Testing for the cointegrating rank of a VAR process with a time trend

Helmut Lütkepohl\textsuperscript{a,*}, Pentti Saikkonen\textsuperscript{b}

\textsuperscript{a}Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-University, Spandauer Str. 1, 10178 Berlin, Germany

\textsuperscript{b}Department of Statistics, University of Helsinki, P.O. Box 54, SF-00014 University of Helsinki, Finland

Received 1 August 1997; received in revised form 1 December 1998; accepted 1 April 1999

Abstract

Standard tests for the cointegrating rank of a vector autoregressive (VAR) process have nonstandard limiting distributions which depend on the characteristics of intercept terms and time trends in the system. In practice, these characteristics are often unknown. Therefore, modified tests are considered which allow for deterministic linear trends in the data generation process (DGP). The tests are based on the Lagrange multiplier (LM) principle and, in contrast to likelihood ratio (LR) tests proposed for this situation, our tests take into account the cointegrating rank specified under the null hypothesis in estimating the trend parameters. The tests are shown to have nonstandard limiting distributions which do not depend on deterministic terms and have better local power and small sample properties than the competing LR tests in many situations. © 2000 Elsevier Science S.A. All rights reserved.

\textit{JEL classification:} C32

\textit{Keywords:} Cointegration; Vector autoregressive process; Asymptotic inference; Rank test

\*Corresponding author. Tel.: + 49-30-2093-5718; fax.: + 49-30-2093-5712.

We thank Katsuto Tanaka, Cheng Hsiao and two anonymous referees for helpful comments on an earlier version of this paper and Martin Schweizer for his advise on Itô integrals. Moreover, we are grateful to Ritva Luukkonen, Kirstin Hubrich and Ralf Brüggemann for performing the computations. The Deutsche Forschungsgemeinschaft, SFB 373, provided financial support.

E-mail addresses: luetke@wiwi.hu-berlin.de (H. Lütkepohl), saikkone@valt.helsinki.fi (P. Saikkonen)
1. Introduction

Determining the number of cointegrating relations of a set of integrated variables is an important part of many empirical studies. A number of tests for cointegration have been proposed in the literature. Some of them are designed for detecting whether there is some cointegrating relation in a given set of variables. These tests are typically based on single equation estimation. The disadvantage of this approach is that it does not answer the question of how many cointegrating relations there are. Having an answer to that question may, however, be of central importance for modelling and interpreting the DGP (data generation process) of a given set of variables.

Therefore tests for determining the number of cointegrating relations have been proposed. Notable examples are Johansen (1988, 1991, 1994), Reinsel and Ahn (1992), Saikkonen and Luukkonen (1997), Stock and Watson (1988), Gregoir and Laroque (1994), Perron and Campbell (1993), Bewley and Yang (1995) and Bierens (1997). A major obstacle in using these tests is the fact that the limiting distributions of most of the test statistics depend on characteristics of the DGP which are often unknown in practice. Specifically they depend on the trend characteristics of the variables. Since these characteristics are commonly unknown in practice a consequence is that the crucial characteristics of the DGP are either assumed or pretests are carried out for determining them. The latter approach effectively results in a procedure with uncertain overall properties. Exceptions are, for instance, the LR (likelihood ratio) tests proposed by Johansen (1994) and Perron and Campbell (1993) which allow for a deterministic linear trend and have a limiting distribution which does not depend on the trend parameters. Unfortunately, it was found by Rahbek (1994) and Saikkonen and Lütkepohl (1999) (henceforth S&L) that these tests may have rather poor local power properties. For instance, in comparing the local power of these tests to that of LR tests which use prior knowledge that there is no deterministic trend component, S&L found that the local power of the latter tests may be more than twice as large as for the tests of Johansen (1994) and Perron and Campbell (1993).

In this study we will propose Lagrange multiplier (LM)-type tests for the cointegrating rank which allow for a deterministic linear trend in the DGP. In contrast to the competing LR tests, they take full advantage of the restrictions implied by the cointegrating rank in estimating the trend parameters. It is shown that, under the null hypothesis, the asymptotic distributions of these tests do not depend on the properties of the deterministic trend and intercept terms. In the univariate case, LM-type tests for unit roots have been proposed and investigated by Nabeya and Tanaka (1990), Schmidt and Phillips (1992) and Ahn (1993). We will show how these tests can be extended to multivariate processes. Since the trend parameters are estimated under the null hypothesis, the tests are expected to be more powerful than tests which do not use this restriction. It
turns out that the local power may indeed be much better than that of the Perron–Campbell and Johansen LR tests. We also report small sample simulation results which show that the small sample power of the LM-type tests is superior to that of the competing LR tests in situations of practical relevance.

The structure of this paper is as follows. In the next section the underlying model and some of its properties are presented. In Section 3, LR tests for cointegration are reviewed which are suitable in the present context. In Section 4, LM-type tests are considered and in Section 5 a local power comparison of the LR and LM-type tests is provided. A small Monte Carlo comparison of the properties of the tests is performed in Section 6 and conclusions follow in Section 7. The proofs are sketched in the appendix.

2. The model

Consider an \( n \)-dimensional time series \( y_t = (y_{1t}, \ldots, y_{nt})' \), \( t = 1, \ldots, T \), generated by

\[
y_t = \mu_0 + \mu_1 t + x_t, \quad t = 1, 2, \ldots,
\]

where \( \mu_0(n \times 1) \) and \( \mu_1(n \times 1) \) are unknown parameters which may, of course, be zero and \( x_t \) is an unobservable error process. Assume that \( x_t \) follows a \( p \)th-order vector autoregressive (VAR) process

\[
x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t, \quad t = 1, 2, \ldots,
\]

where the \( A_j \) are \( (n \times n) \) coefficient matrices. This process can be written in error correction (EC) form as

\[
\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j A x_{t-j} + \varepsilon_t, \quad t = 1, 2, \ldots,
\]

where \( \Pi = -(I_n - A_1 - \cdots - A_p) \) is \( (n \times n) \), the \( \Gamma_j = -(A_{j+1} + \cdots + A_p) \) \( (n \times n) \) are unknown parameters and \( \Delta \) is the usual differencing operator. Here and in the following a sum is zero if the lower bound for the summation index exceeds its upper bound, e.g., if in (2.3) \( p = 1 \). We also assume that the error term \( \varepsilon_t \) is white noise, that is, \( \varepsilon_t \sim (0, \Omega) \) with \( \Omega \) positive definite. Later on we will make more explicit assumptions for the process \( \varepsilon_t \).

Suppose the components of the process \( x_t \) are integrated of order one and cointegrated so that we can write \( \Pi = \pi \beta' \), where \( \pi \) and \( \beta \) are \( (n \times r) \) matrices of full column rank and \( 0 \leq r < n \). Here \( r \) is the cointegrating rank. In order to exclude processes integrated of order two we require that the characteristic equation

\[
det(I_n - A_1 z - \cdots - A_p z^p) = \det \left( I_n - \sum_{j=1}^{p-1} \Gamma_j z^j \right) (1 - z) - \pi \beta' z = 0
\]
has exactly \( n - r \) roots equal to one and all other roots outside the unit circle. As is well known, \( \beta \) is then an (asymptotically) stationary process with a zero mean value (see Engle and Granger, 1987; Johansen, 1991). In fact, if we define \( \Psi = I_n - \Gamma_1 - \cdots - \Gamma_{p-1} = I_n + \sum_{j=1}^{p-1} jA_{j+1} \) we can conclude from Johansen’s (1991) formulation of Granger’s representation theorem that

\[
x_t = C \sum_{i=1}^{r} \varepsilon_i + \zeta_i, \quad t = 1, 2, \ldots, \tag{2.4}
\]

where, apart from the specification of initial values, \( \zeta_i \) is a stationary process and \( C = \beta_1 (x'_1 \Psi \beta_1)^{-1} x'_1 \). Here as well as below, if \( B \) is an \((n \times m)\) matrix of full column rank \((n > m)\) we let \( B_\perp \) stand for an orthogonal complement, that is, \( B_\perp \) is an \((n \times (n-m))\) matrix of full column rank and such that \( B' B_\perp = 0 \). The orthogonal complement of a nonsingular square matrix is zero and the orthogonal complement of zero is an identity matrix of suitable dimension.

