair of hypotheses given in (2.11). We assume now that α and β are $(n \times r_0)$ matrices, that is, their column dimension is equal to the rank of Π under $H(r_0)$. Of course, if the null hypothesis is true, β spans the full cointegration space. We will briefly review the standard LR tests proposed by Johansen & Juselius (1990) and Johansen (1991) and then present alternative tests.

3.1 LR Tests

For a sample $y_1, \ldots, y_T$ the LR test statistics may be obtained as follows. Define $z_t' = (1, \Delta y_{t-1}', \ldots, \Delta y_{t-p+1}')$ and

$$M_T = (T-p)^{-1} \left[\sum_{t=p+1}^{T} y_{t-1}y_{t-1}' - \sum_{t=p+1}^{T} y_{t-1}z_t' \left(\sum_{t=p+1}^{T} z_t z_t' \right)^{-1} \sum_{t=p+1}^{T} z_t y_{t-1}' \right] \tag{3.1}$$

and let $\hat{\Pi}$ be the LS estimator of the matrix Π in the model (2.1). Moreover, denote the corresponding LS residuals by $\hat{\varepsilon}_t$ and define

$$\hat{\Omega} = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t' \tag{3.2}$$

Denoting by $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$ the ordered generalized eigenvalues obtained as solutions of

$$\det(\hat{\Pi} M_T \hat{\Pi}' - \lambda \hat{\Omega}) = 0, \tag{3.3}$$

the LR statistic for testing the pair of hypotheses (2.11) is given by

$$LR_{trace}(r_0) = (T-p)\sum_{j=r_0+1}^{n} \log(1+\hat{\lambda}_j). \tag{3.4}$$

If $\mu_1 \neq 0$ the asymptotic distribution of this test statistic under the null hypothesis is known to be

$$LR_{trace}(r_0) \xrightarrow{d} \text{tr}\left\{\left(\int_0^1 \bar{\mathbf{G}}(s)\mathrm{d}\mathbf{W}(s)'\right)'\left(\int_0^1 \bar{\mathbf{G}}(s)\bar{\mathbf{G}}(s)'\mathrm{d}s\right)^{-1}\left(\int_0^1 \bar{\mathbf{G}}(s)\mathrm{d}\mathbf{W}(s)'\right)\right\}, \tag{3.5}$$

where $\mathbf{W}(s)$ is an $(n-r_0)$-dimensional standard Brownian motion, $\mathbf{G}(s) = [s, W_1(s), \ldots, W_{n-r_0-1}(s)]'$ and $\bar{\mathbf{G}}(s) = \mathbf{G}(s) - \int_0^1 \mathbf{G}(u)\mathrm{d}u$. Critical values for this test may be found in Johansen & Juselius (1990, Table A.1), Johansen (1995, Table 15.3) and Osterwald-Lenum (1992, Table 1), among others.

3.2 Tests Based on Prior Trend Adjustment

For the case when a linear deterministic trend of unknown form is present, it was found in L&S that a test which is more powerful for some alternatives than the LR test may be obtained by prior trend removal. A similar approach may be used under the present assumptions as well. To derive the new tests presented in the following we use the definitions of ν and Ψ and write (2.2) as

$$\Delta y_t - \mu_1 = \Pi(y_{t-1} - \mu_0) + \sum_{j=1}^{p-1}\Gamma_j(\Delta y_{t-j} - \mu_1) + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{3.6}$$

Furthermore, using $\beta(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' = I_n$ we can proceed as in L&S and write (3.6) further as

$$\Delta y_t - \mu_1 = \kappa u_{t-1} + \rho v_{t-1} + \sum_{j=1}^{p-1}\Gamma_j(\Delta y_{t-j} - \mu_1) + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{3.7}$$

where $u_t = \beta'(y_t - \mu_0) = \beta' x_t$, $v_t = \beta_\perp'(y_t - \mu_0) = \beta_\perp'\mu_1 t + \beta_\perp' x_t$, $\kappa = \Pi\beta(\beta'\beta)^{-1}$ and $\rho = \Pi\beta_\perp(\beta_\perp'\beta_\perp)^{-1}$. If $H(r_0)$ in (2.11) holds so that $\Pi = \alpha\beta'$, we have $\kappa = \alpha$ and $\rho = 0$. On the other hand, under the alternative, $\rho \neq 0$. Therefore the idea is to test the restriction $\rho = 0$ in a feasible version of (3.7). Note, however, that ρ is an $(n \times (n - r_0))$ dimensional matrix which is seen to be zero if and only if the $((n-r_0) \times (n-r_0))$ matrix $\rho_* := \alpha_\perp'\rho = 0$. Therefore the model (3.7) is premultiplied by $\alpha_\perp'$ and the linear hypothesis $H_0 : \rho_* = 0$ is tested in a feasible version of

$$\alpha_\perp' \Delta x_t = \kappa_* u_{t-1} + \rho_* v_{t-1} + \sum_{j=1}^{p-1} \Gamma_{*j} \Delta x_{t-j} + \eta_t^*, \qquad t = p+1, p+2, \dots, \tag{3.8}$$

where $\kappa_* := \alpha_\perp' \kappa$, $\Gamma_{*j} = \alpha_\perp' \Gamma_j$ and $\eta_t^* = \alpha_\perp' \varepsilon_t$. Here $\Delta y_t - \mu_1 = \Delta x_t$ has been used. Thus we have to test a set of linear restrictions in a linear model. For this purpose the three asymptotically equivalent LR, LM and Wald tests are available.

The actual test statistic is determined by first obtaining estimators $\tilde{\mu}_0, \tilde{\mu}_1, \tilde{\alpha}, \tilde{\beta}, \tilde{\alpha}_\perp, \tilde{\beta}_\perp$ and $\tilde{\Omega}$ and then computing the test statistic on the basis of the feasible model

$$\tilde{\alpha}_\perp' \Delta \tilde{x}_t = \kappa_* \tilde{u}_{t-1} + \rho_* \tilde{v}_{t-1} + \Gamma_* \tilde{z}_t + e_t^*, \qquad t = p+1, p+2, \dots, T, \tag{3.9}$$

where $\Delta \tilde{x}_t = \Delta y_t - \tilde{\mu}_1$, $\tilde{u}_t = \tilde{\beta}'(y_t - \tilde{\mu}_0)$, $\tilde{v}_t = \tilde{\beta}_\perp'(y_t - \tilde{\mu}_0)$, $\tilde{z}_t' = [\Delta \tilde{x}_{t-1}', \dots, \Delta \tilde{x}_{t-p+1}']$ and $\Gamma_* = [\Gamma_{*1} : \dots : \Gamma_{*p-1}]$. Different tests will be obtained by using Wald, LM or LR formulations as well as using different estimators for the parameters μ_0 and μ_1 and/or α, β and Ω. For illustrative purposes we use the statistic

$$LM^*(r_0) = \operatorname{tr}\left\{ \tilde{\rho}_* \tilde{M}_{vv \cdot q} \tilde{\rho}_*' (\tilde{\alpha}_\perp' \tilde{\Omega} \tilde{\alpha}_\perp)^{-1} \right\}, \tag{3.10}$$

where $\tilde{\rho}_*$ is the LS estimator of ρ_* from (3.9) and

$$\tilde{M}_{vv \cdot q} = \left[\sum_{t=p+1}^{T} \tilde{v}_{t-1} \tilde{v}_{t-1}' - \sum_{t=p+1}^{T} \tilde{v}_{t-1} \tilde{q}_t' \left(\sum_{t=p+1}^{T} \tilde{q}_t \tilde{q}_t' \right)^{-1} \sum_{t=p+1}^{T} \tilde{q}_t \tilde{v}_{t-1}' \right] \tag{3.11}$$

with $\tilde{q}_t = [\tilde{u}_{t-1}', \tilde{z}_t']'$. If the estimators of α, β and Ω are based on the usual RR regression of (2.2) which takes the restrictions specified under the null hypothesis into account, this test statistic may be thought of as an LM type test statistic.

An LR type test may be obtained from (3.6) in the usual way by replacing μ_0 and μ_1 by estimators and considering the ordered generalized eigenvalues $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_n$ obtained as solutions of $\det(\tilde{\Pi} \tilde{M}_T \tilde{\Pi}' - \lambda \tilde{\Omega}) = 0$, where $\tilde{\Pi}$ is the unrestricted LS estimator of Π from

$$\Delta y_t - \tilde{\mu}_1 = \Pi(y_{t-1} - \tilde{\mu}_0) + \sum_{j=1}^{p-1} \Gamma_j (\Delta y_{t-j} - \tilde{\mu}_1) + \tilde{\varepsilon}_t, \qquad t = p+1, p+2, \dots,$$

$\tilde{\Omega}$ is the corresponding residual covariance estimator and

$$\tilde{M}_T = \frac{1}{T-p} \left[\sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)(y_{t-1} - \tilde{\mu}_0)' - \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0) \tilde{z}_t' \left(\sum_{t=p+1}^{T} \tilde{z}_t \tilde{z}_t' \right)^{-1} \sum_{t=p+1}^{T} \tilde{z}_t (y_{t-1} - \tilde{\mu}_0)' \right].$$

Using this notation the 'LR statistic' becomes

$$LR^{\sim}_{trace}(r_0) = (T-p) \sum_{j=r_0+1}^{n} \log(1+\tilde{\lambda}_j). \tag{3.12}$$

The limiting distribution of the test statistics LM^* and $LR^{\sim}_{trace}$ under the null hypothesis is given in the following theorem which is proven in the appendix.

Theorem 1. If $H(r_0)$ in (2.11) is true and the assumptions of the previous section hold, then

$$LR^{\sim}_{trace}(r_0), LM^*(r_0) \xrightarrow{d} \operatorname{tr}\left\{ \left(\int_0^1 \bar{\mathbf{G}}(s)\mathrm{d}\mathbf{W}(s)'\right)' \left(\int_0^1 \mathbf{G}(s)\mathbf{G}(s)'\mathrm{d}s\right)^{-1} \left(\int_0^1 \bar{\mathbf{G}}(s)\mathrm{d}\mathbf{W}(s)'\right)\right\}, \tag{3.13}$$

where, as before, $\mathbf{W}(s)$ is $(n-r_0)$-dimensional standard Brownian motion, $\mathbf{G}(s) = [s, W_1(s), \ldots, W_{n-r_0-1}(s)]'$ and $\bar{\mathbf{G}}(s) = \mathbf{G}(s) - \int_0^1 \mathbf{G}(u)\mathrm{d}u$. □

Thus the limiting null distribution of the test statistics is free of unknown nuisance parameters so that the percentiles can be readily found by simulation. Following Johansen (1995) we have done so and present the results in Table 1. The difference between this limiting distribution and the one obtained for the corresponding LR test given in (3.5) is that in the latter the process $\mathbf{G}(s)$ on the right hand side of (3.13) is replaced by its demeaned version $\bar{\mathbf{G}}(s)$. This demeaned version appears in the first and third integrals on the right hand side of (3.13) where it results as a consequence of replacing the unknown parameter vector μ_1 in the terms $\Delta y_{t-j} - \mu_1$ $(j = 0, \ldots, p-1)$ by the estimator $\tilde{\mu}_1$. In the next section the local power properties of our new tests are explored.

4 Local Power Analysis

4.1 Assumptions and Notation

In this section we assume for simplicity that y_t and hence x_t is a Gaussian VAR(1) process, that is, $\varepsilon_t \sim NID(0, \Omega)$. As before, the initial value $x_0 = 0$. We continue to assume that α and β have column dimension r_0 as specified in the null hypothesis. We consider local alternatives of the form

$$H_T(r_0) : \Pi = \Pi_T = \alpha\beta' + T^{-1}\alpha_1\beta_1', \tag{4.1}$$

Table 1. Percentage Points of the Distribution of

$$\mathrm{tr}\left\{\left(\int_0^1 \bar{\mathbf{G}}(s)\mathrm{d}\mathbf{W}(s)'\right)'\left(\int_0^1 \mathbf{G}(s)\mathbf{G}(s)'\mathrm{d}s\right)^{-1}\left(\int_0^1 \bar{\mathbf{G}}(s)\mathrm{d}\mathbf{W}(s)'\right)\right\}.$$

Dimension	$n-r_0$	2	3	4	5
	90%	8.03	18.19	31.35	48.06
Percentage point	95%	9.79	20.66	33.64	52.06
	99%	14.02	26.20	38.25	56.96

where α and β are fixed $(n \times r_0)$ matrices of rank r_0 and α_1 and β_1 are fixed $(n \times (r-r_0))$ matrices of rank $r-r_0$ and such that the matrices $[\alpha : \alpha_1]$ and $[\beta : \beta_1]$ have full column rank r. We also assume that the eigenvalues of $I_{r_0} + \beta'\alpha$ and $I_r + [\beta : \beta_1]'[\alpha : \alpha_1]$ are less than one in absolute value so that the assumptions from Johansen (1995) and Rahbek (1994) are satisfied.

