Hubrich, Kirstin; Lütkepohl, Helmut; Saikkonen, Pentti

Working Paper

A review of systemscointegration tests

SFB 373 Discussion Paper, No. 1998,101

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Hubrich, Kirstin; Lütkepohl, Helmut; Saikkonen, Pentti (1998) : A review of systemscointegration tests, SFB 373 Discussion Paper, No. 1998,101, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10060720

This Version is available at:
https://hdl.handle.net/10419/61256

A REVIEW OF SYSTEMS COINTEGRATION TESTS

Kirstin Hubrich, Helmut Lütkepohl
Institut für Statistik und Ökonometrie
Humboldt–Universität
Spandauer Str. 1
10178 Berlin
GERMANY
Tel.: +49-30-2093-5705/5718
Fax: +49-30-2093-5712
email: kirstin@wiwi.hu-berlin.de, luetke@wiwi.hu-berlin.de

and

Pentti Saikkonen
Department of Statistics
P.O. Box 54
SF-00014 University of Helsinki
FINLAND
Tel.: +358-9-1918867
Fax: +358-9-1918872
email: pentti.saikkonen@helsinki.fi

October 20, 1998

Abstract

The literature on systems cointegration tests is reviewed and the various sets of assumptions for the asymptotic validity of the tests are compared within a general unifying framework. The comparison includes likelihood ratio tests, Lagrange multiplier and Wald type tests, lag augmentation tests, tests based on canonical correlations, the Stock-Watson tests and Bierens' nonparametric tests. Asymptotic results regarding the power of these tests and previous small sample simulation studies are discussed. Further issues and proposals in the context of systems cointegration tests are also considered briefly. New simulations are presented to compare the tests under uniform conditions. Special emphasis is given to the sensitivity of the test performance with respect to the trending properties of the DGP.

Keywords: systems cointegration tests, LR tests, nonparametric tests, asymptotic power, small sample simulations

[1]We are grateful to Christian Müller for helping with the computations and we thank Jesus Gonzalo and Jürgen Wolters for commenting on an earlier version of the paper. Part of this research was carried out while the third author was visiting the Humboldt University in Berlin. Financial support by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213) is gratefully acknowledged.

Contents

1 Introduction

Since the invention of cointegration by Granger (1981) and Engle & Granger (1987) one branch of the associated literature has been concerned with finding the number of cointegrating relations in a system of time series variables. Even for a single equation analysis it may be important to know the number of cointegrating relations among the variables involved so that the issue is an important one even in that context. Early tests for the number of cointegration relations are due to Johansen (1988) and Stock & Watson (1988). These early tests are derived under special assumptions which are not always realistic in applied work. Moreover, these tests are based on asymptotic considerations and it was found in simulation studies that their small sample properties may not be very satisfactory in some situations of interest from the point of view of applied work.

Therefore, it is not surprising that there has been a substantial growth in the literature on systems cointegration tests with many proposals for improvements of existing tests and suggestions of alternative procedures. By now a wide range of procedures is available which work under various sets of assumptions. Although there are small sample comparisons of some of these tests it may be useful to review the alternative sets of assumptions under which these testing procedures are supposed to work and to compare them in a general unifying framework. We will attempt to do so in the present study.

The study has the following structure. In the next section the general framework is described in which the various tests can be analyzed and compared. Likelihood ratio, Lagrange multiplier and Wald type testing procedures are reviewed in Section 3. We also discuss the different sets of assumptions for the asymptotic validity of these tests. Several further issues and modifications of this group of tests are discussed. An important property of many of these tests is their dependence on deterministic terms in the data generation process. Therefore we have structured the presentation according to the properties of deterministic terms. An applied researcher who wishes to get information on the tests suitable for his or her purposes should decide on the deterministic terms first, e.g., is a deterministic trend term necessary? Once that decision has been made it is possible to go to the appropriate section and check out the suitable tests. In Section 4 some other tests are considered. These include tests based on limiting χ^2 distributions and the Stock & Watson tests, tests based on canonical correlations as well as Bierens' nonparametric tests and choosing the cointegrating

rank by model selection criteria. Some further proposals are also discussed. In Section 5 we first review previous simulation studies which explore and compare the size and power properties of some of the tests in small samples and then we present own simulations to compare the full range of tests on a common footing. Section 6 concludes.

Since the subject of this review is still a lively research area many of the articles cited in the following are not even published yet. Also, there may be important papers related to the topic of which we are not aware. We apologize to the authors for the omission. Some tests and ideas came to our attention only after finishing the simulations and therefore we have not been able to include them in our small sample comparison. Again we apologize to those who proposed them.

Notation

The following general notation is used. y_t denotes an observable n-dimensional vector of time series variables. The normal distribution with mean (vector) μ and variance (covariance matrix) Σ is denoted by $N(\mu, \Sigma)$. A set of independently identically distributed random variables or vectors is said to be *iid*. The symbol $\mathbf{B}_k = (B_1, \ldots, B_k)'$ denotes a k-dimensional standard Brownian motion. Moreover, $\xrightarrow{p}$ and $\xrightarrow{d}$ signify convergence in probability and distribution, respectively, and a.s. is short for almost surely. $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for the order of convergence and convergence in probability, respectively, of a sequence.

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 or trend stationary) after differencing d times while it is still nonstationary after differencing just $d-1$ times.

The symbols $\lambda_{max}(A)$, $\mathrm{rk}(A)$ and $\mathrm{tr}(A)$ signify the maximal eigenvalue, the rank and the trace of the matrix A, respectively. 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. For an $(m \times n)$ matrix A and an $(m \times s)$ matrix B, $[A : B]$ is the $(m \times (n+s))$ matrix whose first n columns are the columns of A and whose last s columns are the columns

of B. The $(n \times n)$ identity matrix is denoted by I_n.

LS, GLS and ML are short for least squares, generalized least squares and maximum likelihood, respectively. LR and LM abbreviate likelihood ratio and Lagrange multiplier, respectively, and RR means reduced rank. EC is a short form for error correction and DGP stands for data generation process. A vector autoregressive process is abbreviated as VAR process or as VAR(p) process if p is the lag order. A vector autoregressive moving average process is briefly called VARMA process. As a general convention, a sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

2 The Testing Framework

The general framework in which the tests for the cointegrating rank are considered in the following assumes an observable n-dimensional time series $y_t = (y_{1t}, \ldots, y_{nt})'$, $t = 1, \ldots, T$, generated as

$$y_t = \mu_0 + \mu_1 t + x_t, \qquad t = 1, 2, \ldots, \tag{2.1}$$

where μ_0 $(n \times 1)$ and μ_1 $(n \times 1)$ are parameters which may, of course, be zero. The first of these parameter vectors will sometimes be referred to as the mean term or mean parameter and μ_1 will be called trend term or trend parameter. The quantity x_t is an unobservable process with zero mean. The different tests make different assumptions regarding μ_0, μ_1, and x_t. For instance, some tests assume that x_t is a general integrated process whereas others require that it follows a pth order VAR process. We will discuss the differences in the assumptions in the following. In all the tests we assume that the components of x_t are at most $I(1)$, that is, $\Delta x_t = x_t - x_{t-1}$ is stationary. This implies that y_t is also at most $I(1)$. Of course, there may be cointegration among the variables of x_t. In other words, there may be stationary $(I(0))$ linear combinations of the elements of x_t. In fact, x_t may be $I(0)$ altogether. If x_t is cointegrated some linear combinations of y_t are trend stationary and y_t is also said to be cointegrated.

For some tests we assume that x_t has a finite order EC representation

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

where Π and the Γ_j are $(n \times n)$ matrices of unknown parameters, $\varepsilon_t \sim (0, \Omega)$ is white noise and the initial values x_t, $t \leq 0$, are assumed to be zero for convenience. Alternatively, it may be assumed that they are fixed constants or that they are stochastic with some fixed distribution. It is important, however, that the initial conditions do not depend on the sample size. As mentioned in the introduction, a sum is zero if the lower bound for the summation index exceeds its upper bound, e.g., if $p = 1$ in (2.2).

The number of linearly independent cointegration vectors of y_t (that is, the dimension of the cointegration space) is called the cointegrating rank of y_t. It will be denoted by r in the following. The tests we consider in this review are designed for checking either the pair of hypotheses

$$H_0(r_0) : r = r_0 \qquad \text{vs.} \qquad H_1(r_0) : r > r_0 \tag{2.3}$$

or the pair

$$H_0(r_0) : r = r_0 \qquad \text{vs.} \qquad \bar{H}_1(r_0) : r = r_0 + 1. \tag{2.4}$$

If $H_0(r_0)$ in (2.3) or (2.4) is true and x_t has the representation (2.2), the matrix Π in the EC form is zero if $r_0 = 0$ or, if $r_0 > 0$, it can be written as a product $\Pi = \alpha\beta'$, where α and β are $(n \times r_0)$ matrices of rank r_0. Thus, we may write the null hypothesis as $H_0(r_0) : \Pi = \alpha\beta'$ with α, β $(n \times r_0)$ and $\text{rk}(\alpha) = \text{rk}(\beta) = r_0$. The matrix β is the cointegration matrix whereas α is sometimes referred to as loading matrix. For later purposes it is useful to note that the matrix $\alpha'_\perp(I_n - \sum_{j=1}^{p-1}\Gamma_j)\beta_\perp$ is nonsingular under the present assumptions because we require that all series are at most $I(1)$ (see Johansen (1995, p. 49)).

If $\text{rk}(\Pi) = r > r_0$, there exist $(n \times r)$ matrices $[\alpha : \alpha_1]$ and $[\beta : \beta_1]$ of rank r, such that

$$\Pi = [\alpha : \alpha_1] \begin{bmatrix} \beta' \\ \beta_1' \end{bmatrix} = \alpha\beta' + \alpha_1\beta_1'.$$

Hence, the alternative hypothesis in (2.3) can be written as

$$H_1(r_0) \ : \ \Pi = \alpha\beta' + \alpha_1\beta_1' \tag{2.5}$$

and analogously for $\bar{H}_1(r_0)$ in (2.4). The tests considered in the following are consistent against fixed alternatives of the type (2.5), that is, if α_1, β_1 are fixed $(n \times (r - r_0))$ matrices of rank $r - r_0 > 0$, the probability of rejecting $H_0(r_0)$ approaches 1 if the sample size T goes to infinity.

For some tests local power studies have been performed where the alternative hypotheses depend on the sample size, e.g.,

$$H_T(r_0) \ : \ \Pi = \alpha\beta' + \frac{1}{T^\delta}\alpha_1\beta_1', \tag{2.6}$$

where δ is a suitable positive constant which is chosen such that the asymptotic rejection probability is between the size of the test and 1. For $\delta = 0$, we just have the alternative in (2.5) and, hence, the rejection probability of the tests tends to 1 as the sample size goes to infinity. On the other hand, if δ is chosen too large, the alternative will not be sufficiently different from the null hypothesis for large T and, hence, the tests cannot discriminate between the two hypotheses so that the power is equal to the size. Obviously, among two tests having asymptotic power greater than the size against a local alternative $H_T(r_0)$ the one with the greater δ is to be preferred from an asymptotic point of view. We will now discuss a number of alternative tests in this framework.

3 Likelihood Ratio, Lagrange Multiplier and Wald Type Tests

In a series of papers Johansen (1988, 1991a, 1992, 1994) has proposed likelihood ratio (LR) tests for (2.3) and (2.4) under the assumption that x_t is a Gaussian VAR(p) process,

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t, \tag{3.1}$$

where the A_j are $(n \times n)$ coefficient matrices and ε_t is Gaussian white noise, that is, $\varepsilon_t \sim iid\ N(0, \Omega)$. The process can be written in EC form as in (2.2) by subtracting x_{t-1} on both sides of (3.1) and rearranging terms so that $\Pi = -(I_n - A_1 - \cdots - A_p)$ and $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ for $j = 1, \ldots, p-1$. For this DGP we will now consider suitable tests under different assumptions for the mean and trend parameters μ_0 and μ_1.

3.1 Zero Mean and Trend Parameters

3.1.1 The DGP

Suppose first that the DGP in (2.1) has zero mean and zero trend term, i.e. $\mu_0 = \mu_1 = 0$, so that $y_t = x_t$. Although this case is not very realistic from a practical point of view and some tests have obtained their motivation only from deterministic trend terms or other deterministic components, it is useful to consider the case without deterministic terms first because this makes it particularly easy to present the different testing concepts. Also, for comparison purposes, it will turn out to be useful to know the properties of the tests under these restrictive assumptions.

3.1.2 The Tests

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

$$M_T = (T-p)^{-1} \left[\sum_{t=p+1}^{T} y_{t-1} y_{t-1}' - \sum_{t=p+1}^{T} y_{t-1} z_t' \left(\sum_{t=p+1}^{T} z_t z_t' \right)^{-1} \sum_{t=p+1}^{T} z_t y_{t-1}' \right] \tag{3.2}$$

and let $\hat{\Pi}$ be the least squares (LS) estimator of the matrix Π in the model (2.2). Moreover, denote the corresponding LS residuals by $\hat{\varepsilon}_t$ and define

$$\hat{\Omega} = (T-p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t' \tag{3.3}$$

Denoting by $\hat{\lambda}_1^0 \geq \cdots \geq \hat{\lambda}_n^0$ the ordered generalized eigenvalues obtained as solutions of

$$\det(\hat{\Pi} M_T \hat{\Pi}' - \lambda \hat{\Omega}) = 0, \tag{3.4}$$

Johansen's LR trace statistic for testing the pair of hypotheses in (2.3) is given by

$$LR_{trace}^0(r_0) = (T-p) \sum_{j=r_0+1}^{n} \log(1+\hat{\lambda}_j^0) \tag{3.5}$$

and the so-called maximum eigenvalue statistic for testing (2.4) is given by

$$LR_{max}^0(r_0) = (T-p) \log(1+\hat{\lambda}_{r_0+1}^0). \tag{3.6}$$

Note that these statistics are identical to those given by Johansen although they are presented in a slightly different way. Under $H_0(r_0)$, the statistics have the following limiting distributions: $LR_{trace}^0(r_0) \xrightarrow{d} \mathrm{tr}(\mathcal{D})$ and $LR_{max}^0(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D})$, where

$$\mathcal{D} = \left(\int_0^1 \mathbf{B}_{n-r_0} \mathrm{d}\mathbf{B}_{n-r_0}' \right)' \left(\int_0^1 \mathbf{B}_{n-r_0} \mathbf{B}_{n-r_0}' \mathrm{d}s \right)^{-1} \left(\int_0^1 \mathbf{B}_{n-r_0} \mathrm{d}\mathbf{B}_{n-r_0}' \right). \tag{3.7}$$

Hence, the limiting distributions are nonstandard and depend on the dimension of the process and the cointegrating rank r_0. Critical values are tabulated in Johansen (1988, Table 1; 1995, Table 15.1) and Osterwald-Lenum (1992, Table 0). It may be worth noting that the asymptotic distribution remains valid if the ε_t are not Gaussian, but have other suitable distributions.

A number of modifications of these tests have been proposed to account for small sample distortions of the test size. For instance, Reinsel & Ahn (1992) and Yap & Reinsel (1995) proposed small sample corrections of the LR trace statistics. Specifically, they consider the statistics

$$LR_{RA}^0(r_0) = (T-(n+1)p) \sum_{j=r_0+1}^{n} \log(1+\hat{\lambda}_j^0) \tag{3.8}$$

and

$$LR_{YR}^0(r_0) = (T-(n+1)p+r_0-1) \sum_{j=r_0+1}^{n} \log(1+\hat{\lambda}_j^0), \tag{3.9}$$

respectively. Yap & Reinsel (1995) also modify the test statistics so as to allow for moving average (MA) residuals of the VAR DGP, that is, they consider VARMA processes. In a slightly different context (see Section 3.2 for details), Saikkonen & Luukkonen (1997) allow the VAR order to be infinite. In that case, if again a finite order VAR is fitted to the data, the asymptotic distribution of the test statistics is shown to remain unchanged provided the order is assumed to approach infinity with the sample size at an appropriate rate. Specifically it is assumed that the order of the fitted process is $o(T^{1/3})$.

Because for $\hat{\lambda}_j^0$ close to zero we have $\hat{\lambda}_j^0 \approx \log(1+\hat{\lambda}_j^0)$, it can be shown that the LR statistics have Wald counterparts, e.g. $W_{trace}^0(r_0) = (T-p)\sum_{j=r_0+1}^n \hat{\lambda}_j^0$ and $W_{max}^0(r_0) = (T-p)\hat{\lambda}_{r_0+1}^0$ with identical asymptotic null distributions as $LR_{trace}^0(r_0)$ and $LR_{max}^0(r_0)$, respectively. Gonzalo & Pitarakis (1997b) also consider the Hotelling-Lawley version of the trace statistic,

$$HL_{trace}^0(r_0) = (T-p)\sum_{j=r_0+1}^{n} \hat{\lambda}_j^0/(1-\hat{\lambda}_j^0)$$

which is asymptotically equivalent to $W_{trace}^0(r_0)$ and $LR_{trace}^0(r_0)$ because $\hat{\lambda}_j^0 \approx \hat{\lambda}_j^0/(1-\hat{\lambda}_j^0)$ for $\hat{\lambda}_j^0$ near zero. Of course, a corresponding maximum eigenvalue statistic may also be set up.

For VAR processes with a deterministic linear trend Lütkepohl & Saikkonen (1997b) use ideas associated with the Lagrange multiplier (LM) principle to construct tests for the cointegrating rank. In the present context these tests may be derived by recalling that if the cointegrating rank is r, then the matrix Π in the EC form (2.2) can be decomposed as $\Pi = \alpha\beta'$, where α and β are $(n \times r)$. Hence, using that $\beta(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' = I_n$, (2.2) can be written as

$$\Delta x_t = \kappa u_{t-1} + \rho v_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{3.10}$$

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 $H_0(r_0)$ holds so that $r = r_0 = \mathrm{rk}(\Pi)$, we have $\rho = 0$. On the other hand, under the alternative, some columns of $\beta_\perp$ will be cointegrating vectors so that $\rho \neq 0$. Thus, a test procedure for the hypothesis $H_0(r_0)$ may be based on testing the restriction $\rho = 0$ in a feasible version of (3.10) by noting that under the present assumptions $x_t = y_t$. Because, under the null hypothesis, κ is equal to α, Lütkepohl & Saikkonen (1997b) propose to premultiply (3.10) by $\alpha_\perp'$ and test the

hypothesis $H_0 : \rho^* = 0$ in a feasible version of

$$\alpha_\perp^{'} \Delta x_t = \rho^* v_{t-1} + \sum_{j=1}^{p-1} \Gamma_j^* \Delta x_{t-j} + \eta_t, \qquad t = p+1, p+2, \ldots, \tag{3.11}$$

where $\Gamma_j^* = \alpha_\perp^{'} \Gamma_j$ and $\eta_t = \alpha_\perp^{'} \varepsilon_t$. For this purpose Lütkepohl & Saikkonen (1997b) suggest to obtain estimators $\tilde{\nu}, \tilde{\alpha}, \tilde{\beta}, \tilde{\tau}, \tilde{\Gamma}_j$ and $\tilde{\Omega}$ by applying RR regression techniques to (2.2). Then a feasible version of v_t is $\tilde{v}_t = \tilde{\beta}_\perp' x_t$. This results in a test statistic

$$LM_{(1)}^0(r_0) = (T-p)\mathrm{tr}\left\{\hat{\rho}^* \tilde{M}_{vv\cdot\Delta X} \hat{\rho}^{*'} (\tilde{\alpha}_\perp' \tilde{\Omega} \tilde{\alpha}_\perp)^{-1}\right\}, \tag{3.12}$$

where $\hat{\rho}^*$ is the LS estimator of ρ^* from (3.11) with v_t replaced by $\tilde{v}_t$, $\tilde{\alpha}_\perp' \tilde{\Omega} \tilde{\alpha}_\perp$ is the residual covariance estimator of the error term in (3.11) and

$$\tilde{M}_{vv\cdot\Delta X} = (T-p)^{-1} \left[\sum_{t=p+1}^{T} \tilde{v}_{t-1} \tilde{v}_{t-1}' - \sum_{t=p+1}^{T} \tilde{v}_{t-1} \Delta X_{t-1}' \left(\sum_{t=p+1}^{T} \Delta X_{t-1} \Delta X_{t-1}' \right)^{-1} \sum_{t=p+1}^{T} \Delta X_{t-1} \tilde{v}_{t-1}' \right] \tag{3.13}$$

with $\Delta X_{t-1}' = [\Delta x_{t-1}', \ldots, \Delta x_{t-p+1}']$. The asymptotic distribution of the test statistic is the same as that of the LR statistic.

Alternatively, ρ^* may be estimated from

$$\tilde{\alpha}_\perp^{'} \Delta x_t = \kappa^* \tilde{u}_{t-1} + \rho^* \tilde{v}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j^* \Delta x_{t-j} + \tilde{\eta}_t, \qquad t = p+1, p+2, \ldots, \tag{3.14}$$

where $\kappa^* = \alpha_\perp' \kappa$ is estimated unrestrictedly together with the other parameters and $\tilde{u}_t = \tilde{\beta}' x_t$. Denoting the corresponding estimator of ρ^* by $\tilde{\rho}^*$ results in a test statistic

$$LM_{(2)}^0(r_0) = (T-p)\mathrm{tr}\left\{\tilde{\rho}^* \tilde{M}_{vv\cdot\Delta X}^* \tilde{\rho}^{*'} (\tilde{\alpha}_\perp' \tilde{\Omega} \tilde{\alpha}_\perp)^{-1}\right\}, \tag{3.15}$$

which has the same asymptotic distribution as $LM_{(1)}^0(r_0)$ under the null hypothesis. Here

$$\tilde{M}_{vv\cdot\Delta X}^* = (T-p)^{-1} \left[\sum_{t=p+1}^{T} \tilde{v}_{t-1} \tilde{v}_{t-1}' - \sum_{t=p+1}^{T} \tilde{v}_{t-1} \Delta \tilde{X}_{t-1}^{*'} \left(\sum_{t=p+1}^{T} \Delta \tilde{X}_{t-1}^* \Delta \tilde{X}_{t-1}^{*'} \right)^{-1} \sum_{t=p+1}^{T} \Delta \tilde{X}_{t-1}^* \tilde{v}_{t-1}' \right] \tag{3.16}$$

with $\Delta \tilde{X}_{t-1}^{*'} = [\tilde{u}_{t-1}', \Delta x_{t-1}', \ldots, \Delta x_{t-p+1}']$. It is well-known that LR and LM tests are asymptotically equivalent. Therefore, it is not difficult to show that $LM_{(1)}^0(r_0)$ and $LM_{(2)}^0(r_0)$ have the same limiting null distribution as $LR_{trace}^0(r_0)$ (Lütkepohl & Saikkonen (1997b)). Note also that the LR_{trace}^0 statistic can be derived alternatively from the auxiliary regression model (3.14) (see Saikkonen & Lütkepohl (1999)). In the following we will discuss the local power of the tests.

3.1.3 Local Power

Assuming a VAR order of $p = 1$, Johansen (1991b, 1995, Chapter 14) has derived the asymptotic distribution of LR^0_{trace} under local alternatives $H_T(r_0)$ as given in (2.6) with $\delta = 1$. We use the following notation to state his result. The symbol $\mathbf{W}(u)$ is used to denote a Brownian motion with covariance matrix Ω and $\mathbf{K}(u)$ denotes the Ornstein-Uhlenbeck process defined by the integral equation

$$\mathbf{K}(u) = \alpha'_\perp \mathbf{W}(u) + \alpha'_\perp \alpha_1 \beta'_1 \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \int_0^u \mathbf{K}(s)\mathrm{d}s \quad (0 \le u \le 1) \tag{3.17}$$

or, equivalently, the stochastic differential equation

$$\mathrm{d}\mathbf{K}(u) = \alpha'_\perp \mathrm{d}\mathbf{W}(u) + \alpha'_\perp \alpha_1 \beta'_1 \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \mathbf{K}(u)\mathrm{d}u \quad (0 \le u \le 1)$$

(see e.g. Johansen (1995, Chapter 14)). Furthermore, $\mathbf{N}(s)$ is the Ornstein-Uhlenbeck process defined by $\mathbf{N}(s) = (\alpha'_\perp \Omega \alpha_\perp)^{-1/2} \mathbf{K}(s)$ or, equivalently,

$$\mathbf{N}(s) = \mathbf{B}_{n-r_0}(s) + ab' \int_0^s \mathbf{N}(u)\mathrm{d}u, \tag{3.18}$$

where $\mathbf{B}_{n-r_0}(s)$ is again an $(n-r_0)$-dimensional standard Brownian motion and the quantities a and b are given by $a = (\alpha'_\perp \Omega \alpha_\perp)^{-1/2} \alpha'_\perp \alpha_1$ and $b = (\alpha'_\perp \Omega \alpha_\perp)^{1/2} (\beta'_\perp \alpha_\perp)^{-1} \beta'_\perp \beta_1$ [cf. Johansen (1995, pp. 207-208)]. In the following the argument of the Ornstein-Uhlenbeck processes is occasionally dropped when no confusion is possible. With these notational conventions Johansen (1991b, 1995) shows that, under local alternatives $H_T(r_0) : \Pi = \alpha\beta' + \frac{1}{T}\alpha_1\beta'_1$, that is, in (2.6) we have $\delta = 1$,

$$LR^0_{trace}(r_0) \xrightarrow{d} \mathrm{tr}\left\{ \left(\int_0^1 \mathbf{N}\mathrm{d}\mathbf{N}' \right)' \left(\int_0^1 \mathbf{N}\mathbf{N}'\mathrm{d}s \right)^{-1} \left(\int_0^1 \mathbf{N}\mathrm{d}\mathbf{N}' \right) \right\}. \tag{3.19}$$

Hence, the asymptotic distribution under local alternatives is obtained from the one under the null hypothesis by replacing the Brownian motions in the latter by Ornstein-Uhlenbeck processes. The same result is easily seen to hold for $W^0_{trace}(r_0)$, $LM^0_{(1)}(r_0)$ and $LM^0_{(2)}(r_0)$. Thus all these tests are asymptotically equivalent under local alternatives. The asymptotic distribution in (3.19) will be used later for comparison purposes. In the next section we consider the case of a nonzero mean.