Note that the present form of the DGP, where the trend is added to the stochastic part, has several advantages for our purposes. Among these advantages is its emphasis on the fact that the trend is at most linear. Starting from the DGP (2.1)/(2.2) implies that the observations \( y_t \) also have a VAR(\( p \)) representation

\[
y_t = y_0 + v_t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \quad t = p + 1, p + 2, \ldots, \tag{2.5}
\]

where \( y_0 = -\Pi \mu_0 + (\Psi + \Pi) \mu_1 \) and \( v_t = -\Pi \mu_1 \). Alternatively, (2.5) may be written in EC form

\[
A y_t = v_0 + v_t + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j A y_{t-j} + \varepsilon_t, \quad t = p + 1, p + 2, \ldots, \tag{2.6}
\]

or

\[
A y_t = v + \alpha(\beta' y_{t-1} - \tau(t-1)) + \sum_{j=1}^{p-1} \Gamma_j A y_{t-j} + \varepsilon_t, \quad t = p + 1, p + 2, \ldots, \tag{2.7}
\]

where \( v = -\Pi \mu_0 + \Psi \mu_1 \) and \( \tau = \beta' \mu_1 \).

In the above model we are interested in testing for a specific cointegrating rank \( r_0 \), that is, a test of the null hypothesis

\[
H(r_0): \ rk(\Pi) = r_0 \quad \text{vs.} \quad \bar{H}(r_0): \ rk(\Pi) > r_0 \tag{2.8}
\]

is desired. In the following sections we will discuss different classes of suitable tests.

3. Previously considered LR tests

For a sample \( y_1, \ldots, y_T \), LR tests for the pair of hypotheses in (2.8) are available under different assumptions regarding the deterministic terms. For
instance, if no prior knowledge regarding $\mu_0$ and $\mu_1$ is available and $v_0$ and $v_1$ are left unrestricted, the LR test may be computed from (2.6) by using reduced rank (RR) regression. We denote the resulting test statistic by $LR^{PC}(r_0)$, because its asymptotic distribution was derived by Perron and Campbell (1993). Its advantage is that it does not depend on the actual values of $\mu_0$ and $\mu_1$. In fact, the asymptotic distribution remains unchanged if the actual parameter vectors are zero. Note, however, that leaving the parameters $v_0$ and $v_1$ unconstrained means that the model (2.6) can in principle even generate quadratic trends. In the above estimation procedure no restrictions are imposed on $v_0$ and $v_1$. Also the cointegrating rank specified under the null hypothesis is not taken into account in estimating the trend parameters. Critical values for the resulting $LR^{PC}$ test are given in Table 1 of Perron and Campbell (1993).

As an alternative, Johansen (1994) proposes to base the LR test on the EC model (2.7) and thereby imposes the restriction that the deterministic trend is at most linear. We will denote the resulting test statistic by $LR^+(r_0)$. Its asymptotic distribution differs from that of $LR^{PC}(r_0)$. Critical values are given, for instance, in Table 15.4 of Johansen (1995).

S&L show that the local power of $LR^+$ is better than that of $LR^{PC}$ in a large part of the parameter space. Clearly this is an advantage of the former test. Still the local power of both tests is poor relative to other tests for the cointegrating rank which use more restrictive assumptions for the deterministic terms, e.g., by ruling out linear trends when appropriate. Therefore it is of interest to investigate whether more powerful tests can be obtained. We will explore one possible approach for constructing alternative tests in the following.

4. LM-type tests

4.1. Preliminaries

The following derivation of LM-type tests for the cointegrating rank is implicitly based on Gaussian assumptions where necessary. For instance, when we talk about maximum likelihood (ML) estimation, the Gaussian likelihood function is assumed. The asymptotic distribution of the test statistic obtained in this way holds under more general conditions, however. They are stated explicitly in Theorem 1 below. The idea is to take into account all restrictions available under the null hypothesis in estimating the trend parameters. In particular, it is taken into account that the trend can be at most linear and not quadratic. Moreover, the cointegrating rank $r_0$, as assumed under $H(r_0)$, is imposed in estimating $\mu_0$ and $\mu_1$. It is hoped that a test which respects the restrictions has more power than tests which do not incorporate all possible restrictions. In Section 5 it will be shown that the LM-type tests presented in the following have indeed better local power than the competing LR tests.
Recalling that, under $H(r_0)$, $\Pi = z\beta'$, where $z$ and $\beta$ are $(n \times r_0)$, and using that $\beta(\beta' \beta)^{-1} \beta' + \beta\beta_\perp(\beta_\perp' \beta_\perp)^{-1}\beta_\perp = I_n$ we can write (2.3) as

$$Ax_t = \kappa u_{t-1} + \rho v_{t-1} + \sum_{j=1}^{p-1} \Gamma_j Ax_{t-j} + \varepsilon_t, \quad t = 1, 2, \ldots,$$

(4.1)

where $u_t = \beta'x_t$, $v_t = \beta_\perp'x_t$, $\kappa = \Pi \beta(\beta' \beta)^{-1}$ and $\rho = \Pi \beta_\perp(\beta_\perp' \beta_\perp)^{-1}$. If $r = r_0 = rk(\Pi)$ and thus the null hypothesis in (2.8) holds, we clearly have $\rho = 0$. However, under the alternative some columns of $\beta_\perp$ can be identified with cointegrating vectors so that $\rho \neq 0$. A further simplification is obtained by premultiplying (4.1) by $\chi_\perp$, considering

$$\chi_\perp Ax_t = \rho_\ast v_{t-1} + \sum_{j=1}^{p-1} \Gamma_{\ast j} Ax_{t-j} + \varepsilon_{\ast t}, \quad t = 1, 2, \ldots,$$

(4.2)

and testing the null hypothesis that $\rho_\ast = \chi_\perp \rho = 0$. Therefore, one of our test procedures for the hypothesis $H(r_0)$ is based on testing the restriction $\rho_\ast = 0$ in a feasible version of (4.2). It may be worth noting that LR tests can also be formulated in such a way that a similar transformation as in (4.2) is used for the EC model on which the tests are based (see S&L). A difference between the LM-type tests considered in the following and a corresponding LR test is that the parameters $z$ and $\beta$ as well as their orthogonal complements are estimated in a different way.

In setting up a feasible version of (4.2) the first step is to trend-adjust the observations $y_t$. For that purpose we estimate the trend parameters $\mu_0$ and $\mu_1$ and then consider $\hat{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t$. As mentioned earlier, the approach for estimating $\mu_0$ and $\mu_1$ assumes that $r = r_0$ as specified in $H(r_0)$. It has some similarities to the procedure employed by Saikkonen and Luukkonen (1997) in the special case where (2.1) involves the a priori restriction $\mu_1 = 0$. It was shown in that paper that a generalized least-squares (GLS) estimator of the level parameter $\mu_0$ is consistent of order $O_p(T^{-1/2})$ in the direction of cointegrating vectors whereas in other directions the estimator is not consistent but bounded in probability. A similar situation appears in the present context. Suppose the hypothesis $H(r_0)$ holds and that the parameters in (2.7) are estimated by Johansen’s (1994) RR method. Instead of estimating the complete trend parameter $\mu_1$ we thereby get an estimator of the parameter $\tau = \beta'\mu_1$ only, that is, $\mu_1$ in the direction of the cointegrating vectors can be estimated. However, the considered model does not contain $\mu_1$ in the direction of $\beta_\perp$ so that this component of $\mu_1$ cannot be estimated directly in this context. Therefore, estimation of the complete trend parameter $\mu_1$ is problematic.

In this situation one may consider estimating $\mu_0$ and $\mu_1$ simply by regressing $y_t$ on a constant and the time variable $t$ as proposed by Stock and Watson (1988) for computing their test statistics. Such a procedure is inefficient here, however, because it ignores not only the short-term dynamics but also the long-term
dynamics and the cointegrating rank assumed under the null hypothesis. Moreover, in Stock and Watson (1988) it is seen to have an impact on the asymptotic distribution of the test statistics. Therefore a different procedure is used in the following.