If we want to ensure that the basic model of the DGP is of the form (2.1) with an intercept and without a trend term in the levels representation even under the alternative we have to assume that $\Pi\mu_1 = 0$ and, hence,

$$\beta_1'\mu_1 = 0. \tag{4.2}$$

Unless otherwise stated we will henceforth assume that our local alternatives are given by (4.1) and (4.2) jointly. Thereby our set-up differs from that of Rahbek (1994). The differences will be discussed in more detail later. Note that the fact that the $(n \times r)$ matrix $[\beta : \beta_\perp]$ is orthogonal to μ_1 also means that r has to be less than n.

We also assume that suitable estimators of the parameters α, β and Ω are available which satisfy (2.16) – (2.18) even under the local alternatives. For the usual RR estimators based on (2.2) these properties follow from arguments similar to those used by Johansen (1995, Chapter 13) and Paruolo (1997).

In addition we assume that estimators $\tilde{\mu}_0$ and $\tilde{\mu}_1$ of μ_0 and μ_1, respectively, are used which satisfy (2.12) – (2.14) under the local alternatives and, moreover,

$$T^{1/2}\beta_\perp'(\tilde{\mu}_1 - \mu_1) \xrightarrow{d} \mathbf{K}(1), \tag{4.3}$$

where $\mathbf{K}(u)$ is the Ornstein-Uhlenbeck process defined by the integral equation

$$\mathbf{K}(u) = \alpha_\perp' \mathbf{B}(u) + \alpha_\perp' \alpha_1 \beta_1' \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} \int_0^u \mathbf{K}(s) \mathrm{d}s \quad (0 \leq u \leq 1) \tag{4.4}$$

or, equivalently, by the stochastic differential equation

$$\mathrm{d}\mathbf{K}(u) = \alpha_\perp' \mathrm{d}\mathbf{B}(u) + \alpha_\perp' \alpha_1 \beta_1' \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} \mathbf{K}(u) \mathrm{d}u \quad (0 \leq u \leq 1).$$

Again estimators $\tilde{\mu}_0$ and $\tilde{\mu}_1$ with suitable properties under local alternatives may, e.g., be found in L&S. Note that we may assume that $\alpha_\perp = \beta_\perp + c\beta$ for a suitable matrix c, so that $\beta_\perp' \beta_\perp = \beta_\perp' \alpha_\perp$. Hence, if the null hypothesis holds and therefore $\alpha_1 \beta_1' = 0$, then using the definition of C and the fact that $\Gamma_i = 0$ $(i = 1, \ldots, p-1)$ for the presently considered VAR(1) case, it can be shown that $\mathbf{K}(1) = \alpha_\perp' \mathbf{B}(1) = \beta_\perp' C \mathbf{B}(1)$ and thus, (4.3) reduces to (2.15).

In the following we will also use the $(n - r_0)$-dimensional Ornstein-Uhlenbeck process defined by

$$\mathbf{N}(u) = \mathbf{W}(u) + \bar{a}\bar{b}' \int_0^u \mathbf{N}(s) \mathrm{d}s \quad (0 \leq u \leq 1), \tag{4.5}$$

where $\mathbf{W}(s)$ is $(n - r_0)$-dimensional standard Brownian motion, as before,

$$\bar{a} = (g' \alpha_\perp' \Omega \alpha_\perp g)^{-1/2} g' \alpha_\perp' \alpha_1 \qquad \text{and} \qquad \bar{b}' = \beta_1' [0 : \gamma] (g' \alpha_\perp' \Omega \alpha_\perp g)^{1/2}$$

with g' being an $((n-r_0) \times (n-r_0))$ matrix which is the inverse of $\beta_\perp' [\mu_1 : \gamma]$. Furthermore we need the process $\mathbf{Z}(s) = [s, \mathbf{N}_2(s)']'$ and its demeaned counterpart $\bar{\mathbf{Z}}(s) = \mathbf{Z}(s) - \int_0^1 \mathbf{Z}(u) \mathrm{d}u$. Here $\mathbf{N}_2(s)$ consists of the last $n - r_0 - 1$ components of $\mathbf{N}(s)$. Now we are ready to consider the local power properties of the LR and LM type tests. We will investigate them in turn in the following and then perform a comparison.

4.2 LR Tests

We first give the asymptotic distribution of the LR test under the local alternatives in (4.1)/(4.2).

Theorem 2. Under the conditions spelled out in Section 4.1,

$$LR_{trace}(r_0) \xrightarrow{d} \operatorname{tr} \left\{ \left(\int_0^1 \bar{\mathbf{Z}}(s) \mathrm{d}\mathbf{N}(s)' \right)' \left(\int_0^1 \bar{\mathbf{Z}}(s) \bar{\mathbf{Z}}(s)' \mathrm{d}s \right)^{-1} \left(\int_0^1 \bar{\mathbf{Z}}(s) \mathrm{d}\mathbf{N}(s)' \right) \right\},$$

where the stochastic integrals may be defined by using (4.5) so that $\int_0^1 \bar{\mathbf{Z}}(s) \mathrm{d}\mathbf{N}(s)' = \int_0^1 \bar{\mathbf{Z}}(s) \mathrm{d}\mathbf{W}(s)' + \int_0^1 \bar{\mathbf{Z}}(s) \mathrm{d}s \bar{b} \bar{a}'$. □

As in Saikkonen & Lütkepohl (1998) (henceforth S&L), the limiting nonnull distribution is obtained from the limiting null distribution of $LR_{trace}(r_0)$ by replacing the involved Brownian motion by an Ornstein-Uhlenbeck process. The definition of this Ornstein-Uhlenbeck process is also similar to those in the cases considered in S&L although there are some differences in the definitions of the parameters $\bar{a}$ and $\bar{b}$. For instance, now these parameters depend on the matrix g which did not appear in the previous counterparts of $\bar{a}$ and $\bar{b}$. The zero matrix in $\bar{b}$ is due to the assumption $\beta_1'\mu_1 = 0$. Note, however, that we necessarily have $\beta_1'\gamma \neq 0$ (because $\beta_1 \in \text{span}(\beta_\perp) = \text{span}([\mu_1 : \gamma])$) so that $\bar{b} = 0$ is not possible.

A proof of Theorem 2 is given in the Appendix. Here we will just provide the main ideas. Before we present them it may be useful to discuss the difference of our Theorem 2 to results of Rahbek (1994) who also considered the local power of the LR tests.

Rahbek (1994) obtained different results because his assumptions were different. He considered two types of local alternatives. The first type is given by $\Pi = \Pi_T = \alpha\beta' + T^{-3/2}\alpha_1\beta_1'$ which is an order of magnitude smaller than we have used. However, from Rahbek's (1994) Theorem 3.1 it can be seen that the limiting distribution obtained in this case reduces to the limiting null distribution if our condition (4.2), $\beta_1'\mu_1 = 0$, (or $\beta_1'C\mu = 0$ in Rahbek's notation) holds. We have employed assumption (4.2) because we wish to consider the case where there is just an intercept in the levels VAR form as in (2.1) and not a linear time trend. This implies that a linear time trend is a priori excluded from the cointegrating relations. To see this more clearly, consider the model

$$y_t = \mu_0 + \mu_1 t + x_t \quad \text{with} \quad x_t = A_1 x_{t-1} + \varepsilon_t$$

which, under the null hypothesis, can be written alternatively as

$$\begin{aligned} \Delta y_t &= \nu + \Pi(y_{t-1} - \mu_1(t-1)) + \varepsilon_t \\ &= \nu + \alpha(\beta' y_{t-1} - \beta'\mu_1(t-1)) + \varepsilon_t, \end{aligned} \tag{4.6}$$

where $\nu = -\Pi\mu_0 + \mu_1$. If we use $\Pi = \alpha\beta' + T^{-3/2}\alpha_1\beta_1'$ in (4.6) it can be seen that Rahbek's (1994) better power result is solely obtained by testing the null hypothesis that a linear time trend should be included in the cointegration relations which is an irrelevant case from the point of view of our assumptions because we have excluded this possibility a priori for the reasons discussed earlier.

The other local alternatives considered by Rahbek (1994) are given by (4.1) combined with $\beta_1'\mu_1 = T^{-1/2}\delta$ $(0 \leq \|\delta\| < \infty)$ (see (3.2) of Rahbek (1994) and note that his $C\mu$ is our μ_1). When $\delta = 0$ we get our local alternatives (see (4.2)). If $\delta \neq 0$ power gains are achieved but, as explained earlier, they are achieved by testing an 'irrelevant' inclusion of a time trend in the cointegrating relations. Since we wish to exclude this feature we only consider the power of the LR test obtained from Theorem 2.

The proof of Theorem 2 follows from a general result given in S&L. The $LR_{trace}(r_0)$ statistic considered in Theorem 2 is identical to the statistic $LR^{i0}(r_0)$ of S&L. Thus, it is based on a RR regression of $\overline{\Delta y}_t$ on $\bar{y}_{t-1}$ where the bar signifies ordinary mean correction. Further, as explained in S&L, an asymptotically equivalent test is obtained by testing the null hypothesis $\rho_T^* = 0$ by conventional likelihood based methods in the auxiliary regression model

$$\tilde{\alpha}_\perp' \overline{\Delta y}_t = \kappa_T^* \tilde{\bar{u}}_{t-1} + \rho_T^* \tilde{\bar{v}}_{t-1} + \tilde{\bar{e}}_t^* \tag{4.7}$$

where $\tilde{\bar{u}}_t = \tilde{\beta}'\bar{y}_t$, $\tilde{\bar{v}}_t = \tilde{\beta}_\perp'\bar{y}_t$ and

$$\tilde{\bar{e}}_t^* = \tilde{\alpha}_\perp'\bar{\varepsilon}_t - \tilde{\alpha}_\perp'\kappa_T(\tilde{\beta} - \beta)'\bar{y}_{t-1} - \tilde{\alpha}_\perp'\rho_T(\tilde{\beta}_\perp - \beta_\perp)'\bar{y}_{t-1}. \tag{4.8}$$

Since any estimator of Ω with the property (2.18) can be used here we may define the test statistic

$$LR^*(r_0) = \operatorname{tr}\left\{\bar{\rho}_T^* \bar{M}_{vv\cdot u} \bar{\rho}_T^{*\prime} (\tilde{\alpha}_\perp' \tilde{\Omega} \tilde{\alpha}_\perp)^{-1}\right\},$$

where $\bar{\rho}_T^*$ is the LS estimator of ρ_T^* in (4.7) and

$$\bar{M}_{vv\cdot u} = \left[\sum_{t=2}^{T} \tilde{\bar{v}}_{t-1}\tilde{\bar{v}}_{t-1}' - \sum_{t=2}^{T} \tilde{\bar{v}}_{t-1}\tilde{\bar{u}}_{t-1}' \left(\sum_{t=2}^{T} \tilde{\bar{u}}_{t-1}\tilde{\bar{u}}_{t-1}'\right)^{-1} \sum_{t=2}^{T} \tilde{\bar{u}}_{t-1}\tilde{\bar{v}}_{t-1}'\right].$$

The test statistic $LR^*(r_0)$ is asymptotically equivalent to $LR_{trace}(r_0)$ and, hence, it suffices to prove Theorem 2 for $LR^*(r_0)$ which is done in the Appendix. We will now turn to the local power of the tests based on trend adjusted data.

4.3 Tests Based on Prior Trend Adjustment

The asymptotic distribution of the tests based on the trend adjusted data under the local alternatives (4.1)/(4.2) is given in the next theorem.