3.2 Nonzero Mean and no Trend Term

3.2.1 The DGP

We now assume that μ_0 is possibly nonzero and there is still no trend, that is, $\mu_1 = 0$. Again we assume that x_t is a finite order VAR(p) process with EC representation (2.2). Then y_t is a VAR(p) process with EC representation

$$\Delta y_t = \nu_0 + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{3.20}$$

where $\nu_0 = -\Pi\mu_0$. Hence, the model can be written alternatively in the form

$$\Delta y_t = \Pi^* y_{t-1}^* + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{3.21}$$

where $\Pi^* = \Pi[I_n : \mu_0]$ and $y_{t-1}^* = [y_{t-1}' : 1]'$. The relevant ML estimators and hence the LR statistics may be obtained easily from these representations using RR regression techniques as in Section 3.1.

3.2.2 The Tests

One possible approach is to obtain the LS estimators $\hat{\Pi}$ and $\hat{\Omega}$ from (3.20). An LR test version which ignores the restrictions on the intercept is obtained by plugging $z_t' = (1, \Delta y_{t-1}', \ldots, \Delta y_{t-p+1}')$ into (3.2), using the resulting matrix M_T^{i0}, say, in (3.4) and considering the corresponding generalized eigenvalues $\hat{\lambda}_1^{i0} \geq \cdots \geq \hat{\lambda}_n^{i0}$. The resulting test statistics will be denoted by $LR_{trace}^{i0}(r_0)$ and $LR_{max}^{i0}(r_0)$, where the superscript i stands for the intercept term and 0 indicates that the trend parameters are zero. The asymptotic null distributions are $LR_{trace}^{i0}(r_0) \xrightarrow{d} \mathrm{tr}(\bar{\mathcal{D}})$ and $LR_{max}^{i0}(r_0) \xrightarrow{d} \lambda_{max}(\bar{\mathcal{D}})$, where

$$\bar{\mathcal{D}} = \left(\int_0^1 \bar{\mathbf{B}} \mathrm{d}\mathbf{B}_{n-r_0}'\right)' \left(\int_0^1 \bar{\mathbf{B}}\bar{\mathbf{B}}' \mathrm{d}s\right)^{-1} \left(\int_0^1 \bar{\mathbf{B}} \mathrm{d}\mathbf{B}_{n-r_0}'\right)$$

with $\bar{\mathbf{B}}(s) = \mathbf{B}_{n-r_0}(s) - \int_0^1 \mathbf{B}_{n-r_0}(u)\mathrm{d}u$ (see Johansen (1991a)). The corresponding critical values are, for example, given in Johansen & Juselius (1990, Table A.2).

In this formulation of the LR tests the intercept term in (3.20) is estimated unrestrictedly. In principle a model of this type with unconstrained intercept ν_0 can also generate a linear trend in the variables. Thus, this possibility is not ruled out in estimating the parameters. In other words, the zero trend restriction is not properly taken into account. To enforce

that restriction in estimating the parameters the model form (3.21) may be used. That is, we may compute LS estimators $\hat{\Pi}^*$ and $\hat{\Omega}^*$ from (3.21) and consider the ordered generalized eigenvalues $\hat{\lambda}_1^* \geq \cdots \geq \hat{\lambda}_n^*$ obtained from $\det(\hat{\Pi}^* M_T^* \hat{\Pi}^{*\prime} - \lambda \hat{\Omega}^*) = 0$, where M_T^* is defined as M_T in (3.2) with y_{t-1} replaced by y_{t-1}^* and $z_t' = (\Delta y_{t-1}', \ldots, \Delta y_{t-p+1}')$ as before. Now statistics similar to (3.5) and (3.6) may be defined as

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

and

$$LR_{max}^*(r_0) = (T-p)\log(1+\hat{\lambda}_{r_0+1}^*),$$

respectively. The asymptotic null distributions are $LR_{trace}^*(r_0) \xrightarrow{d} \mathrm{tr}(\mathcal{D}^*)$ and $LR_{max}^*(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D}^*)$, where

$$\mathcal{D}^* = \left(\int_0^1 \mathbf{B}^* \mathrm{d}\mathbf{B}_{n-r_0}'\right)' \left(\int_0^1 \mathbf{B}^* \mathbf{B}^{*\prime} \mathrm{d}s\right)^{-1} \left(\int_0^1 \mathbf{B}^* \mathrm{d}\mathbf{B}_{n-r_0}'\right)$$

and $\mathbf{B}^*(s) = [\mathbf{B}_{n-r_0}(s)' : 1]'$ (see Johansen (1995)). These distributions are obviously different from the previous ones. Critical values are, for instance, given in Johansen & Juselius (1990, Table A.3), Johansen (1995, Table 15.2) and Osterwald-Lenum (1992, Table 1*).

Saikkonen & Luukkonen (1997) suggest yet another possibility to proceed in the present case without a trend in the DGP. They consider the EC representation

$$\Delta y_t = \Pi(y_{t-1} - \mu_0) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{3.22}$$

and propose to estimate the mean parameter μ_0 in a first step, substitute the estimator, $\hat{\mu}_0$ say, in (3.22) and then proceed as in computing the LR^0 statistics with y_{t-1} replaced by $y_{t-1} - \hat{\mu}_0$.

In order to obtain an appropriate estimator for μ_0 the VAR form $y_t = \nu_0 + A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t$ is initially estimated by LS. The estimators of ν_0 and the A_j will be denoted by $\hat{\nu}_0$ and $\hat{A}_j$, respectively. Then the following quantities are computed:

$$\hat{d}_t = y_t - \sum_{j=1}^{p} \hat{A}_j y_{t-j} \quad \text{and} \quad \hat{C}_t = I_n - \sum_{j=1}^{t-1} \hat{A}_j, \quad t = 1, \ldots, T,$$

with $y_t = 0$ for $t \leq 0$, $\hat{C}_1 = I_n$ and $\hat{A}_j = 0$ for $j > p$. The estimator of μ_0 is then obtained as

$$\hat{\mu}_0 = \left(\sum_{t=1}^{T} \hat{C}_t' \hat{\Omega}^{-1} \hat{C}_t\right)^{-1} \sum_{t=1}^{T} \hat{C}_t' \hat{\Omega}^{-1} \hat{d}_t.$$

Here $\hat{\Omega}$ is the estimator for the residual covariance matrix based on the LS estimator of the VAR(p) process in levels. Note that the estimator $\hat{\mu}_0$ may be regarded as a GLS estimator for μ_0 and it can be motivated as an approximate Gaussian ML estimator if $x_t = 0$, $t \leq 0$.

The test statistics are determined by considering the model

$$\Delta y_t = \Pi(y_{t-1} - \hat{\mu}_0) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t^*, \qquad t = p+1, p+2, \ldots. \tag{3.23}$$

That is, they are based on the generalized eigenvalues $\hat{\lambda}_1^{SL} \geq \cdots \geq \hat{\lambda}_n^{SL}$, say, from $\det(\tilde{\Pi} M_T^{SL} \tilde{\Pi}' - \lambda \tilde{\Omega}) = 0$, where $\tilde{\Pi}$ and $\tilde{\Omega}$ are obtained from LS estimation of (3.23) and M_T^{SL} is defined as in (3.2) with y_{t-1} replaced by $y_{t-1} - \hat{\mu}_0$ and $z_t' = (\Delta y_{t-1}', \ldots, \Delta y_{t-p+1}')$ as before. The trace statistic for testing the pair of hypotheses (2.3) then becomes

$$LR_{trace}^{SL}(r_0) = (T-p) \sum_{j=r_0+1}^{n} \log(1 + \hat{\lambda}_j^{SL}) \tag{3.24}$$

and the statistic for testing (2.4) is

$$LR_{max}^{SL}(r_0) = (T-p) \log(1 + \hat{\lambda}_{r_0+1}^{SL}).$$

These test statistics have the same asymptotic null distributions as $LR_{trace}^0(r_0)$ and $LR_{max}^0(r_0)$, respectively. Hence, the same tables may be used for critical values.

Saikkonen & Luukkonen (1997) actually work under more general assumptions than the narrow finite order VAR(p) framework considered in the foregoing. They allow the DGP to be of infinite order VAR type and show that their modified tests remain asymptotically valid under these more general assumptions if a finite order VAR is fitted to the data and the order is assumed to go to infinity with the sample size at the rate $o(T^{1/3})$. Saikkonen (1992) shows that this result even holds if the short-term dynamics are not taken into account in the scaling matrix M_T. Because the resulting test was found to have poor small sample properties if the short-term dynamics are neglected, we will not give the details of this modification here.

For completeness we mention that tests corresponding to $LR_{YR}^0(r_0)$, $LR_{RA}^0(r_0)$, $W_{trace}^0(r_0)$ and $W_{max}^0(r_0)$ may be set up in the obvious way for the tests considered so far in this subsection. These tests are again asymptotically equivalent to the corresponding LR tests. Moreover, we can use the LM forms of the test statistics in (3.12) and (3.15) if we simply replace x_t by $\hat{x}_t = y_t - \hat{\mu}_0$. These forms of the LM statistics will be denoted by $LM_{ma(1)}(r_0)$ and $LM_{ma(2)}(r_0)$, respectively. Again their asymptotic properties are identical to those of $LR_{trace}^0(r_0)$.

3.2.3 Local Power

Since we have a number of different tests now which are applicable under the same set of assumptions, power considerations and small sample behaviour should be used as a basis for deciding on which test to use in a particular situation. Therefore it is useful to consider the local power of the different tests. Saikkonen & Lütkepohl (1999) have derived the local power of the trace statistics in a VAR(1) setting for alternatives $H_T(r_0) : \Pi = \alpha\beta' + \frac{1}{T}\alpha_1\beta_1'$. Using the notation from Section 3.1.3, they give the following results:

$$LR^{i0}_{trace}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \bar{\mathbf{N}}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \bar{\mathbf{N}}\bar{\mathbf{N}}'\mathrm{d}s\right)^{-1}\left(\int_0^1 \bar{\mathbf{N}}\mathrm{d}\mathbf{N}'\right)\right\},$$

where $\bar{\mathbf{N}}(s) = \mathbf{N}(s) - \int_0^1 \mathbf{N}(u)\mathrm{d}u$,

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

where $\mathbf{N}^*(s) = [\mathbf{N}(s)' : 1]'$ and

$$LR^{SL}_{trace}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}\mathbf{N}'\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}\mathrm{d}\mathbf{N}'\right)\right\}.$$

The corresponding Wald and LM statistics have the same local power properties, of course. Saikkonen & Lütkepohl (1999) make the following observations based on these asymptotic results. None of the limiting distributions depends on the dimension and cointegrating rank of the process directly but just on $n - r_0$, the number of common trends under the null hypothesis. Moreover, it follows from (3.18) and (3.19) that the limiting distributions depend on α, β, Ω, α_1 and β_1 only through $a = (\alpha_\perp'\Omega\alpha_\perp)^{-1/2}\alpha_\perp'\alpha_1$ and $b = (\alpha_\perp'\Omega\alpha_\perp)^{1/2}(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1$. This implies, for instance, for the case $r - r_0 = 1$, where α_1 and β_1 are $(n \times 1)$ vectors, that the limiting distributions only depend on the two parameters

$$l^2 = a'ab'b \quad \text{and} \quad d^2 = (b'a)^2/(a'ab'b). \tag{3.25}$$

A different parameterization is used in Johansen (1995) and Saikkonen & Lütkepohl (1999). The present one is used here because it simplifies the interpretation of the results. Note that $l^2 = 0$ if and only if the null hypothesis holds. Hence, $l = \sqrt{l^2}$ may be thought of as the distance of the local alternative from the null hypothesis. Moreover, $0 < d^2 \leq 1$, where $d^2 = 0$ implies a process with $I(2)$ components which is excluded by our assumptions. The quantity $d = \sqrt{d^2}$ may be interpreted as the direction of the local alternative. We will use

the quantities l and d later in presenting simulations for comparing the local power of the tests.

The local power of none of the test statistics depends on the actual value of the mean term. Moreover, note that the limiting distribution of $LR^{SL}_{trace}(r_0)$ is the same as that of $LR^{0}_{trace}(r_0)$ given in (3.19). This result means that prior knowledge of a zero mean is not helpful for improving the asymptotic local power of the test for the cointegrating rank. In other words, the same local power can be achieved with and without such prior knowledge. For the univariate case, a similar result was also obtained by Elliott, Rothenberg & Stock (1996). These authors also show that other asymptotic distributions result if the initial values are dependent on the sample size, e.g., by making them dependent on the local alternative. In that respect our zero initial value assumption is a simplification with substance.

Since the local power functions of the tests involve nonstandard distributions the relative efficiencies of the various tests are not obvious. Therefore, we present some simulation results similar to those in Saikkonen & Lütkepohl (1999) in Figure 1. The results are based on the values of l and d given in the figure. Details on how the simulations were performed may be found in Saikkonen & Lütkepohl (1999), where different parameter values are used but otherwise the simulations are done in the same way.

A few interesting features can be seen from Figure 1. First of all, it is striking how much local power can be gained from estimating the mean term in the "right way" relative to just including an intercept term in the RR regression as in $LR^{*}_{trace}(r_0)$ and $LR^{i0}_{trace}(r_0)$. For some combinations of l and d the rejection probabilities of $LR^{SL}_{trace}(r_0)$ are seen to be about twice as large as those of $LR^{*}_{trace}(r_0)$ and $LR^{i0}_{trace}(r_0)$.

It is also interesting to see that, for a large part of the parameter space considered in our study, $LR^{*}_{trace}(r_0)$ has smaller local power than $LR^{i0}_{trace}(r_0)$. Recall that $LR^{*}_{trace}(r_0)$ uses the additional correct restriction that the intercept term may be confined to the cointegration relations whereas such a restriction is not used in $LR^{i0}_{trace}(r_0)$. Obviously, in this case imposing the extra restriction in $LR^{*}_{trace}(r_0)$ may result in a loss of asymptotic local power although the restriction holds for the DGP. This result is in line with simulations by Horvath & Watson (1995) who compare the local power of $LR^{i0}_{trace}(r_0)$ and $LR^{*}_{trace}(r_0)$ in a more restrictive setting and find the same result. In fact, in Horvath & Watson's study $LR^{*}_{trace}(r_0)$ was always inferior to $LR^{i0}_{trace}(r_0)$. In Figure 1 it is seen that in part of our parameter space

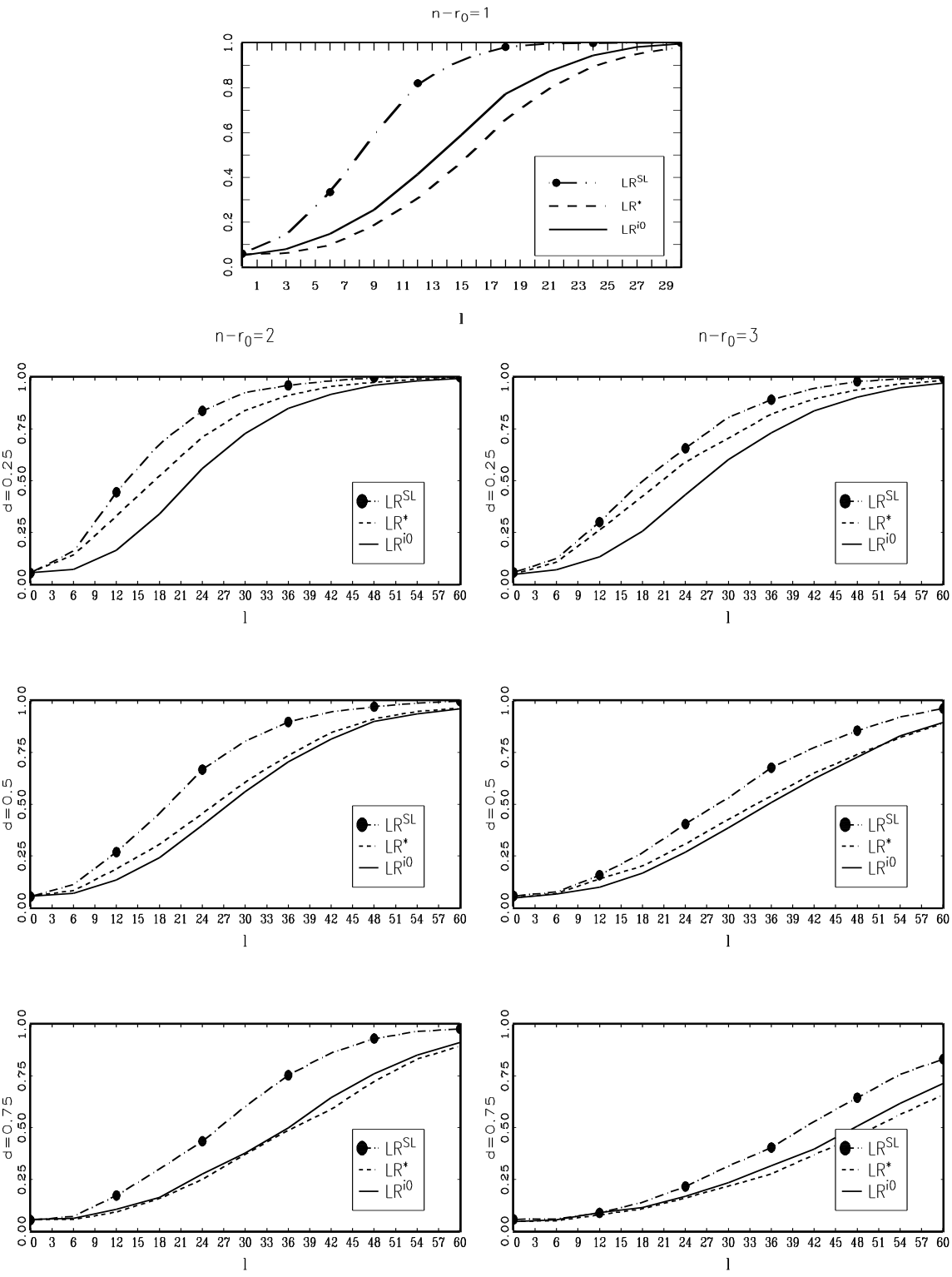

Figure 1: Local Power of LR tests

the opposite may be true. Of course, from the point of view of local power maximization, neither $LR^{*}_{trace}(r_0)$ nor $LR^{i0}_{trace}(r_0)$ should be used. Clearly, $LR^{SL}_{trace}(r_0)$ is the better choice under the present assumptions.

Another issue of practical importance is the dependence of the power on $n - r_0$, the number of stochastic trends under $H_0(r_0)$. In Figure 1 it can also be seen that increasing $n - r_0$ results in a loss of power for all the tests. This behaviour was also observed by Johansen (1995) in studying the local power of $LR^{0}_{trace}(r_0)$.

3.3 Unrestricted Mean and Trend Term

3.3.1 The DGP

We now consider the situation where both μ_0 and μ_1 may be nonzero. In that case the DGP has an EC representation

$$\Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{3.26}$$

where $\nu_0 = -\Pi\mu_0 + \left(\sum_{j=1}^{p} jA_j\right)\mu_1$ and $\nu_1 = -\Pi\mu_1$. Note that, for instance, Johansen (1995) starts from this model and thereby allows y_t to have a quadratic trend because, without constraints on ν_0 and ν_1, the model can generate quadratic trends. Of course, our DGP (2.1) can have at most linear deterministic trends. Since higher order deterministic trends seem to be less important in applied work we will continue to use (2.1) as our basic model and thereby rule out quadratic trends.

If the trend is at most linear, using $\Pi = \alpha\beta'$, (3.26) can be written alternatively as

$$\begin{aligned} \Delta y_t &= \nu + \alpha(\beta' y_{t-1} - \tau(t-1)) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t \\ &= \nu + \Pi^{+} y^{+}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \end{aligned} \tag{3.27}$$

where $\nu = -\Pi\mu_0 + (I_n - \Gamma_1 - \cdots - \Gamma_{p-1})\mu_1$, $\tau = \beta'\mu_1$, $\Pi^{+} = \alpha[\beta' : \tau] = \Pi[I_n : \mu_1]$ and $y^{+}_{t-1} = [y'_{t-1} : t-1]$. The relevant ML estimators and hence the LR statistics may be obtained easily from these representations using RR regression.

3.3.2 The Tests

An obvious idea is to base an LR type test on the representation (3.26) by estimating ν_0 and ν_1 unrestrictedly. The resulting tests were proposed by Perron & Campbell (1993).

They use $z_t' = (1, t, \Delta y_{t-1}', \ldots, \Delta y_{t-p+1}')$ in (3.2). The test statistics are then computed as in (3.5) and (3.6) and will be denoted by LR_{trace}^{PC} and LR_{max}^{PC}, respectively. The limiting null distributions are again nonstandard and differ from those of the previous tests. Specifically we have $LR_{trace}^{PC}(r_0) \xrightarrow{d} \mathrm{tr}(\mathcal{D}^{PC})$ and $LR_{max}^{PC}(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D}^{PC})$, where

$$\mathcal{D}^{PC} = \left(\int_0^1 \mathbf{B}^{PC} \mathrm{d}\mathbf{B}_{n-r_0}'\right)' \left(\int_0^1 \mathbf{B}^{PC}\mathbf{B}^{PC\prime}\mathrm{d}s\right)^{-1} \left(\int_0^1 \mathbf{B}^{PC} \mathrm{d}\mathbf{B}_{n-r_0}'\right)$$

and $\mathbf{B}^{PC}$ is a trend adjusted version of $\mathbf{B}_{n-r_0}$. Critical values for these tests are given in Perron & Campbell (1993, Table 1) (see also Rahbek (1994, Table 1)). It should be noted that the null distribution is unchanged if in fact there is no trend and intercept term in the true DGP ($\nu_0 = \nu_1 = 0$). Again Gaussian assumptions are not necessary for the ε_t as long as their distributions are sufficiently well-behaved.

The foregoing test does not take into account the restrictions for ν_0 and ν_1 which result from excluding quadratic trends in the DGP. This restriction may be enforced by using (3.27) as a basis for estimation. The resulting estimators and weighting matrix M_T^+, say, are then substituted in the determinantal equation (3.4) and, denoting the generalized eigenvalues obtained in this way by $\hat{\lambda}_1^+ \geq \cdots \geq \hat{\lambda}_n^+$, we get test statistics

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

and

$$LR_{max}^{+}(r_0) = (T-p) \log(1 + \hat{\lambda}_{r_0+1}^+).$$

The limiting null distributions are $LR_{trace}^{+}(r_0) \xrightarrow{d} \mathrm{tr}(\mathcal{D}^{+})$ and $LR_{max}^{+}(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D}^{+})$, where

$$\mathcal{D}^{+} = \left(\int_0^1 \mathbf{B}^{+} \mathrm{d}\mathbf{B}_{n-r_0}'\right)' \left(\int_0^1 \mathbf{B}^{+}\mathbf{B}^{+\prime}\mathrm{d}s\right)^{-1} \left(\int_0^1 \mathbf{B}^{+} \mathrm{d}\mathbf{B}_{n-r_0}'\right)$$

and $\mathbf{B}^{+}(s) = [\bar{\mathbf{B}}(s)', s - \frac{1}{2}]'$ where $\bar{\mathbf{B}}(s) = \mathbf{B}_{n-r_0}(s) - \int_0^1 \mathbf{B}_{n-r_0}(u)\mathrm{d}u$ is $(n-r_0)$-dimensional (see Johansen (1994, 1995)). Critical values for these distributions are given in Johansen (1994, Table V; 1995, Table 15.4) .

Yet another set of tests may be obtained by trend adjusting the data first and then computing LR or LM type tests on the basis of the trend adjusted data. Some alternative methods for trend estimation were proposed by Lütkepohl & Saikkonen (1997b) and Saikkonen & Lütkepohl (1997). Their unifying characteristic is that they are all based on parameter estimators computed under the null hypothesis. Using this additional restriction

is hoped to result in more powerful tests. We will briefly describe the methods for estimating μ_0 and μ_1 in the following assuming that estimators $\tilde{\nu}, \tilde{\alpha}, \tilde{\beta}, \tilde{\tau}, \tilde{\Gamma}_j$ are available from a RR regression of (3.27).