The first step of our test procedure consists of using the RR method to estimate the parameters in (2.7). The resulting estimators will be indicated by the symbol “~” and, under the null hypothesis $H(r_0)$, they are consistent. Based on these estimators we will then obtain suitable estimators of $k_0$ and $k_1$ and use the resulting detrended $y_t$ in a feasible version of (4.2) to construct an ‘LM-type test’ of the null hypothesis $\rho^* = 0$. We use the term LM-type test here primarily because the trend parameters are estimated under the null hypothesis which distinguishes our procedure from the LR tests discussed in Section 3.

### 4.2. Estimation of trend parameters

Our test procedures require estimators of the trend parameters $k_0$ and $k_1$ with appropriate asymptotic properties under the null hypothesis. As already pointed out above, a major problem is to obtain reasonable estimators in the direction of $\beta$. Thus, we shall first discuss the estimation of the parameters $\delta_\# = \beta'_1 \mu_0$ and $\tau_\# = \beta'_1 \mu_1$. Our estimation procedures are based on the initial value assumption $x_t = 0, t \leq 0$, and similarly $y_t = 0, t \leq 0$. It should be noted, however, that our asymptotic results are also valid with other initial value assumptions provided the distribution of the initial values is independent of the sample size.

One possible estimation method is based on the fact that the expectation of $\beta'_1 \delta_\#$ equals $\delta_\#$ for $t = 1$ and $\tau_\#$ for $t > 1$. Thus, we consider estimating $\delta_\#$ and $\tau_\#$ by LS from the regression model

$$\tilde{\beta}'_1 y_t = \tau_\# + \delta_\# x_t + e_t, \quad t = 1, \ldots, T,$$

(4.3)

where $t_1 = 1$ and $x_t = 0$ for $t > 1$. Here any matrix $\tilde{\beta}'_1$ satisfying $\tilde{\beta}'_1 \tilde{\beta}'_1 = I_{n-r_0}$ and $\tilde{\beta}'_1 \beta = 0$ may be used with $\tilde{\beta}$, the RR estimator from (2.7). The error term equals $e_t = \beta'_1 x_t + \tilde{\beta}'_1 y_t - \beta'_1 y_t$. Of course, since $x_t = 0$ for $t > 1$, $\delta_\#$ is essentially just estimated from the first observation in (4.3) and therefore it cannot be estimated consistently. On the other hand, although the parameter $\tau_\#$ is not contained in model (2.7), it can be identified as the expectation of $\beta'_1 y_t$ and, hence, it can be estimated consistently. The precise properties of the estimators $\hat{\tau}_\#$ and $\hat{\delta}_\#$ resulting from (4.3) are given in Lemma 1 in the appendix. These considerations lead to an estimator of the parameters $\mu_1$ given by

$$\hat{\mu}_1 = \tilde{\beta}(\beta' \beta)^{-1} \hat{\tau}_\# + \tilde{\beta}'_1 (\beta'_1 \beta'_1)^{-1} \hat{\tau}_\#.$$  

(4.4)

Note that the right-hand side is invariant to normalizations of $\tilde{\beta}$ and $\tilde{\beta}'_1$.

Constructing an estimator for $\mu_0$ is slightly more complicated because, unlike $\mu_1$, an estimator of $\mu_0$ in the direction of $\beta$ is not directly obtained from the estimation of the null model. Denote $\delta = \beta' \mu_0$ and notice that $v = -\delta + \Psi \mu_1$, 

$$\mu_0 = \tilde{\beta}(\beta' \beta)^{-1} \hat{\tau}_\# + \tilde{\beta}'_1 (\beta'_1 \beta'_1)^{-1} \hat{\tau}_\#.$$  

(4.4)
where \( v, \alpha \) and \( \Psi \) can be estimated using (2.7). Thus, using the estimators \( \tilde{\mu}_1 \) and \( \tilde{\Psi} = I_n - \tilde{\Gamma}_1 - \cdots - \tilde{\Gamma}_{p-1} \) we introduce the estimator

\[
\delta = (\tilde{z}'\tilde{\Theta}^{-1}\tilde{z})^{-1}\tilde{z}'\tilde{\Theta}^{-1}I(\tilde{\Psi}\tilde{\mu}_1 - \tilde{\nu}).
\]  

(4.5)

This estimator is motivated by the analogous estimator of the intercept term discussed in Saikkonen (1992). It is seen in the appendix that \( \tilde{\mu}_1 = \mu_1 + O_p(T^{-1/2}) \) so that, since all the other estimators in the definition of \( \delta \) are known to be consistent of order \( O_p(T^{-1/2}) \), we have \( \delta = \delta + O_p(T^{-1/2}) \). Similarly to (4.4) we now define

\[
\tilde{\mu}_0 = \tilde{\mu}(\tilde{\beta}'\tilde{\beta})^{-1}\delta + \tilde{\beta}_1(\tilde{\beta}_1'\tilde{\beta}_1)^{-1}\tilde{\zeta}.
\]  

(4.6)

Thereby we have a first method to obtain estimators of the trend parameters.

As is well known, the LS estimator of (4.3) is asymptotically as efficient as optimal GLS estimation. From an asymptotic point of view the estimators \( \tilde{\tau}_* \) and \( \tilde{\delta}_* \) are thus quite reasonable. However, since they ignore all short-run dynamics of the process \( \beta'\Delta y_t \), it may be preferable to consider alternative estimators which also try to allow for the short-run dynamics. One possibility is to fit an autoregressive model to the error term in (4.3) and apply a feasible GLS estimation method. Although in this situation a finite order autoregressive model will only be an approximation, this approach might still have useful finite sample properties. Below we shall consider an alternative approach in which an autoregressive approximation of this kind is not involved. The ideas are somewhat similar to those employed in the GLS estimation of Saikkonen and Luukkonen (1997).

First, note that the levels parameters of the VAR process (2.2) may be obtained as \( A_1 = I_n + \alpha \beta' + \Gamma_1, \ A_j = \Gamma_j - \Gamma_{j-1} \) (\( j = 2, \ldots, p - 1 \)), \( A_p = - \Gamma_{p-1} \). The corresponding sample analogs \( \tilde{A}_1, \ldots, \tilde{A}_p \) are defined similarly by using the estimators \( \tilde{z}, \tilde{\beta} \) and \( \tilde{\Gamma}_j \) (\( j = 1, \ldots, p - 1 \)). Hence, the \( \tilde{A}_j \) satisfy the restrictions implied by the cointegrating rank.

To determine the estimators \( \tilde{\mu}_0 \) and \( \tilde{\mu}_1 \) proposed in the following, consider the trend adjusted series \( \tilde{\xi}_t = y_t - \tilde{\mu}_0 - \tilde{\mu}_1 t \) for \( t \geq 1 \) and \( \tilde{\xi}_t = 0 \) for \( t \leq 0 \) and define \( \tilde{\xi}_t = (I_n - \tilde{A}_1 L - \cdots - \tilde{A}_p L^p)\tilde{\xi}_t = \tilde{A}(L)\tilde{\xi}_t \), where \( L \) is the lag operator and \( \tilde{A}(L) \) is defined in the obvious way. Now consider the identity

\[
\tilde{\xi}_t = \tilde{A}(L)(\tilde{\xi}_t - \chi_t) + (\tilde{A}(L) - A(L))\chi_t + \varepsilon_t
\]