Theorem 3. Under the conditions of Section 4.1,

$$LR^{\sim}_{trace}(r_0), LM^*(r_0) \xrightarrow{d} \operatorname{tr}\left\{ \left(\int_0^1 \bar{\mathbf{Z}}(s)\mathrm{d}\mathbf{N}(s)' \right)' \left(\int_0^1 \mathbf{Z}(s)\mathbf{Z}(s)'\mathrm{d}s \right)^{-1} \left(\int_0^1 \bar{\mathbf{Z}}(s)\mathrm{d}\mathbf{N}(s)' \right) \right\}.$$

□

Comparing this limiting distribution to that of $LR_{trace}(r_0)$ given in Theorem 2 shows that the nonnull distributions differ in the same way as the limiting distributions under the null hypothesis. The proof of Theorem 3 is also given in the Appendix. It is again based on Theorem 1 of S&L. In this case it is applied to the model

$$\tilde{\alpha}'_{\perp}\Delta\tilde{x}_t = \kappa^*_T\tilde{u}_{t-1} + \rho^*_T\tilde{v}_{t-1} + \tilde{e}^*_t \tag{4.9}$$

where $\Delta\tilde{x}_t = \Delta y_t - \tilde{\mu}_1$, $\tilde{u}_t = \tilde{\beta}'(y_t - \tilde{\mu}_0)$, $\tilde{v}_t = \tilde{\beta}'_{\perp}(y_t - \tilde{\mu}_0)$, $\kappa^*_T = \tilde{\alpha}'_{\perp}\kappa_T$, $\rho^*_T = \tilde{\alpha}'_{\perp}\rho_T$, and $e^*_t = \tilde{\alpha}'_{\perp}e_t$ with

$$\kappa_T = \alpha + T^{-1}\alpha_1\beta'_1\beta(\beta'\beta)^{-1}, \qquad \rho_T = T^{-1}\alpha_1\beta'_1\beta_{\perp}(\alpha'_{\perp}\beta_{\perp})^{-1},$$

$$e_t = \tilde{e}_t - \kappa_T(\tilde{\beta} - \beta)'(y_{t-1} - \tilde{\mu}_0) - \rho_T(\tilde{\beta}_{\perp} - \beta_{\perp})'(y_{t-1} - \tilde{\mu}_0) \tag{4.10}$$

and

$$\tilde{e}_t = \varepsilon_t + \kappa_T\beta'(\tilde{\mu}_0 - \mu_0) - (\tilde{\mu}_1 - \mu_1) + \rho_T\beta'_{\perp}(\tilde{\mu}_0 - \mu_0).$$

4.4 Local Power Comparison

Since the asymptotic distributions of the LR tests and the tests based on trend adjusted data differ we have simulated the resulting local power in a similar way as in S&L in order to compare the local power functions. In the following, we consider the case where $r - r_0 = 1$, so that $\bar{a}$ and $\bar{b}$ are $(n \times 1)$ vectors and simulate the discrete time counterpart of the $(n - r_0)$-dimensional Ornstein-Uhlenbeck process $\mathbf{N}(s)$ as

$$\Delta\mathbf{N}_t = \frac{1}{T}ab'\mathbf{N}_{t-1} + \epsilon_t, \qquad t = 1, \ldots, T = 1000, \tag{4.11}$$

with $\epsilon_t \sim NID(0, I_{n-r_0})$, $\mathbf{N}_0 = 0$,

$$b' = \begin{cases} (0,1) & \text{for } n - r_0 = 2 \\ (0,1,0) & \text{for } n - r_0 = 3 \end{cases}$$

and

$$a' = \begin{cases} (a_1, a_2) & \text{for } n - r_0 = 2 \\ (a_1, a_2, 0) & \text{for } n - r_0 = 3 \end{cases}.$$

These choices are motivated as follows. We assume that parameters are chosen such that $g'\alpha'_\perp \Omega \alpha_\perp g = I_{n-r_0}$. This is not very restrictive because suitable normalizations of the parameters may be selected. Then $\bar{b}' = \beta_1'[0 : \gamma]$, where γ is essentially equal to β_1 for $n - r_0 = 2$. Hence, $b' = (0, 1)$ may be viewed as the relevant part of $\bar{b}'$. If $n - r_0 = 3$, one of the columns of γ may be set equal to β_1 while the other one may be made orthogonal to β_1. Assuming that the first column of γ equals β_1 justifies $b' = (0, 1, 0)$ for $n - r_0 = 3$. The vector a' is simply given a fully flexible form. We choose $a_1 = \sqrt{(1 - d^2)l^2}$ and $a_2 = -\sqrt{d^2 l^2}$ and report local power results as a function of d and l. Note that $l^2 = a'ab'b$ and $d^2 = (b'a)/(a'ab'b)$. Consequently, $l^2 = 0$ if and only if the null hypothesis holds so that $l = \sqrt{l^2}$ may be thought of as the distance of the local alternative from the null hypothesis. Moreover, $0 < d^2 \leq 1$ and $d = \sqrt{d^2}$ may be interpreted as the direction of the local alternative. It can be shown that values of d close to zero correspond to processes close to being $I(2)$. The specific values used for l and d will be given later.

From the $\mathbf{N}_t$ we get $\mathbf{Z}'_t = [t, \mathbf{N}'_{2,t}]$ and $\bar{\mathbf{Z}}_t = \mathbf{Z}_t - T^{-1}\sum_{s=1}^{T-1} \mathbf{Z}_s$, where $\mathbf{N}_{2,t}$ consists of the last $n - r_0 - 1$ components of $\mathbf{N}_t$. Then we compute

$$G_T = T^{-2}\sum_{t=1}^{T} \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1}, \qquad \bar{G}_T = T^{-2}\sum_{t=1}^{T} \bar{\mathbf{Z}}_{t-1}\bar{\mathbf{Z}}'_{t-1} \qquad \text{and} \qquad S_T = T^{-1}\sum_{t=1}^{T} \bar{\mathbf{Z}}_{t-1}\Delta\mathbf{N}'_t.$$

These quantities converge weakly to $\int_0^1 \mathbf{Z}(s)\mathbf{Z}(s)'\mathrm{d}s$, $\int_0^1 \bar{\mathbf{Z}}(s)\bar{\mathbf{Z}}(s)'\mathrm{d}s$ and $\int_0^1 \bar{\mathbf{Z}}(s)\mathrm{d}\mathbf{N}(s)'$, respectively. Hence we obtain the desired asymptotic distributions of $LR_{trace}(r_0)$ and $LM^*(r_0)$ as $\mathrm{tr}(S'_T\bar{G}_T^{-1}S_T)$ and $\mathrm{tr}(S'_T G_T^{-1} S_T)$, respectively. The resulting rejection frequencies for different values of l and d are plotted in Figure 1 together with the local power functions of Johansen's (1992, 1994) LR tests which allow for a general linear trend. Hence, they do not impose the restriction $\beta'\mu_1 = 0$. The local power of these tests is also considered by S&L and is given here for comparison purposes. To distinguish the tests from the previously considered LR tests we denote them by LR^+_{trace}.

Note that the present local power study is similar but not identical to the one reported by S&L for some other tests. A major difference is that now the dependence of the power function on the parameters of the process is more complicated due to the fact that $\bar{b}'$ involves the matrices $[0 : \gamma]$ and g and therefore does not reduce to the simple form used in S&L

even for $r - r_0 = 1$. Moreover, now the local power is given as a function of the quantities l and d. The advantage is that thereby we also get a measure of the distance from the $I(2)$ case. Therefore the results of the present study are not immediately comparable to the local power of other tests for the cointegrating rank that were, e.g., considered by S&L. We have computed the local power of the LR^{+}_{trace} tests analogously to S&L.

It is obvious from the figure that prior knowledge regarding the trend being not in the cointegration relations is helpful for improving the local power of the tests. The LR^{+}_{trace} test which does not use this information is overall inferior to LM^{*} and LR_{trace} which use the information. Comparing the local power of the latter two tests it is apparent that LM^{*} is in general considerably more powerful than LR_{trace}. In parts of the parameter space (for d close to 1 and moderate values of l) the former test has about twice the local power of the latter test. The power gains are less impressive for processes close to being $I(2)$ (d close to zero). Thus, especially for processes which are not close to $I(2)$ processes, substantial power gains are possible by using our new tests. It is also seen in the figure that the local power of all tests tends to decline with increasing number of common trends under the alternative. In other words, the local power tends to be lower for $n - r_0 = 3$ than for $n - r_0 = 2$. This behaviour was also observed for other tests for the cointegrating rank (see S&L).

It should be understood, however, that local power properties are informative about the performance of the tests in large samples when alternatives close to the null hypothesis are of interest. In small samples the situation may be different. Therefore we present some small sample simulations in the following section.

5 Small Sample Comparison

A limited Monte Carlo experiment was performed to study the small sample properties of our tests and to compare them to other tests for the cointegrating rank. A three-dimensional VAR(1) DGP from Toda (1995) of the form

$$y_t = \begin{bmatrix} 0 \\ 0 \\ \delta \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim NID\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta_1 & \theta_2 \\ \theta_1 & 1 & 0 \\ \theta_2 & 0 & 1 \end{bmatrix} \right) \tag{5.1}$$

is used to compare the tests. Toda shows that this type of process may be regarded as a "canonical form" for investigating the properties of LR tests for the cointegrating rank. Other VAR(1) processes of interest in practice can be obtained by linear transformations of y_t which leave the tests invariant. If $\psi_1 = \psi_2 = 1$ the cointegrating rank is $r = 0$ and the process consists of three nonstationary components. The third component has a deterministic linear trend if $\delta \neq 0$ and there is no linear trend if $\delta = 0$. For $\theta_1 = \theta_2 = 0$ the three components are independent whereas they are instantaneously correlated for $\theta_i \neq 0$. If $\psi_2 = 1$ and $|\psi_1| < 1$ the cointegrating rank is $r = 1$. Again there will be a linear trend in this case if $\delta \neq 0$. If both ψ_1 and ψ_2 are less than 1 in absolute value the process has cointegrating rank 2 ($r = 2$).

We have generated 1000 samples of sizes $T = 100$ and 200 plus 50 presample values starting with an initial value of zero. Since we delete the first presample values and thereby use nonzero initial values for the samples in the simulations this set-up differs slightly from the theoretical framework where zero initial values are assumed for the sample. Thereby we can check the robustness of our results with respect to a deviation from the zero initial value assumption. In Tables 2 and 3 we give selected rejection frequencies for $T = 100$, $H(r_0 = 0)$ and $H(r_0 = 1)$ of the tests LR_{trace}, $LR^{\sim}_{trace}$, LM^* and three tests which allow for deterministic linear trends without using the restriction $\beta'\mu_1 = 0$, that is, they do not use the fact that there is no trend in the cointegrating relations. As mentioned earlier, LR^{+}_{trace} is a test proposed by Johansen (1992, 1994) whereas LR^{PC}_{trace} was discussed by Perron & Campbell (1993) and LM^{GLS}_* is due to Saikkonen & Lütkepohl (1997). The estimation of the trend parameters for the tests $LR^{\sim}_{trace}$ and LM^* is also based on the GLS procedure described in the latter article.

The rejection frequencies in the tables are obtained by using asymptotic critical values for a test level of 5%. We have not computed rejection frequencies corrected for the actual small sample sizes because these will also not be available in practice. Hence, a test is useful for applied work only if it respects roughly the nominal significance level. Otherwise the Type I error cannot be controlled in practice and a test which does not bound this error is of limited value even if it has good power properties.

The results for all test statistics are based on the same generated time series. Hence the entries for a given DGP in a single column of the panels in the tables are not independent but can be compared directly. Still it may be worth recalling that for a true rejection probability P the standard error of an estimator based on 1000 replications of the experiment is $s_P = \sqrt{P(1-P)/1000}$. For example, $s_{0.05} = s_{0.95} = 0.007$. Note also that the tests are not performed sequentially. Thus, the test results for $r_0 = 1$ are not conditioned on the outcome of the test of $r_0 = 0$.

In the first, fourth, sixth and last columns of Table 2 it is seen that the sizes of the tests differ quite a bit. The LM type tests LM^* and LM_*^{GLS} tend to be conservative. As a consequence their power is lower than that of the corresponding LR type tests in the same group. Generally it is obvious in both tables that the tests which use the restriction $\beta'\mu_1 = 0$ (i.e., LR_{trace}, $LR_{trace}^{\sim}$ and LM^*) are considerably more powerful than the other tests for many parameter values. Thus, generally it pays to use the restriction if it is valid. Among the first three tests, $LR_{trace}^{\sim}$, which is based on prior trend adjustment, is overall better than the LR_{trace} test, in particular for alternatives close to the null hypothesis if the residual correlation is small or moderate (see, e.g., $\psi_1 = 0.9$, $\theta_1 = \theta_2 = 0$ and $\theta_1 = \theta_2 = 0.4$ in Tables 2 and 3). Clearly, among the tests based on prior trend adjustment the LR version performs better than the LM version. Hence, based on these results, the use of the $LR_{trace}^{\sim}$ test can be recommended. In none of the cases it is markedly inferior to LR_{trace} and sometimes it is much better. Similar results were also obtained for sample sizes $T = 200$.