In developing the first set of estimators Lütkepohl & Saikkonen (1997b) define

$$\delta = \beta'\mu_0, \quad \delta_* = \beta_\perp'\mu_0, \quad \tau = \beta'\mu_1, \quad \tau_* = \beta_\perp'\mu_1$$

and note that

$$\mu_0 = \beta(\beta'\beta)^{-1}\delta + \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\delta_* \tag{3.28}$$

and

$$\mu_1 = \beta(\beta'\beta)^{-1}\tau + \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\tau_*. \tag{3.29}$$

They construct different suitable estimators $\tilde{\delta}, \tilde{\delta}_*, \tilde{\tau}$ and $\tilde{\tau}_*$ and use them together with the previously obtained estimators $\tilde{\beta}$ and $\tilde{\beta}_\perp$ in (3.28) and (3.29) to compute estimators $\tilde{\mu}_0^{(1)}$ and $\tilde{\mu}_1^{(1)}$, say, of μ_0 and μ_1, respectively. Details may be found in Lütkepohl & Saikkonen (1997b). Trend adjusted data are computed as $\tilde{x}_t^{(1)} = y_t - \tilde{\mu}_0^{(1)} - \tilde{\mu}_1^{(1)}t$. These may be used to construct test statistics like in Section 3.1.1 where zero mean and trend terms are assumed. The resulting test statistics will be denoted by $LR_{trace(ta)}(r_0)$, $LR_{max(ta)}(r_0)$, $LM_{ta(1)}(r_0)$ and $LM_{ta(2)}(r_0)$.[1] They have the following asymptotic distributions under the null hypothesis: $LR_{trace(ta)}(r_0), LM_{ta(1)}(r_0), LM_{ta(2)}(r_0) \xrightarrow{d} \mathrm{tr}(\mathcal{D}_*)$ and $LR_{max(ta)}(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D}_*)$, where

$$\mathcal{D}_* = \left(\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\right)' \left(\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'\mathrm{d}s\right)^{-1} \left(\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\right) \tag{3.30}$$

and $\mathbf{B}_*(s) = \mathbf{B}_{n-r_0}(s) - s\mathbf{B}_{n-r_0}(1)$ is an $(n-r_0)$-dimensional Brownian bridge and $\mathrm{d}\mathbf{B}_*(s) = \mathrm{d}\mathbf{B}_{n-r_0}(s) - \mathrm{d}s\mathbf{B}_{n-r_0}(1)$. Obviously, this distribution is free of unknown nuisance parameters and does not depend on the actual trending properties of the process. Critical values are given in Lütkepohl & Saikkonen (1997b, Table 1) and Saikkonen & Lütkepohl (1997, Table 1).

Another estimator of the mean and trend parameters proposed by Saikkonen & Lütkepohl (1997) is based on a GLS approach applied to the model $y_t = \mu_0 + \mu_1 t + x_t$. Using the RR

[1] $LM_{ta(1)}(r_0)$ and $LM_{ta(2)}(r_0)$ correspond to $LM(r_0)$ and $LM_*(r_0)$, respectively, in Lütkepohl & Saikkonen (1997b).

estimators from (3.27) we get estimators

$$\begin{aligned}\tilde{A}_1 &= I_n + \tilde{\alpha}\tilde{\beta}' + \tilde{\Gamma}_1,\\ \tilde{A}_j &= \tilde{\Gamma}_j - \tilde{\Gamma}_{j-1}, \quad j = 2, \ldots, p-1,\\ \tilde{A}_p &= -\tilde{\Gamma}_{p-1}\end{aligned}$$

of the levels VAR coefficients and define $\tilde{A}(L) = I_n - \tilde{A}_1 L - \cdots - \tilde{A}_p L^p$. Moreover, define $\tilde{G}_t = \tilde{A}(L)a_t$ and $\tilde{H}_t = \tilde{A}(L)b_t$, with

$$a_t = \begin{cases} 1 & \text{for} \quad t \geq 1 \\ 0 & \text{for} \quad t \leq 0 \end{cases}, \qquad b_t = \begin{cases} t & \text{for} \quad t \geq 1 \\ 0 & \text{for} \quad t \leq 0 \end{cases}$$

and estimate μ_0 and μ_1 by multivariate LS from the regression model

$$\tilde{Q}\tilde{A}(L)y_t = \tilde{Q}\tilde{G}_t\mu_0 + \tilde{Q}\tilde{H}_t\mu_1 + \eta_t^*, \quad t = p+1, \ldots, T, \tag{3.31}$$

where

$$\tilde{Q} = \begin{bmatrix} (\tilde{\alpha}'\tilde{\Omega}^{-1}\tilde{\alpha})^{-1/2}\tilde{\alpha}'\tilde{\Omega}^{-1} \\ (\tilde{\alpha}_\perp'\tilde{\Omega}\tilde{\alpha}_\perp)^{-1/2}\tilde{\alpha}_\perp' \end{bmatrix}$$

is a weight matrix which ensures that the error term roughly has a unit covariance matrix. Note that $\tilde{Q}'\tilde{Q} = \tilde{\Omega}^{-1}$. Thus, estimating the transformed model (3.31) by LS amounts to GLS estimation of the untransformed model $y_t = \mu_0 + \mu_1 t + x_t$. We will denote the resulting estimators of μ_0 and μ_1 by $\tilde{\mu}_0^{GLS}$ and $\tilde{\mu}_1^{GLS}$, respectively. These estimators can also be motivated as approximate Gaussian ML estimators under the assumption of zero initial values. The resulting test statistics will be based on trend adjusted data $\tilde{x}_t = y_t - \tilde{\mu}_0^{GLS} - \tilde{\mu}_1^{GLS} t$. Using these data in the same way as in Section 3.1.2 gives test statistics which will be denoted by $LR_{trace(ta)}^{GLS}(r_0)$, $LR_{max(ta)}^{GLS}(r_0)$, $LM_{ta(1)}^{GLS}(r_0)$ and $LM_{ta(2)}^{GLS}(r_0)$. They have the same limiting distributions under H_0 as the corresponding $LR_{(ta)}$, $W_{(ta)}$ and $LM_{ta(1)}$ statistics. Of course, Wald versions or modifications similar to those proposed by Reinsel & Ahn (1992) or Yap & Reinsel (1995) may be used here as well without affecting the asymptotic properties. We will now compare the local power of these tests.

3.3.3 Local Power

Using again a VAR(1) framework, Saikkonen & Lütkepohl (1997, 1997a) and Lütkepohl & Saikkonen (1997b) derive the asymptotic power of these tests under the local alternatives in

(2.6) with $\delta = 1$, that is, $H_T(r_0) \ : \ \Pi = \alpha\beta' + \frac{1}{T}\alpha_1\beta_1'$, where, as usual, 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. Using the notation from Section 3.1.3 we have

$$LR^{PC}_{trace}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}^{PC}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}^{PC}\mathbf{N}^{PC\prime}\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}^{PC}\mathrm{d}\mathbf{N}'\right)\right\},$$

where $\mathbf{N}^{PC}(s)$ is a trend adjusted version of $\mathbf{N}(s)$, that is, $\mathbf{N}(s)$ is corrected for mean and linear trend. Moreover,

$$LR^{+}_{trace}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}^{+}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}^{+}\mathbf{N}^{+\prime}\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}^{+}\mathrm{d}\mathbf{N}'\right)\right\},$$

where $\mathbf{N}^{+}(s) = [\bar{\mathbf{N}}(s)' : s - \frac{1}{2}]'$. Furthermore,

$$\begin{aligned} &LR_{trace(ta)}(r_0), LM_{ta(1)}(r_0), LM_{ta(2)}(r_0), LR^{GLS}_{trace(ta)}(r_0), LM^{GLS}_{ta(1)}(r_0), LM^{GLS}_{ta(2)}(r_0) \\ &\xrightarrow{d} \mathrm{tr}\left\{\left(\textstyle\int_0^1 \mathbf{N}_*(s)\mathrm{d}\mathbf{N}_*(s)'\right)'\left(\int_0^1 \mathbf{N}_*(s)\mathbf{N}_*(s)'\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}_*(s)\mathrm{d}\mathbf{N}_*(s)'\right)\right\}, \end{aligned}$$

where $\mathbf{N}_*(s) = \mathbf{N}(s) - s\mathbf{N}(1)$ and $\mathrm{d}\mathbf{N}_*(s) = \mathrm{d}\mathbf{N}(s) - \mathrm{d}s\mathbf{N}(1)$. Thus, all the tests based on prior trend adjustment of the data are asymptotically equivalent. It may be worth emphasizing again that these asymptotic distributions are obtained under our assumption that the initial values do not depend on the sample size T. As mentioned earlier, it follows from Elliott et al. (1996) that different asymptotic distributions result if the initial values depend on the sample size.

In the present framework, the local power again 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 μ_0 and μ_1. These results are analogous to those for tests based on processes without trend. As in Section 3.2.3, it can be seen that the limiting distributions depend on α, β, Ω, α_1 and β_1 through $a = (\alpha_\perp'\Omega\alpha_\perp)^{-1/2}\alpha_\perp'\alpha_1$ and $b = (\alpha_\perp'\Omega\alpha_\perp)^{1/2}(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1$ only. It follows from Johansen (1995, Corollary 14.5) that for the case $r - r_0 = 1$, where α_1 and β_1 are just vectors, the limiting distribution only depends on $l^2 = a'ab'b$ if $n - r_0 = 1$ and on l^2 and $d^2 = (b'a)^2/(a'ab'b)$ for $n - r_0 > 1$. Again l may be viewed as the distance from the null space and values of d close to zero indicate processes close to being $I(2)$.

Lütkepohl & Saikkonen (1997b) simulate the local power based on the asymptotic distributions given in the foregoing. We have also done so using a different parameterization of the local power curves and considering a wider range of parameter values. Some of the results

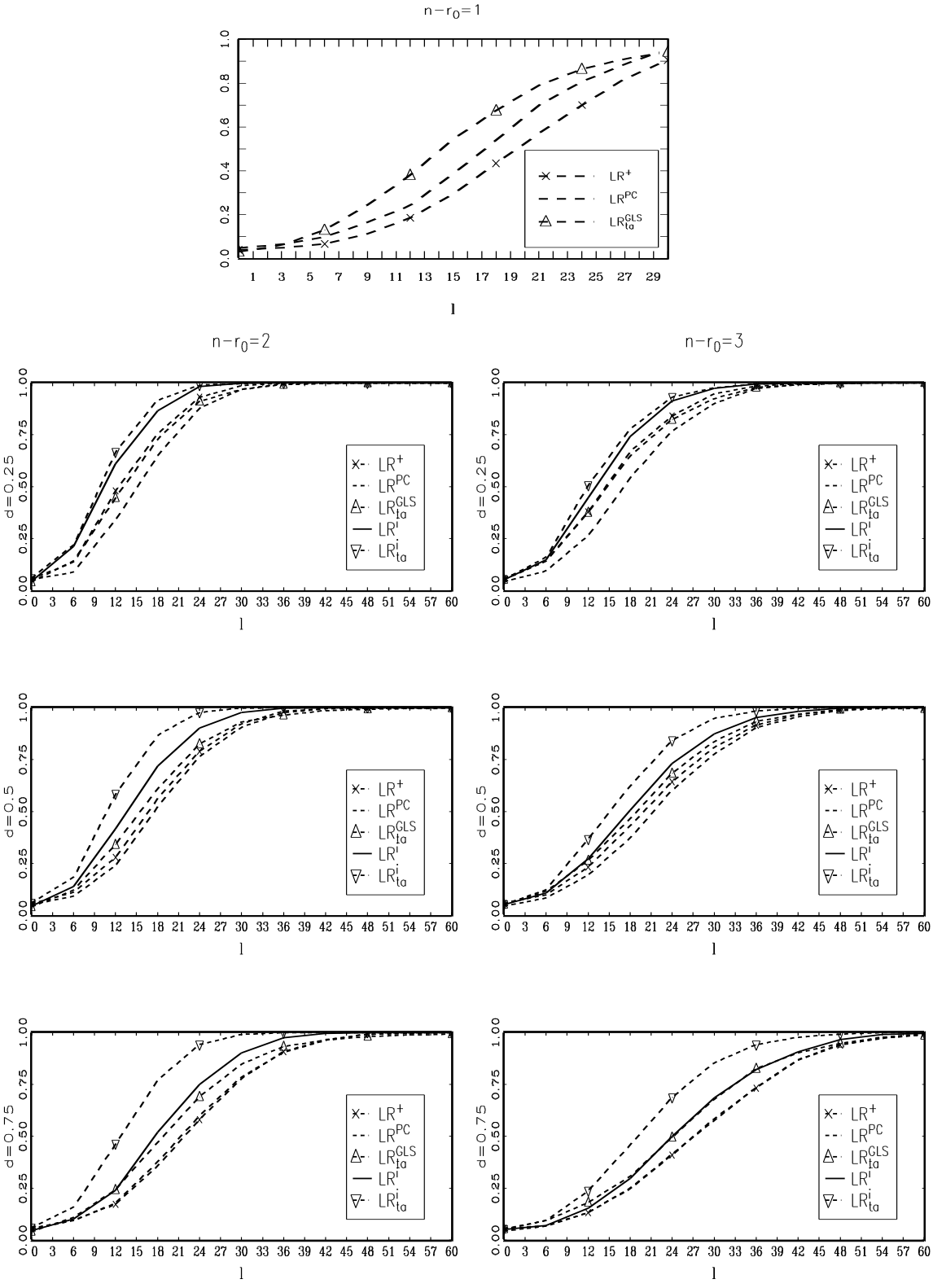

Figure 2: Local Power of LR- and LM-type tests

are given in Figure 2 together with local power functions of tests which will be discussed in Section 3.4. The following results emerge from the simulations. Because the test based on $LR^{+}_{trace}(r_0)$ imposes the restriction that the estimated trend is at most linear whereas this restriction is not used in $LR^{PC}_{trace}(r_0)$, the local powers of the two tests differ. It turns out, however, that $LR^{+}_{trace}(r_0)$ is not always superior to $LR^{PC}_{trace}(r_0)$ despite the additional correct restriction used in the former test. The tests based on prior trend adjustment have generally more local power than LR^{PC}_{trace}. The power gains can be substantial especially for d close to 1, that is, for processes not close to being $I(2)$. In many situations the tests based on prior trend adjustment also outperform $LR^{+}_{trace}(r_0)$ in terms of asymptotic local power. Unlike the case without trend, there is no clear winner under the presently considered conditions where a trend is potentially present. Thus, for a particular data set one may want to try different tests or base a decision on which test to use on the small sample performance which is discussed in Section 5. Note, however, that the local power presented in Figure 1 for the tests which assume $\mu_1 = 0$ is sometimes substantially better than the local powers in Figure 2. Hence, it pays to spend some effort on specifying the deterministic term properly.

In the next section we will consider a particular situation which has received some attention in the theoretical literature and is also often assumed in applied work. It is assumed that a trend appears in the variables and not in the cointegrating relations.

3.4 Trending Variables and No Trend in the Cointegration Relations

3.4.1 The DGP

If it is known that a trend is present so that $\mu_1 \neq 0$ but it is confined to the variables and does not appear in the cointegrating relations, that is, $\beta'\mu_1 = \tau = 0$, tests for the cointegrating rank may be constructed so as to take that additional restriction into account. We will turn to that situation now. Although having this kind of prior knowledge may not be very realistic in many situations, considering the corresponding cointegration tests is useful from a conceptual point of view. Moreover, the resulting tests are perhaps the most frequently used ones for determining the cointegrating rank in applied work. It may be worth emphasizing, however, that for the $(n \times r)$ matrix β to satisfy $\beta'\mu_1 = 0$, $\mu_1 \neq 0$ implies that $r < n$. Hence,

even under the alternative the rank must be smaller than n in the present setting, a fact which is sometimes ignored in empirical applications. As a consequence, it is, for example, unreasonable to test $H_0 : r = 1$ vs. $H_1 : r > 1$ in a bivariate process in this context.

3.4.2 The Tests

If it is known that there is a deterministic trend ($\mu_1 \neq 0$), but μ_1 is orthogonal to β, that is, $\beta'\mu_1 = 0$ and, hence, $\nu_1 = 0$, the LS estimators $\hat{\Pi}$ and $\hat{\Omega}$ may be obtained from (3.20). The corresponding test statistics are obtained by plugging $z_t' = (1, \Delta y_{t-1}', \ldots, \Delta y_{t-p+1}')$ into (3.2), using the resulting M_T^{i0} matrix in (3.4) and considering the corresponding generalized eigenvalues. The resulting LR statistics will be denoted by $LR^i_{trace}(r_0)$ and $LR^i_{max}(r_0)$, where the i stands for 'intercept'. The asymptotic null distributions of these test statistics are as follows: $LR^i_{trace}(r_0) \xrightarrow{d} \mathrm{tr}(\mathcal{D}^i)$ and $LR^i_{max}(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D}^i)$, where

$$\mathcal{D}^i = \left(\int_0^1 \bar{\mathbf{B}}^i \mathrm{d}\mathbf{B}_{n-r_0}'\right)' \left(\int_0^1 \bar{\mathbf{B}}^i \bar{\mathbf{B}}^{i\prime} \mathrm{d}s\right)^{-1} \left(\int_0^1 \bar{\mathbf{B}}^i \mathrm{d}\mathbf{B}_{n-r_0}'\right), \tag{3.32}$$

$\mathbf{B}^i(s) = [\mathbf{B}_{n-r_0-1}(s)' : s]'$ and $\bar{\mathbf{B}}^i(s) = \mathbf{B}^i(s) - \int_0^1 \mathbf{B}^i(u)\mathrm{d}u$ (see Johansen (1995)). Critical values for these test statistics may be found in Johansen & Juselius (1990, Table A.1), Johansen (1995, Table 15.3) and Osterwald-Lenum (1992, Table 1), among others. Unfortunately, the null distributions change if the true μ_1 turns out to be zero. In that case, the tests are precisely the ones based on $LR^{i0}_{trace}(r_0)$ and $LR^{i0}_{max}(r_0)$, respectively. In Section 3.2.2 these tests were seen to have limiting distributions quite different from those of $LR^i_{trace}(r_0)$ and $LR^i_{max}(r_0)$ although the test statistics are identical.

In practice, the restriction $\mu_1 = 0$ will usually be unknown. In that case, Johansen (1992) suggests to test null hypotheses for increasing ranks $r_0 = 0, 1, ...$, with both types of tests, $LR^*(r_0)$ and $LR^i(r_0)$, using the critical values obtained under the assumption $\mu_1 \neq 0$ for the latter. If both tests reject for a given r_0 we proceed to $H_0(r_0 + 1)$ etc.. If at some stage $H_0(r_0)$ is rejected by $LR^*(r_0)$ and not by $LR^i(r_0)$, then r_0 is chosen as cointegrating rank and the process is assumed to have $\mu_1 \neq 0$. If $H_0(r_0)$ is not rejected by $LR^*(r_0)$ for some r_0, we conclude that $r = r_0$ and the process does not have a linear trend. The same holds, of course, if both $LR^i(r_0)$ and $LR^*(r_0)$ do not reject.

Assuming again that $\mu_1 \neq 0$ and $\beta'\mu_1 = 0$ is known so that there is a deterministic linear

trend in the variables but not in the cointegration relations, the EC form may be written as

$$\Delta y_t - \mu_1 = \Pi(y_{t-1} - \mu_0) + \sum_{j=1}^{p-1} \Gamma_j(\Delta y_{t-j} - \mu_1) + \varepsilon_t, \qquad t = p+1, p+2, \ldots \tag{3.33}$$

Saikkonen & Lütkepohl (1998a) propose to base LR and LM type tests for the cointegrating rank on this model form. The resulting test statistics will be denoted by $LR^i_{ta}(r_0)$, $LM^i_{ta(1)}(r_0)$ and $LM^i_{ta(2)}(r_0)$, where the former statistic is the trace version and the latter two statistics correspond to the two different versions of LM type tests considered in Section 3.1. Saikkonen & Lütkepohl (1998a) show that the two LM type tests and $LR^i_{ta}(r_0)$ have identical asymptotic properties. More precisely,

$$LM^i_{ta(1)}(r_0), LM^i_{ta(2)}(r_0), LR^i_{ta}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \bar{\mathbf{B}}^i \mathrm{d}\mathbf{B}'_{n-r_0}\right)' \left(\int_0^1 \mathbf{B}^i \mathbf{B}^{i\prime} \mathrm{d}s\right)^{-1} \left(\int_0^1 \bar{\mathbf{B}}^i \mathrm{d}\mathbf{B}'_{n-r_0}\right)\right\}, \tag{3.34}$$

where $\mathbf{B}^i(s)$ denotes the same process as in (3.32). We will now turn to a local power comparison of these tests.

3.4.3 Local Power

We consider again local alternatives of the form (2.6) with $\delta = 1$. Moreover, to ensure that we remain in the presently considered model class where a trend appears in the variables and not in the cointegrating relations we have to assume that $\Pi\mu_1 = 0$ and, hence,

$$\beta_1'\mu_1 = 0. \tag{3.35}$$

Saikkonen & Lütkepohl (1998a) derive the limiting distributions of the test statistics considered previously under these assumptions. In other words, it is assumed that there is no trend in the cointegration relations even under the alternative. Note that the fact that the $(n \times r)$ matrix $[\beta : \beta_1]$ is orthogonal to μ_1 also means that r has to be less than n.

Using the notation from Section 3.1.2, we now define the $(n - r_0)$-dimensional Ornstein-Uhlenbeck process

$$\tilde{\mathbf{N}}(u) = \mathbf{B}(u) + \tilde{a}\tilde{b}' \int_0^u \tilde{\mathbf{N}}(s)\mathrm{d}s \quad (0 \le u \le 1), \tag{3.36}$$

where

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

with g' being an $((n - r_0) \times (n - r_0))$ matrix which is the inverse of $\beta_\perp'[\mu_1 : \gamma]$ and γ any $(n \times (n - r_0 - 1))$ matrix orthogonal to β and μ_1 such that $[\beta : \mu_1 : \gamma]$ is of full

rank. Furthermore we need the process $\mathbf{Z}(s) = [\tilde{\mathbf{N}}_2(s)' : s]'$ and its demeaned counterpart $\bar{\mathbf{Z}}(s) = \mathbf{Z}(s) - \int_0^1 \mathbf{Z}(u)\mathrm{d}u$. Here $\tilde{\mathbf{N}}_2(s)$ consists of the last $n - r_0 - 1$ components of $\tilde{\mathbf{N}}(s)$.

Assuming again that the initial values are not dependent on the sample size, the limiting distributions derived by Saikkonen & Lütkepohl (1998a) for VAR(1) processes under the presently considered local alternatives are:

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

and

$$LM^i_{ta(1)}(r_0), LM^i_{ta(2)}(r_0), LR^i_{ta}(r_0)$$
$$\xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \bar{\mathbf{Z}}(s)\mathrm{d}\tilde{\mathbf{N}}(s)'\right)' \left(\int_0^1 \mathbf{Z}(s)\mathbf{Z}(s)'\mathrm{d}s\right)^{-1} \left(\int_0^1 \bar{\mathbf{Z}}(s)\mathrm{d}\tilde{\mathbf{N}}(s)'\right)\right\}.$$

Comparing the latter two limiting distributions shows that the nonnull distributions differ in the same way as the limiting distributions under the null hypothesis.

The limiting distributions are similar to those obtained in previous sections under local alternatives. An important difference results from the definitions of the parameters $\tilde{a}$ and $\tilde{b}$, however. These parameters now depend on the matrix g which did not appear in the previous counterparts of $\tilde{a}$ and $\tilde{b}$. The zero matrix in $\tilde{b}$ is due to the assumption $\beta_1'\mu_1 = 0$. Note, however, that we necessarily have $\beta_1'\gamma \neq 0$ (because $\beta_1 \in \mathrm{span}(\beta_\perp) = \mathrm{span}([\mu_1 : \gamma])$) so that $\tilde{b} = 0$ is not possible.

For $LR^i_{trace}(r_0)$, Rahbek (1994) obtained the limiting distribution under different local alternatives. He considered two types of local alternatives. The first type is given by $\Pi = \alpha\beta' + T^{-3/2}\alpha_1\beta_1'$ which is an order of magnitude smaller than the one used here. However, Rahbek's limiting distribution reduces to the limiting null distribution if condition (3.35), $\beta_1'\mu_1 = 0$, holds. In the present analysis we have employed assumption (3.35) because we wish to consider the case where a linear time trend is a priori excluded from the cointegrating relations. Another local alternative considered by Rahbek (1994) is given by $\Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1'$ combined with $\beta_1'\mu_1 = T^{-1/2}\delta$ $(0 \leq \delta'\delta < \infty)$. When $\delta = 0$ we get our local alternatives. If $\delta \neq 0$ power gains are achieved but they are achieved by testing an inclusion of a time trend in the cointegrating relations which is excluded by assumption in the present framework.