(4.7)

where \( \varepsilon_t^{aux} = (\tilde{A}(L) - A(L))\chi_t + \varepsilon_t, \tilde{G}_t = \tilde{A}(L)a_t \) and \( \tilde{H}_t = \tilde{A}(L)b_t \) with

\[
a_t = \begin{cases} 1 & \text{for } t \geq 1, \\ 0 & \text{for } t \leq 0 \end{cases} \quad b_t = \begin{cases} t & \text{for } t \geq 1, \\ 0 & \text{for } t \leq 0. \end{cases}
\]
One might estimate the differences $\mu_0 - \tilde{\mu}_0$ and $\mu_1 - \tilde{\mu}_1$ from the auxiliary regression model in (4.7) by appropriate feasible GLS and use the resulting estimators to correct the estimators $\tilde{\mu}_0$ and $\tilde{\mu}_1$. The corrected estimators can be interpreted as two-step versions of ML estimators based on (2.1), (2.3), $rk(II) = r_0$ and the initial value assumption $x_t = 0$ for $t \leq 0$. However, since we wish to estimate the parameters $\mu_0$ and $\mu_1$ in the direction of $\beta_1$ we shall apply this approach after a further transformation given by

$$K_t = \tilde{K}_t\phi_0 + \tilde{K}_2t\phi_1 + e_t^{(1)}, \quad t = 1, \ldots, T,$$

where $\tilde{K}_t = \tilde{z}_t'\tilde{e}_t$, $\tilde{K}_1 = \tilde{z}_t'\tilde{G}_1\tilde{p}(\tilde{p}_1)_{\tilde{L}}^{-1} \tilde{K}_2 = \tilde{z}_1'\tilde{H}_1\tilde{p}(\tilde{p}_1)_{\tilde{L}}^{-1}$, $\phi_0 = \tilde{p}_1(\mu_0 - \tilde{\mu}_0)$ and $\phi_1 = \tilde{p}_1(\mu_1 - \tilde{\mu}_1)$. Here $\tilde{z}_1$ is any $(n \times (n - r_0))$ matrix of rank $n - r_0$ satisfying $\tilde{z}_1'\tilde{z} = 0$ and note that it follows from the definitions that $\tilde{K}_1 = 0$ and $\tilde{K}_2 = \tilde{z}_1'\tilde{\Phi}\tilde{p}(\tilde{p}_1)_{\tilde{L}}^{-1}$ for $t \geq p + 1$. The precise definition of the error term $e_t^{(1)}$ is of limited importance here. It may be worth noting, however, that we have approximately $e_t^{(1)} \approx \tilde{z}_t'\tilde{e}_t$. Thus, it seems reasonable to estimate the parameter vector $\phi = [\phi_0', \phi_1']'$ in (4.8) by the GLS estimator

$$\hat{\phi} = \left[\tilde{z}_t'\tilde{G}_1\tilde{p}(\tilde{p}_1)_{\tilde{L}}^{-1}\tilde{K}_1\right]^{-1} \sum_{t=1}^{T} \tilde{z}_t'\tilde{G}_1\tilde{p}(\tilde{p}_1)_{\tilde{L}}^{-1}\tilde{e}_t,$$

where $\tilde{K}_t = [\tilde{K}_1, \tilde{K}_2]$. Then new estimators of $\delta_*$ and $\tau_*$ can finally be defined by

$$\tilde{\delta}_*^{(1)} = \tilde{\delta}_* + \hat{\phi}_0 \quad \text{and} \quad \tilde{\tau}_*^{(1)} = \tilde{\tau}_* + \hat{\phi}_1.$$

Thus, the estimators $\tilde{\delta}_*^{(1)}$ and $\tilde{\tau}_*^{(1)}$ are obtained by correcting the initial estimators $\tilde{\delta}_*$ and $\tilde{\tau}_*$. Asymptotically, this correction has no effect (see Lemma 2 of the appendix). This result is quite reasonable because, as already pointed out, the asymptotic properties of the estimators $\tilde{\delta}_*$ and $\tilde{\tau}_*$ cannot be improved by taking the covariance structure of the error term $e_t$ in (4.3) into account. In finite samples the situation may be different, however, and the fact that the error term $e_t^{(1)}$ in (4.8) is approximately white noise may make the two-step estimator preferable to $\tilde{\delta}_*$ and $\tilde{\tau}_*$. Of course, it is also possible to use the estimators $\tilde{\delta}_*^{(1)}$ and $\tilde{\tau}_*^{(1)}$ as initial estimators and obtain similar asymptotically equivalent three-step estimators or even higher-order iterates. In what follows, we shall use the notation $\tilde{\delta}_*$ and $\tilde{\tau}_*$ to signify any of these asymptotically equivalent estimators. The same notational convention applies to the estimators $\tilde{\mu}_0$ and $\tilde{\mu}_1$ defined in (4.6) and (4.4).

We close this subsection by noting that the above estimators can be straightforwardly modified to the case where it is assumed that $\beta_1'\mu_1 = 0$ and, hence, no deterministic trend term appears in model (2.5), i.e. $v_1 = 0$. Then we have $\tau = 0$ and in the above formulas we simply define $\tilde{\tau} = 0$. It is easy to check that the derivations used in the proofs in the appendix remain valid under this assumption.
4.3. Test statistics

As discussed in Section 4.1, a reasonable approach for testing the hypothesis $H(r_0)$ is to test the constraint $\rho_* = 0$ in a feasible version of the regression model (4.2). Hence, we consider

$$\tilde{z}_{t, i} A \tilde{x}_t = \rho_* \tilde{u}_{t-1} + \sum_{j=1}^{p-1} \Gamma_{*j} A \tilde{x}_{t-j} + \eta_*, \quad t = p + 1, \ldots, T,$$

(4.9)

where $\tilde{x}_t = y_t - \tilde{\mu}_0 - \tilde{\mu}_1 t$ with $\tilde{\mu}_0$ and $\tilde{\mu}_1$ being any of the estimators considered in the previous subsection. Conventional test statistics for the hypothesis $\rho_* = 0$ may be based on an LS estimation of (4.9). For instance, the resulting LM statistic is

$$LM(r_0) = tr\{\hat{\rho}_* \hat{M}_{tv, AX} \hat{\rho}_* (\tilde{z}_{t, i} \tilde{z}_{t, i})^{-1}\},$$

(4.10)

where $\hat{\rho}_*$ is the LS estimator of $\rho_*$ from (4.9), $\tilde{z}_{t, i} \tilde{z}_{t, i} \approx \hat{\sigma}_n$ is the residual covariance estimator of the error term in (4.9) and

$$\hat{M}_{tv, AX} = \left[ \sum_{t=p+1}^{T} \tilde{u}_{t-1} \tilde{u}_{t-1}' - \sum_{t=p+1}^{T} \tilde{u}_{t-1} A \tilde{X}_{t-1}' \right.$$

$$\times \left( \sum_{t=p+1}^{T} A \tilde{X}_{t-1} \tilde{X}_{t-1}' \right)^{-1} \sum_{t=p+1}^{T} A \tilde{X}_{t-1} \tilde{u}_{t-1}' \]^{-1}

(4.11)

with $A \tilde{X}_{t-1} = (A \tilde{x}_{t-1, 1}, \ldots, A \tilde{x}_{t-1, p+1})'$.

An asymptotically equivalent test statistic results by estimating $\rho_*$ from

$$\tilde{z}_{t, i} A \tilde{x}_t = \kappa_\beta \tilde{u}_{t-1} + \rho_* \tilde{v}_{t-1} + \sum_{j=1}^{p-1} \Gamma_{*j} A \tilde{x}_{t-j} + \tilde{\eta}_*, \quad t = p + 1, \ldots, T,$$

(4.12)

where $\tilde{u}_t = \beta \tilde{x}_t$ and $\kappa_\beta = \tilde{z}'_\beta \kappa$ is zero under the null hypothesis (see (4.1)). In the present formulation of the model $\kappa_\beta$ is estimated unrestrictedly together with the other parameters. Denoting the corresponding estimator of $\rho_*$ by $\hat{\rho}_*$ gives a test statistic

$$LM_\beta(r_0) = tr\{\hat{\rho}_* \hat{M}'_{tv, AX} \hat{\rho}_* (\tilde{z}_{t, i} \tilde{z}_{t, i})^{-1}\}. $$

(4.13)

Here $\hat{M}'_{tv, AX}$ is defined analogously to (4.11) with $A \tilde{X}_{t-1} = (\tilde{u}_{t-1}, A \tilde{x}_{t-1, 1}, \ldots, A \tilde{x}_{t-1, p+1})'$ replacing $A \tilde{X}_{t-1}$ in that formula.