6 Conclusions

We have considered tests for the cointegrating rank of a VAR process when the DGP has a deterministic linear trend component, whereas there is no such trend in the cointegrating relations. In this case the DGP has a VAR representation with an intercept and no additional linear trend term. This appears to be the most popular case in applied work. We have investigated the asymptotic and small sample properties of Johansen's LR tests proposed for this situation. Moreover, we have discussed alternative tests which are based on prior trend adjustment of the data. For this purpose, the parameters of the mean and trend terms are estimated under the null hypothesis by means of a GLS procedure. The estimated mean and trend terms are subtracted from the observations and LR and LM type tests are then applied to the adjusted data.

A comparison of the local power of the different types of tests shows that the tests based on prior trend adjustment have considerably more local power than standard LR tests. Moreover, taking into account the fact that there is no trend in the cointegrating relations results in more powerful tests than allowing for a general linear trend. The latter result is also confirmed in a limited simulation study. Furthermore, in that study it is found that prior trend adjustment in the way proposed here is helpful for improving on power in particular when the actual DGP is close to one which satisfies the null hypothesis. LM type versions of the tests are somewhat conservative in small samples and, hence, have smaller power than their LR counterparts. Thus, overall the use of LR type tests based on prior trend adjustment is recommended in practice. As a final note we emphasize, however, that the tests we have considered here are not suitable for testing a null hypothesis against a stationary alternative, that is, for an n-dimensional system we cannot test the null hypothesis that the cointegrating rank is $n-1$.

Appendix

In the following proofs we use some well-known limit theorems for stationary and integrated variables which may, e.g., be found in Sims, Stock & Watson (1990). We shall assume appropriate initial values for the process x_t so that $\beta' x_t$ and Δx_t are (jointly) stationary. This assumption occasionally simplifies the notation without affecting the results. For convenience and without loss of generality we also assume that the estimators $\tilde{\beta}$ and $\tilde{\beta}_\perp$ satisfy the (infeasible) normalizations $\beta'\tilde{\beta} = I_r$ and $\beta_\perp'\tilde{\beta}_\perp = I_{n-r}$, respectively. Notice that this also means that β and $\beta_\perp$ are normalized similarly so that $\beta'\beta = I_r$ and $\beta_\perp'\beta_\perp = I_{n-r}$.

A.1 Proof of Theorem 1

First notice that, under the assumptions of Theorem 1, $r = r_0$, that is, r is the true cointegrating rank which at the same time is the rank specified under the null hypothesis. To prove Theorem 1 first note that it is well-known that $LM^*(r_0)$ and $LR^{\sim}_{trace}(r_0)$ are asymptotically equivalent. Hence, it suffices to prove Theorem 1 for the former statistic. The proof is essentially based on deriving the limiting distribution of the estimator

$$\hat{\rho}_* = \sum_{t=p+1}^{T} e_t^* \left[\tilde{v}'_{t-1} - \tilde{q}'_t \left(\sum_{t=p+1}^{T} \tilde{q}_t \tilde{q}'_t \right)^{-1} \sum_{t=p+1}^{T} \tilde{q}_t \tilde{v}'_{t-1} \right] \tilde{M}_{vv \cdot q}^{-1}, \qquad (A.1)$$

where the equality follows from standard LS theory. Define the nonsingular $((n-r) \times (n-r))$ matrix

$$\underset{\sim}{D}_T = \beta'_\perp D_T = \beta'_\perp [\mu_1 : \gamma] \begin{bmatrix} T^{3/2} & 0 \\ 0 & T \end{bmatrix}.$$

The inverse of $\underset{\sim}{D}_T$ can be written as

$$\underset{\sim}{D}_T^{-1} = \begin{bmatrix} T^{-3/2} h' \\ T^{-1} H' \end{bmatrix},$$

where h is $((n-r) \times 1)$ and such that $h' \beta_\perp \mu_1 = 1$ and H is $((n-r) \times (n-r-1))$ such that $H' \beta_\perp \mu_1 = 0$. The matrix $\underset{\sim}{D}_T^{-1}$ will in particular be used with the process v_t defined in (3.7) and to this end it is useful to note the identity

$$\underset{\sim}{D}_T^{-1} v_t = \begin{bmatrix} T^{-3/2} t \\ 0 \end{bmatrix} + \begin{bmatrix} T^{-3/2} h' \beta'_\perp x_t \\ T^{-1} H' \beta'_\perp x_t \end{bmatrix}. \qquad (A.2)$$

To derive the asymptotic properties of the estimator $\hat{\rho}_*$, we first show that

$$\underset{\sim}{D}_T^{-1} \tilde{M}_{vv \cdot q} \underset{\sim}{D}_T'^{-1} = \underset{\sim}{D}_T^{-1} \sum_{t=p+1}^{T} v_{t-1} v'_{t-1} \underset{\sim}{D}_T'^{-1} + o_p(1). \qquad (A.3)$$

Define $q_t = [u'_{t-1}, z'_t]'$, where $z_t = [\Delta x'_{t-1}, \ldots, \Delta x'_{t-p+1}]$ and notice that q_t is a zero mean stationary process with positive definite covariance matrix. Thus, from the definition of $\tilde{M}_{vv \cdot q}$ it can be seen that $(A.3)$ follows from the results given in the next lemma.

Lemma A.1

$$T^{-1} \sum_{t=p+1}^{T} \tilde{q}_t \tilde{q}'_t = E(q_t q'_t) + o_p(1), \qquad (A.4)$$

$$T^{-1/2} \underset{\sim}{D}_T^{-1} \sum_{t=p+1}^{T} \tilde{v}_{t-1} \tilde{q}'_t = o_p(1) \qquad (A.5)$$

and

$$\underset{\sim}{D}_T^{-1} \sum_{t=p+1}^{T} \tilde{v}_{t-1} \tilde{v}'_{t-1} \underset{\sim}{D}_T'^{-1} = \underset{\sim}{D}_T^{-1} \sum_{t=p+1}^{T} v_{t-1} v'_{t-1} \underset{\sim}{D}_T'^{-1} + o_p(1). \qquad (A.6)$$

Proof: To prove $(A.4)$, note first that

$$\Delta\tilde{x}_t = \Delta x_t - (\tilde{\mu}_1 - \mu_1), \tag{A.7}$$

where here as well as in subsequent similar equalities, it is assumed that $t \geq p+1$. From this and (2.7) – (2.10) it follows, for $i, j = 1, \ldots, p-1$, that

$$T^{-1} \sum_{t=p+1}^{T} \Delta\tilde{x}_{t-i}\Delta\tilde{x}'_{t-j} = T^{-1} \sum_{t=p+1}^{T} \Delta x_{t-i}\Delta x'_{t-j} + o_p(1). \tag{A.8}$$

Next note that, since $\beta'\mu_1 = 0$ by assumption, one obtains from (2.16),

$$\tilde{\beta}'\mu_1 = (\tilde{\beta} - \beta)'\mu_1 = O_p(T^{-3/2}) \tag{A.9}$$

while (2.12), (2.13) and (2.16) imply

$$\tilde{\beta}'(\tilde{\mu}_0 - \mu_0) = O_p(T^{-1/2}). \tag{A.10}$$

Since

$$\tilde{u}_t = \tilde{\beta}'x_t - \tilde{\beta}'(\tilde{\mu}_0 - \mu_0) + \tilde{\beta}'\mu_1 t, \tag{A.11}$$

these results and (2.16) readily yield

$$T^{-1} \sum_{t=p+1}^{T} \tilde{u}_{t-1}\tilde{u}'_{t-1} = T^{-1} \sum_{t=p+1}^{T} u_{t-1}u'_{t-1} + o_p(1), \tag{A.12}$$

where we have also used the fact that $u_t = \beta' x_t$. Arguments used for $(A.8)$ and $(A.12)$ also readily show that, for $i = 1, \ldots, p-1$,

$$T^{-1} \sum_{t=p+1}^{T} \Delta\tilde{x}_{t-i}\tilde{u}'_{t-1} = T^{-1} \sum_{t=p+1}^{T} \Delta x_{t-i}u'_{t-1} + o_p(1). \tag{A.13}$$

Combining $(A.8)$, $(A.12)$ and $(A.13)$ proves $(A.4)$.

Next consider $(A.5)$ and recall that $\tilde{v}_t = \tilde{\beta}'_\perp(y_t - \tilde{\mu}_0)$. Thus, we have for $j = 1, \ldots, p-1$,

$$\begin{aligned} T^{-1/2}\mathrm{D}_T^{-1} \sum_{t=p+1}^{T} \tilde{v}_{t-1}\Delta\tilde{x}'_{t-j} &= T^{-1/2}\mathrm{D}_T^{-1}\tilde{\beta}'_\perp \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)\Delta\tilde{x}'_{t-j} \\ &= T^{-1/2}\mathrm{D}_T^{-1}(\tilde{\beta}_\perp - \beta_\perp)'\beta\beta' \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)\Delta\tilde{x}'_{t-j} \\ &\quad + T^{-1/2}\mathrm{D}_T^{-1}\beta'_\perp \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)\Delta\tilde{x}'_{t-j}, \end{aligned} \tag{A.14}$$

where the latter equality makes use of the normalizations $\beta_\perp'\tilde\beta_\perp = \beta_\perp'\beta_\perp = I_{n-r}$ and $\beta'\beta = I_r$ and the identity $\beta\beta' + \beta_\perp\beta_\perp' = I_n$. Notice that

$$\beta'(y_t - \tilde\mu_0) = u_t - \beta'(\tilde\mu_0 - \mu_0) \tag{A.15}$$

and recall that u_t and Δx_t are zero mean stationary processes. Hence, since $\underline{D}_T^{-1} = O(T^{-1})$ one obtains from these facts, $(A.7)$, (2.12) and (2.14) – (2.16) that the first term in the last expression of $(A.14)$ is of order $o_p(1)$. From $(A.7)$ and (2.14) – (2.16) it can similarly be seen that the sample mean of $\Delta\tilde x_{t-j}$ is of order $O_p(T^{-1/2})$. This fact in conjunction with (2.13) implies that in the latter term of the last expression of $(A.14)$ replacing $\tilde\mu_0$ by μ_0 causes an error of order $o_p(1)$. Thus, we can conclude that

$$\begin{aligned} T^{-1/2}\underline{D}_T^{-1}\sum_{t=p+1}^{T}\tilde v_{t-1}\Delta\tilde x_{t-j}' &= T^{-1/2}\underline{D}_T^{-1}\sum_{t=p+1}^{T} v_{t-1}\Delta\tilde x_{t-j}' + o_p(1) \\ &= T^{-1/2}\underline{D}_T^{-1}\sum_{t=p+1}^{T} v_{t-1}\Delta x_{t-j}' \\ &\quad - T^{-1/2}\underline{D}_T^{-1}\sum_{t=p+1}^{T} v_{t-1}(\tilde\mu_1 - \mu_1) + o_p(1) \\ &= o_p(1). \end{aligned} \tag{A.16}$$

Here the second equality follows from $(A.7)$ and the third one can be justified by using (2.14), (2.15) and well-known limit theorems together with the identity $(A.2)$. To complete the proof of $(A.5)$ we have to show that a result similar to $(A.16)$ also holds with $\Delta\tilde x_{t-j}$ in the first expression replaced by $\tilde u_{t-1}$. We have

$$\begin{aligned} T^{-1/2}\underline{D}_T^{-1}\sum_{t=p+1}^{T}\tilde v_{t-1}\tilde u_{t-1}' &= T^{-1/2}\underline{D}_T^{-1}\tilde\beta_\perp'\sum_{t=p+1}^{T}(y_{t-1}-\tilde\mu_0)(y_{t-1}-\tilde\mu_0)'\tilde\beta \\ &= T^{-1/2}\underline{D}_T^{-1}\beta_\perp'\sum_{t=p+1}^{T}(y_{t-1}-\tilde\mu_0)(y_{t-1}-\tilde\mu_0)'\beta \\ &\quad + T^{-1/2}\underline{D}_T^{-1}(\tilde\beta_\perp-\beta_\perp)'\sum_{t=p+1}^{T}(y_{t-1}-\tilde\mu_0)(y_{t-1}-\tilde\mu_0)'\beta \\ &\quad + T^{-1/2}\underline{D}_T^{-1}\beta_\perp'\sum_{t=p+1}^{T}(y_{t-1}-\tilde\mu_0)(y_{t-1}-\tilde\mu_0)'(\tilde\beta-\beta) \\ &\quad + T^{-1/2}\underline{D}_T^{-1}(\tilde\beta_\perp-\beta_\perp)'\sum_{t=p+1}^{T}(y_{t-1}-\tilde\mu_0)(y_{t-1}-\tilde\mu_0)'(\tilde\beta-\beta) \\ &\overset{def}{=} A_{1T} + A_{2T} + A_{3T} + A_{4T}. \end{aligned} \tag{A.17}$$