Saikkonen & Lütkepohl (1998a) also report simulated local power functions for the case where $r - r_0 = 1$, so that $\tilde{a}$ and $\tilde{b}$ are $(n \times 1)$ vectors, using $\tilde{a}' = (\sqrt{(1-d^2)l^2}, -\sqrt{d^2l^2})$

and $\tilde{b}' = (0,1)$ for $n - r_0 = 2$ as well as $\tilde{a}' = (\sqrt{(1-d^2)l^2}, -\sqrt{d^2 l^2}, 0)$ and $\tilde{b}' = (0,1,0)$ for $n - r_0 = 3$. These choices are motivated as follows. It is assumed 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 $\tilde{b}' = \beta_1'[0 : \gamma]$, where γ is essentially equal to β_1 for $n - r_0 = 2$. Hence, $\tilde{b}' = (0,1)$ is a suitable choice if $\tilde{a}$ is fully flexible. 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 $\tilde{b}' = (0,1,0)$ for $n - r_0 = 3$. The first two components of the vector $\tilde{a}'$ are simply given a fully flexible form. The values used for l and d are given in Figure 2 which also summarizes the local power results for the presently considered tests.

It is obvious from the figure that prior knowledge regarding the trend being not in the cointegration relations is quite helpful for improving the local power of the tests. In parts of the parameter space, the tests which do not use this information are considerably inferior to $LR^i_{trace}(r_0)$, $LM^i_{ta(1)}(r_0)$ and $LM^i_{ta(2)}(r_0)$ which use the information. Comparing the local power of the latter types of tests it is clear that $LR^i_{ta}(r_0)$, $LM^i_{ta(1)}(r_0)$ and $LM^i_{ta(2)}(r_0)$ are generally substantially superior to $LR^i_{trace}(r_0)$. Moreover, a comparison with the results in Figure 1 shows that an improvement in local power is possible if there is no deterministic linear trend term and $\mu_1 = 0$ is known a priori.

3.5 Summary of Models and Test Statistics

A summary of the model types in a VAR(1) setting and the corresponding LR type test statistics is given in Table 1. The associated asymptotic distributions under local alternatives are given in Table 2. The corresponding null distributions are obtained from these distributions by replacing the Ornstein-Uhlenbeck processes by the associated Brownian motions.

Based on the foregoing theoretical results it is seen to be advantageous in applied work to pay attention to the specification of the deterministic terms. If a linear trend term in the variables and the cointegration relations cannot be excluded a priori the LR^+, LR^{PC} or LR^{GLS}_{ta} tests or their LM and Wald relatives have to be used. Substantial power gains are possible, however, if a deterministic trend can be excluded from the variables and/or the cointegration relations. Based on the local power comparison LR^{SL} tests should be used if $\mu_1 = 0$ and LR^i_{ta} type tests are the preferred candidates if $\mu_1 \neq 0$ whereas $\beta'\mu_1 = 0$.

Table 1: Models for Lag order 1 and LR Type Tests

Assumption for deterministic term	Model	Notation for LR statistic	Reference
$\mu_0 = \mu_1 = 0$	$\Delta y_t = \Pi y_{t-1} + \varepsilon_t$	LR^0	Johansen (1988, 1995)
μ_0 arbitrary	$\Delta y_t = \nu_0 + \Pi y_{t-1} + \varepsilon_t$	LR^{i0}	Johansen (1991a)
$\mu_1 = 0$	$\Delta y_t = [\Pi : \eta] \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix} + \varepsilon_t$	LR^*	Johansen (1991a)
	$\Delta y_t = \Pi(y_{t-1} - \tilde{\mu}_0) + \tilde{\varepsilon}_t$	LR^{SL}	Saikkonen & Luukkonen (1997)
μ_0, μ_1 arbitrary	$\Delta y_t = \nu + [\Pi : \delta] \begin{bmatrix} y_{t-1} \\ t-1 \end{bmatrix} + \varepsilon_t$	LR^+	Johansen (1992, 1994, 1995)
	$\Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \varepsilon_t$	LR^{PC}	Perron & Campbell (1993)
	$\Delta y_t - \hat{\mu}_1 = \Pi(y_{t-1} - \hat{\mu}_0 - \hat{\mu}_1(t-1)) + \hat{\varepsilon}_t$	LR^{GLS}_{ta}	Saikkonen & Lütkepohl (1997)
μ_0 arbitrary	$\Delta y_t = \nu_0 + \Pi y_{t-1} + \varepsilon_t$	LR^i	Johansen (1991a)
$\mu_1 \neq 0, \beta'\mu_1 = 0$	$\Delta y_t - \hat{\mu}_1 = \Pi(y_{t-1} - \hat{\mu}_0) + \hat{\varepsilon}_t$	LR^i_{ta}	Saikkonen & Lütkepohl (1998a)

3.6 Further Issues

3.6.1 Incorporating Prior Information on the Cointegration Vectors

Horvath & Watson (1995) consider the power of LR and Wald tests under the assumption that prior knowledge on the cointegrating vectors is available. For instance, if interest rates are analyzed, the interest rate spread may represent a cointegration relation with known parameters. In that case some cointegration vectors may be fixed and the likelihood for a given cointegrating rank is optimized given the known cointegration vectors. In this case the asymptotic distributions of the LR statistics will also depend on the number of known cointegration vectors. Tests with deterministic mean and trend terms can be derived as well. Details and critical values may be found in Horvath & Watson (1995). They conclude that taking into account prior knowledge of this kind may result in improved asymptotic power. Thus, generally speaking, taking into account more information of this kind results in more powerful tests, at least asymptotically. As we have seen earlier, this result is not generally true with respect to information on the deterministic terms of the process. Also, applying these tests may be problematic, of course, if the prior information regarding the cointegrating vectors is incorrect. If there are no known cointegrating vectors but identifying information is available for β the approach of Kleibergen & van Dijk (1994) may be used. They discuss LR, LM and Wald tests for the cointegrating rank in a model where prior knowledge allows

Table 2: Limiting Distributions of Tests

Notation:

$\mathbf{B}(s)$ $(n-r_0)$-dimensional standard Brownian motion,

$\mathbf{N}(s) = \mathbf{B}(s) + ab' \int_0^s \mathbf{N}(u)\mathrm{d}u$, $a = (\alpha_\perp'\Omega\alpha_\perp)^{-1/2}\alpha_\perp'\alpha_1$, $b = (\alpha_\perp'\Omega\alpha_\perp)^{1/2}(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1$,

$\bar{\mathbf{N}}(s) = \mathbf{N}(s) - \int_0^1 \mathbf{N}(u)\mathrm{d}u$, $\mathbf{N}^*(s) = [\mathbf{N}(s)' : 1]'$, $\mathbf{N}^+(s) = [\bar{\mathbf{N}}(s)' : s - \frac{1}{2}]'$,

$\mathbf{N}^{PC}(s)$ trend adjusted version of $\mathbf{N}(s)$,

$\mathbf{N}_*(s) = \mathbf{N}(s) - s\mathbf{N}(1)$,

$\tilde{\mathbf{N}}(s) = \mathbf{B}(s) + \tilde{a}\tilde{b}' \int_0^s \tilde{\mathbf{N}}(u)\mathrm{d}u$, $\tilde{a} = (g'\alpha_\perp'\Omega\alpha_\perp g)^{-1/2} g'\alpha_\perp'\alpha_1$, $\tilde{b}' = \beta_1'[0:\gamma](g'\alpha_\perp'\Omega\alpha_\perp g)^{1/2}$,

$g' = \{\beta_\perp'[\mu_1 : \gamma]\}^{-1}$,

γ any $(n \times (n - r_0 - 1))$ matrix orthogonal to β and μ_1 such that $[\beta : \mu_1 : \gamma]$ is of full rank,

$\tilde{\mathbf{N}}_2(s)$ the last $n - r_0 - 1$ components of $\tilde{\mathbf{N}}(s)$,

$\mathbf{Z}(s) = [\tilde{\mathbf{N}}_2(s)' : s]'$, $\bar{\mathbf{Z}}(s) = \mathbf{Z}(s) - \int_0^1 \mathbf{Z}(u)\mathrm{d}u$.

Test statistic	Limiting distribution
LR^0	$\mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}\mathbf{N}'\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}\mathrm{d}\mathbf{N}'\right)\right\}$
LR^{i0}	$\mathrm{tr}\left\{\left(\int_0^1 \bar{\mathbf{N}}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \bar{\mathbf{N}}\bar{\mathbf{N}}'\mathrm{d}s\right)^{-1}\left(\int_0^1 \bar{\mathbf{N}}\mathrm{d}\mathbf{N}'\right)\right\}$
LR^*	$\mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}^*\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}^*\mathbf{N}^{*\prime}\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}^*\mathrm{d}\mathbf{N}'\right)\right\}$
LR^{SL}	$\mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}\mathbf{N}'\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}\mathrm{d}\mathbf{N}'\right)\right\}$
LR^+	$\mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}^+\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}^+\mathbf{N}^{+\prime}\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}^+\mathrm{d}\mathbf{N}'\right)\right\}$
LR^{PC}	$\mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}^{PC}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}^{PC}\mathbf{N}^{PC\prime}\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}^{PC}\mathrm{d}\mathbf{N}'\right)\right\}$
LR^{GLS}	$\mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}_*\mathrm{d}\mathbf{N}_*'\right)'\left(\int_0^1 \mathbf{N}_*\mathbf{N}_*'\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}_*\mathrm{d}\mathbf{N}_*'\right)\right\}$
LR^i	$\mathrm{tr}\left\{\left(\int_0^1 \bar{\mathbf{Z}}\mathrm{d}\tilde{\mathbf{N}}'\right)'\left(\int_0^1 \bar{\mathbf{Z}}\bar{\mathbf{Z}}'\mathrm{d}s\right)^{-1}\left(\int_0^1 \bar{\mathbf{Z}}\mathrm{d}\tilde{\mathbf{N}}'\right)\right\}$
LR^i_{ta}	$\mathrm{tr}\left\{\left(\int_0^1 \bar{\mathbf{Z}}\mathrm{d}\tilde{\mathbf{N}}'\right)'\left(\int_0^1 \mathbf{Z}\mathbf{Z}'\mathrm{d}s\right)^{-1}\left(\int_0^1 \bar{\mathbf{Z}}\mathrm{d}\tilde{\mathbf{N}}'\right)\right\}$

to place identifying restrictions on the cointegrating vectors.

3.6.2 Choosing the VAR Order

In the foregoing theoretical discussion it has been assumed that the DGP is a finite order VAR(p) process and that the order p is known. Of course, these assumptions are unrealistic in practice for various reasons. For instance, the true DGP may not be a finite order process.

If it is a finite order VAR process then the true order is not likely to be known. Therefore it is of interest to know the consequences of a violation of the assumption that the DGP is a VAR process with known finite order.

Lütkepohl & Saikkonen (1997a) consider this problem by extending asymptotic results of Ng & Perron (1995) for univariate unit root tests. In their analysis they assume a process without deterministic trend term, that is, regarding the deterministic term they work under the assumptions of Sections 3.1 and 3.2. Two alternative ways for specifying the VAR order are discussed. The first possibility is to use a deterministic rule for choosing the order. It is shown that if the DGP is a finite or infinite order VAR and the order p used in computing LR test statistics is chosen as a function of the sample size T in such a way that $p \to \infty$ and $p^3/T \to 0$ as $T \to \infty$, then the limiting null distributions of the LR tests for the cointegrating rank are the same as in the case of a known finite order.

The second and perhaps more realistic approach considered by Lütkepohl & Saikkonen (1997a) assumes that the order is chosen by some data-dependent rule. Specifically, it is shown that if a model selection criterion of the form

$$\log\det(\tilde{\Omega}_p) + (p+1)C_T/T, \tag{3.37}$$

with

$$p \leq p_T = o(T^{1/3}), \quad C_T > n^2, \quad \text{and} \quad C_T/T \to 0 \quad \text{as} \quad T \to \infty$$

is applied for choosing an order $\tilde{p}$, say, for the model on which the cointegration tests are based, then the limiting null distributions of the LR tests are again the same as in the case of a known finite order. In (3.37), $\tilde{\Omega}_p$ is the usual estimator of the residual covariance matrix Ω based on a model of order p (see (3.3)). The result regarding the limiting distribution remains true if the DGP is of finite or infinite order. It follows that the order may be chosen, e.g., by the Akaike (1974, 1973) information criterion which minimizes

$$AIC(p) = \log\det(\tilde{\Omega}_p) + 2(p+1)n^2/T,$$

so that $C_T = 2n^2$, or by the Schwarz (1978) criterion,

$$SC(p) = \log\det(\tilde{\Omega}_p) + (p+1)n^2\log T/T,$$

so that $C_T = n^2 \log T$, or by other similar criteria (see, e.g., Lütkepohl (1991, Chapter 4)). If any of these criteria is used for order selection prior to testing for the cointegrating rank, the

asymptotic properties of the LR tests remain unchanged. Note that the SC criterion will choose the order consistently if the DGP is actually of finite order whereas AIC overestimates the order with positive probability in this case. The aforementioned result for the LR tests remains valid in either case despite the differences in the properties of the order selection criteria. Thus, the asymptotic theory for the cointegration tests does not provide help in discriminating between the criteria.

In contrast, Reimers (1992), Haug (1996), Lütkepohl & Saikkonen (1997a), Boswijk & Franses (1992), Cheung & Lai (1993), Yap & Reinsel (1995), Ho & Sørensen (1996), Saikkonen & Luukkonen (1997), Gonzalo & Pitarakis (1997b), Bierens (1997b) and Gonzalo & Lee (1998) have found by simulation that in small samples the cointegration tests are strongly dependent on the choice of the VAR order. Choosing a very small lag length which results in a poor approximation of the true DGP may equally well result in major size distortions and reduced power of the tests. A similar result is also obtained if a large lag length is used which introduces substantial sampling uncertainty into the estimated model.

Generally, increasing the lag length eventually results in size distortions and loss in power. Therefore, most studies recommend to use an order selection criterion to choose the lag order. However, the different studies provide different recommendations regarding the order selection criterion. Apparently, the recommendations depend on the DGP used for the simulations. If processes with one lag are used, the parsimonious SC criterion performs best, as would have been expected (see e.g. Reimers (1992) and Haug (1996)). Otherwise, if higher order processes are employed in the simulations, the Hannan-Quinn criterion or AIC turn out to be advantageous (see, e.g., Reimers (1992) and Lütkepohl & Saikkonen (1997a)).

In general, however, choosing the lag length with order selection criteria which tend to find a balance between a good approximation of the DGP and an efficient use of the sample information appears to be a good strategy for applied work (see Lütkepohl & Saikkonen (1997a)). A universally optimal criterion for this purpose does not seem to exist, however.

3.6.3 Subset and Conditional Models

Although we have presented the likelihood based cointegration tests in terms of full VARs or EC models where no restrictions are placed on the short-run parameters Γ_j, $j = 1, \ldots, p-1$, it should be understood that such restrictions have no implications for the asymptotic

properties of the tests. On the other hand, the computation of the test statistics may become a bit more difficult if there are restrictions on the short-run coefficients, e.g., if lags of some variables appear in one equation but not in others. This situation is fairly common in practice and implies that there are zero restrictions on the short-run parameters. In that case optimization of the likelihood function and, hence, computation of the test statistics may require numerical optimization algorithms. Alternatively, one may use some kind of GLS approach or one may simply ignore the restrictions. Although this does not change the asymptotic properties of the cointegration tests, it may have an impact on the small sample properties.

The situation is different if there are restrictions on the loading coefficients α. If in some of the equations of the full system none of the cointegrating relations is actually present, that is, some rows of α are restricted to zero, the corresponding variables are weakly exogenous for the cointegration parameters. In that case a conditional model can be set up for those variables which are not weakly exogenous. The number of cointegrating relations can then be tested within the conditional model. Harbo, Johansen, Nielsen & Rahbek (1998) consider LR tests for this purpose and Kauppi (1997) extends the LM type tests for application in conditional models. He also shows by way of simulation that tests for the cointegrating rank in conditional models may be considerably more powerful than the corresponding LR tests based on full models. These results presume that the conditioning is done correctly, i.e., all unmodelled variables are indeed weakly exogenous.

Rahbek & Mosconi (1998) and Seo (1998) discuss the case where additional stationary regressors appear in a VAR model. They show that these variables have an impact on the asymptotic distribution of LR tests for the cointegrating rank and they propose modifications which result in asymptotically valid tests.

3.6.4 Seasonal Dummies and Structural Shifts

In the foregoing discussion we have focussed on the specific case where the deterministic part of the DGP consists of mean and linear trend terms because these terms are of particular importance from a practical point of view and they are also important for the asymptotic theory. In practice there are sometimes additional candidates for deterministic terms. Important examples are seasonal dummy variables and shift dummies which may be used to

account for structural breaks in the DGP. Whereas the inclusion of centered seasonal dummies has no consequences for the asymptotic properties of the tests (see Johansen (1995)) the same is not true for shift dummies of the form

$$d_t = \begin{cases} 0, & t < T_1 \\ 1, & t \geq T_1 \end{cases},$$

that is, the DGP has the form

$$y_t = \mu_0 + \mu_1 t + \delta d_t + x_t, \quad t = 1, 2, \ldots, \tag{3.38}$$

where x_t is as in Section 2. LR tests for the cointegrating rank in this situation have been discussed by Johansen & Nielsen (1993). They show that the asymptotic distribution under the null hypothesis depends on the timing of the shift or, more precisely, it depends on the fraction of the sample before and after the structural shift. If the fraction is known, that is, T_1 is a known break point, then critical values can be generated.

An alternative proposal is due to Saikkonen & Lütkepohl (1998b). They extend the approach described in Section 3.3.2 and propose to estimate the deterministic term in (3.38) by GLS type procedures first. Then the estimated deterministic term is subtracted from y_t and the cointegration tests are applied to the adjusted data. The resulting test statistics have the same limiting null distributions as the $LR^{GLS}_{trace(ta)}$ statistics in Section 3.3.2.

In these proposals the break date is assumed to be known. For applied work procedures for unknown break dates are also of interest. Moreover, it would be useful to have tests which allow for more general types of structural changes.

3.6.5 Sequential Testing

Often it is of interest to choose the cointegrating rank from the full range of possibilities without prior knowledge or a good guess of what the actual rank may be. In that case one may want to test various different ranks before a decision can be made. A typical sequence of tests proceeds by checking $H_0(0)$ and, if that null hypothesis is rejected, $H_0(1)$, etc. until the null hypothesis cannot be rejected for the first time. Obviously the tests in such a sequence are not independent. In particular, for $r_0 > 0$, a test of $H_0(r_0)$ is performed only if $H_0(0), \ldots, H_0(r_0 - 1)$ are rejected previously. Therefore Johansen (1995, Chapter 12) considers the joint probability $Pr\{LR(0) > c_{n,\gamma}, \ldots, LR(r) > c_{n-r,\gamma}\}$, where $LR(r_0)$ is

the suitable LR statistic for the DGP at hand and $c_{n-r_0,\gamma}$ is the corresponding asymptotic critical value for a test with significance level γ. Johansen shows that the asymptotic rejection probability is bounded by γ,

$$\lim_{T\to\infty} Pr\{LR(0) > c_{n,\gamma}, \ldots, LR(r) > c_{n-r,\gamma}\} \leq \gamma$$

with equality if r is the true cointegrating rank, that is, $r = \mathrm{rk}(\Pi)$. Thus, asymptotically the testing sequence has the correct size. In small samples the test decision will strongly depend on the properties of the individual tests, however, which will be investigated in more detail in Section 5.

3.6.6 Bootstrapping

In many situations it is possible to improve the small sample performance of asymptotically valid statistical procedures by using the bootstrap. Therefore it is not surprising that the bootstrap was also considered in the context of testing for cointegration. For example, Van Giersbergen (1996) and Harris & Judge (1998) investigate the possibility to generate more accurate critical values for LR cointegration tests by resampling the residuals of a VAR or EC model. Harris & Judge consider a standard parametric bootstrap for that purpose whereas van Giersbergen also investigates a nonparametric variant, the so-called stationary bootstrap, which samples blocks of residuals of random length. A possible advantage of the latter resampling scheme is that it may capture remaining residual autocorrelation in case of underspecified lag order.

In Monte Carlo experiments mixed evidence is found regarding the potential for improving cointegration tests by using bootstrap methods. In other words, it is by no means clear that the extra computational burden associated with bootstrapping methods is justified in the case of interest here. Moreover, a formal proof does not seem to be available that bootstrapping cointegration tests actually improves the convergence rate towards the asymptotic distribution.

3.6.7 Nonnormal Processes

The likelihood based tests considered in the previous sections are derived under Gaussian assumptions, that is, the DGP is assumed to be normally distributed. Although the Gaussian

assumptions can be relaxed for much of the asymptotic theory, the Gaussian likelihood is the basis for deriving the explicit form of the test statistics. In practice, many time series and especially financial data contain outliers and volatility clusters which are not consistent with Gaussian assumptions. Therefore Lucas (1997, 1998) considers other than Gaussian likelihood functions and derives LR and LM type tests for the cointegrating rank from them. Specifically he considers (pseudo) likelihood functions which are essentially of the form

$$\prod_{t=1}^{T} \det(\Omega)^{-1/2} \exp\{-\psi(\Omega^{-1/2}\varepsilon_t)\}, \tag{3.39}$$

where $\psi(\cdot)$ is some suitable function. Of course, for $\psi(u) = u'u/2$ the Gaussian likelihood is obtained. Using this set-up Lucas shows that considerable power gains are possible for cointegration tests if the distribution of the DGP is far from normal. Such gains may also result if (3.39) is not the true but just a pseudo likelihood. On the other hand, the level and the power of the tests strongly depend on the true DGP. If it is in fact Gaussian or close to normality some power may be lost by using the pseudo likelihood approach. Therefore Boswijk & Lucas (1997) propose a seminonparametric approach where the true density is estimated rather than assumed and, hence, can adapt to the actual data properties.

Another approach to deal with outliers is used by Caner (1998) who considers the case where the residuals may have a distribution for which the variances do not exist. He shows that considerable size distortion may result for the standard LR type tests in that case.

3.6.8 Possible Pitfalls

A number of studies consider situations where specific problems may arise for cointegration tests. We already mentioned the case of nonnormal distributions which also falls into this category. Other examples are processes with long memory, variables which are actually $I(2)$ or a singular or almost singular residual covariance matrix. These and other possible problems are analyzed by Gonzalo & Lee (1998) in the present context. They find that the aforementioned problems may lead to finding spurious cointegration relations by standard LR tests for the cointegrating rank. An LR test in the presence of variables with long memory is discussed by Lyhagen (1998). Ho & Sørensen (1996) consider large dimensional systems and find that the variability of the number of cointegrating relations found by LR tests gets larger if the number of observations is small and the dimension gets larger and/or the VAR order is misspecified.

Table 3: Other Tests

Assumptions for the deterministic terms	Test statistic	Section	References
$\mu_0 = \mu_1 = 0$	LR_{trace}^{aug}	Sec. 4.1	Lütkepohl & Saikkonen (1997b)
	LR_{trace}^{FM}, W_{trace}^{FM}	Sec. 4.1	Quintos (1998b)
	SW^{VAR}, SW^{NP}	Sec. 4.3	Stock & Watson (1988)
μ_0 arbitrary, $\mu_1 = 0$	LR_{trace}^{aug}	Sec. 4.1	Lütkepohl & Saikkonen (1997b)
	LR_{trace}^{FM}, W_{trace}^{FM}	Sec. 4.1	Quintos (1998b)
	BY_{min}, BY_{trace}, BY^{ρ}, BY^{t}	Sec. 4.2	Bewley & Yang (1995)
	SW_{ma}^{VAR}, SW_{ma}^{NP}	Sec. 4.3	Stock & Watson (1988)
μ_0, μ_1 arbitrary	LR_{trace}^{aug}	Sec. 4.1	Lütkepohl & Saikkonen (1997b)
	LR_{trace}^{FM}, W_{trace}^{FM}	Sec. 4.1	Quintos (1998b)
	SW_{ta}^{VAR}, SW_{ta}^{NP}	Sec. 4.3	Stock & Watson (1988)
	BI_{min}, BI_{trace}	Sec. 4.4	Bierens (1997b)
μ_0 arbitrary $\mu_1 \neq 0, \beta'\mu_1 = 0$	BY_{min}^{i}, BY_{trace}^{i}, BY_{max}	Sec. 4.2	Yang & Bewley (1996)

4 Other Tests

An overview of the tests discussed in this section is given in Table 3. Detailed explanations of the tests and the notation will be given in the following.

4.1 Tests Based on Limiting χ^2 Distributions

A couple of tests have been proposed based on statistics with standard limiting χ^2 distributions under H_0. Lütkepohl & Saikkonen (1997b) obtain this result by augmenting the VAR DGP by an additional lag whereas Quintos (1998a, b) uses the fully modified estimation approach of Phillips (1995) as the basis for her tests. We will discuss the two approaches in turn.