The following theorem gives the limiting distribution of the test statistics under the null hypothesis. We use the symbol $B(s)$ to signify an $(n - r_0)$-dimensional standard Brownian motion and $B_*(s) = B(s) - sB(1)$ $(0 \leq s \leq 1)$ denotes an $(n - r_0)$-dimensional standard Brownian bridge.
Theorem 1. Suppose that the $e_t$ in (2.2) are a martingale difference sequence with $E(e_t|e_s, s < t) = 0$, $E(e_t^4|e_s, s < t) = \Omega$ is positive definite and the fourth moments are bounded. Then, under the null hypothesis,

$$LM(r_0), LM_{\#}(r_0) \overset{d}{\to} tr\left\{ \left( \int_0^1 B_{\#}(s) dB_{\#}(s)' \right)^{-1} \left( \int_0^1 B_{\#}(s)B_{\#}(s)' ds \right)^{-1} \times \left( \int_0^1 B_{\#}(s) dB_{\#}(s)' \right)^{-1} \left( \int_0^1 B_{\#}(s) dB_{\#}(s)' \right)^{-1} \right\},$$

where $dB_{\#}(s) = dB(s) - ds B(1)$ and an integral with respect to the semimartingale $B_{\#}(s)$ is understood to be an Itô integral.

Note that the integral $\int_0^1 B_{\#}(s) dB_{\#}(s)'$ may be interpreted as a short-hand notation for $\int_0^1 B(s) dB(s)' - B(1)|_s B(s)' - \int_0^1 B(s) dB(1)' + \frac{1}{2} B(1) B(1)'$. The proof of the theorem is given in the appendix. Notice that the limiting distribution of the test statistics is free of unknown nuisance parameters. This means, in particular, that it does not depend on the actual trending properties of the process. Note also that we have not assumed that $\mu_0$ and $\mu_1$ are actually nonzero. The percentiles of the asymptotic distributions can be readily found by simulation. We will provide the details in Section 4.4. In the special case where $n - r_0 = 1$ the limiting distribution simplifies, as shown in Schmidt and Phillips (1992), because in that case the stochastic integral in Theorem 1 equals $-\frac{1}{2}$. The following corollary deals with this special case.

Corollary 1. If $n - r_0 = 1$ then, under the null hypothesis,

$$LM(r_0), LM_{\#}(r_0) \overset{d}{\to} \left( 4 \int_0^1 B_{\#}(s)^2 ds \right)^{-1}.$$

The limiting distribution in Corollary 1 was previously obtained by Ahn (1993, Theorem 1) who developed LM tests for testing the univariate autoregressive unit root hypothesis in the presence of a time trend. He also simulated asymptotic critical values of the test statistic in this special case.

Now consider the case where $\nu_1 = 0$ holds and, thus, a deterministic trend term does not appear in (2.5). Then $\tau = 0$ and the estimator $\hat{\mu}_1$ is defined accordingly. It can be seen from the proof of Theorem 1 that in this case the same proof applies and we get the same limiting distribution of the resulting test statistics in this case. However, we do not pursue this case here.

One might also augment the auxiliary regression equations (4.9) and (4.12) by including an intercept term. In the related univariate case this approach has been used by Schmidt and Phillips (1992) and Ahn (1993). In the same way as in these special cases, the limiting distribution of the test statistics will then change and involve a demeaned Brownian bridge.
4.4. Percentiles of the asymptotic distributions

In Table 1 simulated percentage points of the asymptotic distribution given in Theorem 1 are presented for different values of $d = n - r_0$. These percentage points are simulated in the following manner. We have generated $T = 1000$ $d$-dimensional independent standard normal variates $\varepsilon_t \sim \text{N}(0, I_d)$ and have computed the quantities

$$A_T = \frac{1}{T^2} \sum_{t=1}^{T} \left[ \sum_{k=1}^{t-1} (\varepsilon_k - \bar{\varepsilon}) \right] \left[ \sum_{k=1}^{t-1} (\varepsilon_k - \bar{\varepsilon})' \right],$$

and

$$B_T = \frac{1}{T} \sum_{t=1}^{T} \left[ \sum_{k=1}^{t-1} (\varepsilon_k - \bar{\varepsilon}) \right] (\varepsilon_t - \bar{\varepsilon})',

$$

where $\bar{\varepsilon} = T^{-1} \sum_{t=1}^{T} \varepsilon_t$. The quantity in (4.14) converges weakly to $\int_0^1 \text{B}_d(s) \text{B}_d(s)' ds$ whereas (4.15) converges to $\int_0^1 \text{d} \text{B}_d(s) \text{d} \text{B}_d(s)'$. Hence, the quantity $\text{tr} (B_T A_T^{-1} B_T)$ has the desired asymptotic distribution. The percentiles in Table 1 are obtained from 20,000 replications of this experiment. Independent drawings are used for each dimension $d$.

5. Local power analysis

In this section we will give the limiting distributions of the LM-type tests under local alternatives and compare the resulting local power to that of the LR tests $LR^+$ and $LR^{PC}$. We consider local alternatives of the form

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

$$

where $\alpha$ and $\beta$ are fixed $(n \times r_0)$ matrices of rank $r_0$ and $\alpha_1$ and $\beta_1$ 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 $r$. In the following we use the assumptions from Johansen (1995) and Rahbek (1994) that the VAR order is $p = 1$ and the

<table>
<thead>
<tr>
<th>Dimension</th>
<th>$d$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage point</td>
<td>90%</td>
<td>5.43</td>
<td>13.89</td>
<td>25.90</td>
<td>42.03</td>
<td>61.81</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>6.83</td>
<td>15.92</td>
<td>28.47</td>
<td>45.13</td>
<td>65.69</td>
</tr>
<tr>
<td></td>
<td>99%</td>
<td>10.19</td>
<td>20.37</td>
<td>33.54</td>
<td>51.27</td>
<td>73.57</td>
</tr>
</tbody>
</table>


eigenvalues of the matrices $I_{r_0} + \beta\alpha$ and $I_r + [\beta; \beta_1][\alpha; \alpha_1]$ are less than 1 in modulus.

The power of our LM-type tests against alternatives of the form (5.1) may be obtained from a general result given in S&L. To apply that result we consider the RR regression model

$$\Delta \tilde{x}_t = \alpha\beta' \tilde{x}_{t-1} + \tilde{e}_t, \quad t = 1, \ldots, T,$$

(5.2)

where $\tilde{x}_t = y_t - \tilde{\mu}_0 - \tilde{\mu}_1 t$, $\tilde{e}_t = e_t - \alpha\beta'(\tilde{\mu}_0 - \mu_0) + \alpha\beta'(\tilde{\mu}_1 - \mu_1)(t - 1) + T^{-1}\alpha_1\beta_1'(\tilde{\mu}_0 - \mu_0) + T^{-1}\alpha_1\beta_1'(\tilde{\mu}_1 - \mu_1)(t - 1) - (\tilde{\mu}_1 - \mu_1) + T^{-1}\alpha_1\beta_1' \tilde{x}_{t-1}$ and $\tilde{\mu}_0$ and $\tilde{\mu}_1$ are any of the estimators from Section 4.2. Choosing $Y_t = \Delta \tilde{x}_t$, $X_t = \tilde{x}_{t-1}$, $\tilde{e}_t = \tilde{e}_t$, $A = \alpha$ and $B = \beta$, we may obtain the asymptotic distribution of our LM-type statistics from Theorem 1 of S&L. It can be shown that for this particular case the limiting distribution reduces to the form given in the next theorem. The result is similar to the corresponding results in S&L in that under local alternatives an Ornstein–Uhlenbeck process replaces the Brownian motion used in the null distribution in our Theorem 1. The proof can be obtained in a similar way as in S&L. The details are given in Lütkepohl and Saikkonen (1997).

