

# Maximum Eigenvalue Versus Trace Tests for the Cointegrating Rank of a VAR Process\*

by

Helmut Lütkepohl

Humboldt University Berlin

Pentti Saikkonen

University of Helsinki

and

Carsten Trenkler

Humboldt University Berlin

Address for correspondence: Helmut Lütkepohl, Institut für Statistik und Ökonometrie, Humboldt-Universität, Spandauer Str. 1, D-10178 Berlin, Germany, email: luetke@wiwi.hu-berlin.de

## Abstract

The properties of a range of maximum eigenvalue and trace tests for the cointegrating rank of a vector autoregressive process are compared. The tests are all likelihood ratio type tests and operate under different assumptions regarding the deterministic part of the data generation process. The asymptotic distributions under local alternatives are given and the local power is derived. It is found that the local power of corresponding maximum eigenvalue and trace tests is very similar. A Monte Carlo comparison shows, however, that there may be slight differences in small samples. The trace tests tend to have more distorted sizes whereas their power is in some situations superior to that of the maximum eigenvalue tests.

---

\*We thank Ralf Brüggemann for helpful comments and we are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213), for financial support. The second author also thanks the Yrjö Jahsson Foundation and the Alexander von Humboldt Foundation for financial support under a Humboldt research award. Part of this research was done while he was visiting the Humboldt University in Berlin.

# 1 Introduction

In empirical studies of systems of economic time series, the number of cointegrating relations is often of major interest because it affects the model setup and inference procedures at other stages of the analysis. Therefore, the cointegrating rank of a system is usually investigated at an early stage. If a vector autoregressive (VAR) model is an adequate description of the data generation process (DGP), likelihood ratio (LR) type tests as proposed by Johansen (1988, 1995) are the most commonly used inference tools in this context. Two variants of these tests are available, the so-called maximum eigenvalue tests and the trace tests. Both types of tests are frequently applied in empirical studies. Given the long-time coexistence of the two tests it is surprising that only little is known about the relative performance of the two types of LR tests.

Toda (1994) reports on a Monte Carlo experiment comparing the small sample properties of the two types of tests and finds that neither of the tests is uniformly superior but the trace tests perform better in some situations where the power is low. His simulation study is limited in a number of respects, however. First, he considers bivariate DGPs only. For these processes, maximum eigenvalue and trace tests differ only when the null hypothesis states that the cointegrating rank is zero which is clearly a special situation. Second, Toda considers only tests which allow for a deterministic linear trend in the DGP. The performance of tests which do not allow for a linear trend but just a mean term is known to be quite different.

In the present study we will perform a comparison between maximum eigenvalue and trace tests taking into account some further aspects. First of all, we will compare tests based on different assumptions regarding the deterministic part, that is, we consider tests for DGPs with and without linear trend terms. Also different treatments of the mean and deterministic trend term which have been proposed in the literature will be considered in the present study. More precisely, a number of LR type tests reviewed in Hubrich, Lütkepohl & Saikkonen (2001) will be included in the comparison. Moreover, we do not only focus on small sample simulations but also compare the local power of the different tests. Finally, our Monte Carlo study takes into account higher dimensional processes as well.

The study is structured as follows. In the next section the model framework and the hypotheses of interest are discussed. The tests and their asymptotic properties are presented

in Section 3. A local power comparison is performed in Section 4 and the results of a small sample simulation study are presented in Section 5. Conclusions are drawn in Section 6. Some mathematical derivations are deferred to the Appendix.