To analyze A_{1T}, notice that

$$\beta_\perp'(y_t - \tilde\mu_0) = v_t - \beta_\perp'(\tilde\mu_0 - \mu_0). \tag{A.18}$$

Using this, $(A.15)$, (2.12), (2.13) and well-known limit theorems in conjunction with the identity $(A.2)$ one can see that $A_{1T} = o_p(1)$. Next note that

$$A_{2T} = T^{-1/2} \underline{D}_T^{-1} (\tilde{\beta}_\perp - \beta_\perp)' \beta\beta' \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)(y_{t-1} - \tilde{\mu}_0)' \beta,$$

where we have again made use of the chosen normalizations. From $(A.15)$ and (2.12) it follows that the second sample moments of $\beta'(y_{t-1} - \tilde{\mu}_0)$ are of order $O_p(1)$ which immediately gives $A_{2T} = o_p(1)$. For A_{3T} we similarly use the chosen normalizations and write

$$A_{3T} = T^{-1/2} \underline{D}_T^{-1} \beta_\perp' \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)(y_{t-1} - \tilde{\mu}_0)' \beta_\perp \beta_\perp' (\tilde{\beta} - \beta).$$

Recall the definition of the matrix $\underline{D}_T$ and notice that $\beta_\perp \underline{D}_T = \beta_\perp \beta_\perp' D_T = D_T$, where the latter equality follows from the fact that $D_T \in \text{span}(\beta_\perp)$. Thus, we can write $\beta_\perp'(\tilde{\beta} - \beta) = \underline{D}_T'^{-1} D_T'(\tilde{\beta} - \beta)$ so that, using (2.16), (2.13), $(A.18)$ and well-known limit theorems in conjunction with the identity $(A.2)$, we find that $A_{3T} = o_p(1)$. By similar arguments it can also be shown that $A_{4T} = o_p(1)$. Thus, we have shown that $(A.17)$ is of order $o_p(1)$ which together with $(A.16)$ implies $(A.5)$.

Finally consider $(A.6)$. We apply a decomposition similar to that in $(A.17)$ and write

$$\underline{D}_T^{-1} \sum_{t=p+1}^{T} \tilde{v}_{t-1} \tilde{v}_{t-1}' \underline{D}_T'^{-1} \stackrel{def}{=} A_{5T} + A_{6T} + A_{6T}' + A_{7T}$$

where the quantities on the r.h.s. will be described below. First we have

$$\begin{aligned} A_{5T} &= \underline{D}_T^{-1} \beta_\perp' \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)(y_{t-1} - \tilde{\mu}_0)' \beta_\perp \underline{D}_T'^{-1} \\ &= \underline{D}_T^{-1} \sum_{t=p+1}^{T} v_{t-1} v_{t-1}' \underline{D}_T'^{-1} + o_p(1) \end{aligned}$$

by (2.13), $(A.18)$ and arguments similar to those used for $(A.17)$. Thus, we need to show that $A_{6T} = o_p(1)$ and $A_{7T} = o_p(1)$. For the former we have

$$\begin{aligned} A_{6T} &= \underline{D}_T^{-1} (\tilde{\beta}_\perp - \beta_\perp)' \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)(y_{t-1} - \tilde{\mu}_0)' \beta_\perp \underline{D}_T'^{-1} \\ &= \underline{D}_T^{-1} (\tilde{\beta}_\perp - \beta_\perp)' \beta\beta' \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0)(y_{t-1} - \tilde{\mu}_0)' \beta_\perp \underline{D}_T'^{-1} \end{aligned}$$

again by the chosen normalizations. Thus, $(A.15)$ and $(A.18)$ in conjunction with arguments already used earlier show that $A_{6T} = o_p(1)$. Finally, note that A_{7T} is defined by replacing $\beta_\perp \underline{D}_T'^{-1}$ in the above definition of A_{6T} by $(\tilde{\beta}_\perp - \beta_\perp) \underline{D}_T'^{-1}$ so that in a similar fashion it can be seen that $A_{7T} = o_p(1)$. Thus, we have estabished $(A.6)$ and thereby Lemma A.1. □

Now we consider the first factor on the r.h.s. of $(A.1)$. When the null hypothesis holds, the error term e_t^* in (3.9) can be expressed as $e_t^* = \tilde{\alpha}'_\perp e_t$, where

$$\begin{aligned} e_t &= \varepsilon_t - (\tilde{\mu}_1 - \mu_1) + \sum_{j=1}^{p-1} \Gamma_j(\tilde{\mu}_1 - \mu_1) - \alpha(\tilde{u}_{t-1} - u_{t-1}) \\ &= \varepsilon_t - \Psi(\tilde{\mu}_1 - \mu_1) - \alpha(\tilde{\beta} - \beta)' y_{t-1} + \alpha(\tilde{\beta}'\tilde{\mu}_0 - \beta'\mu_0) \\ &= \varepsilon_t - \Psi(\tilde{\mu}_1 - \mu_1) - \alpha(\tilde{\beta} - \beta)'(y_{t-1} - \tilde{\mu}_0) + \alpha\beta'(\tilde{\mu}_0 - \mu_0) \\ &= \varepsilon_t - \Psi(\tilde{\mu}_1 - \mu_1) - \alpha(\tilde{\beta} - \beta)'\beta_\perp\beta'_\perp(y_{t-1} - \tilde{\mu}_0) + \alpha\beta'(\tilde{\mu}_0 - \mu_0). \end{aligned} \tag{A.19}$$

Here the last equality makes use of the normalization $\beta'\tilde{\beta} = \beta'\beta = I_r$ and the identity $\beta\beta' + \beta_\perp\beta'_\perp = I_n$. We shall demonstrate first that

$$\sum_{t=p+1}^{T} e_t^* \tilde{q}'_t \left(\sum_{t=p+1}^{T} \tilde{q}_t \tilde{q}'_t \right)^{-1} \sum_{t=p+1}^{T} \tilde{q}_t \tilde{v}'_{t-1} \mathrm{D}_T'^{-1} = o_p(1). \tag{A.20}$$

By $(A.4)$ and $(A.5)$ it suffices to show that

$$T^{-1/2} \sum_{t=p+1}^{T} e_t^* \tilde{q}'_t = \tilde{\alpha}'_\perp T^{-1/2} \sum_{t=p+1}^{T} e_t \tilde{q}'_t = O_p(1). \tag{A.21}$$

Using the above expression of e_t, $(A.7)$, $(A.18)$, and properties of the involved estimators given in (2.12) – (2.17) it can be seen that $(A.21)$ holds with $\tilde{q}_t$ replaced by $\Delta\tilde{x}_{t-i}$ $(i = 1, \ldots, p-1)$. Using $(A.11)$ instead of $(A.7)$ one can similarly see that $(A.21)$ holds with $\tilde{q}_t$ replaced by $\tilde{u}_{t-1}$. Details of these derivations are straightforward but rather lengthy and therefore omitted. Since $\tilde{q}_t$ is composed of $\Delta\tilde{x}_{t-i}$ $(i = 1, \ldots, p-1)$ and $\tilde{u}_{t-1}$ we have thus established $(A.21)$ and it suffices to consider

$$\begin{aligned} \sum_{t=p+1}^{T} e_t^* \tilde{v}'_{t-1} \mathrm{D}_T'^{-1} &= \tilde{\alpha}'_\perp \sum_{t=p+1}^{T} \varepsilon_t \tilde{v}'_{t-1} \mathrm{D}_T'^{-1} - \tilde{\alpha}'_\perp \Psi(\tilde{\mu}_1 - \mu_1) \sum_{t=p+1}^{T} \tilde{v}'_{t-1} \mathrm{D}_T'^{-1} \\ &\quad - \tilde{\alpha}'_\perp \alpha(\tilde{\beta} - \beta)' \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0) \tilde{v}'_{t-1} \mathrm{D}_T'^{-1} \\ &\quad + \tilde{\alpha}'_\perp \alpha\beta'(\tilde{\mu}_0 - \mu_0) \sum_{t=p+1}^{T} \tilde{v}'_{t-1} \mathrm{D}_T'^{-1}. \end{aligned} \tag{A.22}$$

We wish to show that the last two terms on the r.h.s are of order $o_p(1)$. First consider the latter and note that, in the same way as in $(A.14)$, we have

$$\begin{aligned} T^{-1/2} \mathrm{D}_T^{-1} \sum_{t=p+1}^{T} \tilde{v}_{t-1} &= T^{-1/2} \mathrm{D}_T^{-1} (\tilde{\beta}_\perp - \beta_\perp)' \beta\beta' \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0) \\ &\quad + T^{-1/2} \mathrm{D}_T^{-1} \beta'_\perp \sum_{t=p+1}^{T} (y_{t-1} - \tilde{\mu}_0) \\ &= O_p(1), \end{aligned} \tag{A.23}$$

where the latter equality follows from $(A.15)$ and $(A.18)$ by arguments similar to those used for $(A.14)$ and $(A.17)$. Thus, since (2.12) and (2.17) imply that $\tilde{\alpha}_\perp' \alpha\beta'(\tilde{\mu}_0 - \mu_0) = o_p(T^{-1/2})$ the last term on the r.h.s. of $(A.22)$ is of order $o_p(1)$. A decomposition similar to that used in $(A.23)$ can also be applied to the sum in the third term on the r.h.s. of $(A.22)$ after which arguments essentially the same as those used for A_{3T} and A_{4T} in $(A.17)$ show that this term is of order $o_p(1)$. Thus, we can write

$$\begin{aligned}\sum_{t=p+1}^{T} e_t^* \tilde{v}_{t-1}' \mathrm{D}_T'^{-1} &= \tilde{\alpha}_\perp' \sum_{t=p+1}^{T} \varepsilon_t \tilde{v}_{t-1}' \mathrm{D}_T'^{-1} - \tilde{\alpha}_\perp' \Psi(\tilde{\mu}_1 - \mu_1) \sum_{t=p+1}^{T} \tilde{v}_{t-1}' \mathrm{D}_T'^{-1} + o_p(1) \\ &= \tilde{\alpha}_\perp' \sum_{t=p+1}^{T} \varepsilon_t \tilde{v}_{t-1}' \mathrm{D}_T'^{-1} - \tilde{\alpha}_\perp' \Psi \beta_\perp \beta_\perp' (\tilde{\mu}_1 - \mu_1) \sum_{t=p+1}^{T} \tilde{v}_{t-1}' \mathrm{D}_T'^{-1} + o_p(1),\end{aligned}$$

where the latter equality makes use of the identity $\beta\beta' + \beta_\perp\beta_\perp' = I_n$, (2.14), (2.17) and properties of the series $\tilde{v}_t$ already used several times. Furthermore, since $(\tilde{\mu}_1 - \mu_1) = O_p(T^{-1/2})$ by (2.14) and (2.15), arguments similar to those used for $(A.14)$ and $(A.16)$ show that in the last expression $\tilde{v}_{t-1}$ can be replaced by v_{t-1} and further that $\tilde{\alpha}_\perp$ can be replaced by $\alpha_\perp$. Hence, we can write

$$\begin{aligned}\sum_{t=p+1}^{T} e_t^* \tilde{v}_{t-1}' \mathrm{D}_T'^{-1} &= \alpha_\perp' \sum_{t=p+1}^{T} \varepsilon_t v_{t-1}' \mathrm{D}_T'^{-1} \\ &\quad - \alpha_\perp' \Psi \beta_\perp \beta_\perp' (\tilde{\mu}_1 - \mu_1) \sum_{t=p+1}^{T} v_{t-1}' \mathrm{D}_T'^{-1} + o_p(1).\end{aligned} \tag{A.24}$$