**Theorem 2.** Let $W(u)$ denote a Brownian motion with covariance matrix $\Omega$ and $K(t)$ the Ornstein–Uhlenbeck process defined by the integral equation

$$K(u) = \int_0^u K(s) \, ds \quad (0 \leq u \leq 1).$$

(5.3)

Furthermore, define $N(s) = (\alpha_1' \Omega \alpha_1)^{-1/2} K(s)$, $N_0(s) = N(s) - sN(1)$ and $dN_0(s) = dN(s) - ds N(1)$. With this notation, if $\Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1'$ in (2.3),

$$LM(r_0) \overset{d}{=} tr\left\{\left(\int_0^1 N_0(s) dN_0(s)\right)\left(\int_0^1 N_0(s) N_0(s)' \, ds\right)^{-1}\right\} \times \left(\int_0^1 N_0(s) dN_0(s)\right).$$

(5.4)

Obviously, the local power depends on $n - r_0$ only and not on the dimension $n$ and $r_0$ separately. Moreover, it does not depend on the actual values of the mean and trend parameters $\mu_0$ and $\mu_1$ as in the case of the LR tests studied in S&L. Furthermore, it is straightforward to check that

$$N(s) = B(s) + ab' \int_0^s N(u) \, du,$$

(5.5)

where $B(s)$ is again an $(n - r_0)$-dimensional standard Brownian motion and the quantities $a$ and $b$ are given by $a = (\alpha_1' \Omega \alpha_1)^{-1/2} \alpha_1' \alpha_1$ and $b = (\alpha_1' \Omega \alpha_1)^{1/2} (\beta_1' \alpha_1)^{-1}\beta_1' \beta_1$ (cf. Johansen, 1995, pp. 207–208). Hence, the limiting distribution depends on $\alpha$, $\beta$, $\Omega$, $\alpha_1$ and $\beta_1$ only through $a$ and $b$. This implies, for instance,
for the case \( r - r_0 = 1 \), where \( \alpha_1 \) and \( \beta_1 \) are just vectors, that the limiting distribution only depends on the two parameters \( f = b' \alpha \) and \( g^2 = d'a'b' - (a'b)^2 \) (see Johansen, 1995, Corollary 14.5).

Since the asymptotic distribution of the LM-type test given in Theorem 2 differs from those of the LR tests for the cointegrating rank given in Corollary 1 of S&L we have simulated the resulting local power in the same way as in S&L. In other words, we consider the case where \( \alpha_1 \) and \( \beta_1 \) are \((n \times 1)\) vectors and simulate the discrete time counterpart of the \((n - r_0)\)-dimensional Ornstein–Uhlenbeck process \( \mathcal{N}(s) \) as \( \Delta \mathcal{N}_t = (1/T) \alpha_1 \beta_1' \mathcal{N}_{t-1} + \epsilon_t \), \( t = 1, \ldots, T = 1000 \), with \( \epsilon_t \sim \text{iid } N(0, I_{n-r_0}) \), \( \mathcal{N}_0 = 0 \). From the \( \mathcal{N}_t \) we compute

\[
A_T = \frac{1}{T^2} \sum_{t=1}^T \left[ \sum_{k=1}^{t-1} (\Delta \mathcal{N}_k - \bar{\Delta \mathcal{N}}) \right] \left[ \sum_{k=1}^{t-1} (\Delta \mathcal{N}_k - \bar{\Delta \mathcal{N}}) \right]',
\]

and

\[
B_T = \frac{1}{T} \sum_{t=1}^T \left[ \sum_{k=1}^{t-1} (\Delta \mathcal{N}_k - \bar{\Delta \mathcal{N}}) \right] (\Delta \mathcal{N}_t - \bar{\Delta \mathcal{N}})',
\]

where \( \bar{\Delta \mathcal{N}} = T^{-1} \sum_{t=1}^T \Delta \mathcal{N}_t \). The quantity in (5.6) converges weakly to \( \int_0^1 \mathcal{N}_s(\mathcal{N}_s)' \) and (5.7) converges to \( \int_0^1 \mathcal{N}_s(\mathcal{N}_s)' \). Hence, the quantity \( \text{tr}(B_T A_T^{-1} B_T) \) has the desired asymptotic distribution. The resulting rejection frequencies for the case \( n - r_0 = 1 \) with \( \alpha_1 = -f(f = 0, 3, 6, \ldots, 30) \) and \( \beta_1 = 1 \) are plotted in Fig. 1 together with the corresponding rejection frequencies of the competing LR tests \( LR^+ \) and \( LR^{PC} \). The rejection frequencies of the latter tests are generated in a similar way as in S&L. It is seen that the LM-type tests have substantially more local power than \( LR^+ \) and \( LR^{PC} \). For instance, for \( f = 15 \), \( LR^{PC} \) rejects in 36% of the replications whereas the rejection frequency for \( LM \) is 57%. We have also performed simulations for \( n - r_0 > 1 \) and found that the power gains are less dramatic if the parameter space has more than one

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{local_power.pdf}
\caption{Local power for \( n - r_0 = 1 \). (--- \( LR^+ \), --- \( LR^{PC} \), \( \cdot \cdot \cdot \) \( LM \)).}
\end{figure}
additional dimension under the alternative. Still some gains were found in these situations as well (see Lütkepohl and Saikkonen, 1997).

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. Given the substantial gains in local power for \( n - r_0 = 1 \) we expect the LM-type tests to perform well also in small samples, however. This issue is investigated in the following section.

6. Small sample properties of the tests

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 special case of a two-dimensional VAR(1) DGP from Toda (1994, 1995) of the form

\[
y_t = \begin{bmatrix} 0 \\ \delta \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix} y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid } N(0, I_2)
\]  

(6.1)
is used to compare the tests. Other forms of VAR(1) processes of interest in practice, e.g. EC forms, 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 two nonstationary components. In that case the second component has a deterministic linear trend if \( \delta \neq 0 \) and there is no linear trend if \( \delta = 0 \). 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 \( \psi_1 \) and \( \psi_2 \) are less than 1 in absolute value the process is stationary \( (r = 2) \). In that case a nonzero \( \delta \) cannot generate a linear trend. The parameter \( \delta \) will be set to zero in the following because the test statistics of interest here are invariant with respect to the value of \( \delta \).

We have generated samples of sizes \( T = 100 \) and 200 plus 50 presample values starting with an initial value of zero. The last presample values are used for estimation purposes. This means, of course, that we effectively use nonzero initial values in the actual samples on which the simulations are based. Thereby we deviate from the assumptions underlying our asymptotic theory and, hence, we can check the robustness of our results with respect to the initial value assumptions made in Section 4.2. Selected rejection frequencies of the tests \( LR^*, LR^{PC}, LM \) and \( LM_\delta \) are given in Tables 2 and 3. For the LM-type tests the trend parameters are estimated by the two-step procedure of Section 4.2, that is, \( \bar{\beta}^{(1)} = \bar{\beta} \bar{\beta}^{-1} \bar{\delta} + \bar{\beta} (\bar{\beta} \bar{\beta})^{-1} \bar{\delta}^{(1)}(\bar{\beta}) \) and \( \bar{\delta}^{(1)} = \bar{\beta} \bar{\beta}^{(1)} \bar{\beta}^{-1} \bar{\delta} + \bar{\beta} (\bar{\beta} \bar{\beta})^{-1} \bar{\delta}^{(1)} \) are used (see Section 4.2).

The rejection frequencies in the tables are based on asymptotic critical values for a test level of 5%. We have not computed rejection frequencies corrected for
Table 2
Relative rejection frequencies of test statistics for DGP with cointegrating rank $r = 0$ ($\psi_1 = \psi_2 = 1$), $\delta = 0$, nominal significance level 0.05

<table>
<thead>
<tr>
<th>Test statistic</th>
<th>$T = 100$</th>
<th></th>
<th></th>
<th>$T = 200$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_0 = 0$</td>
<td>$r_0 = 1$</td>
<td></td>
<td>$r_0 = 0$</td>
<td>$r_0 = 1$</td>
<td></td>
</tr>
<tr>
<td>$LR^+$</td>
<td>0.057</td>
<td>0.007</td>
<td></td>
<td>0.045</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>$LR^{PC}$</td>
<td>0.049</td>
<td>0.006</td>
<td></td>
<td>0.044</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>$LM$</td>
<td>0.038</td>
<td>0.098</td>
<td></td>
<td>0.042</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>$LM_a$</td>
<td>0.038</td>
<td>0.058</td>
<td></td>
<td>0.042</td>
<td>0.037</td>
<td></td>
</tr>
</tbody>
</table>

Table 3
Relative rejection frequencies of test statistics for DGP with cointegrating rank $r = 2$, $\psi_2 = 0.5$, $\delta = 0$, nominal significance level 0.05