The following terminology and notation is used throughout. The differencing operator is denoted by  $\Delta$ , that is, for a time series or stochastic process  $y_t$  we have  $\Delta y_t = y_t - y_{t-1}$ . The symbol  $I(d)$  denotes an integrated process of order  $d$ , that is, the purely stochastic part of the process is stationary or asymptotically stationary after differencing  $d$  times while it is still nonstationary after differencing just  $d - 1$  times. Convergence in distribution or weak convergence is signified by  $\xrightarrow{d}$ . The trace, the rank and the maximal eigenvalue of the matrix  $A$  are denoted by  $\text{tr}(A)$ ,  $\text{rk}(A)$  and  $\lambda_{\max}(A)$ , respectively. If  $A$  is an  $(n \times m)$  matrix of full column rank ( $n > m$ ), we denote an orthogonal complement by  $A_{\perp}$  so that  $A_{\perp}$  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 a zero matrix is an identity matrix of suitable dimension. An  $(n \times n)$  identity matrix is denoted by  $I_n$ . GLS is used to abbreviate generalized least squares and r.h.s. is short for right hand side. VECM abbreviates vector error correction model. A sum is defined to be zero if the lower bound of the summation index exceeds its upper bound.

## 2 Models and Hypotheses of Interest

### 2.1 The Models

An observable  $n$ -dimensional time series  $y_t = (y_{1t}, \dots, y_{nt})'$  ( $t = 1, \dots, T$ ) is considered which is generated by the DGP

$$y_t = \mu_0 + \mu_1 t + x_t, \quad t = 1, 2, \dots, \quad (2.1)$$

where  $\mu_0$  and  $\mu_1$  are  $(n \times 1)$  parameter vectors of the deterministic term. The first one,  $\mu_0$ , will be referred to as mean term and  $\mu_1$  is called trend parameter. The process  $x_t$  represents the stochastic part which is in general unobservable and has zero mean. Thus, the deterministic and stochastic parts are simply added in (2.1). It is assumed that the components of  $x_t$  are at most  $I(1)$ . This implies that  $y_t$  is also at most  $I(1)$ . Of course, there may be cointegration among the variables of  $x_t$  and, hence, of  $y_t$ . Moreover,  $x_t$  is assumed

to be generated by a VAR process. For convenience, we write the process in VECM form,

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \quad t = 1, 2, \dots, \quad (2.2)$$

where  $\Pi$  and the  $\Gamma_j$  are  $(n \times n)$  matrices of unknown parameters and  $\varepsilon_t \sim \text{iid}(0, \Omega)$  is independently, identically distributed white noise with zero mean and nonsingular covariance matrix  $\Omega$ . Moreover, in setting up LR type tests it is assumed that the  $\varepsilon_t$  are Gaussian, that is,  $\varepsilon_t \sim \text{iid}N(0, \Omega)$  and, hence,  $x_t$  and  $y_t$  are also Gaussian processes. The sum on the r.h.s. of (2.2) represents the short-term dynamics of the process. It disappears if  $p = 1$ . The initial values  $x_t$  ( $t = -p + 1, \dots, 0$ ) are assumed to be zero for convenience. The asymptotic analysis remains valid if they are assumed to be any random variables which do not depend on the sample size.

Defining  $Y_{t-1}^{t-p} = (y'_{t-1}, \dots, y'_{t-p})'$ , it is easy to see that the process  $y_t$  also has a VECM representation which can be written in the following alternative ways:

$$\begin{aligned} \Delta y_t &= \nu_0 + \nu_1 t + \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t \\ &= \nu + [\nu_1 : \Pi] \begin{bmatrix} t-1 \\ y_{t-1} \end{bmatrix} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t \\ &= \nu + \Pi^+ y_{t-1}^+ + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t, \quad t = p+1, p+2, \dots, \end{aligned} \quad (2.3)$$

where  $\nu_0 = -\Pi \mu_0 + (I_n + \Pi - \sum_{j=1}^{p-1} \Gamma_j) \mu_1$ ,  $\nu_1 = -\Pi \mu_1$ ,  $\nu = \nu_0 + \nu_1$ ,  $\Pi^+ = [\nu_1 : \Pi]$ ,  $y_{t-1}^+ = [t-1 : y'_{t-1}]'$ ,  $\Gamma = [\Gamma_1 : \dots : \Gamma_{p-1}]$  and  $\Delta Y_{t-1}^{t-p+1} = Y_{t-1}^{t-p+1} - Y_{t-2}^{t-p}$ . Depending on the assumptions for  $\mu_0$  and  $\mu_1$ , different restrictions will be imposed on  $\nu$ ,  $\nu_0$  and  $\nu_1$ . A brief discussion of the different cases will be given in Section 3, where the cointegration tests are presented. The hypotheses of interest will be considered next.