4.1.1 Lag Augmentation Tests

Lütkepohl & Saikkonen (1997b) point out that a small modification of Johansen's LR tests results in a χ^2 limiting null distribution which does not depend on the deterministic trend

terms. Assuming again that the true DGP is a finite order VAR(p) process, it suffices to augment the regressor vector z_t by y_{t-1-p} and to estimate a VAR($p+1$) instead of the VAR(p). Let us denote the resulting M_T matrix by M_T^{aug}. Notice, that we still have $\hat{\Pi} = -(I_n - \hat{A}_1 - \ldots - \hat{A}_p)$, where the $\hat{A}_j$ are now LS estimators of the coefficient matrices of a VAR($p+1$) model. Equivalently, $\hat{\Pi}$ may be obtained as the LS estimator of the matrix Π in the model

$$\Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + \Xi y_{t-p-1} + \varepsilon_t.$$

The test statistic obtained in this way is denoted by $LR_{trace}^{aug}(r_0)$, that is, $LR_{trace}^{aug}(r_0) = (T-p)\sum_{j=r_0+1}^{n} \log(1+\hat{\lambda}_j^{aug})$, where the $\hat{\lambda}_j^{aug}$ are now obtained from the lag augmented model. It has a $\chi^2((n-r_0)^2)$ limiting distribution under the null hypothesis which does not depend on the properties of the deterministic terms. The same asymptotic χ^2 distributions are also obtained if ν_1 and/or ν_0 are known to be zero and are thus set to zero in the estimation. Of course, asymptotically equivalent statistics are obtained by using small sample corrections as in the Reinsel-Ahn and Yap-Reinsel variants of the LR statistics in (3.8) and (3.9). Moreover, the Wald version has the same asymptotic properties as the LR version.

One would expect that the standard LR type tests have better power properties because an inefficiency is built into the modified statistics by adding an extra lag in the estimation procedure. Although adding one extra lag may seem like a minor modification, in this case it has rather severe consequences for the properties of the estimators and, hence, for the tests of the cointegrating rank. Consider, for instance, a univariate $I(1)$ process. Then the estimator $\hat{\Pi}$ is superconsistent ($\hat{\Pi} - 1 = O_p(T^{-1})$) if the true AR order is used whereas it is just $\sqrt{T}$-consistent ($\hat{\Pi} - 1 = O_p(T^{-1/2})$) if an extra lag is added (see Choi (1993)). Hence, a unit root test based on the latter estimator is asymptotically inefficient relative to one based on the former estimator. More precisely, the local power of the lag augmentation test of the hypothesis $H_0 : \Pi = 1$ against local alternatives of the form $\bar{H}_T : \Pi = 1 - \theta/T$ will be equal to the size of the test whereas the tests of the previous section were seen to have larger power against such alternatives.

In a multivariate situation the difference between the convergence properties of the two estimators may not be quite so dramatic. In general there will be just some linear combinations of elements of Π which will be estimated superconsistently if there is no overfitting, whereas estimators of individual elements have the usual $\sqrt{T}$-convergence rate. As a con-

sequence, in small samples the loss in power may not be quite as severe in multivariate cointegration tests as in univariate unit root tests. The small sample properties of the lag augmentation tests will also be explored in Section 5.

4.1.2 Tests Based on Fully Modified Estimation

In presenting the tests of Quintos (1998b) we use the following notation. For time series vectors v_t, w_t define

$$\Sigma_{vw} = E(v_t w_t'),$$
$$\Lambda_{vw} = \sum_{j=1}^{\infty} E(v_t w_{t+j}'),$$
$$\Omega_{vw} = \Sigma_{vw} + \Lambda_{vw} + \Lambda_{vw}'$$

and

$$\Delta_{vw} = \Sigma_{vw} + \Lambda_{vw},$$

with estimators $S_{vw} = \frac{1}{T}\sum_{t=1}^{T} v_t w_t'$ of Σ_{vw} and a kernel based estimator $\hat{\Lambda}_{vw}$ of Λ_{vw} (see, e.g., Phillips (1995)). The resulting estimators of Ω_{vw} and Δ_{vw} are indicated by hats as well.

Now we focus on a model without deterministic terms,

$$\Delta y_t = \Pi y_{t-1} + u_t,$$

where $u_t = \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t$. Although the lag order p is not assumed known in the present setup there are some assumptions that make the model class more restrictive than in Section 3. Perhaps the most restrictive of these assumptions is that $\beta' u_t$ is not serially correlated, where β is as usual the cointegration matrix (this is an implication of Assumption 2 of Quintos (1998b)). Of course, this assumption restricts the Γ_j matrices in the definition of u_t. In fact, if $r = n$ is not excluded as a possibility, the condition implies that u_t has to be white noise and, hence, only first order VAR models are considered. Another assumption made by Quintos (1998b) is that the lag polynomial $I_r - (I_r + \beta'\alpha)L$ is invertible which implies that $\beta'\alpha$ is nonsingular. This assumption may not be as restrictive as the previous one and it is, in fact, satisfied for VAR(1) processes under the assumption that all variables are at most $I(1)$. For higher order processes the assumption does exclude parts of the parameter space which are permitted in the framework of Section 3. For instance, if in a bivariate process β represents a spread (i.e., $\beta' = (1, -1)$), the loadings of the cointegration relations cannot be the same for both equations, that is, $\alpha' = (1, 1)$ is not a possible loading vector.

For presenting the test statistics we define

$$\Delta y_t^+ = \Delta y_t - \hat{\Omega}_{u\Delta y}\hat{\Omega}_{\Delta y\Delta y}^{-1}\Delta y_{t-1}$$

and

$$\hat{\Delta}_{\Delta yu}^+ = \hat{\Delta}_{\Delta yu} - \hat{\Delta}_{\Delta y\Delta y}^+\hat{\Omega}_{\Delta y\Delta y}^{-1}\hat{\Omega}_{\Delta yu},$$

where $\hat{u}_t = \Delta y_t - \hat{\Pi} y_{t-1}$ is substituted for u_t with $\hat{\Pi}$ being the LS estimator of Π. Furthermore, we let $\hat{\lambda}_1^{FM} \geq \cdots \geq \hat{\lambda}_n^{FM}$ be the ordered eigenvalues of

$$\hat{\Omega}_{u\Delta y}^{-1/2}(S'_{y\Delta y^+} - \hat{\Delta}_{\Delta yu}^{+\prime})S_{yy}^{-1}(S_{y\Delta y^+} - \hat{\Delta}_{\Delta yu}^+)\hat{\Omega}_{u\Delta y}^{-1/2}.$$

Then the test statistic proposed by Quintos (1998b) is

$$W_{trace}^{FM}(r_0) = T\sum_{j=r_0+1}^{n}\hat{\lambda}_j^{FM}$$

which has an asymptotic $\chi^2((n-r_0)^2)$ distribution under $H_0(r_0) : \mathrm{rk}(\Pi) = r_0 > 0$. For $r_0 = 0 = \mathrm{rk}(\Pi)$ it has a degenerate distribution, i.e., $W_{trace}^{FM}(0) \xrightarrow{p} 0$. Thus it may be used with critical values from a χ^2 distribution for $r_0 = 0, \ldots, n-1$, although it will be conservative for $r_0 = 0$. Quintos also points out that a test with asymptotically correct size can be obtained if suitable instrument variables are available that can be used in the estimation of the matrix Π and hence in the rank test. An asymptotically equivalent alternative to $W_{trace}^{FM}(r_0)$ is the LR type statistic

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

Quintos (1998b) mentions that modifications with intercept term and deterministic trend are possible without changing the χ^2 limiting distribution. A similar test statistic to $W_{trace}^{FM}(r_0)$ based on fully modified generalized method of moments estimation is proposed by Quintos (1998a) for the case where suitable instruments are available. This test also has a χ^2 limiting distribution if $r = 0$.

Although the local power properties are not given by Quintos (1998b), it seems plausible that the tests have local power of a similar rate as that of the LR tests discussed in Section 3. Hence, the loss in local power may be less than for the lag augmentation tests. Of course, the price is that more restrictive assumptions are required for the DGP and that additional instrument variables may be required.

4.2 Tests Based on Canonical Correlations of Levels

Bewley & Yang (1995) and Yang & Bewley (1996) propose tests for cointegration which may be thought of as being based on the canonical correlations of the levels y_t and y_{t-1} if the DGP is a VAR(1) process. Thereby they differ from Johansen's LR tests which, in the special VAR(1) case, may be regarded as a canonical correlation analysis for the levels and first differences. Consider a VAR(p) model in EC form,

$$\Delta y_t = \nu_0 + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots \tag{4.1}$$

and suppose (2.3) is the pair of hypotheses of interest. Bewley & Yang propose to compute residuals g_t and h_t from regressions of y_t and y_{t-1} on $[1, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}]$, define $G = [g_{p+1}, \ldots, g_T]$, $H = [h_{p+1}, \ldots, h_T]$ and solve the eigenvalue problem

$$\det[GH'(HH')^{-1}HG' - \gamma GG'] = 0.$$

Let $\gamma_1 \leq \cdots \leq \gamma_n$ be the ordered eigenvalues and $q_1, \ldots, q_n$ the corresponding eigenvectors satisfying

$$GH'(HH')^{-1}HG'q_j = \gamma_j GG'q_j, \quad j = 1, \ldots, n.$$

Note that for $p = 1$ the γ_j are just the canonical correlations of y_t and y_{t-1}. The first r_0 of the q_j are used to estimate the cointegration relations, i.e., $\hat{\beta} = [q_1, \ldots, q_{r_0}]$. Estimators of the other parameters in (4.1) are then obtained by replacing β with $\hat{\beta}$ and applying multivariate LS estimation to (4.1). The resulting estimators will be denoted by $\hat{\nu}_0, \hat{\alpha}, \hat{\Gamma}_j$ and the residuals are $\hat{\varepsilon}_t$. From these quantities we also obtain estimators $\hat{\alpha}_\perp$ and $\hat{\beta}_\perp$ in the usual way. Based on these estimators the following definitions are made:

$$\hat{C} = \hat{\beta}_\perp[\hat{\alpha}'_\perp(I_n - \sum_{j=1}^{p-1} \hat{\Gamma}_j)\hat{\beta}_\perp]^{-1}\hat{\alpha}'_\perp,$$

$$h_t^{r_0} = \hat{\alpha}'_\perp h_t \quad \text{and} \quad H_{r_0} = [h_{p+1}^{r_0}, \ldots, h_T^{r_0}],$$

$$g_t^{r_0} = \hat{\alpha}'_\perp(h_t + \hat{C}\hat{\varepsilon}_t) \quad \text{and} \quad G_{r_0} = [g_{p+1}^{r_0}, \ldots, g_T^{r_0}].$$

Denoting the ordered eigenvalues obtained as solutions of

$$\det[G_{r_0}H'_{r_0}(H_{r_0}H'_{r_0})^{-1}H_{r_0}G'_{r_0} - \gamma G_{r_0}G'_{r_0}] = 0 \tag{4.2}$$

by $\hat{\gamma}_1 \leq \cdots \leq \hat{\gamma}_{n-r_0}$, the following statistics result:

$$BY_{min}(r_0) = (T-p)(1-\hat{\gamma}_1) \tag{4.3}$$

and

$$BY_{trace}(r_0) = (T-p)\sum_{j=1}^{n-r_0}(1-\hat{\gamma}_j). \tag{4.4}$$

When $\hat{\gamma}_1$ in $BY_{min}(r_0)$ is small and, hence, the statistic is large, this indicates that there are fewer than $n - r_0$ common trends. Thus, $H_0(r_0)$ is rejected when the statistic assumes a large value. The asymptotic null distributions of these statistics are functionals of Brownian motions and depend on $n - r_0$ and the trending properties of the process. Some critical values for the case without deterministic trend ($\mu_1 = 0$) are given in Table 1 of Bewley & Yang (1995)[2]. Critical values of the test statistics BY^i_{trace} and BY^i_{min} for the case $\mu_1 \neq 0$, $\beta'\mu_1 = 0$, may be found in Table 1 of Yang & Bewley (1996).

In the latter article the authors also note that for a given cointegrating rank r_0, the pair of hypotheses

$$H_0 : \mu_1 = 0 \quad \text{vs.} \quad H_1 : \mu_1 \neq 0$$

may be tested with the statistic

$$BY_{max}(r_0) = (T-p)(1-\hat{\gamma}_{n-r_0}) \tag{4.5}$$

where H_0 is rejected if $BY_{max}(r_0)$ is small. Critical values appear in Table 1 of their article.

Bewley & Yang (1995) also propose two other tests for the case $\mu_1 = 0$. They suggest to perform Dickey-Fuller type tests for the parameter ρ in

$$\hat{q}_1' g_t^{r_0} = \rho \hat{q}_1' h_t^{r_0} + v_t,$$

where $\hat{q}_1$ is the eigenvector associated with $\hat{\mu}_1$ from the eigenvalue problem (4.2). Denoting the regression estimator of ρ by $\hat{\rho}$ and its t-statistic by $t_{\hat{\rho}}$, the resulting statistics are

$$BY^{\rho}(r_0) = (T-p)(\hat{\rho}-1) \tag{4.6}$$

and

$$BY^{t}(r_0) = t_{\hat{\rho}}. \tag{4.7}$$

Critical values for these tests for the driftless case ($\mu_1 = 0$) are also given in Table 1 of Bewley & Yang (1995). The tests have some similarity to tests based on the residuals of a

[2]The table has to be used cautiously because the legend is not fully clear and the user is expected to add minus signs to some of the entries.

cointegrating regression as proposed by Engle & Granger (1987), Engle & Yoo (1987) and Phillips & Ouliaris (1990).

In assessing the tests proposed here it may be worth pointing out that Bewley & Yang work under similar conditions as the corresponding tests presented in Section 3 except that they use the additional assumption that $\alpha'_\perp\beta_\perp$ is nonsingular. This assumption is not restrictive for first order VAR processes because for these processes it follows from the fact that $\alpha'_\perp(I_n - \sum_{j=1}^{p-1}\Gamma_j)\beta_\perp$ has to be invertible if $I(2)$ processes are to be excluded (see Section 2). However, for higher order processes the assumption excludes processes which are permitted in the likelihood framework of Section 3. Given that the likelihood based approaches often have superior properties to more ad hoc approaches and given that a likelihood ratio approach results in a canonical correlation analysis of first differences and lagged levels one may wonder about the properties of the approach of this section especially as it requires more restrictive assumptions in some sense.

Bewley & Yang (1995) and Yang & Bewley (1996) mention that the tests are consistent. In the former article some small sample simulations are reported which show that the power of the BY_{trace} test is not very good. Therefore its use is not recommended. On the other hand, in some cases the other tests proposed by Bewley and Yang have better small sample power than the corresponding LR tests. The small sample properties of these tests relative to other tests will be explored in more detail in Section 5.

4.3 Stock–Watson Tests

We present the tests proposed by Stock & Watson (1988) for the case where x_t is observable first, that is, we assume $\mu_0 = \mu_1 = 0$. Afterwards, the necessary modifications for processes with deterministic terms are discussed. The Stock–Watson tests are suitable for the pair of hypotheses given in (2.3), that is, $H_0(r_0) : r = r_0$ is tested against the alternative $H_1(r_0) : r > r_0$. The idea underlying the tests is as follows. If there are r_0 cointegrating relationships for x_t then there must be $n - r_0$ common trends. Suppose the original data have been transformed such that the last $n - r_0$ components of x_t are just the common trends. In practice this is usually not the case. Therefore Stock & Watson suggest to perform a principal components analysis and order the principal components from smallest to largest. Denoting by Q the orthogonal matrix whose rows correspond to the principal components, $z_t = Qy_t$ where the y_t

are the original data. The common trends τ_t are then chosen to be the last $n-r_0$ components of z_t. Under $H_0(r_0)$ this process has $n - r_0$ unit roots (i.e., $\Delta\tau_t$ is stationary with possibly infinite order VAR representation). Thus, computing the multivariate LS estimator of Φ in $\tau_t = \Phi\tau_{t-1} + u_t$ results in a matrix

$$\tilde{\Phi} = \left(\sum_{t=2}^{T} \tau_t \tau_{t-1}'\right) \left(\sum_{t=2}^{T} \tau_{t-1}\tau_{t-1}'\right)^{-1} \tag{4.8}$$

whose eigenvalues approach 1 in large samples. Therefore the test rejects $H_0(r_0)$ if the eigenvalue of $\tilde{\Phi}$ with smallest modulus is much smaller than 1.

In order to get a test statistic which does not depend on unknown nuisance parameters adjustments for the short-term dynamics are made. Two types of adjustments were suggested by Stock & Watson (1988). The first one assumes that τ_t has a VAR(p) representation. In that case $\Delta\tau_t$ has a VAR($p-1$) representation, say

$$\Delta\tau_t = \Xi_1 \Delta\tau_{t-1} + \cdots + \Xi_{p-1}\Delta\tau_{t-p+1} + v_t.$$

Denoting the LS estimators of the Ξ_i by $\hat{\Xi}_i$ we may compute residuals

$$\hat{\xi}_t = \tau_t - \hat{\Xi}_1\tau_{t-1} - \cdots - \hat{\Xi}_{p-1}\tau_{t-p+1}$$

and use them to compute the estimator

$$\hat{\Phi} = \left(\sum_{t=p+1}^{T} \hat{\xi}_t \hat{\xi}_{t-1}'\right) \left(\sum_{t=p+1}^{T} \hat{\xi}_{t-1}\hat{\xi}_{t-1}'\right)^{-1}.$$

The test statistic proposed by Stock & Watson is then

$$SW^{VAR}(r_0) = (T-p)(\text{real}(\hat{\lambda}_{min}) - 1),$$

where real($\hat{\lambda}_{min}$) denotes the real part of the eigenvalue of $\hat{\Phi}$ with smallest modulus. $H_0(r_0)$ is rejected if $SW^{VAR}(r_0)$ is small. The asymptotic null distribution of this test statistic is nonstandard. Its critical values are tabulated in Table 1 of Stock & Watson (1988).

This test may be regarded as nonparametric because the VAR(p) may just be an approximation to a more general DGP of the common trends τ_t. Note that, if y_t is a VAR(p) process, τ_t will in general not have a finite order VAR representation. Therefore Stock & Watson (1988) also propose another nonparametric procedure to account for the short-term dynamics and thereby eliminate nuisance parameters in the limiting distribution of their test

statistics. They suggest to compute residuals $\hat{u}_t = \tau_t - \tilde{\Phi}\tau_{t-1}$, where $\tilde{\Phi}$ is defined in (4.8). Then an estimator of the covariance matrix of u_t is determined as

$$\hat{\Omega}_u = \sum_{s=1}^{T} K\left(\frac{s}{h_T}\right) \sum_{t=s+1}^{T} \hat{u}_t \hat{u}'_{t-s},$$

where h_T is the bandwidth and $K(\cdot)$ is a kernel function. For instance, the quadratic kernel may be used:

$$K(z) = \frac{25}{12\pi^2 z^2}\left[\frac{\sin(6\pi z/5)}{6\pi z/5} - \cos(6\pi z/5)\right].$$

Haug (1996) suggests to use a procedure of automatic bandwidth choice as proposed by Andrews & Monahan (1992) and Andrews (1991). The nonparametric estimator of Φ is then

$$\hat{\Phi}_{NP} = \left(T^{-2}\sum_{t=1}^{T}\tau_t\tau'_{t-1} - T^{-1}\hat{\Omega}_u\right)\left(T^{-2}\sum_{t=1}^{T}\tau_{t-1}\tau'_{t-1}\right)^{-1}.$$

Denoting the real part of the eigenvalue with smallest modulus of this matrix by $\text{real}(\hat{\lambda}^{NP}_{min})$ the test statistic becomes

$$SW^{NP}(r_0) = (T-p)(\text{real}(\hat{\lambda}^{NP}_{min}) - 1).$$

Under H_0 it has the same asymptotic distribution as $SW^{VAR}(r_0)$. Of course, the two statistics will differ in small samples (see Section 5.2).

Modifications are necessary if μ_0 and μ_1 are nonzero. If μ_0 is arbitrary and $\mu_1 = 0$, z_t is replaced by $z_t - \bar{z}$, where z_t is the transformed vector where the last $n - r_0$ components are noncointegrated $I(1)$ processes, i.e. the common trends. The transformation is obtained by a principal components analysis as above. If both μ_0 and μ_1 are nonzero, z_t is replaced by $z_t - \hat{\mu}_0 - \hat{\mu}_1 t$, where $\hat{\mu}_0$ and $\hat{\mu}_1$ are obtained by regressing z_t on a linear trend. The resulting test statistics will be denoted by $SW^{VAR}_{ma}(r_0), SW^{NP}_{ma}(r_0), SW^{VAR}_{ta}(r_0)$ and $SW^{NP}_{ta}(r_0)$, respectively. The limiting distributions of these test statistics differ from those of $SW^{VAR}(r_0)$ and $SW^{NP}(r_0)$. The critical values are listed in Tables 2 and 3 of Stock & Watson (1988).

4.4 Bierens' Nonparametric Tests

Bierens (1997a, b) develops nonparametric tests for the cointegrating rank. In his approach y_t may have a quite general DGP which does not have to be a finite order VAR process. It is assumed that y_t is of the form (2.1) and x_t is such that Δx_t is stationary and ergodic.

Bierens considers weighted means of y_t and Δy_t and exploits the difference in their asymptotic behaviour to construct a test for the cointegrating rank. He defines matrices

$$D_m = \frac{8\pi^2}{T}\sum_{k=1}^{m} k^2 \left(\frac{1}{T}\sum_{t=1}^{T}\cos(2k\pi(t-0.5)/T)y_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\cos(2k\pi(t-0.5)/T)y_t\right)'$$

and

$$F_m = 2T\sum_{k=1}^{m} \left(\frac{1}{T}\sum_{t=1}^{T}\cos(2k\pi(t-0.5)/T)\Delta y_t\right)\left(\frac{1}{T}\sum_{t=1}^{T}\cos(2k\pi(t-0.5)/T)\Delta y_t\right)'.$$

Other weights than $\cos(2k\pi(t-0.5)/T)$ may be chosen here. The present choice is recommended, however, to ensure invariance of the test statistics to drift terms.

Denoting by $\hat{\lambda}_1^B \geq \cdots \geq \hat{\lambda}_n^B$ the generalized eigenvalues obtained from

$$\det[D_m - \lambda(F_m + T^{-2}D_m^{-1})^{-1}] = 0,$$

Bierens proposes to use $BI(r_0) = \hat{\lambda}_{n-r_0}^B$ as a test statistic for testing the hypotheses in (2.3). The null hypothesis $H_0(r_0)$ is rejected if $BI(r_0)$ becomes too small, that is, the test is left-sided. Again the asymptotic null distribution is nonstandard and critical values may be found in Bierens (1997b).

The power of the test will depend on the number of terms m. For a test with 5% significance level Bierens (1997a) suggests using the following rule for choosing this quantity if $n \leq 5$ and $r \leq 4$:

$$m = \begin{cases} n & \text{if } r_0 > 0 \text{ or } n < 2 \\ n+1 & \text{if } r_0 = 0 \text{ and } n \geq 2 \end{cases}.$$

Bierens also points out that the tests are not scale invariant, that is, for a nonsingular matrix Q, in small samples, the test decision may differ if the test is based on Qy_t rather than y_t. Asymptotically the tests are invariant under linear transformations of the data, however.

4.5 Choosing the Cointegrating Rank by Model Selection Criteria

It was pointed out, for instance, by Lütkepohl & Poskitt (1998) and Gonzalo & Pitarakis (1997b) that statistics for testing the cointegrating rank of a system can also be used to construct criteria for estimating the rank consistently. Suppose that $Q_T(r_0)$ is a test statistic for testing $H_0(r_0)$ with the properties

$$\text{plim}\, Q_T(r)/T > 0 \quad \text{for} \quad r < r_0 \tag{4.9}$$

and

$$Q_T(r) = O_p(1) \quad \text{if} \quad r \geq r_0. \tag{4.10}$$

These conditions hold for most of the test statistics considered in the foregoing. Furthermore, let $\varphi(r)$ be any strictly monotonically increasing function of r and let c_T be a sequence of numbers which is $o(T)$. Then, defining $Q_T(n)$ to be zero and choosing the cointegrating rank such that a criterion of the form

$$C_T(r) = Q_T(r) + \varphi(r)c_T, \quad r = 0, \ldots, n, \tag{4.11}$$

is minimized, results in a consistent estimator $\hat{r}$ of the cointegrating rank r. More precisely, in that case, plim $\hat{r} = r$, that is, $Pr[\hat{r} = r] \to 1$ (see Lütkepohl & Poskitt (1998, Therorem 1), Gonzalo & Pitarakis (1997b) and Poskitt (1996) for a result on strongly consistent estimation of the cointegrating rank).