<table>
<thead>
<tr>
<th>Test statistic</th>
<th>$T = 100$</th>
<th></th>
<th></th>
<th></th>
<th>$T = 200$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_0 = 0$</td>
<td>$r_0 = 1$</td>
<td></td>
<td>$r_0 = 0$</td>
<td>$r_0 = 1$</td>
<td></td>
<td>$r_0 = 0$</td>
</tr>
<tr>
<td>$LR^+$</td>
<td>0.995</td>
<td>0.077</td>
<td>0.999</td>
<td>0.160</td>
<td>1.000</td>
<td>0.574</td>
<td>1.000</td>
</tr>
<tr>
<td>$LR^{PC}$</td>
<td>0.999</td>
<td>0.084</td>
<td>0.999</td>
<td>0.186</td>
<td>1.000</td>
<td>0.635</td>
<td>1.000</td>
</tr>
<tr>
<td>$LM$</td>
<td>0.895</td>
<td>0.187</td>
<td>0.931</td>
<td>0.354</td>
<td>0.977</td>
<td>0.738</td>
<td>0.995</td>
</tr>
<tr>
<td>$LM_a$</td>
<td>0.895</td>
<td>0.148</td>
<td>0.931</td>
<td>0.322</td>
<td>0.977</td>
<td>0.711</td>
<td>0.995</td>
</tr>
</tbody>
</table>

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 control 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 tables are not independent but can be compared directly. Still it may be worth recalling that for a true rejection probability $\gamma$ the standard error of an estimator based on 1000 replications of the experiment is $s_\gamma = \sqrt{\gamma(1-\gamma)/1000}$. For example, $s_{0.05} = s_{0.95} = 0.007$, and $s_{0.5} = 0.016$. 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 Table 2 results for processes with true cointegrating rank \( r = 0 \) (\( \psi_1 = \psi_2 = 1 \)) are presented. It is seen that both versions of the LM tests are somewhat conservative in this case and thus their size is bounded by the desired 5%. In the previous sections we have not derived asymptotic results for testing \( r_0 = 1 \) when the true rank is \( r = 0 \). Therefore we cannot speculate on what to expect in the present situation when testing \( H(1) \). For the present processes the \( LR^+ \) and \( LR^{PC} \) tests turn out to be conservative in this case whereas the LM tests reject more often than the nominal 5%.

In Table 3 results for DGPs with cointegrating rank \( r = 2 \) (\( \psi_2 = 0.5 \) and \( \psi_1 \) varying) are given. The LM tests are clearly more powerful than \( LR^+ \) and \( LR^{PC} \) for testing \( r_0 = 1 \) if \( \psi_1 \) is close to 1, that is, if the alternative is close to the null. This may be a reflection of the superior local power of the tests. Note that testing \( H(1) \) in a bivariate process with true cointegrating rank \( r = 2 \) corresponds precisely to the situation for which the local power is depicted in Fig. 1. On the other hand, the LR tests are more powerful for \( H(0) \). Here the conservativeness of the LM-type tests observed in Table 2 is a problem. Clearly, in comparing tests with different sizes, it is not surprising that the test with the smaller size rejects less frequently. In other words, the LR tests have an advantage here due to their larger rejection probability if \( H(0) \) is true. From this small simulation experiment it seems that LM-type tests are preferable when higher cointegrating ranks are tested. This observation is also in line with further simulation results which we obtained for other parameter values.

In both tables \( LM \) and \( LM_* \) perform similarly. Overall \( LM \) is slightly superior to \( LM_* \) in terms of power in situations in which the two tests differ. The overall conclusion from our simulations is that in parts of the parameter space the LM tests have better performance than \( LR^+ \) and \( LR^{PC} \) and in other parts the opposite is true. Therefore, in practice one may not want to rely on just one of these tests but decide on the cointegrating rank on the basis of the results from LM and LR tests.

7. Conclusions

Standard tests for the cointegrating rank of a VAR process have nonstandard null distributions which depend on the usually unknown properties of deterministic terms. Therefore, we have considered tests which allow for deterministic linear trend terms. The tests are based on ideas suggested by the LM principle. Unlike standard LR tests, our LM-type tests use all available restrictions under the null hypothesis in estimating the trend parameters. In particular, they use the knowledge that the trend is at most linear and they are based on
the cointegrating rank which is specified under the null hypothesis. Their asymptotic null distributions are nonstandard. However, they do not depend on the actual properties of deterministic terms. It is shown that in parts of the parameter space the LM-type tests have local power superior to LR tests which operate under comparable assumptions.

In a simulation study the LM-type tests were found to have not only attractive local power but also outperform the competing LR tests for processes with linear trends in terms of small sample power in situations when the alternative is close to the null hypothesis, that is, in the region where good power is of particular importance. Moreover, they tend to be conservative and, hence, observe the significance level chosen if the null hypothesis is correct. Since in some cases the LR tests were found to be superior in terms of power it is recommended to perform both types of tests in practice and use the information from all of them in deciding on the cointegrating rank.

Appendix A. Proofs

In the proofs we make use of the fact that the RR (ML) estimators obtained from (2.7) are consistent and, in particular, that $\hat{\beta} = \beta + O_p(T^{-1})$ and $\hat{\tau} = \tau + O_p(T^{-1})$ (see Johansen, 1991, 1994). Note that here we have implicitly assumed that the parameter matrix $\beta$ as well as the estimator $\hat{\beta}$ have been made unique by a suitable normalization. A general formulation of a normalization of this kind is given by $\beta_c = \beta (c' \beta)^{-1}$ where the $(n \times r)$ matrix $c$ is such that $c' \beta$ is nonsingular (see Johansen, 1991; Paruolo, 1997). From the definition of our test statistic $LM(r_0)$ in (4.10) it follows that our test procedure does not require the specification of normalizing restrictions which are only needed in the theoretical derivations and even there the existence of a suitable normalization is all that is needed. Note, however, that the normalization of the cointegrating vectors also imposes restrictions on the parameters $\alpha$ and $\tau$. For simplicity, we shall not make these restrictions explicit in the notation and the same notational convention applies to other quantities like $\beta_1$ or $\tilde{\beta}_1$ whose normalized versions will sometimes be needed. Finally, note that $\tilde{\beta}_1 = \beta_1 + O_p(T^{-1})$ (cf. Paruolo, 1997).

We will first give three lemmas which imply important properties of the estimators of the trend parameters $\mu_0$ and $\mu_1$. Then we will turn to a proof of Theorem 1. Unless otherwise stated, all limits are taken as the sample size $T \to \infty$. In the following the proofs are sketched only. Details are given in Lütkepohl and Saikkonen (1997).

A.1. Intermediate results

The following intermediate results are stated only. They are proven in Lütkepohl and Saikkonen (1997).
Lemma A.1. Consider the regression model (4.3) where the estimator $\tilde{\beta}_I$ has been made unique by a suitable normalization and let $\tilde{\delta}_*$ and $\tilde{\tau}_*$ be the LS estimators of the parameters $\delta_*$ and $\tau_*$, respectively. Then, under the null hypothesis $H(r_0)$,

$$\tilde{\delta}_* = \delta_* + O_p(1)$$  \hspace{1cm} (A.1)

and

$$T^{1/2}(\tilde{\tau}_* - \tau_*) \overset{d}{\rightarrow} N(0, \beta_0' \Omega \beta_0),$$  \hspace{1cm} (A.2)

where it is assumed that $\beta_0$ has been normalized in the same way as $\tilde{\beta}_I$.

The following lemma states that the asymptotic properties of $\tilde{\delta}_*^{(1)}$ and $\tilde{\tau}_*^{(1)}$ obtained by correcting $\tilde{\delta}_*$ and $\tilde{\tau}_*$ are the same as those of the latter estimators in Lemma A.1.

Lemma A.2. Consider the estimators $\tilde{\delta}_*^{(1)}$ and $\tilde{\tau}_*^{(1)}$ and suppose the estimator $\tilde{\beta}_I$ in their definitions has been made unique by a suitable normalization. Then, the result of Lemma A.1 holds with $\tilde{\delta}_*$ and $\tilde{\tau}_*$ replaced by $\tilde{\delta}_*^{(1)}$ and $\tilde{\tau}_*^{(1)}$, respectively.

In the following lemma we summarize some properties of the estimators $\mu_0$ and $\mu_1$ under the local alternatives in (5.1). As a consequence of the previous lemma, any of the estimators from Section 4.2 may be used here. The results of the lemma are basic for proving both Theorems 1 and 2.