To obtain the limiting distribution of this quantity and that in $(A.3)$, we conclude from $(A.2)$ and well-known limit theorems that

$$T^{1/2} \mathrm{D}_T^{-1} v_{[Ts]} \xrightarrow{d} [s, \mathbf{F}_*(s)']' \stackrel{def}{=} \mathbf{F}(s)', \tag{A.25}$$

where $\mathbf{F}_*(s) = H'\beta_\perp' C\mathbf{B}(s)$ with C and $\mathbf{B}(s)$ as in (2.8) and (2.9), respectively. Thus, $\mathbf{F}_*(s)$ is an $(n-r-1)$-dimensional Brownian motion with covariance matrix $H'\beta_\perp' C\Omega C'\beta_\perp H \stackrel{def}{=} \Sigma$. Next note that due to the normalization $\beta_\perp'\beta_\perp = I_{n-r}$ we have $\alpha_\perp' \Psi \beta_\perp \beta_\perp' C = \alpha_\perp'$ [see the definition of C below (2.8)]. Thus, applying (2.15), $(A.25)$ and the usual limit theorems to $(A.24)$ yields

$$\sum_{t=p+1}^{T} e_t^* \tilde{v}_{t-1}' \mathrm{D}_T'^{-1} \xrightarrow{d} \alpha_\perp' \left(\int_0^1 \mathrm{d}\mathbf{B}(s)\mathbf{F}(s)' - \mathbf{B}(1) \int_0^1 \mathbf{F}(s)'\mathrm{d}s \right) = \alpha_\perp' \int_0^1 \mathrm{d}\mathbf{B}(s)\bar{\mathbf{F}}(s)', \tag{A.26}$$

where $\bar{\mathbf{F}}(s)$ is the demeaned version of $\mathbf{F}(s)$. From $(A.3)$ one similarly obtains

$$\mathrm{D}_T^{-1} \tilde{M}_{vv\cdot q} \mathrm{D}_T'^{-1} \xrightarrow{d} \int_0^1 \mathbf{F}(s)\mathbf{F}(s)'\mathrm{d}s. \tag{A.27}$$

Because the weak convergencies in $(A.26)$ and $(A.27)$ hold jointly, it follows from the consistency of the estimators $\tilde{\alpha}$ and $\tilde{\Omega}$ and the definition of the test statistic $LM^*(r_0)$ that

$$LM^*(r_0) \xrightarrow{d} \operatorname{tr}\left\{(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' \int_0^1 \mathrm{d}\mathbf{B}(s)\bar{\mathbf{F}}(s)' \left(\int_0^1 \mathbf{F}(s)\mathbf{F}(s)'\mathrm{d}s\right)^{-1} \int_0^1 \bar{\mathbf{F}}(s)\mathrm{d}\mathbf{B}(s)'\right\}.$$

Theorem 1 is obtained from this result by noting that $\mathbf{W}(s) = (\alpha_\perp'\Omega\alpha_\perp)^{-1/2}\alpha_\perp'\mathbf{B}(s)$ and $\mathbf{G}(s) = \operatorname{diag}[1 : \Sigma^{1/2}]\mathbf{F}(s)$.

A.2 Proof of Theorem 2

Note that α and β are now $(n \times r_0)$ matrices and analogous dimensions hold for other quantities. Since $r = r_0$ in Section A.1, r now has to be replaced by r_0. For instance, h is $((n - r_0) \times 1)$ and H is $((n - r_0) \times (n - r_0 - 1))$ etc.. Some results shown to hold under the null hypothesis now have to be generalized for our special process of Section 4.1 to the case of local alternatives. For this purpose, some of the previously used arguments can be used directly because they only assume suitable rates of convergence and we assume that (2.16)–(2.18) and (2.12)–(2.14) still hold. We begin with an important intermediate result.

Lemma A.2. Under the assumptions of Section 4.1,

$$T^{1/2}\mathrm{D}_T^{-1}v_{[Ts]} \xrightarrow{d} \begin{bmatrix} s \\ H'\mathbf{K}(s) \end{bmatrix} \stackrel{def}{=} \mathbf{R}(s)$$

and, denoting by $\bar{v}_t$ the mean-adjusted version of v_t,

$$T^{1/2}\mathrm{D}_T^{-1}\bar{v}_{[Ts]} \xrightarrow{d} \bar{\mathbf{R}}(s) \stackrel{def}{=} \mathbf{R}(s) - \int_0^1 \mathbf{R}(u)\mathrm{d}u.$$

Proof: By $(A.2)$ it is obvious that it suffices to show that $T^{-1/2}\beta_\perp' x_{[Ts]} \xrightarrow{d} \mathbf{K}(s)$. This, however, follows from Johansen (1997, Theorem 14.1) and the fact that we may assume that $\alpha_\perp = \beta_\perp + \beta c$ for a suitable matrix c. □

The Proof of Lemma A.1 made use of some known convergence results of the processes $\beta' x_t$ and $\beta_\perp' x_t$ under the null hypothesis. For later purposes it is useful to note that from Johansen (1995, Chapter 14) and Rahbek (1994) we can further conclude that the first and second sample moments of these processes converge weakly and the rate of convergence under our local alternatives is the same as under the null hypothesis.

Since $\overline{\Delta y}_t$ and $\bar{y}_t$ do not depend on the value of μ_0 we may assume that $\mu_0 = 0$. Then we have $\beta' y_t = \beta' x_t$. Next note that

$$\begin{aligned}\tilde{\bar{u}}_t - \bar{u}_t &= (\tilde{\beta} - \beta)' \bar{y}_t \\ &= (\tilde{\beta} - \beta)' \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \beta_\perp' \bar{y}_t \\ &= (\tilde{\beta} - \beta)' \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \underline{D}_T \underline{D}_T^{-1} \beta_\perp' \bar{y}_t \\ &= (\tilde{\beta} - \beta)' D_T \underline{D}_T^{-1} \bar{v}_t,\end{aligned}$$

where the last equality follows because

$$\beta_\perp (\beta_\perp' \beta_\perp)^{-1} \underline{D}_T = \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \beta_\perp' [\mu_1 : \gamma] \begin{bmatrix} T^{3/2} & 0 \\ 0 & T \end{bmatrix} = D_T.$$

From Lemma A.2 it follows that $\max_{1 \le t \le T} \|\underline{D}_T^{-1} \bar{v}_t\| = O_p(T^{-1/2})$. Hence, since $(\tilde{\beta} - \beta)' D_T = O_p(1)$ by (2.16), we conclude that

$$\max_{1 \le t \le T} \|\tilde{\bar{u}}_t - \bar{u}_t\| = O_p(T^{-1/2}). \qquad (A.28)$$

Similarly,

$$\begin{aligned}\tilde{\bar{v}}_t - \bar{v}_t &= (\tilde{\beta}_\perp - \beta_\perp)' \bar{y}_t \\ &= (\tilde{\beta}_\perp - \beta_\perp)' \beta (\beta' \beta)^{-1} \beta' \bar{y}_t \\ &= (\tilde{\beta}_\perp - \beta_\perp)' \beta (\beta' \beta)^{-1} \beta' \bar{x}_t.\end{aligned}$$

Here $T^{-1} \max_{1 \le t \le T} \|\beta' \bar{x}_t\| = O_p(T^{-1/2})$. It is well-known that this last result holds under the null hypothesis when $\beta' x_t$ is (asymptotically) stationary and it follows from Johansen (1995, Theorem 14.1) that it also holds under our local alternatives. Thus, since $(\tilde{\beta}_\perp - \beta_\perp)' \beta = O_p(T^{-1})$, we have

$$\max_{1 \le t \le T} \|\tilde{\bar{v}}_t - \bar{v}_t\| = O_p(T^{-1/2}). \qquad (A.29)$$

Now the proof proceeds in a similar way as that of Theorem 1. Thus, we consider

$$\bar{\rho}_T^* - \rho_T^* = \sum_{t=2}^{T} \tilde{\bar{e}}_t^* \left[\tilde{\bar{v}}_{t-1}' - \tilde{\bar{u}}_{t-1}' \left(\sum_{t=2}^{T} \tilde{\bar{u}}_{t-1} \tilde{\bar{u}}_{t-1}' \right)^{-1} \sum_{t=2}^{T} \tilde{\bar{u}}_{t-1} \tilde{\bar{v}}_{t-1}' \right] \bar{M}_{vv \cdot u}^{-1}. \qquad (A.30)$$

First we note that

$$\underline{D}_T^{-1} \bar{M}_{vv \cdot u} \underline{D}_T'^{-1} = \underline{D}_T^{-1} \sum_{t=2}^{T} \bar{v}_{t-1} \bar{v}_{t-1}' \underline{D}_T'^{-1} + o_p(1). \qquad (A.31)$$

To justify this equality, use $(A.28)$ and $(A.29)$ to show first that in the definition of $\bar{M}_{vv \cdot u}$ the series $\tilde{\bar{u}}_{t-1}$ and $\tilde{\bar{v}}_{t-1}$ can be replaced by $\bar{u}_{t-1}$ and $\bar{v}_{t-1}$, respectively. After this, $(A.31)$ may be obtained from the results given in Johansen (1995, Chapter 14) and S&L.

The next step is to show that an analog of $(A.21)$ holds, that is,

$$T^{-1/2}\sum_{t=2}^{T}\tilde{\tilde{e}}_t^{*}\tilde{\tilde{u}}_{t-1}' = O_p(1).$$

This can be established by using the definition of $\tilde{\tilde{e}}_t^{*}$ in (4.8), (2.16) and $(A.28)$ in conjunction with the results in Johansen (1995, Chapter 14) and S&L. In a similar way as in the proof of Theorem 1 we can now conclude from the above discussion that

$$\begin{aligned}\sum_{t=2}^{T}\tilde{\tilde{e}}_t^{*}\left[\tilde{\tilde{v}}_{t-1}' - \tilde{\tilde{u}}_{t-1}'\left(\sum_{t=2}^{T}\tilde{\tilde{u}}_{t-1}\tilde{\tilde{u}}_{t-1}'\right)^{-1}\sum_{t=2}^{T}\tilde{\tilde{u}}_{t-1}\tilde{\tilde{v}}_{t-1}'\right]\mathrm{D}_T'^{-1} &= \sum_{t=2}^{T}\tilde{\tilde{e}}_t^{*}\tilde{\tilde{v}}_{t-1}'\mathrm{D}_T'^{-1} + o_p(1) \\ &= \alpha_\perp'\sum_{t=2}^{T}\varepsilon_t\bar{v}_{t-1}'\mathrm{D}_T'^{-1} + o_p(1).\end{aligned} \tag{A.32}$$

Here the second equality can be established by using the definition of $\tilde{\tilde{e}}_t^{*}$, (2.17) and $(A.29)$ in conjunction with the results in Johansen (1995, Chapter 14) and S&L.