Note that (4.9) says that a test based on $Q_T(r)$ has to be consistent and (4.10) requires basically that the test statistic has an asymptotic distribution if the null hypothesis is true. Thus most of the statistics considered in the foregoing may be used for $Q_T(\cdot)$. Furthermore, the conditions for $\varphi(r)$ and c_T are quite general. For example, for the sequence c_T one may choose $c_T = T$ for every finite sample size and pretend that c_T gets smaller than T eventually for large samples. A simple choice for $\varphi(\cdot)$ would be the identity function, $\varphi(r) = r$. Another choice that has been used in this context is $\varphi(r) = r(2n - r)$, that is, $\varphi(r)$ represents the number of freely varying parameters in the $(n \times n)$ matrix Π when its rank is r. Obviously, both choices of $\varphi(r)$ are strictly monotonically increasing in r and, hence, qualify here. With the latter choice and $c_T = 2$ we get the well-known AIC criterion, the Hannan-Quinn criterion results for $c_T = 2 \log\log T$, and the Schwarz criterion is obtained for $c_T = \log T$ (see, e.g., Lütkepohl (1991, Chapter 4) for a discussion of these criteria). Note that instead of $Q_T(r)$ usually the more common $T \log\det\tilde{\Omega}(r)$ is used, where $\tilde{\Omega}(r)$ is the Gaussian ML estimator of the residual covariance matrix Ω for a cointegrating rank r. If $Q_T(r)$ is the LR statistic based on Gaussian assumptions, $Q_T(r)$ and $T \log\det\tilde{\Omega}(r)$ differ only by a constant which is independent of r (see Lütkepohl (1991, Chapter 11)). Hence, replacing the latter term by the former results in equivalent criteria. Using $Q_T(r)$ in the present context is more convenient from a computational point of view.

HQ and SC are both consistent criteria whereas AIC may not be consistent. On the other hand, even for the SC criterion which has the largest penalty term, the sequence c_T is

still much smaller than T which may be regarded as an upper bound. Hence, from this point of view, even much more parsimonious criteria could be justified. In fact, in a simulation study by Reimers (1992) the criteria did overall quite poorly in finding the true cointegrating rank. Even the most parsimonious criterion SC frequently overestimated the cointegration rank. This result suggests that more parsimonious criteria may be justified not only on the basis of asymptotic theory but also from a small sample point of view.

Phillips (1996) discusses a criterion based on Bayesian principles which can also be used for consistent selection of the cointegrating rank of a VAR process. Another Bayesian variant is proposed by Kleibergen & Paap (1997). Aznar & Salvador (1997) consider the determination of the cointegrating rank, the lag order and the properties of the deterministic terms by model selection criteria. In particular, they discuss the Schwarz criterion for that purpose.

4.6 Further Proposals

There are numerous other proposals for determining the number of cointegration relations in a system of variables. Some of them are quite different from those considered in this survey. Although we do not intend to present them here in detail it may be worth mentioning some of the related literature in order to help the reader in locating other contributions to the topic. Since we are not striving for completeness we apologize for any omissions.

First of all, as mentioned earlier, we have focused our survey on systems cointegration tests and therefore we have neglected the extensive literature on testing for cointegration in conditional single equation models. Examples of publications in this area are Engle & Granger (1987), Phillips & Ouliaris (1990), Banerjee, Dolado, Galbraith & Hendry (1993), Choi (1994), Shin (1994), Haug (1996), to name just a few.

Within the systems framework considered in the present paper, instead of testing the null hypothesis that the cointegrating rank is r_0 against the alternative that the true rank is greater than r_0, it is also possible to test against a smaller rank than r_0. In the univariate case this corresponds to testing the null of stationarity against the $I(1)$ alternative. This type of tests has been considered, e.g., by Kwiatkowski, Phillips, Schmidt & Shin (1992), Choi (1994) and Saikkonen & Luukkonen (1993) in the univariate case. For the multivariate case Harris (1997) proposes a test for $H_0 : r = r_0$ against $H_1 : r < r_0$, $r_0 > 0$, and Snell (1999) considers testing $H_0 : r = r_0$ against $H_1 : r = r_0 - 1$. Choi & Ahn (1995, 1998)

develop tests for the null of stationarity of multiple time series and of the residual vector of a system of equations. The latter tests can be used to check whether the dependent variables are cointegrated with the regressors.

A Bayesian analysis of the number of cointegration vectors is discussed by Kleibergen & Paap (1997) and tests for seasonal cointegration are treated by Johansen & Schaumburg (1997).

5 Small Sample Properties

With respect to the application of the tests in empirical analyses it is important not only to compare their asymptotic properties but also to know how these tests perform in small samples. Therefore, small sample comparisons presented in the literature are summarized in Section 5.1. Since the results of small sample simulations depend to some extent on the DGPs used, new simulations are presented in Section 5.2 which compare a wide range of systems cointegration tests under uniform conditions.

5.1 Previous Studies

A number of small sample simulation studies have been performed to learn more about the properties of the different tests for the cointegrating rank of a set of variables in samples of the size usually encountered in applied work. In some cases, authors who have proposed modifications of existing tests or who have suggested new testing procedures have also provided limited small sample evidence in favour of their proposals (e.g., Stock & Watson (1988), Saikkonen & Luukkonen (1997), Yap & Reinsel (1995), Quintos (1998b)). In other studies a specific class of tests is compared only (e.g., Toda (1994, 1995)), and in yet another group of studies broad ranges of tests are compared (e.g., Reimers (1991, 1992), Haug (1996)). Some of the simulation studies arrive at different conclusions which is partly due to the specific DGPs and simulation designs used. Moreover, conflicting results may arise from different subjective interpretations of the simulation evidence.

Most DGPs used in previous studies of systems cointegration tests are bivariate or three-dimensional processes with lag orders one or two. In some studies higher dimensional processes have also been used. For instance, a four-dimensional VAR(2) process is used in a small simulation study by Reinsel & Ahn (1992) and a five-dimensional process is considered by Bewley & Yang (1995). Gonzalo & Pitarakis (1997a) base their simulations on DGPs of higher dimensions ($n = 2, 5$ and 10). Four- to nine-dimensional data-based DGPs are employed by Ho & Sørensen (1996). Toda (1994, 1995) shows that for comparing LR tests using rather simple DGPs is justified because these tests are invariant to nonsingular linear transformations of the data and, hence, more complicated processes can be reduced to equivalent simple versions. In some studies DGPs derived from real economic data are used (e.g.,

Reimers (1992), Ho & Sørensen (1996) and Bierens (1997b)). Moreover, DGPs with moving average components are employed in a number of simulation studies (Boswijk & Franses (1992), Cheung & Lai (1993), Yap & Reinsel (1995), Saikkonen & Luukkonen (1997), Gonzalo & Pitarakis (1997b) and Lütkepohl & Saikkonen (1997a)). The sample sizes considered in the Monte Carlo studies cover a range of 30 to 10 000 observations. In many cases they are concentrated on sample sizes around 200, however. Hence, most studies cover situations often encountered in macroeconometric applications.

Johansen's LR tests or their small sample modifications are included in almost all simulation studies because they may be regarded as benchmark tests with comparatively good performance and wide application in empirical research. Of the other tests discussed in the previous sections, only the Stock & Watson (SW) tests have been included in several simulations, e.g., by Haug (1996), Reimers (1992) and Yap & Reinsel (1995). The simulation evidence for the other tests considered in this paper is rather limited.

Table 4 summarizes a large number of simulation studies with respect to the specific test statistics considered, the dimension and lag order of the DGPs as well as the sample sizes on which the simulations have been based. Also, it includes the information whether sequential testing procedures have been carried out as well as the main topics analysed in these simulation studies. Some of the results have been discussed in the previous sections where the corresponding theoretical results are considered. For our own simulations the following issues are important:

- In none of the studies all the tests have been compared simultaneously. Therefore, we will include a wide range of tests that cannot be excluded on the basis of previous simulation evidence or some other criterion (see below).

- The performance picture becomes more diffuse for more complicated processes of higher order or larger dimension. Therefore, we will focus on simple processes only. A test or class of tests which does not perform well in simple situations under ideal conditions cannot be expected to do well for complicated DGPs in general even if there is simulation evidence to that effect in a previous simulation study. In other words, in this case we assume that the favorable performance is at least partly a function of the special simulation design in other studies.

Table 4: Some Previous Simulation Studies

Author(s)	Test	DGP	Sample Size	Seq. Tests	Topics analysed
Stock & Watson (1988)	SW_{ma}, SW_{ta}	VAR(2), 3-dim., VARMA(1,1), 3-dim.	200	no	size, power
Reinsel & Ahn (1992)	LR^0, LR^{i0}	VAR(2), 4-dim., $r = 2$	200	no	size, finite sample correction
Reimers (1991, 1992)	LR^{i0}, LR^{i0}_{RA}, SW^{NP}_{ma}, SW^{VAR}_{ma}	random walks (2/3-dim.), VAR(1), 2-dim., VAR(2), 3-dim., databased (German interest rates)	50, 100, 200	no	size, power, lag order, dimension
Boswijk & Franses (1992)	LR^i	VARX, 2-dim., VARMAX, 2-dim.	50	no	size, power, lag order
Cheung & Lai (1993)	LR^{i0}, LR^{i0}_{mod}	VAR(1)^H, 2-dim., VAR(2), 2-dim., VARMA(1,1), 2-dim.	50-300 50-200 200	no	size, power, lag order, dimension, finite order correction, non-normality, MA component
Toda (1994)	LR^i, LR^{PC}	VAR(1), 2-dim.	100, 200, 300, 400	yes	size, power, initial values, innov. corr., determ. terms
Toda (1995)	LR^i, LR^{i0}, LR^*, LR^i & LR^*	VAR(1), 2-dim., with drift parameters	100, 200, 300	yes	size, power, initial values, innovation correlation, deterministic terms
Yap & Reinsel (1995)	LR^{i0}_{YP}, SW^{NP}_{ma}, SW^{VAR}_{ma}	VAR(2), 3-dim., VARMA(1,1), 3-dim.	200	no	size, power, lag order, MA component
Bewley & Yang (1995)	LR^{i0}, BY	VAR(1)^G, 2-dim., VAR(1), 5-dim.	75, 100, 200	no	size, size-adjusted power, innovation correlation
Haug (1996)	LR^i, SW^{NP}_{ma}, SW^{VAR}_{ma}	VAR(1)^G, 2-dim.	100	no	size, (not) size-adjusted power, lag order
Ho & Sørensen (1996)	LR^i	4- to 9-dim. data-based DGP (sectorial data), 7-dim. data-based DGP, (exchange rates) VAR(p), p=1,2,3 , p_{AIC}, p_{SC},	40	yes	lag order, dimension, finite sample correction, ratio no. of parameters to no. of observations,
Lütkepohl & Poskitt (1998)	LR^i_{RA}, AIC, HQ, SC, $CIR1$, $CIR2$	VAR(1), 3-dim. VAR(1)^G, 2-dim.	50, 100, 200	no	size, power, lag order
Saikkonen & Luukkonen (1997)	LR^{i0}, LR^{SL}, W^{i0}, W^{SL}	VAR(2)^{YR}, 3-dim., VARMA(1,1)^{YR}, 3-dim.	200	no	size, size-adjusted power, lag order, MA component
Lütkepohl & Saikkonen (1997b)	LR^+, LR^{PC}, LR^{aug}, LM^{LS}	VAR(1)^T	100, 200	no	size, power, deterministic terms
Saikkonen & Lütkepohl (1997)	LR^+, LR^{PC}, LM^{LS}, LM^{GLS}, LR^{LS}_{ta}, LR^{GLS}_{ta}	VAR(1)^T, 2-dim.	100, 200	no	size, power, innovation correlation, deterministic terms
Saikkonen & Lütkepohl (1998a)	LR^i, LM^i, LR^i_{ta}, LR^+, LR^{PC}, LM^{GLS}	VAR(1)^T, 2-dim.	100, 200	no	size, power, deterministic terms, innovation correlation
Lütkepohl & Saikkonen (1997a)	LR^0, LR^i, AIC, HQ, SC	VAR(2)^{YR}, 3-dim. VARMA(1,1)^{YR}, 3-dim.	100, 200	yes	size, power, lag order, MA component
Bierens (1997a, 1997b)	LR^+, BI	VAR(8), data based, (wages, GNP)	80	yes	size, power, lag order $\alpha = 10\%$
Gonzalo & Pitarakis (1997b)	LR^0, AIC, SC, HQ, $LCIC$,	VAR(1), 3-dim., VAR(2), 2-dim., VARMA(0,1), 3-dim.	150-160	no	size, lag order
Gonzalo & Pitarakis (1997a)	LR^0, AIC, SC, HQ, LCT, $RALR$	VAR(1), n=2,5,10	30, 90, 150, 400	no	size, dimension
Gonzalo & Lee (1998)	LR^i, LR^*, LR^i_{trace}, LR^i_{max}	diff. VAR(1)^G $n = 2$ with det. term	100 (1000, 10000)	no	size, lag order, deterministic terms, order of integration, singularity of error covariance, random unit root coefficient
Quintos (1998b)	LR^0, W^{FM}	VARMA	100, 200, 300	no	size, power

Note: H: DGP of Haug (1996), T: DGP of Toda (1994, 1995) , YR: DGP of Yap & Reinsel (1995), G: DGP of Gonzalo (1994)

- Small sample modifications for LR tests such as those suggested by Reinsel & Ahn (1992) and Yap & Reinsel (1995) (see Section 3.1.2) were found to improve the size or power performance of some tests in specific situations. However, it was found in Cheung & Lai (1993) that these modifications are inadequate in parts of the parameter space. Therefore, we will generally focus on the unmodified, original versions of the statistics. Thereby, situations may be identified where small sample modifications are useful.

- Previous simulation studies have shown that both the trace and the maximum eigenvalue statistic of the LR tests perform similarly (see, for example, Cheung & Lai (1993) and Toda (1994, 1995)). However, Toda (1994) finds that the trace test performs better in some cases where the power of both statistics is quite low. Therefore, only the trace statistics are considered in our simulations.

- Saikkonen & Luukkonen (1997) concluded that the Wald versions of the likelihood based tests are inferior to the LR versions. Therefore, we exclude the Wald versions from our simulation comparison.

- In a series of papers Lütkepohl & Saikkonen propose and compare different versions of LM type tests and LR tests with prior trend adjustment. In the simulations presented below the test statistics that were found to perform best have been used.

- In preliminary simulations we found that the nonparametric SW test performed rather similarly to SW^{VAR}. In contrast to Haug (1996) it was found that neither the SW^{VAR} test nor his variant of SW^{NP} outperforms the other. Our version of the SW^{NP} test performs somewhat worse than the others. Therefore, we report only results for SW^{VAR}.

- According to Bewley & Yang (1995), their minimum test statistic performs similarly to BY^{ρ} and is more robust to variations in the innovation correlation than BY_{trace} and BY^{t}. Furthermore, their trace statistic is outperformed by their other tests. Therefore, only results for BY^{i}_{min} will be reported.

- For the Bierens test the minimum statistic was advocated by Bierens (1997a) and is, therefore, included in the simulations.

- The choice of the lag order was found to have an important impact on the test performance. This problem was studied theoretically in Section 3.6.2 where also some simulation evidence is reviewed. To avoid distortions resulting from inadequate lag order selection we focus on the properties of the tests when the lag order is specified correctly. Again the objective is to single out tests which do not work well under ideal conditions.

- Initial values were found to have an impact on the test performance (see Toda (1994, 1995)). Since in practice the initial values are unknown, we choose them randomly in our simulations by starting the simulated series at zero and discarding the first few observations.

- The deterministic terms of a DGP have a severe impact on the performance of the tests. Therefore, we will structure our results according to the assumptions for the deterministic terms.

5.2 New Simulations

5.2.1 Simulation Details

Tables 1 to 3 (see Sections 3 and 4) summarize the tests reviewed and compared in this paper. The assumptions for the deterministic terms of the DGP, $y_t = \mu_0 + \mu_1 t + x_t$, under which the tests are supposed to work are presented together with the references for the tests. For the LR tests the respective model to be used for estimation is presented, whereas for the other tests the reader is referred to the relevant section of the paper. The tables are also meant as a reference for the notation used in the paper.

Different groups of tests are considered in the following according to their assumptions regarding the deterministic terms with respect to their performance for alternative DGPs. Tests with the underlying assumption $\mu_0 = \mu_1 = 0$ are not included in the simulations as these tests are not relevant for applied research.

The size of the respective tests is analysed first because, in order to control the type I error, a test should not exceed the nominal significance level considerably to be useful in applied work. If the actual size is much smaller than the nominal size, the test is conservative and tends to have little or no power. This is a particular problem in the present case

since the cointegrating rank is usually determined in a sequential procedure starting with the null hypothesis $r = 0$, which will often terminate too early and, hence, indicate fewer cointegrating relations than the true number. On the other hand, if the actual size is much larger than the nominal size, the test procedure tends to overestimate the cointegrating rank which can also result in severe problems for the interpretation. Strictly speaking in that situation we do not have a valid test because bounding the type I error is a fundamental property of a test. In order to evaluate the results with respect to their random variation the standard error of the true rejection probability P based on 1000 replications is given by $s = \sqrt{P(1-P)/1000}$. In the case of a 5% nominal significance level, as in the simulations presented below, the standard deviation is $s = 0.007$ and, thus, the two-standard-error confidence interval for P is $[0.036, 0.064]$. A power comparison is performed for those tests which have roughly a correct size.

The performance of the tests will be evaluated on the basis of their behaviour in terms of size and power in response to variations of the key parameters influencing the DGPs considered (such as the autoregressive eigenvalues, the innovation correlation, the cointegrating rank and the deterministic part). DGPs of different dimensions will be employed in the simulations and the impact of the length of the time series on the test performance will be studied.

In the simulations, the correct lag specification is taken as the most favourable situation for the tests with respect to dynamic specification. We have generated samples of sizes $T = 50, 100, 200$ and 400 plus 50 presample values. The initial value is set to 0 and 50 presample values are discarded to eliminate the influence of the zero initial value. The first sample values are used for estimation purposes. Simulation results are presented for $T = 100$ only as economists often have to work with a similar sample size. Results for other sample sizes are commented on if they provide additional insights.[3]

All programmes are written in GAUSS. The Johansen and Stock & Watson routine are partly adapted from Sam Ouliaris' COINT GAUSS program.[4] The RNDNS function in GAUSS with a fixed seed has been used to generate the pseudo-normal variates. All simulations were performed on a Pentium MMX PC.

[3]Further results on different sample sizes, variations in the parameters of the DGPs and results on further processes are available from the first author upon request.

[4]Version 1, 1991.

The rejection frequencies in the tables are obtained on the basis of asymptotic critical values for a nominal level of 5%. The simulation results presented in Tables 5 to 8 and Figures 3 to 6 are based on unconditional tests for the respective cointegrating rank, that is, none of the tests is conditioned on the outcome of some other tests, as would be the case in sequential testing procedures.

5.2.2 The DGPs

Because one of the main objectives of this study is to analyse the impact of the trending properties of the DGP on the performance of the tests, a DGP of the general form

$$w_t = \begin{bmatrix} 0 \\ \delta e_{n-r} \end{bmatrix} + \begin{bmatrix} \Psi & 0 \\ 0 & I_{n-r} \end{bmatrix} w_{t-1} + u_t, \qquad u_t \sim iidN\left(0, \begin{bmatrix} I_r & \Theta \\ \Theta' & I_{n-r} \end{bmatrix}\right) \tag{5.1}$$

is employed, as in Toda (1994). The quantity δ is a nonnegative scalar, e_{n-r} is an $(n-r)$-vector of the form $e_{n-r} = (0, ..., 0, 1)'$, Ψ is a diagonal matrix with diagonal elements inside or on the unit circle. The diagonal elements of this matrix are the eigenvalues which determine the number of cointegrating relations and common trends. 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 nonsingular linear transformations of w_t and, hence, leave the LR tests for the cointegrating rank invariant. The simplicity of the process makes it attractive for our purposes.

If $n = 2$, the corresponding two-dimensional VAR(1) DGP is of the following form:

$$w_t = \begin{bmatrix} 0 \\ \delta \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix} w_{t-1} + u_t, \qquad u_t \sim iid\, N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}\right). \tag{5.2}$$

If $\psi_1 = \psi_2 = 1$ in DGP (5.2), the cointegrating rank is $r = 0$ and the process consists of two nonstationary components. In this case, the second component has a deterministic linear trend if $\delta \neq 0$ and no linear trend is present 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 $|\psi_1| < 1$ and $|\psi_2| < 1$, the process is stationary ($r = 2$). In that case, a nonzero δ cannot generate a linear trend.

Since the properties of some tests depend partly on whether or not μ_1 is zero, but are otherwise independent of the particular values of the deterministic parameters it suffices to

vary the deterministic part by varying δ. If $\delta = 0$ the deterministic trend term of the DGP in the representation (2.1) is $\mu_1 = 0$. In this case, all tests may be applied that allow for arbitrary values of μ_0 and are derived under the assumption $\mu_1 = 0$. Moreover, tests for arbitrary values of μ_0 as well as μ_1 are applicable. For $\delta \neq 0$ and $r = 0$ or 1 with $\psi_2 = 1$, the mean and trend parameters for the w_t process are

$$\mu_0^w = \begin{bmatrix} 0 \\ w_0 \end{bmatrix}, \qquad \mu_1^w = \begin{bmatrix} 0 \\ \delta \end{bmatrix}$$

and $\beta'\mu_1^w = 0$. In this case, tests which require zero trend parameters are inappropriate whereas all tests may be applied which allow for a nonzero trend parameter and possibly require $\beta'\mu_1^w = 0$. In the simulations we only report results for $\delta = 0$ and 1. The choice of δ has no impact on those tests which allow for a linear trend and are invariant with respect to the values of the trend parameters. Hence, for these tests the value of $\delta \neq 0$ is of no consequence. In studying the small sample behaviour of some tests for which a distinction between $\delta = 0$ and $\delta \neq 0$ is necessary, δ values close to zero give similar results to the case where $\delta = 0$. Hence, focusing only on results where δ is equal to 0 and where it is markedly distinct from zero ($\delta = 1$) provides a good picture of the differences in the behaviour of the tests for $\delta = 0$ and $\delta \neq 0$.

The parameter θ in the residual covariance matrix of the process (5.2) controls the correlation between the two components of the process. For $\theta = 0$ the two components are independent whereas they are instantaneously correlated for $\theta \neq 0$. In the simulations either no innovation correlation ($\theta = 0$) or high innovation correlation ($\theta = 0.8$) is assumed. The simulations are only carried out for nonnegative values of θ because Toda (1995) shows that the LR tests are invariant to the sign of θ.

To understand the relation between the VAR representation (5.2) and the EC representation it may be useful to consider the transformation

$$\begin{aligned} y_{1t} &= w_{1t} + w_{2t} \\ y_{2t} &= w_{2t} \end{aligned} \tag{5.3}$$

Hence,

$$\begin{aligned} \Delta y_{1t} &= (\psi_1 - 1)(y_{1,t-1} - y_{2,t-1}) + (\psi_2 - 1)y_{2,t-1} + \epsilon_{1t} \\ \Delta y_{2t} &= \nu_{20} + (\psi_2 - 1)y_{2,t-1} + \epsilon_{2t} \end{aligned} \tag{5.4}$$

where $\nu_{20} = \delta$ and

$$\begin{aligned} \epsilon_{1t} &= u_{1t} + u_{2t} \\ \epsilon_{2t} &= u_{2t} \end{aligned} \quad \text{so that} \quad \Sigma_\epsilon = \begin{bmatrix} 2+2\theta & 1+\theta \\ 1+\theta & 1 \end{bmatrix}.$$

Thus, if $\psi_2 = 1$ and $|\psi_1| < 1$ (i.e., $r = 1$), $\psi_1 - 1$ determines the loading coefficient of the cointegrating relation in the system. Moreover, θ determines the instantaneous residual correlation in the EC form. A value of θ close to zero indicates intermediate correlation and a value of θ close to one results in large instantaneous correlation. Furthermore, using the transformation (5.3),

$$\mu_0 = \begin{bmatrix} w_0 \\ w_0 \end{bmatrix} \quad \text{and} \quad \mu_1 = \begin{bmatrix} \delta \\ \delta \end{bmatrix}.$$

The three-dimensional version of Toda's DGP used in the simulations has the form

$$w_t = \begin{bmatrix} 0 \\ 0 \\ \delta \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} w_{t-1} + u_t, \qquad u_t \sim iid\, N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta_1 & \theta_2 \\ \theta_1 & 1 & 0 \\ \theta_2 & 0 & 1 \end{bmatrix} \right). \tag{5.5}$$

Here, δ generates the drift in the third component. The third eigenvalue is unity whereas ψ_1 and ψ_2 vary. If $|\psi_1| < 1$ and $|\psi_2| < 1$, the cointegrating rank is $r = 2$. If $|\psi_1| < 1$ and $\psi_2 = 1$, the cointegrating rank is $r = 1$, and the rank is zero if $\psi_1 = \psi_2 = \psi_3 = 1$. The quantities θ_1 and θ_2 are the correlations between the first component of the process with the second and third components, respectively.

We have also used other DGPs and we will comment on the results where additional insights were obtained. In the next subsection the results from the bivariate DGPs are presented and discussed and thereafter results from three-dimensional DGPs are considered.