## 2.2 Hypotheses of Interest

It is well-known (see, e.g., Johansen (1995)) that the number of linearly independent cointegrating relations of  $y_t$  is equal to the rank of  $\Pi$  which will be denoted by  $r$  in the following. This number is the quantity of interest in the tests considered in this study. We discuss tests designed for checking the pairs of hypotheses

$$H(r_0) : \text{rk}(\Pi) = r_0 \quad \text{versus} \quad \bar{H}(r_0) : \text{rk}(\Pi) > r_0 \quad (2.4)$$

and

$$H(r_0) : \text{rk}(\Pi) = r_0 \quad \text{versus} \quad H(r_0 + 1) : \text{rk}(\Pi) = r_0 + 1. \quad (2.5)$$

Tests of the former pair of hypotheses are often referred to as trace tests and those for the latter pair are known as maximum eigenvalue tests.

We are also interested in the local power of the tests. If  $H(r_0)$  is true, the matrix  $\Pi$  in the VECM form (2.3) can be written as a product  $\Pi = \alpha\beta'$ , where  $\alpha$  and  $\beta$  are  $(n \times r_0)$  matrices of rank  $r_0$ . Thus, we may write the null hypothesis as  $H(r_0) : \Pi = \alpha\beta'$  with  $\alpha, \beta$   $(n \times r_0)$  and  $\text{rk}(\alpha) = \text{rk}(\beta) = r_0$ . Furthermore, 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

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

For local power analyses the following local alternative hypothesis will be considered:

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

where the factor  $1/T$  in the second term on the r.h.s. ensures that a proper asymptotic distribution of the test statistics is obtained for  $T \rightarrow \infty$ .

### 3 The Tests

In this section we will consider LR type tests for the hypotheses in (2.4)/(2.5) under alternative assumptions for the deterministic mean and trend terms. An overview of the assumptions for  $\mu_0$  and  $\mu_1$  and the associated models and tests is given in Table 1 which is adopted from Hubrich, Lütkepohl & Saikkonen (2001). Under Gaussian assumptions, the LR type statistics related to the models in Table 1 may be obtained by concentrating out the short-run dynamics and performing a reduced-rank regression. The LR statistics of interest can then be obtained by solving a suitable eigenvalue problem. For example, for the case where no deterministic terms appear in the model ( $\mu_0 = \mu_1 = 0$ ) the LR statistics can be obtained as follows. Denote the residuals from regressing  $\Delta y_t$  and  $y_{t-1}$  on  $\Delta Y_{t-1}^{t-p+1}$  by  $R_{0t}$  and  $R_{1t}$ , respectively, and define  $S_{ij} = T^{-1} \sum_{t=1}^T R_{it}R'_{jt}$  ( $i, j = 0, 1$ ). Moreover, denote the ordered (generalized) eigenvalues from solving  $\det(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}) = 0$  by  $\lambda_1 \geq \dots \geq \lambda_n$ . Then the trace statistic for testing the pair of hypotheses in (2.4) can be shown to be

$$LR_{trace}^0(r_0) = -T \sum_{j=r_0+1}^n \log(1 - \lambda_j) \quad (3.1)$$

and the maximum eigenvalue statistic for testing (2.5) is

$$LR_{max}^0(r_0) = -T \log(1 - \lambda_{r_0+1}) \quad (3.2)$$

(see Johansen (1995)).