It follows from $(A.32)$, Lemma A.2 and Theorem 2.1 of Hansen (1992) that

$$\alpha_\perp'\sum_{t=2}^{T}\varepsilon_t\bar{v}_{t-1}'\mathrm{D}_T'^{-1} \xrightarrow{d} \alpha_\perp'\left(\int_0^1\bar{\mathbf{R}}(s)\mathrm{d}\mathbf{B}(s)'\right)'. \tag{A.33}$$

Moreover, from $(A.31)$,

$$\mathrm{D}_T^{-1}\bar{M}_{vv\cdot u}\mathrm{D}_T'^{-1} \xrightarrow{d} \int_0^1\bar{\mathbf{R}}(s)\bar{\mathbf{R}}(s)'\mathrm{d}s. \tag{A.34}$$

Furthermore,

$$\begin{aligned}\rho_T^{*}\mathrm{D}_T &= \tilde{\alpha}_\perp'\rho_T\mathrm{D}_T \\ &= T^{-1}\tilde{\alpha}_\perp'\alpha_1\beta_1'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'[\mu_1:\gamma]\begin{bmatrix}T^{3/2} & 0\\ 0 & T\end{bmatrix} \\ &= T^{-1}\tilde{\alpha}_\perp'\alpha_1\beta_1'[\mu_1:\gamma]\begin{bmatrix}T^{3/2} & 0\\ 0 & T\end{bmatrix} \\ &= \tilde{\alpha}_\perp'\alpha_1\beta_1'[0:\gamma] \\ &= \alpha_\perp'\alpha_1\beta_1'[0:\gamma] + o_p(1).\end{aligned} \tag{A.35}$$

Here the first and second equalities follow from definitions, the third one is obtained because $[\mu_1:\gamma]\in\mathrm{span}(\beta_\perp)$ and the fourth one is a consequence of the assumption $\beta_1'\mu_1 = 0$. Combining $(A.33)-(A.35)$ yields

$$\mathrm{D}_T'\bar{\rho}_T^{*\prime} \xrightarrow{d} \left(\int_0^1\bar{\mathbf{R}}(s)\bar{\mathbf{R}}(s)'\mathrm{d}s\right)^{-1}\left(\int_0^1\bar{\mathbf{R}}(s)\mathrm{d}\mathbf{B}(s)'\alpha_\perp + \int_0^1\bar{\mathbf{R}}(s)\bar{\mathbf{R}}(s)'\mathrm{d}s\begin{bmatrix}0\\ \gamma'\end{bmatrix}\beta_1\alpha_1'\alpha_\perp\right). \tag{A.36}$$

Now using $g = [h : H]$ and, thus,

$$\mathbf{R}(s)' \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} = [s, \mathbf{K}(s)'H] \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} = \mathbf{K}(s)'[h : H] \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} = \mathbf{K}(s)'g \begin{bmatrix} 0 \\ \gamma' \end{bmatrix}, \qquad (A.37)$$

we get

$$\left(\int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{B}(s)' \alpha_\perp g + \int_0^1 \bar{\mathbf{R}}(s) \bar{\mathbf{R}}(s)' \mathrm{d}s \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} \beta_1 \alpha_1' \alpha_\perp g \right) g^{-1}$$

$$= \left(\int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{B}(s)' \alpha_\perp g + \int_0^1 \bar{\mathbf{R}}(s) \mathbf{R}(s)' \mathrm{d}s \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} \beta_1 \alpha_1' \alpha_\perp g \right) g^{-1}$$

$$= \left(\int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{B}(s)' \alpha_\perp g + \int_0^1 \bar{\mathbf{R}}(s) \mathbf{K}(s)' \mathrm{d}s g \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} \beta_1 \alpha_1' \alpha_\perp g \right) g^{-1}$$

$$= \int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{Q}(s)' g^{-1}, \qquad (A.38)$$

where by the definition of $\mathbf{K}(s)$, the Ornstein-Uhlenbeck process

$$\mathbf{Q}(s) = g'\mathbf{K}(s) = \begin{bmatrix} h'\mathbf{K}(s) \\ H'\mathbf{K}(s) \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1(s) \\ \mathbf{Q}_2(s) \end{bmatrix} \qquad (A.39)$$

satisfies the stochastic differential equation

$$\begin{aligned} \mathrm{d}\mathbf{Q}(s) &= g'\alpha_\perp' \mathrm{d}\mathbf{B}(s) + g'\alpha_\perp' \alpha_1 \beta_1' \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} g'^{-1} \mathbf{Q}(s) \mathrm{d}s \\ &= g'\alpha_\perp' \mathrm{d}\mathbf{B}(s) + g'\alpha_\perp' \alpha_1 \beta_1' [0 : \gamma] \mathbf{Q}(s) \mathrm{d}s. \end{aligned}$$

Here the latter equality is obtained by observing that $g'^{-1} = \beta_\perp' [\mu_1 : \gamma]$ and using an argument similar to that in $(A.35)$. Hence, we have

$$\mathrm{D}_T' \bar{\rho}_T^{*\prime} \xrightarrow{d} \left(\int_0^1 \bar{\mathbf{R}}(s) \bar{\mathbf{R}}(s)' \mathrm{d}s \right)^{-1} \int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{Q}(s)' g^{-1} \qquad (A.40)$$

and, by the definition of the test statistic $LR_{trace}(r_0)$,

$$\begin{aligned} LR_{trace}(r_0) &\xrightarrow{d} \mathrm{tr} \left\{ (\alpha_\perp' \Omega \alpha_\perp)^{-1} g'^{-1} \left(\int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{Q}(s)' \right)' \left(\int_0^1 \bar{\mathbf{R}}(s) \bar{\mathbf{R}}(s)' \mathrm{d}s \right)^{-1} \int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{Q}(s)' g^{-1} \right\} \\ &= \mathrm{tr} \left\{ (g'\alpha_\perp' \Omega \alpha_\perp g)^{-1} \left(\int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{Q}(s)' \right)' \left(\int_0^1 \bar{\mathbf{R}}(s) \bar{\mathbf{R}}(s)' \mathrm{d}s \right)^{-1} \int_0^1 \bar{\mathbf{R}}(s) \mathrm{d}\mathbf{Q}(s)' \right\}. \end{aligned}$$

Since $\mathbf{R}(s)' = [s, \mathbf{Q}_2(s)']'$ and $\mathbf{N}(s) = (g'\alpha_\perp' \Omega \alpha_\perp g)^{-1/2} \mathbf{Q}(s)$, this is the desired result. Hence, the proof is complete.

A.3 Proof of Theorem 3

The proof proceeds in the same way as that of Theorem 2. As in the previous subsection we consider the limiting behaviour of

$$\hat{\rho}_T^* - \rho_T^* = \sum_{t=2}^T e_t^* \left[\tilde{v}'_{t-1} - \tilde{u}'_{t-1} \left(\sum_{t=2}^T \tilde{u}_{t-1} \tilde{u}'_{t-1} \right)^{-1} \sum_{t=2}^T \tilde{u}_{t-1} \tilde{v}'_{t-1} \right] \tilde{M}_{vv \cdot u}^{-1}. \qquad (A.41)$$

We first note that

$$\underline{D}_T^{-1} \tilde{M}_{vv \cdot u} \underline{D}_T'^{-1} = \underline{D}_T^{-1} \sum_{t=2}^T v_{t-1} v'_{t-1} \underline{D}_T'^{-1} + o_p(1). \qquad (A.42)$$

This can be shown by demonstrating that

$$T^{-1} \sum_{t=2}^T \tilde{u}_{t-1} \tilde{u}'_{t-1} = T^{-1} \sum_{t=2}^T u_{t-1} u'_{t-1} + o_p(1), \qquad (A.43)$$

$$T^{-1/2} \underline{D}_T^{-1} \sum_{t=2}^T \tilde{v}_{t-1} \tilde{u}'_{t-1} = o_p(1) \qquad (A.44)$$

and

$$\underline{D}_T^{-1} \sum_{t=2}^T \tilde{v}_{t-1} \tilde{v}'_{t-1} \underline{D}_T'^{-1} = \underline{D}_T^{-1} \sum_{t=2}^T v_{t-1} v'_{t-1} \underline{D}_T'^{-1} + o_p(1). \qquad (A.45)$$

These results can be established in exactly the same way as their counterparts in the proof of Theorem 2.

Next consider the first factor on the r.h.s. of $(A.41)$. The error term e_t in (4.10) can be written as

$$e_t = e_t^{(0)} + e_t^{(1)},$$

where $e_t^{(0)}$ is essentially the error term under the null hypothesis and $e_t^{(1)}$ is the remaining part of e_t. Specifically, using (4.10), it can be shown that

$$e_t^{(0)} = \varepsilon_t + \kappa_T \beta'(\tilde{\mu}_0 - \mu_0) - (\tilde{\mu}_1 - \mu_1) - \kappa_T (\tilde{\beta} - \beta)' \beta_\perp \beta'_\perp (y_{t-1} - \tilde{\mu}_0)$$

and

$$e_t^{(1)} = \rho_T \beta'_\perp (\tilde{\mu}_0 - \mu_0) - \rho_T (\tilde{\beta}_\perp - \beta_\perp)' \beta \beta' (y_{t-1} - \tilde{\mu}_0).$$

It can be shown with similar arguments as for $(A.21)$ that

$$T^{-1/2} \tilde{\alpha}'_\perp \sum_{t=2}^T e_t^{(i)} \tilde{u}'_{t-1} = O_p(1), \qquad i = 1, 2,$$

and, hence, by $(A.43)$ and $(A.44)$,

$$\sum_{t=2}^T e_t^* \tilde{u}'_{t-1} \left(\sum_{t=2}^T \tilde{u}_{t-1} \tilde{u}'_{t-1} \right)^{-1} \sum_{t=2}^T \tilde{u}_{t-1} \tilde{v}'_{t-1} \underline{D}_T'^{-1} = o_p(1). \qquad (A.46)$$

Therefore we may consider

$$\sum_{t=2}^{T} e_t^* \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1} = \tilde{\alpha}_\perp' \sum_{t=2}^{T} e_t^{(0)} \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1} + \tilde{\alpha}_\perp' \sum_{t=2}^{T} e_t^{(1)} \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1}. \qquad (A.47)$$

Here

$$\tilde{\alpha}_\perp' \sum_{t=2}^{T} e_t^{(0)} \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1} = \alpha_\perp' \sum_{t=2}^{T} \varepsilon_t v_{t-1}' \underline{\mathrm{D}}_T'^{-1} - \alpha_\perp' (\tilde{\mu}_1 - \mu_1) \sum_{t=2}^{T} v_{t-1}' \underline{\mathrm{D}}_T'^{-1} + o_p(1) \qquad (A.48)$$

which can be established in the same way as $(A.24)$ because $e_t^{(0)}$ is the error term e_t under the null hypothesis and the involved quantities converge at the same rates as under the null hypothesis. Now consider the second term on the r.h.s. of $(A.47)$. By the definition of $e_t^{(1)}$ and since $\rho_T = O(T^{-1})$, we also have

$$\begin{aligned} & \tilde{\alpha}_\perp' \sum_{t=2}^{T} e_t^{(1)} \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1} \\ & = \tilde{\alpha}_\perp' \rho_T \beta_\perp' (\tilde{\mu}_0 - \mu_0) \sum_{t=2}^{T} \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1} - \tilde{\alpha}_\perp' \rho_T (\tilde{\beta}_\perp - \beta_\perp)' \beta \sum_{t=2}^{T} \beta' (y_{t-1} - \tilde{\mu}_0) \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1} \qquad (A.49) \\ & = o_p(1) \end{aligned}$$

by arguments already used earlier. From $(A.47) - (A.49)$ we can now conclude that

$$\sum_{t=2}^{T} e_t^* \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1} = \alpha_\perp' \sum_{t=2}^{T} \varepsilon_t v_{t-1}' \underline{\mathrm{D}}_T'^{-1} - \beta_\perp' (\tilde{\mu}_1 - \mu_1) \sum_{t=2}^{T} v_{t-1}' \underline{\mathrm{D}}_T'^{-1} + o_p(1), \qquad (A.50)$$

where we have replaced $\alpha_\perp$ in the second term on the r.h.s. by $\beta_\perp$. This is easily seen to be justified by (2.14) and the fact that $\alpha_\perp = \beta_\perp + \beta c$ may be assumed for a suitable matrix c, as mentioned earlier.

The next step is to obtain the weak limits of $(A.42)$ and $(A.50)$ and combine the results in the same way as for the LR statistic in the previous subsection. First consider $(A.42)$ and note that Lemma A.2 and a standard application of the continuous mapping theorem give

$$\underline{\mathrm{D}}_T^{-1} \tilde{M}_{vv \cdot u} \underline{\mathrm{D}}_T'^{-1} \xrightarrow{d} \int_0^1 \mathbf{R}(s) \mathbf{R}(s)' \mathrm{d}s. \qquad (A.51)$$

Using (4.3) and proceeding as in the proof of Theorem 2 we get

$$\sum_{t=2}^{T} e_t^* \tilde{v}_{t-1}' \underline{\mathrm{D}}_T'^{-1} \xrightarrow{d} \alpha_\perp' \left(\int_0^1 \mathbf{R}(s) \mathrm{d}\mathbf{B}(s)' \right)' - \mathbf{K}(1) \int_0^1 \mathbf{R}(s)' \mathrm{d}s. \qquad (A.52)$$

From $(A.51)$ and $(A.52)$ one can derive the limiting distribution of $(\hat{\rho}_T^* - \rho_T^*) \underline{\mathrm{D}}_T$. Combining

$(A.51)$, $(A.52)$, $(A.35)$ and $(A.37)$ yields

$$\underset{=}{\mathrm{D}}_T' \hat{\rho}_T^{*\prime} \xrightarrow{d} \left(\int_0^1 \mathbf{R}(s)\mathbf{R}(s)'\mathrm{d}s \right)^{-1}$$
$$\times \left(\int_0^1 \mathbf{R}(s)\mathrm{d}\mathbf{B}(s)'\alpha_\perp - \int_0^1 \mathbf{R}(s)\mathrm{d}s\mathbf{K}(1)' + \int_0^1 \mathbf{R}(s)\mathbf{K}(s)'\mathrm{d}sg \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} \beta_1\alpha_1'\alpha_\perp \right). \tag{A.53}$$

With similar arguments as for $(A.38)$ we get

$$\left(\int_0^1 \mathbf{R}(s)\mathrm{d}\mathbf{B}(s)'\alpha_\perp g - \int_0^1 \mathbf{R}(s)\mathrm{d}s\mathbf{K}(1)'g + \int_0^1 \mathbf{R}(s)\mathbf{K}(s)'\mathrm{d}sg \begin{bmatrix} 0 \\ \gamma' \end{bmatrix} \beta_1\alpha_1'\alpha_\perp g \right) g^{-1}$$

$$= \left(\int_0^1 \mathbf{R}(s)\mathrm{d}\mathbf{Q}(s)' - \int_0^1 \mathbf{R}(s)\mathrm{d}s\mathbf{Q}(1)' \right) g^{-1}$$

$$= \int_0^1 \bar{\mathbf{R}}(s)\mathrm{d}\mathbf{Q}(s)'g^{-1}.$$

This means that we can write $(A.33)$ as

$$\underset{=}{\mathrm{D}}_T' \hat{\rho}_T^{*\prime} \xrightarrow{d} \left(\int_0^1 \mathbf{R}(s)\mathbf{R}(s)'\mathrm{d}s \right)^{-1} \int_0^1 \bar{\mathbf{R}}(s)\mathrm{d}\mathbf{Q}(s)'g^{-1}$$

Using these results we get the limiting distribution in Theorem 3 in the same way as in the proof of Theorem 2.