5.2.3 Two-dimensional VAR(1)

5.2.3.1 Sizes of Tests

DGPs without Linear Trend ($\delta = 0$) Table 5 shows empirical sizes of the tests when the cointegrating rank is $r = 0$ ($\psi_1 = \psi_2 = 1$) or $r = 1$ ($\psi_2 = 1, \psi_1 < 1$). The tables are structured according to the underlying assumptions of the tests regarding the deterministic terms. The first part of Table 5 shows the tests that should be applied if restrictions on the deterministic terms were known. These tests take into account that the data exhibit

no linear trend (μ_0 arbitrary, $\mu_1 = 0$). The second part of Table 5 displays the empirical sizes of the tests which can be applied if a linear trend is present. These tests allow for a trend in the variables and in the cointegrating relation. The results are presented for sample size $T = 100$. Simulation results for $T = 200$ (not presented here) show slightly more conservative sizes for all tests.

Table 5: Sizes of Tests for 2-dimensional VAR(1) with $r = 0$ or 1 ($\psi_2 = 1$), $\delta = 0$

Test Statistics	Assumptions for test statistics	$\psi_1 = 1$ ($r = 0$) $H_0 : r_0 = 0$	$\psi_1 = 0.9$ ($r = 1$) $H_0 : r_0 = 1$	$\psi_1 = 0.8$ ($r = 1$) $H_0 : r_0 = 1$	$\psi_1 = 0.7$ ($r = 1$) $H_0 : r_0 = 1$
		$\theta = 0$			
LR^*	μ_0 arbitrary,	0.069	0.017	0.039	0.047
LR^{i0}	$\mu_1 = 0$	0.060	0.023	0.042	0.049
LR^{SL}		0.104	0.075	0.096	0.096
SW_{ma}^{VAR}		0.039	0.191	0.207	0.203
LR^+	μ_0, μ_1 arbitrary	0.060	0.008	0.024	0.041
LR^{PC}		0.060	0.011	0.024	0.038
LR^{aug}		0.432	0.030	0.072	0.100
LR_{ta}^{GLS}		0.052	0.018	0.047	0.047
$LM_{ta(2)}^{GLS}$		0.033	0.018	0.042	0.042
SW_{ta}^{VAR}		0.041	0.111	0.184	0.187
BI		0.040	0.009	0.016	0.029
		$\theta = 0.8$			
LR^*	μ_0 arbitrary,	0.069	0.037	0.055	0.057
LR^{i0}	$\mu_1 = 0$	0.060	0.040	0.053	0.051
LR^{SL}		0.104	0.078	0.074	0.079
SW_{ma}^{VAR}		0.039	0.150	0.252	0.238
LR^+	μ_0, μ_1 arbitrary	0.060	0.040	0.065	0.070
LR^{PC}		0.060	0.033	0.056	0.062
LR^{aug}		0.432	0.044	0.080	0.095
LR_{ta}^{GLS}		0.052	0.017	0.031	0.032
$LM_{ta(2)}^{GLS}$		0.033	0.016	0.030	0.029
SW_{ta}^{VAR}		0.041	0.112	0.242	0.324
BI		0.040	0.022	0.036	0.051

Note: Sample size $T = 100$, nominal significance level 0.05

LR^* and LR^{i0} reveal reasonable size properties, but are rather conservative for ψ_1 close to 1 and innovation correlation $\theta = 0$. LR^{SL} exhibits a tendency to overreject (i.e., the empirical size is outside the confidence interval of a 5% nominal significance level), but can still be considered to have reasonable size properties as the empirical size is below 10 % in most cases. SW_{ma}^{VAR} displays more severe size distortions and is, therefore, not recommended for applied work. For $r = 0$, the residual correlation represented by θ has no effect on the tests. On the other hand, for $r = 1$, in general larger residual correlation θ leads to better size properties. For example, LR^* is quite conservative for $\theta = 0$ and ψ_1 close to unity, but displays an empirical size inside the confidence interval for the rejection frequency of a 5% nominal significance level if $\theta = 0.8$. This outcome has been explained by Toda (1995). He refers to Johansen (1991a) who shows that if the true cointegrating rank is $r = 1$, the eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2$ and the null hypothesis $H_0 : r = 0$ is tested, $\hat{\lambda}_1$ converges to the true value λ_1, while $T\hat{\lambda}_2$ converges to a random variable. Therefore, for Toda's DGP,

$$\hat{\lambda}_1 \xrightarrow{p} \lambda_1 = \frac{1}{1 + (1 - \theta^2)(1 + \psi_1)/(1 - \psi_1)}.$$

Hence, the closer $|\theta|$ is to one, the larger is λ_1. Thus, the null hypothesis of no cointegration is rejected more easily for $|\theta|$ close to one.

The second group of tests considered in this context are the tests that allow for arbitrary mean and trend parameters. These tests are not needed if it were known that $\mu_1 = 0$. However, in applied research it may not always be clear whether a linear deterministic trend is present in the data. $LM_{ta(2)}^{GLS}$ has been chosen to represent the class of LM type tests. A comparison of alternative tests within this class of tests can be found in Lütkepohl & Saikkonen (1997b) and Saikkonen & Lütkepohl (1997) (see also Section 3). These studies indicate that $LM_{ta(2)}^{GLS}$ performs overall best in the class of LM type tests.

The size of the LR tests (LR^+ and LR^{PC}) appears to be reasonable for $r_0 = 0$ (see Table 5), whereas it is somewhat conservative for $r_0 = 1$ if the eigenvalue ψ_1 is close to unity and θ is zero. The LR^{aug} test has distorted size properties for $r_0 = 0$, whereas for $r_0 = 1$ it appears to have a reasonable size. In contrast, the SW_{ta}^{VAR} test shows a correct size for $r_0 = 0$, but a heavily distorted size for $r_0 = 1$. Bierens' nonparametric test has correct size for $r_0 = 0$, but the size is quite conservative for $r_0 = 1$. If ψ_1 is close to unity and θ is zero, then its size performance is similar to the LR tests LR^+ and LR^{PC}.

Overall, LR_{ta}^{GLS} and $LM_{ta(2)}^{GLS}$ are most robust with respect to variations of the parameters

ψ_1 and θ as well as to the sample size, but tend to be somewhat conservative for ψ_1 near unity. The results for $T = 200$ are not presented here to save space.

We have not included the Bewley-Yang tests in the table because appropriate critical values are not available for all cases.

DGP with Linear Trend ($\delta \neq 0$) To illustrate the size performance of the tests that are constructed for a DGP without a linear trend in case of incorrect application, simulations for $\delta = 1.0$ are presented in Table 6. LR^* and LR^{SL} display severe overrejection whereas LR^{i0} and SW_{ma}^{VAR} are extremely conservative. Therefore, these tests should not be applied if the possibility of a linear trend cannot be excluded. If it is not clear whether or not the data analysed have a linear trend, tests from the second group in Table 5, that allow for both deterministic components, have to be applied. Their size properties do not depend on the value of δ.

Table 6: Sizes of Tests for 2-dimensional VAR(1) with r=0 or 1 ($\psi_2 = 1$), $\delta = 1$

Test Statistics	Assumptions for test statistics	$\psi_1 = 1$ ($r = 0$) $H_0 : r_0 = 0$	$\psi_1 = 0.9$ ($r = 1$) $H_0 : r_0 = 1$	$\psi_1 = 0.8$ ($r = 1$) $H_0 : r_0 = 1$	$\psi_1 = 0.7$ ($r = 1$) $H_0 : r_0 = 1$
		$\theta = 0$			
LR^*	μ_0 arbitrary,	1.000	0.395	0.881	0.998
LR^{i0}	$\mu_1 = 0$	0.028	0.003	0.003	0.007
LR^{SL}		1.000	0.751	0.955	0.999
SW_{ma}^{VAR}		0.031	0.000	0.000	0.000
		$\theta = 0.8$			
LR^*	μ_0 arbitrary,	1.000	0.391	0.832	0.986
LR^{i0}	$\mu_1 = 0$	0.028	0.009	0.007	0.006
LR^{SL}		1.000	0.307	0.466	0.760
SW_{ma}^{VAR}		0.031	0.000	0.000	0.000

Note: Sample size $T = 100$, nominal significance level 0.05

5.2.3.2 Power

DGP without Linear Trend The power of the tests for data without linear trend is depicted in Figure 3, Panels A to C. The graphs show that LR^{SL} clearly outperforms the other tests in terms of power. Although this might have been expected because its size is comparatively large, it should be noted that the difference in power is substantial, especially in case of a cointegration rank of $r = 2$ (Panel C). Saikkonen & Luukkonen (1997) also find that LR^{SL} performs better than other LR tests even if size-adjusted power is considered. LR^{i0} outperforms LR^{*} in terms of power except in Panel B where both tests perform equally well. The SW_{ma}^{VAR} displays very low power in case of $\theta = 0.8$ and $r_0 = 0$ (Panel B) in comparison with the other tests, whereas in case of $r = 2$ the SW test has very large power (Panel C) due to its large size .

The power of the tests for arbitrary mean and trend term are depicted in Figure 3, Panel D to F. The performance of the tests is rather similar except for the Bierens test, which is outperformed by all other tests in terms of power.[5] If the null hypothesis $r_0 = 0$ is tested, LR^{+} and LR^{PC} perform better than the other tests (Panels D and E), whereas for $r_0 = 1$, LR_{ta}^{GLS} and $LM_{(2)}^{GLS}$ have larger power (Panel F). However, the LR and LM type tests perform generally rather similarly in terms of power.

A comparison of the best performing tests of both groups of tests considered above is presented in Figure 4. LR^{SL} clearly outperforms the other tests in terms of power. Therefore, this test statistic should be chosen if it is known that there is no deterministic trend in the data although the test has a tendency to overreject somewhat, a fact, that is useful to remember in interpreting the test results. The comparison of the performance of LR^{i0} and LR^{SL} with the tests that allow for unrestricted mean and trend terms, LR^{+} and LR_{ta}^{GLS}, shows that in case of a DGP without a linear trend it is worth applying tests that incorporate an additional restriction on the trend term since this leads to a gain in power of the test.

5.2.4 Three-dimensional VAR(1)

The three-dimensional DGP proposed by Toda (1995) (see Section 5.2.2) was used for further simulations. In this section tests that may be applied in case of trending data are considered.

[5] The SW_{ta}^{VAR} test is not presented because of its severely distorted size.

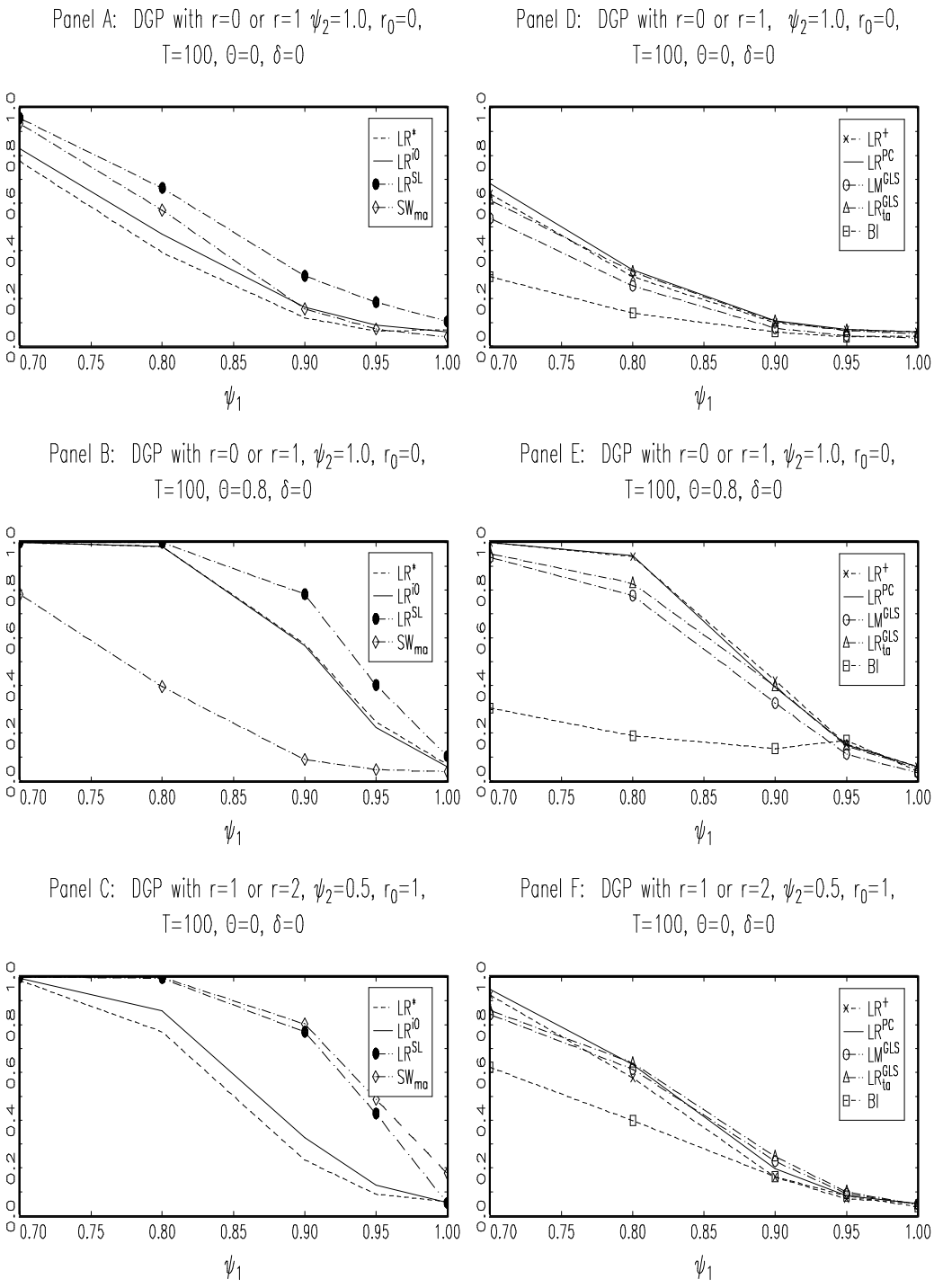

Figure 3: Relative Rejection Frequencies for 2-dimensional VAR(1), Assumptions for Test Statistics: μ_0 arbitrary, $\mu_1 = 0$ (Panel A-C), μ_0, μ_1 arbitrary (Panel D-F)

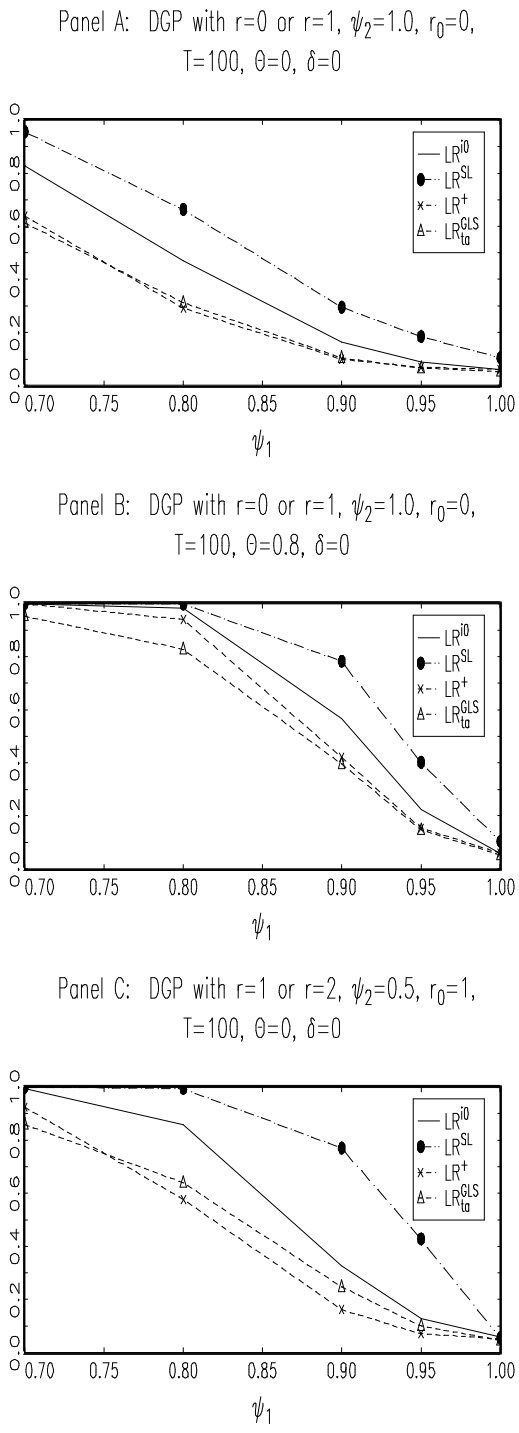

Figure 4: Relative Rejection Frequencies for 2-dimensional VAR(1), Comparison of Best Performing Tests of the Respective Groups of Tests Displayed in Figure 3

As has been discussed in Section 3.4, in case of tests constructed for an arbitrary mean and restricted trend term in the DGP, the cointegrating rank of a system can only be $r < n$ due to the restriction $\beta'\mu_1 = 0$, and, therefore, the rank tested under the null hypothesis should be smaller than $n-1$ ($r_0 < n-1$). Therefore, testing the null hypotheses $r_0 = 0$ and $r_0 = 1$ is considered for the three-dimensional DGP in the following. In addition to the tests that impose a restriction on the trend term (μ_0 arbitrary, $\mu_1 \neq 0, \beta'\mu_1 = 0$), tests that allow for a trend in the cointegrating relation (μ_0, μ_1 arbitrary) are also presented in order to compare the size and power performance of both groups of tests.

The simulation results presented are those for sample size T=100. Further simulations have been carried out for other sample sizes as well as different values of the innovation correlation and values of ψ_2. These simulations are not presented in detail to save space, but are commented on in the respective context.

5.2.4.1 Sizes of Tests

DGP with Linear Trend Term ($\delta = 1$) The simulation results for the empirical sizes of the tests are presented in Table 7. The first group of tests is correctly applied if $\delta = 1$ in the DGP. The LR and LM type tests of this group show reasonable sizes, although for $\theta_1 = 0.4$ and $\theta_2 = 0.8$ the tests have a tendency to overreject. Further simulations (not presented here) have shown that this tendency is less pronounced if only one of the innovation correlations, i.e. the innovation correlation between the two nonstationary components, θ_1, or between the stationary and the nonstationary component, θ_2, is high and the other one is zero. If both correlations are large this leads to size distortions of the tests. In contrast, BY^i displays severe overrejection even if $\theta_1 = \theta_2 = 0$. The size distortions are clearly too strong to recommend these tests for use in practice.

The results for the second group of tests that do not restrict the trend parameter of the DGP confirm the tendency of the foregoing results. The results differ in that the tests are more conservative for $\theta_1 = \theta_2 = 0$. On the other hand, their tendency to overreject is less pronounced than that of the first group of tests if $\theta_1 = 0.4$ and $\theta_2 = 0.8$. The Stock & Watson test (SW_{ta}^{VAR}) again shows unacceptable overrejection. The Bierens test (BI) is very conservative in case of a sample size of $T = 100$. Even for $T = 400$ the size of the BI test turned out to be conservative for eigenvalues ψ_1 near unity (not presented here).

In general, the dependence of the test performance on the innovation correlation is less pronounced in large samples ($T = 400$). In turn, this dependence is even stronger for samples of size $T = 50$, especially LR and LM type tests become very conservative for $\theta_1 = \theta_2 = 0$ even for a value of $\psi_1 = 0.7$.

Table 7: Sizes of Tests for 3-dimensional VAR(1), $r = 0$ or 1 ($\psi_2 = \psi_3 = 1$), $\delta = 1$

Test Statistics	Assumptions for test statistics	$\psi_1 = 1$ ($r = 0$) $H_0 : r_0 = 0$	$\psi_1 = 0.9$ ($r = 1$) $H_0 : r_0 = 1$	$\psi_1 = 0.8$ ($r = 1$) $H_0 : r_0 = 1$	$\psi_1 = 0.7$ ($r = 1$) $H_0 : r_0 = 1$
		$\theta_1 = \theta_2 = 0$			
LR^i	μ_0 arbitrary,	0.062	0.011	0.029	0.046
LR^i_{ta}	$\mu_1 \neq 0$,	0.056	0.029	0.049	0.054
$LM^i_{ta(2)}$	$\beta'\mu_1 = 0$	0.037	0.036	0.045	0.043
BY^i		0.047	0.030	0.152	0.354
LR^+	μ_0, μ_1	0.065	0.009	0.018	0.034
LR^{PC}	arbitrary	0.054	0.010	0.019	0.033
LR^{GLS}_{ta}		0.058	0.007	0.017	0.035
$LM^{GLS}_{ta(2)}$		0.028	0.005	0.010	0.022
SW^{VAR}_{ta}		0.023	0.072	0.172	0.210
BI		0.032	0.004	0.007	0.015
		$\theta_1 = 0.4, \theta_2 = 0.8$			
LR^i	μ_0 arbitrary,	0.063	0.092	0.088	0.084
LR^i_{ta}	$\mu_1 \neq 0$,	0.056	0.131	0.091	0.081
$LM^i_{ta(2)}$	$\beta'\mu_1 = 0$	0.036	0.106	0.079	0.065
BY^i		0.023	0.028	0.142	0.365
LR^+	μ_0, μ_1	0.065	0.075	0.080	0.074
LR^{PC}	arbitrary	0.054	0.064	0.074	0.063
LR^{GLS}_{ta}		0.058	0.031	0.037	0.048
$LM^{GLS}_{ta(2)}$		0.028	0.022	0.024	0.032
SW^{VAR}_{ta}		0.023	0.042	0.112	0.216
BI		0.037	0.009	0.011	0.020

Note: Sample size $T = 100$, nominal significance level 0.05

DGP without Linear Trend Table 8 shows simulation results for $\delta = 0$. In this case, substantial overrejections occur for LR^i and LR^i_{ta}. Because their sizes are too large, it is not

surprising that the power of both tests is higher than in the case of $\delta = 1$. Power results are not presented for this case because of the size distortions of the tests. LR^{+} and LR_{ta}^{GLS} are shown for comparison purposes. They do not depend on the values of μ_0 and μ_1 and, hence, have exactly the same size and power properties as for $\delta = 1$. These results lead to the conclusion that tests based on the intercept version of the VAR or EC model should be used with caution if the trending properties are not clear.

Table 8: Sizes of Tests for 3-dimensional VAR(1), $r = 0$ or 1 ($\psi_2 = \psi_3 = 1$), $\delta = 0$

Test Statistics	Assumptions for test statistics	$\psi_1 = 1$ ($r = 0$) $H_0 : r_0 = 0$	$\psi_1 = 0.9$ ($r = 1$) $H_0 : r_0 = 1$	$\psi_1 = 0.8$ ($r = 1$) $H_0 : r_0 = 1$	$\psi_1 = 0.7$ ($r = 1$) $H_0 : r_0 = 1$
		$\theta_1 = \theta_2 = 0$			
LR^{i}	μ_0 arbitrary,	0.120	0.044	0.080	0.107
LR_{ta}^{i}	$\mu_1 \neq 0$,	0.096	0.173	0.399	0.655
	$\beta'\mu_1 = 0$				
LR^{+}	μ_0, μ_1	0.065	0.009	0.018	0.034
LR_{ta}^{GLS}	arbitrary	0.058	0.007	0.017	0.035
		$\theta_1 = 0.4, \quad \theta_2 = 0.8$			
LR^{i}	μ_0 arbitrary,	0.120	0.150	0.150	0.147
LR_{ta}^{i}	$\mu_1 \neq 0$,	0.096	0.776	0.963	0.981
	$\beta'\mu_1 = 0$				
LR^{+}	μ_0, μ_1	0.065	0.075	0.080	0.074
LR_{ta}^{GLS}	arbitrary	0.058	0.031	0.037	0.048

Note: Sample size $T = 100$, nominal significance level 0.05

5.2.4.2 Power

DGP with Linear Trend Figure 5, Panels A and B, show the power of the first group of tests for $r_0 = 0$ if the true cointegrating rank is $r = 1$. The LR and LM type tests perform rather similarly. However, LR^{i} and LR_{ta}^{i} tend to outperform LM_{ta}^{i}. SW_{ma}^{VAR} is not presented because of heavily distorted size. In the second group of tests (see Panels C and D), LR^{+}, LR^{PC} and LR_{ta}^{GLS} perform best, but exhibit lower power than the tests with restriction on the trend parameter for most values of ψ_1. The SW_{ta}^{VAR} tests appear to have lower power

despite their large empirical sizes. The Bierens test is clearly outperformed by the LR and LM type tests in terms of power. In fact, its performance is so poor that its use cannot be recommended. $LM_{ta(2)}^{GLS}$ is outperformed in terms of power by the original LR tests and the LR tests with prior trend adjustment. Although this result is not surprising given the lower size of the test, it makes this test less useful in applied research. LR^{+} and LR^{PC} perform rather similarly. Each of these tests outperforms the other one in some situations. Therefore, only the original test suggested by Johansen (1994) is considered in the graphs.