Lemma A.3. Under the conditions of Theorem 2,

$$\beta'(\tilde{\mu}_0 - \mu_0) = O_p(T^{-1/2}),$$  \hspace{1cm} (A.3)

$$\beta'_I(\tilde{\mu}_0 - \mu_0) = O_p(1),$$  \hspace{1cm} (A.4)

$$\beta'(\tilde{\mu}_1 - \mu_1) = O_p(T^{-3/2}),$$  \hspace{1cm} (A.5)

$$T^{1/2}\beta'_I(\tilde{\mu}_1 - \mu_1) \overset{d}{\rightarrow} K(1),$$  \hspace{1cm} (A.6)

where the weak convergence in (A.6) holds jointly with other weak convergencies related to the LM-type tests. When the null hypothesis is true, these results hold under the conditions of Theorem 1 and the right-hand side of (A.6) becomes $\beta_0' C \Omega^{1/2} B(1)$.

A.2. Proof of Theorem 1

Now we can turn to the proof of Theorem 1. We will make use of well-known properties of integrated processes which, under our assumptions, follow e.g. from Lemma 1 of Sims et al. (1990). We shall explicitly only consider the test
statistic $LM_\mathbf{u}(r_0)$. It will be obvious from the derivations that the same limiting distribution is obtained for $LM(r_0)$.

The asymptotic null distribution of the test statistics follows from the limiting distribution of the estimator
\[
\tilde{\rho}_* = \sum_{t=p+1}^{T} \tilde{\eta}_{t-1}\mathbf{\tilde{X}}^{t-1}_t (\tilde{M}_{vu\cdot AX}^{-1})^{-1} - \sum_{t=p+1}^{T} \tilde{\eta}_{t-1}\mathbf{\tilde{X}}^{t-1}_t \left( \sum_{t=p+1}^{T} \mathbf{\Delta X}_{t-1}^* \mathbf{\Delta X}_{t-1}^* \right)^{-1} \times \sum_{t=p+1}^{T} \mathbf{\Delta X}_{t-1}^* \mathbf{\tilde{v}}_{t-1} (\tilde{M}_{vu\cdot AX}^{-1})^{-1},
\]
where the equality follows from standard LS theory. Hence, we will derive the asymptotic distribution of $\tilde{\rho}_*$ in the following.

Using the intermediate results from the previous section (see especially Lemma A.3),
\[
T^{-1} \sum_{t=p+1}^{T} \mathbf{\Delta X}_{t-1}^* \mathbf{\Delta X}_{t-1}^* = E(\mathbf{\Delta X}_{t-1}^* \mathbf{\Delta X}_{t-1}^* ) + o_p(1),
\] (A.7)
\[
T^{-3/2} \sum_{t=p+1}^{T} \mathbf{\Delta X}_{t-1}^* \mathbf{\tilde{v}}_{t-1} = o_p(1)
\] (A.8)
and
\[
T^{-2} \sum_{t=p+1}^{T} \mathbf{\tilde{v}}_{t-1} \mathbf{\tilde{v}}_{t-1} = T^{-2} \sum_{t=p+1}^{T} \mathbf{w}_t \mathbf{w}_t + o_p(1),
\] (A.9)
where $\mathbf{w}_t = \mathbf{v}_t - \mathbf{\beta}^\prime \mathbf{\tilde{\mu}}_1 - \mathbf{\mu} \mathbf{1}$, $\mathbf{\Delta X}_{t-1}^* = [\mathbf{\Delta x}_{t-1}^*, \ldots, \mathbf{\Delta x}_{t-p+1}^*]$ and, for ease of exposition, $\mathbf{u}_t$ and $\mathbf{\Delta x}_t$ are treated as stationary. Since $E(\mathbf{\Delta X}_{t-1}^* \mathbf{\Delta X}_{t-1}^* )$ is positive definite it follows from the above results and the definition of $\tilde{M}_{vu\cdot AX}$ that
\[
T^{-2} \tilde{M}_{vu\cdot AX} = T^{-2} \sum_{t=p+1}^{T} \mathbf{w}_t \mathbf{w}_t + o_p(1).
\] (A.10)

By (2.4) and the multivariate invariance principle, $T^{-1/2} \mathbf{v}_{[T\mathbf{s}]} \xrightarrow{d} \mathbf{\beta}^\prime \mathbf{C} \mathbf{\Omega}^{1/2} \mathbf{B}(\mathbf{s})$. As usual $[T\mathbf{s}]$ denotes the integer part of $T\mathbf{s}$. Since this convergence in distribution holds jointly with that of the estimator $\tilde{\mu}_1$, we can conclude that $T^{-1/2} \mathbf{w}_{[T\mathbf{s}]} \xrightarrow{d} \mathbf{\beta}^\prime \mathbf{C} \mathbf{\Omega}^{1/2} \mathbf{B}_\mathbf{s}(\mathbf{s})$. Thus, from (A.10) and a standard application of the continuous mapping theorem we find that
\[
T^{-2} \tilde{M}_{vu\cdot AX} \xrightarrow{d} \mathbf{\beta}^\prime \mathbf{C} \mathbf{\Omega}^{1/2} \left( \int_0^1 \mathbf{B}_\mathbf{s}(\mathbf{s}) \mathbf{B}_\mathbf{s}(\mathbf{s}) \prime ds \right) \mathbf{\Omega}^{1/2} \mathbf{C} \mathbf{\beta}^\prime ,
\] (A.11)
where the r.h.s. is positive definite (a.s.).
Furthermore, it can be shown that
\[
T^{-1/2} \sum_{t=p+1}^{T} \tilde{\eta}_{gt} A \tilde{X}_{t-1} = O_p(1). \tag{A.12}
\]
Combining this with (A.7), (A.8) and (A.11) shows that
\[
T \hat{\rho}_* = T^{-1} \sum_{t=p+1}^{T} \tilde{\eta}_{gt} \tilde{v}_{t-1} (T^{-2} \tilde{M}_{t \perp A})^{-1} + o_p(1). \tag{A.13}
\]
Moreover, using the consistency properties of the estimators from Lemma 3 it can be shown that
\[
T^{-1} \sum_{t=p+1}^{T} \tilde{\eta}_{gt} \tilde{v}_{t-1} = T^{-1} \sum_{t=p+1}^{T} \left[ x_{t}^{'} \beta_{t}' \nu_{t} - F \beta_{t}' (\bar{\mu}_{1} - \mu_{1}) \right] \\
\times \left[ v_{t-1} - \beta_{t}' (\bar{\mu}_{1} - \mu_{1}) (t-1) \right]' + o_p(1), \tag{A.14}
\]
where \( F = x_{t}^{'} \Psi \beta_{t}' (\beta_{t}' \beta_{t})^{-1} \). By Lemmas A.1 or A.2 we have
\[
T^{1/2} \beta_{t}' (\bar{\mu}_{1} - \mu_{1}) \xrightarrow{d} \beta_{t}' C \Omega^{1/2} B(1) \text{ while Eq. (2.4) shows that } \beta_{t}' C \nu_{t} \text{ is the innovation of the martingale that determines the nonstationarity of the process } v_{t}. \]
Thus, since the definitions imply that \( F \beta_{t}' C = x_{t}^{'} \) we can use these facts and well-known properties of integrated processes to conclude from (A.14) that
\[
T^{-1} \sum_{t=p+1}^{T} \tilde{v}_{t-1} \tilde{\eta}_{gt} \xrightarrow{d} \beta_{t}' C \Omega^{1/2} \left\{ \int_{0}^{1} B(s) \frac{d B(s)}{s} - B(1) \right\}^{'} \Omega^{1/2} x_{t}^{'} \\
- \int_{0}^{1} B(s) \frac{d B(1)'}{s} + \int_{0}^{1} B(1) B(1)' \frac{d B(s)}{s} \Omega^{1/2} x_{t}^{'} \\
= \beta_{t}' C \Omega^{1/2} \int_{0}^{1} B_{*}(s) \frac{d B_{*}(s)'}{s} \Omega^{1/2} x_{t}^{'} , \tag{A.15}
\]
where the equality is obtained by direct calculation. From (A.11) and (A.15) we can derive the limiting distribution of the estimator \( \hat{\rho}_* \) which, in conjunction with (A.11), the consistency of the estimators \( \tilde{z} \) and \( \tilde{\Omega} \), and the definition of the test statistic \( LM_{\#}(r_0) \) gives the desired result. □

References