References

Engle, R.F. & C.W.J. Granger (1987), Co-integration and error correction: Representation, estimation and testing, *Econometrica*, 55, 251-276.

Granger, C.W.J. (1981), Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, 16, 121 - 130.

Hansen, B.E. (1992), Convergence to stochastic integrals for dependent heterogeneous processes, *Econometric Theory*, 8, 489 - 500.

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

Johansen, S. (1992), Determination of cointegration rank in the presence of a linear trend, *Oxford Bulletin of Economics and Statistics*, 54, 383-397.

Johansen, S. (1994), The role of the constant and linear terms in cointegration analysis of nonstationary time series, *Econometric Reviews*, 13, 205-231.

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

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

Lütkepohl, H. & P. Saikkonen (1997), Testing for the cointegrating rank of a VAR process with a time trend, Discussion Paper 79, SFB 373, Humboldt-Universität zu Berlin.

Osterwald-Lenum, M. (1992), A note with fractiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics: Four cases, *Oxford Bulletin of Economics and Statistics*, 54, 461-472.

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

Rahbek, A.C. (1994), The power of some multivariate cointegration tests, Discussion Paper.

Reinsel, G.C. & S.K. Ahn (1992), Vector AR models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis*, 13, 353–375.

Saikkonen, P. & H. Lütkepohl (1997), Trend adjustment prior to testing for the cointegrating rank of a VAR process, Discussion Paper 84, SFB 373, Humboldt-Universität zu Berlin.

Saikkonen, P. & H. Lütkepohl (1998), Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory*, forthcoming.

Sims, C.A., J.H. Stock & M.W. Watson (1990), Inference in linear time series models with some unit roots, *Econometrica,* 58, 113-144.

Toda, H.Y. (1994), Finite sample properties of likelihood ratio tests for cointagrating ranks when linear trends are present, *Review of Economics and Statistics*, 76, 66 - 79.

Toda, H.Y. (1995), Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions, *Econometric Theory*, 11, 1015 - 1032.

Table 2. Relative Rejection Frequencies of Test Statistics for DGP (5.1) with Cointegrating Rank $r = 0$ or 1, $\psi_2 = 1$, $\delta = 1$, Sample Size $T = 100$, Nominal Significance Level 0.05.

Test Statistic	$\psi_1 = 1$ $(r = 0)$		$\psi_1 = 0.9$ $(r = 1)$		$\psi_1 = 0.8$ $(r = 1)$		$\psi_1 = 0.7$ $(r = 1)$	
	$r_0 = 0$	$r_0 = 1$	$r_0 = 0$	$r_0 = 1$	$r_0 = 0$	$r_0 = 1$	$r_0 = 0$	$r_0 = 1$
	$\theta_1 = \theta_2 = 0$							
LR_{trace}	0.062	0.007	0.092	0.011	0.281	0.029	0.609	0.046
LM^*	0.037	0.022	0.068	0.036	0.245	0.045	0.524	0.043
$LR^{\sim}_{trace}$	0.056	0.018	0.106	0.029	0.333	0.049	0.624	0.054
LR^{+}_{trace}	0.065	0.007	0.098	0.009	0.193	0.018	0.424	0.034
LR^{PC}_{trace}	0.054	0.003	0.090	0.010	0.201	0.019	0.438	0.033
LM^{GLS}_{*}	0.028	0.005	0.047	0.005	0.116	0.010	0.290	0.022
	$\theta_1 = \theta_2 = 0.4$							
LR_{trace}	0.062	0.007	0.170	0.019	0.523	0.043	0.864	0.060
LM^*	0.036	0.022	0.143	0.052	0.452	0.067	0.793	0.061
$LR^{\sim}_{trace}$	0.059	0.018	0.208	0.053	0.560	0.073	0.858	0.068
LR^{+}_{trace}	0.065	0.007	0.140	0.012	0.362	0.044	0.711	0.052
LR^{PC}_{trace}	0.054	0.003	0.132	0.013	0.368	0.043	0.711	0.052
LM^{GLS}_{*}	0.028	0.005	0.070	0.009	0.245	0.023	0.535	0.028
	$\theta_1 = 0.4, \theta_2 = 0.8$							
LR_{trace}	0.063	0.007	0.781	0.092	0.998	0.088	1.0	0.084
LM^*	0.036	0.020	0.680	0.106	0.965	0.079	0.991	0.065
$LR^{\sim}_{trace}$	0.056	0.018	0.756	0.131	0.975	0.091	0.993	0.081
LR^{+}_{trace}	0.065	0.007	0.638	0.075	0.983	0.080	1.0	0.074
LR^{PC}_{trace}	0.054	0.003	0.552	0.064	0.975	0.074	1.0	0.063
LM^{GLS}_{*}	0.028	0.005	0.448	0.022	0.877	0.024	0.960	0.032

Table 3. Relative Rejection Frequencies of Test Statistics for DGP (5.1) with Cointegrating Rank $r = 2$, $\delta = 1$, Sample Size $T = 100$, Nominal Significance Level 0.05.

Test Statistic	$\psi_1 = 0.9$ $\psi_2 = 0.9$ $r_0 = 0$	$r_0 = 1$	$\psi_1 = 0.8$ $\psi_2 = 0.9$ $r_0 = 0$	$r_0 = 1$	$\psi_1 = 0.7$ $\psi_2 = 0.9$ $r_0 = 0$	$r_0 = 1$
			$\theta_1 = \theta_2 = 0$			
LR_{trace}	0.172	0.022	0.425	0.076	0.749	0.122
LM^*	0.212	0.076	0.436	0.177	0.707	0.209
$LR_{trace}^{\sim}$	0.263	0.085	0.524	0.200	0.785	0.238
LR_{trace}^{+}	0.130	0.014	0.279	0.038	0.528	0.073
LR_{trace}^{PC}	0.143	0.023	0.325	0.051	0.585	0.094
LM_*^{GLS}	0.084	0.013	0.206	0.035	0.403	0.069
			$\theta_1 = \theta_2 = 0.4$			
LR_{trace}	0.228	0.032	0.580	0.105	0.893	0.143
LM^*	0.241	0.109	0.559	0.211	0.858	0.227
$LR_{trace}^{\sim}$	0.305	0.114	0.658	0.239	0.897	0.252
LR_{trace}^{+}	0.150	0.021	0.390	0.060	0.714	0.103
LR_{trace}^{PC}	0.180	0.027	0.426	0.072	0.757	0.112
LM_*^{GLS}	0.099	0.017	0.304	0.043	0.567	0.072
			$\theta_1 = 0.4, \theta_2 = 0.8$			
LR_{trace}	0.787	0.182	0.997	0.205	1.0	0.208
LM^*	0.727	0.284	0.976	0.262	0.991	0.241
$LR_{trace}^{\sim}$	0.801	0.318	0.983	0.306	0.996	0.273
LR_{trace}^{+}	0.595	0.117	0.976	0.143	1.0	0.139
LR_{trace}^{PC}	0.561	0.114	0.969	0.148	0.999	0.143
LM_*^{GLS}	0.452	0.051	0.871	0.060	0.964	0.075

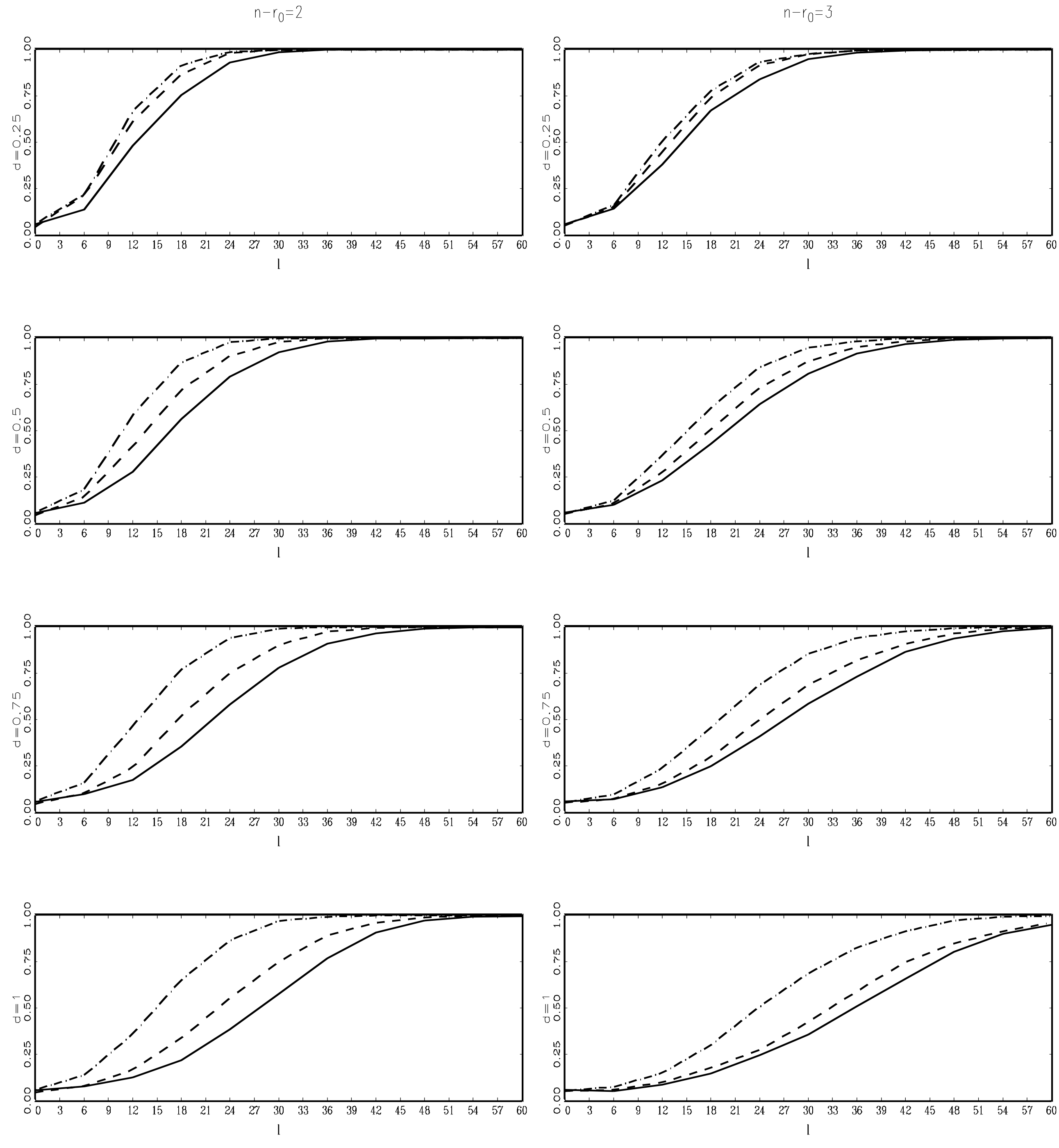

Figure 1: Local Power of LR and LM-type tests

—LR^{+}_{trace}, -·- LM^{*}, LR_{trace}, - - -LR_{trace}