Comparing all tests in Figure 5, it is seen that they display a similar reduction in power if ψ_1 approaches unity. However, tests that are based on the restriction of the trend parameter, $\beta'\mu_1 = 0$ (Panels A and B), have in most cases substantially higher power. Looking at the power curves in Figure 5 clearly reveals that they are steeper than those of the corresponding tests which allow for a general linear trend. Note that both types of tests perform similarly under the null hypothesis in the present case.

Further results of the relative power performance are presented in Figure 6. Here, the power is shown for the case where the cointegrating rank of the DGP is $r = 2$ whereas $r_0 = 0$ (Figure 6, Panels A and B) and $r_0 = 1$ (Figure 6, Panels C and D) are tested. In Figure 6, the results of the tests with major size distortions (BY^i, SW_{ma}^{VAR} and SW_{ta}^{VAR}) are not presented. However, their power has similar characteristics as in the case $r = 1$ (Figure 5).

A similar picture as in Figure 5 emerges in Figure 6 for the DGP with higher cointegrating rank. The power of the tests based on the restriction of the trend parameter ($\beta'\mu_1 = 0$) is higher than that of the tests that impose no restrictions on the deterministic terms. For $r_0 = 0$, the power is larger for $r = 2$ than for $r = 1$, as expected. However, if $r_0 = 1$ is tested, the tests have considerably lower power for all values of θ_1 and θ_2. To generalize, if $r > 1$, testing higher cointegrating ranks leads to a loss of power. LR^i is in most cases outperformed by LR_{ta}^i and LM^i, whereas LR_{ta}^i performs generally better than LM^i. Of the second group of tests allowing for a trend, either LR_{ta}^{GLS} or LR^{+} perform better depending on the respective situation.

5.2.5 Three-dimensional VAR(2)

Simulations for the three-dimensional VAR(2) proposed by Yap & Reinsel (1995) have also been carried out to learn about the performance of the tests for a DGP with higher lag order.

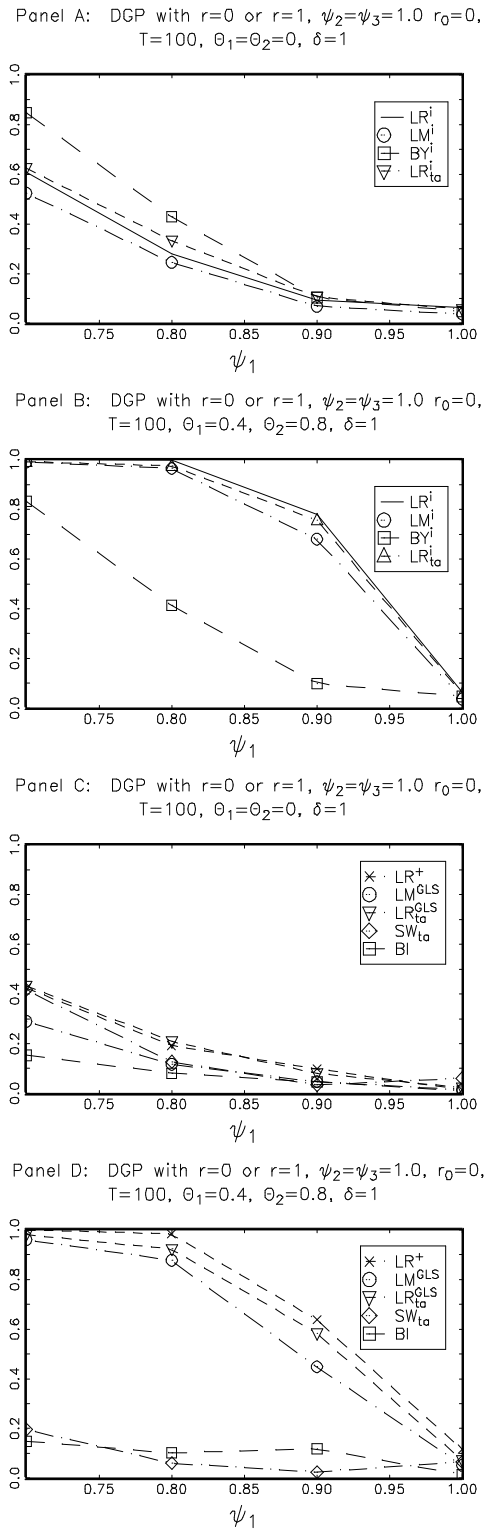

Figure 5: Relative Rejection Frequencies for 3-dimensional VAR(1), Assumptions for Test Statistics: μ_0 arbitrary, $\mu_1 \neq 0$, $\beta'\mu_1 = 0$ (Panel A/B), μ_0, μ_1 arbitrary (Panel C/D)

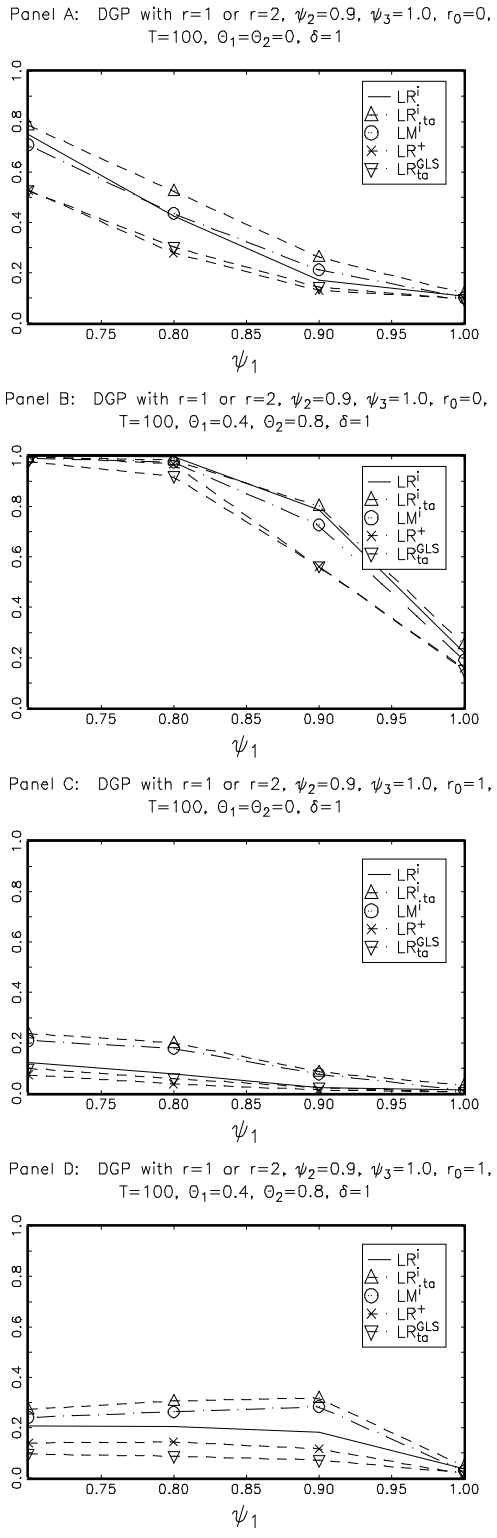

Figure 6: Relative Rejection Frequencies for 3-dimensional VAR(1), Further Results of Best Performing Tests of the Respective Groups of Tests Compared in Figure 5

This DGP has been modified by adding a linear trend in simulations for tests which require a linear trend. The results are not presented in detail to save space. It is perhaps worth noting, however, that the results mainly confirm the findings from the two- and three-dimensional VAR(1) processes. Tests allowing for arbitrary μ_0 and μ_1 (LR^{+}, LR^{PC}, LM_{ta}^{GLS}) have lower power than the tests for $\mu_1 = 0$ in the DGP (LR^{i0}, LR^{SL}) or with $\mu_1 \neq 0$, $\beta'\mu_1 = 0$ (LR^{i}, LR_{ta}^{i}).

5.2.6 Conclusions from Simulations

5.2.6.1 Comparing Different Classes of Tests One central aim of this simulation study has been to compare a wide range of tests on the basis of the same DGPs. From the results it becomes clear that the tests suggested by Stock & Watson (1988) as well as the canonical correlations tests proposed by Yang & Bewley (1996) have empirical sizes considerably in excess of the nominal size in many cases. For the SW tests, this confirms findings by other authors (e.g., Reimers (1992) and Yap & Reinsel (1995)). Therefore, these tests cannot be recommended for applied research, especially as they are also outperformed by LR and LM type tests in terms of power. The Bierens test has correct size or is conservative, but it has very low power, even in quite large samples. Also, it becomes extremely conservative in small samples ($T = 50$ or 100). LR^{aug} has dramatic variation in size and sometimes overrejects in an unacceptable way. These tests are therefore not recommended for applied work. Comparing LR, LM and Wald type tests it has been found that Wald type tests are inferior. Therefore, the relative performance of LR tests, trend adjusted LR tests and LM type tests is of foremost interest and has been considered in some detail.

5.2.6.2 Deterministic Terms One main focus has been on the assumptions regarding the deterministic terms underlying the different tests. As earlier studies have shown, it is very important to specify the deterministic terms correctly (see, e.g., Toda (1995) and Gonzalo & Lee (1998)). Systems cointegration tests can be classified with respect to the deterministic terms assumed for the underlying DGP. One group of tests can be applied if the data do not have a linear trend. Another group of tests is constructed for trending data. Finally, a third group of tests can be applied in either situation and allows for a trend in the cointegration relation.

According to the simulation results it is worth taking additional known restrictions on

deterministic terms into account. The tests that allow for arbitrary deterministic terms have generally lower power than the tests that restrict μ_1. The loss in power can be quite severe. However, if there is uncertainty with respect to the trending properties, tests without restrictions on the deterministic terms have to be applied. In this group of tests, LR tests (original and trend adjusted) outperform all the other tests considered, except that in some situations near the null hypothesis the LM type tests and the trend adjusted LR tests perform better than the original LR tests. Comparing LR^{+} and LR^{PC}, both show superior performance in some situations. Either of these tests may be applied. However, LR^{+} will probably be applied in practice as it has been implemented in frequently used computer packages such as PcGive 9.0 or Eviews.

If data with deterministic trend are analysed and no trend enters the cointegration relations, the application of the LM type test LM^{i}_{ta} as well as the trend adjusted LR test LR^{i}_{ta} proposed by Saikkonen & Lütkepohl (1998a) is recommended as it has higher power in many situations. If it is known that the data do not have a linear trend, the LR^{SL} test suggested by Saikkonen & Luukkonen (1997) is the obvious choice because it outperforms all other tests constructed for such data in terms of power although it has a slight tendency to reject too often under the null hypothesis. Some further topics have been analysed in the simulation study and the results are summarized in the following.

5.2.6.3 Eigenvalues Close to Unity If the DGP has near integrated components, the sizes of all tests become generally more conservative, which leads to lower power of the tests. Exceptions are the LR^{SL}, LR^{i}_{ta}, LM^{i} and the LR^{GLS}_{ta} tests, which even perform quite well if an eigenvalue is near unity. In this respect, the original LR tests are outperformed by these recently suggested tests.

5.2.6.4 Testing Different Cointegrating Ranks In testing higher cointegrating ranks, LR^{GLS}_{ta} and $LM^{GLS}_{ta(2)}$ have a power advantage over LR^{+}, whereas LR^{i}_{ta} and LM^{i} display higher power than LR^{i}, and the power advantage of LR^{SL} over LR^{i0} becomes even more pronounced. According to simulation studies based on sequential testing procedures it is difficult for standard LR tests to find the correct rank if the cointegrating rank is greater than one (see Toda (1994, 1995), Lütkepohl & Saikkonen (1997a)). Therefore, it may be worth applying one of the aforementioned tests for testing large cointegrating ranks as alter-

natives to carrying out standard LR tests. However, in a sequential testing procedure, LM type tests or LR tests with prior trend adjustment may be inferior to the LR tests if the true cointegrating rank is large, as they have lower power if $r_0 = 0$ is tested.

5.2.6.5 Dimensionality and Sample Size A higher dimension of a system leads to a more conservative size of the cointegrating rank tests. Therefore, there is also a tendency to lower power. Simulations for different sample sizes have shown that increasing the sample size from $T = 50$ to $T = 100$ in non-sequential tests does not improve the test performance markedly, whereas most tests (except the BI test) are more reliable for $T = 200$. A direct comparison of the power of, for example, the LR^{+} test for the two- and for the three-dimensional Toda-DGPs in Figure 4, Panel A, and Figure 5, Panel C, shows a power of 0.6 in case of the two-dimensional DGP and of about 0.4 in case of the three-dimensional DGP for an eigenvalue of ψ_1=0.7 if $r_0 = 0$ is tested. However, other studies have found that in sequential testing procedures it is still difficult, even for LR tests, to find a cointegration rank greater than 1 using samples of size $T = 200$. For the Bierens test, even a sample size of $T = 400$ is not large enough to get a reasonable power performance.

5.2.6.6 Recommendations for Applied Research The findings of this simulation study show that the application of LR^{SL}, LR^{i}_{ta}, LR^{+}, LR^{PC} and LR^{GLS}_{ta} in the respective situation can be recommended, as these tests have some strengths in certain situations. Regarding the deterministic terms, the conclusion is that it is important to specify these terms correctly as the tests may reject far too often or have substantially reduced power in case of misspecification.

6 Conclusions

The present study compares a wide range of systems cointegration tests with respect to their asymptotic properties as well as their small sample performance. One objective has been to present the tests in a general unifying framework in order to compare the underlying assumptions of the different test procedures. In applied work, it is very important to know the assumptions under which a test procedure is supposed to work in order to choose the appropriate test. An important aspect in this context is the assumption about the trend parameters of the process and whether the test statistics depend on these parameters. Therefore, one focus of this study is to analyse the sensitivity of the test performance to the trending properties of the data generating process.

This review comprises the discussion of asymptotic power results of the tests considered as well as the analysis of size and power in small samples. The discussion of the asymptotic power results shows that all tests considered in this paper are consistent against fixed alternatives. Different studies on the asymptotic properties of a number of tests are reviewed, including studies which consider a data generating process where a linear trend is present. The results indicate a trade-off between the generality of the assumptions for the validity of a test on the one hand and its local power on the other hand. Previous simulation studies suggest that in small samples some multivariate tests for the cointegrating rank often show size distortions and/or poor power in certain situations. These results are investigated further in new simulations.

In contrast to most previous studies, the scope of this review is very broad as it comprises, besides some well-known test procedures, a number of systems tests for cointegration for which limited knowledge about their properties has been available so far. Modifications of existing tests allowing, for example, for outliers or nonnormal errors are also discussed in this study.

The aim of the simulation study is to explore the strengths and weaknesses of the different systems cointegration tests by carrying out a comparison on the basis of some simple DGPs. The idea is that tests which don't work well for a simple DGP cannot be trusted in more complicated situations in general even if they have been shown to perform well in specific complicated situations. Using this criterion, it is found that the lag augmentation tests, the canonical correlation tests, the Stock & Watson tests as well as the Bierens tests perform

very poorly even under ideal conditions. Therefore their use cannot be recommended for applied work. Generally likelihood based methods are found to perform better but also have deficiencies in some situations.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. Petrov & F. Csáki (eds), *2nd International Symposium on Information Theory*, Académiai Kiadó, Budapest, pp. 267–281.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, Vol. AC-19, pp. 716–723.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**: 817–858.

Andrews, D. & Monahan, J. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica* **60**: 953–966.

Aznar, A. & Salvador, M. (1997). Testing for the rank of the co-integration space and the form of the intercept using the BIC criterion, Paper presented at the ESEM 97 in Toulouse.

Banerjee, A., Dolado, J., Galbraith, J. & Hendry, D. (1993). *Co-integration, Error-Correction, and the Econometric Analysis of Non-stationary Data*, Oxford University Press, Oxford.

Bewley, R. & Yang, M. (1995). Tests for cointegration based on canonical correlation analysis, *Journal of the American Statistical Association* **90**: 990–996.

Bierens, H. (1997a). Cointegration analysis, *in* C. Heij, H. Schumacher, B. Hanzon & C. Praagman (eds), *System Dynamics in Economic and Financial Models*, John Wiley, New Yorck.

Bierens, H. (1997b). Nonparametric cointegration analysis, *Journal of Econometrics* **77**: 379–404.

Boswijk, P. & Franses, P. (1992). Dynamic specification and cointegration, *Oxford Bulletin of Economics and Statistics* **54**(3): 369–381.

Boswijk, P. & Lucas, A. (1997). Semi-nonparametric cointegration testing, *Research Memorandum 1997–41*, Vrije Universiteit Amsterdam.

Caner, M. (1998). Tests for cointegration with infinite variance errors, *Journal of Econometrics* **86**: 155–175.

Cheung, Y.-W. & Lai, K. (1993). Finite-sample sizes of Johansen's likelihood ratio tests for cointegration, *Oxford Bulletin of Economics and Statistics* **55**: 313–328.

Choi, I. (1993). Asymptotic normality of the least-squares estimates for higher order autoregressive integrated processes with some applications, *Econometric Theory* **9**: 263–282.

Choi, I. (1994). Residual-based tests for the null of stationarity with applications to U.S. macroeconomic time series, *Econometric Theory* **10**: 720–749.

Choi, I. & Ahn, B. (1995). Testing for cointegration in a system of equations, *Econometric Theory* **11**: 952–983.

Choi, I. & Ahn, B. (1999). Testing the null of stationarity for multiple time series, *Journal of Econometrics* **88**: 41–77.

Elliott, G., Rothenberg, T. & Stock, J. (1996). Efficient tests for an autoregressive unit root, *Econometrica* **64**(4): 813–836.

Engle, R. & Granger, C. (1987). Cointegration and error correction: Representation, estimation and testing, *Econometrica* **55**: 251–276.

Engle, R. & Yoo, B. (1987). Forecasting and testing in cointegrated systems, *Journal of Econometrics* **35**: 143–159.

Gonzalo, J. (1994). Five alternative methods of estimating long-run equilibrium relationships, *Journal of Econometrics* **60**: 203–233.

Gonzalo, J. & Lee, T.-H. (1998). Pitfalls in testing for long run relationships, *Journal of Econometrics* **86**: 129–154.

Gonzalo, J. & Pitarakis, J.-Y. (1997a). Dimensionality effect in cointegration analysis, *mimeo*, Universidad Carlos III de Madrid.

Gonzalo, J. & Pitarakis, J.-Y. (1997b). Specification via model selection in vector error correction models, *mimeo*, Universidad Carlos III de Madrid.

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

Harbo, I., Johansen, S., Nielsen, B. & Rahbek, A. (1998). Test for cointegration rank in partial systems, *Journal of Business & Economic Statistics, forthcoming.*

Harris, D. (1997). Principal components analysis of cointegrated time series, *Econometric Theory* **13**: 529–557.

Harris, R. & Judge, G. (1998). Small sample testing for cointegration using the bootstrap approach, *Economics Letters* **58**: 31–37.

Haug, A. (1996). Tests for cointegration: A Monte Carlo comparison, *Journal of Econometrics* **71**: 89–115.

Ho, M. & Sørensen, B. (1996). Finding cointegration rank in high dimensional systems using the Johansen test: An illustration using data based Monte Carlo simulations, *Review of Economics and Statistics* **78**(4): 726–732.

Horvath, M. & Watson, M. (1995). Testing for cointegration when some of the cointegrating vectors are prespecified, *Econometric Theory* **11**: 984–1014.

Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* **12**: 231–54.

Johansen, S. (1991a). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* **59**(6): 1551–81.

Johansen, S. (1991b). The power function of the likelihood ratio test for cointegration, *in* J. Gruber (ed.), *Econometric Decision Models: New Methods of Modelling and Applications*, Springer-Verlag, pp. 323–335.

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 University Press, Oxford.

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

Johansen, S. & Nielsen, B. (1993). Manual for the simulation program DisCo. Institute of Mathematical Statistics, University of Copenhagen.

Johansen, S. & Schaumburg, E. (1997). Likelihood analysis of seasonal cointegration, *EUI Working Papers 97/16*, European University Institute.

Kauppi, H. (1997). Testing for cointegration in conditional vector autoregressive models, mimeo, University of Helsinki.

Kleibergen, F. & Paap, R. (1997). Priors, posterior odds and Lagrange multiplier statistics in Bayesian analyses of cointegration, mimeo, Erasmus University Rotterdam.

Kleibergen, F. & van Dijk, H. (1994). Direct cointegration testing in error correction models, *Journal of Econometrics* **63**: 61–103.

Kwiatkowski, D., Phillips, P., Schmidt, P. & Shin, Y. (1992). Testing the null of stationarity against the alternative of a unit root: How sure are we that the economic time series have a unit root?, *Journal of Econometrics* **54**: 159–178.

Lucas, A. (1997). Cointegration testing using pseudo likelihood ratio tests, *Econometric Theory* **13**: 149–169.

Lucas, A. (1998). Inference on cointegrating ranks using LR and LM tests based on pseudo-likelihoods, *Econometric Reviews* **17**(2): 185–214.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin.

Lütkepohl, H. & Poskitt, D. (1998). Consistent estimation of the number of cointegration relations in a vector autoregressive model, *in* R. Galata & H. Küchenhoff (eds), *Econometrics in Theory and Practice. Festschrift for Hans Schneeweiß*, Physica-Verlag, Heidelberg, pp. 87–100.

Lütkepohl, H. & Saikkonen, P. (1997a). Order selection in testing for the cointegrating rank of a VAR process, *Discussion Paper 93*, SFB 373, Humboldt-Universität zu Berlin.

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

Lyhagen, J. (1998). Maximum likelihood estimation of the multivariate fractional cointegrating model, *Working Paper in Economics and Finance 233*, Stockholm School of Economics.

Ng, S. & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag, *Journal of the American Statistical Association* **90**: 268–281.

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**: 461–72.

Perron, P. & Campbell, J. (1993). A note on Johansen's cointegration procedure when trends are present, *Empirical Economics* **18**: 777–789.

Phillips, P. (1995). Fully modified least squares and vector autoregression, *Econometrica* **63**: 1023–1078.

Phillips, P. (1996). Econometric model determination, *Econometrica* **64**: 763–812.

Phillips, P. & Ouliaris, S. (1990). Asymptotic properties of residual-based tests for cointegration, *Econometrica* **58**: 165–193.

Poskitt, D. (1996). On the determination of cointegrating structure via canonical correlations, mimeo, Australian National University.

Quintos, C. (1998a). Analysis of cointegration vectors using the GMM approach, *Journal of Econometrics* **85**: 155–188.

Quintos, C. (1998b). Fully modified vector autoregressive inference in partially nonstationary models, *Journal of the American Statistical Association* **93**(442): 783–795.

Rahbek, A. (1994). The power of some multivariate cointegration tests, *Preprint 6*, Department of Theoretical Statistics, University of Copenhagen.

Rahbek, A. & Mosconi, R. (1998). The role of stationary regressors in the cointegration test, *Preprint 1*, Department of Theoretical Statistics, University of Copenhagen.

Reimers, H. (1991). *Analyse kointegrierter Variablen mittels vektorautoregressiver Modelle*, Physica-Verlag, Heidelberg.

Reimers, H. (1992). Comparisons of tests for multivariate cointegration, *Statistical Papers* **33**: 335–359.

Reinsel, G. & Ahn, S. (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. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation, *Econometric Theory* **8**: 1–27.

Saikkonen, P. & Lütkepohl, H. (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. & Lütkepohl, H. (1998a). Testing for the cointegrating rank of a VAR process with an intercept, *Discussion Paper 51*, SFB 373, Humboldt-Universität zu Berlin.

Saikkonen, P. & Lütkepohl, H. (1998b). Testing for the cointegrating rank of a VAR process with structural shifts, *Discussion Paper 82*, SFB 373, Humboldt-Universität zu Berlin.

Saikkonen, P. & Lütkepohl, H. (1999). Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory, forthcoming.*

Saikkonen, P. & Luukkonen, R. (1993). Testing for a moving average unit root in autoregressive integrated moving average models, *Journal of the American Statistical Association* **88**: 596–601.

Saikkonen, P. & Luukkonen, R. (1997). Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics* **81**: 93–129.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**: 461–464.

Seo, B. (1998). Statistical inference on cointegration rank in error correction models with stationary covariates, *Journal of Econometrics* **85**: 339–385.

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration, *Econometric Theory* **10**: 91–115.

Snell, A. (1999). Testing for r versus $r-1$ cointegrating vectors, *Journal of Econometrics* **88**: 151–191.

Stock, J. & Watson, M. (1988). Testing for common trends, *Journal of the American Statistical Association* **83**: 1097–1107.

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

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

Van Giersbergen, N. (1996). Bootstrapping the trace statistic in VAR models: Monte Carlo results and applications, *Oxford Bulletin of Economics and Statistics* **58**(2): 391–408.

Yang, M. & Bewley, R. (1996). On cointegration tests for VAR models with drift, *Economics Letters* **51**: 45–50.

Yap, S. & Reinsel, G. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model, *Journal of the American Statistical Association* **90**: 253–267.