If  $\mu_1 = 0$  and  $\mu_0$  is unrestricted, a DGP with possibly nonzero mean term is considered whereas a deterministic linear trend term is excluded by assumption. Three variants of LR type tests have been considered in the literature for this situation. The first test is obtained by dropping  $\nu_1 t$  in the first equation of (2.3) and including an intercept term in the VECM. This term has to satisfy restrictions because otherwise the model may generate linear trends. The second test enforces the restriction on the constant term by absorbing the intercept into the cointegration relations. Finally, in the third testing setup the mean term  $\mu_0$  is estimated in a first step and then it is subtracted from  $y_t$ . The tests are performed on the mean adjusted data,  $\tilde{x}_t = y_t - \tilde{\mu}_0$ , noting that  $\Delta \tilde{x}_t = \Delta y_t$ . Suitable estimators  $\tilde{\mu}_0$  are proposed by Saikkonen & Luukkonen (1997) and Saikkonen & Lütkepohl (2000a). The asymptotic distributions of the test statistics do not depend on the specific choice of the two alternative estimators. The situation is different in small samples, however. We use the estimator proposed by Saikkonen & Lütkepohl (2000a) in the simulations reported in Sec. 5 because it resulted in slightly better size properties of the test in preliminary simulations.

If both  $\mu_0$  and  $\mu_1$  are unrestricted a deterministic linear trend is allowed for. Three different LR type tests have been proposed for this situation. Again, these test statistics can be obtained conveniently via the VECMs. The first model is set up in such a way so as to impose the linearity of the trend term. The second model includes the trend term outside the cointegration term. In principle such a model can generate quadratic trends if no restrictions are imposed on the deterministic parameters. Finally, the last test in Table 1 is based on prior trend adjustment via a GLS procedure and estimation of the resulting VECM for the trend adjusted variables. Critical values for all these tests may be found in the references given in Table 1. The notation for the different tests also corresponds to that used in the survey by Hubrich et al. (2001).

It may be worth noting that there is also another group of tests which is applied frequently in practice. They assume that the variables may have deterministic linear trends whereas a trend in the cointegrating relations is excluded. These tests are discussed in detail in Saikkonen & Lütkepohl (2000b). Despite their popularity in practice, it is questionable that

**Table 1.** Models and LR type tests.

Assumptions for deterministic terms	Model	LR type statistic	Reference
$\mu_0 = \mu_1 = 0$	$\Delta y_t = \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	$LR^0$	Johansen (1988, 1995)
$\mu_0$ arbitrary	$\Delta y_t = \nu_0 + \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	$LR^{i0}$	Johansen (1991, 1995)
$\mu_1 = 0$	$\Delta y_t = [\nu_0 : \Pi] \begin{bmatrix} 1 \\ y_{t-1} \end{bmatrix} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	$LR^*$	Johansen & Juselius (1990)
	$\Delta y_t = \Pi(y_{t-1} - \mu_0) + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	$LR^{SL}$	Saikkonen & Luukkonen (1997) Saikkonen & Lütkepohl (2000a)
$\mu_0, \mu_1$ arbitrary	$\Delta y_t = \nu + [\nu_1 : \Pi] \begin{bmatrix} t-1 \\ y_{t-1} \end{bmatrix} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	$LR^+$	Johansen (1992, 1994, 1995)
	$\Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \Gamma \Delta Y_{t-1}^{t-p+1} + u_t$	$LR^{PC}$	Perron & Campbell (1993)
	$\Delta y_t - \mu_1 = \Pi(y_{t-1} - \mu_0 - \mu_1(t-1)) + \sum_{j=1}^{p-1} \Gamma_j(\Delta y_{t-j} - \mu_1) + u_t$	$LR^{GLS}$	Saikkonen & Lütkepohl (2000a) Lütkepohl & Saikkonen (2000)

the necessary assumptions can be justified easily in many situations. Therefore, we do not include them in the present comparison.

The limiting distributions of the LR statistics listed in Table 1, under the local alternatives (2.6), have been derived for DGPs with order  $p = 1$  only. Under  $H_T(r_0)$ , most trace and maximum eigenvalue statistics have limiting distributions  $LR_{trace}(r_0) \xrightarrow{d} \text{tr}(\mathcal{D})$  and  $LR_{max}(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D})$ , respectively, where  $\mathcal{D} = \left( \int_0^1 \mathbf{F} d\mathbf{N}' \right)' \left( \int_0^1 \mathbf{F} \mathbf{F}' ds \right)^{-1} \left( \int_0^1 \mathbf{F} d\mathbf{N}' \right)$ ,  $\mathbf{F}(s)$  is an  $(n - r_0)$ -dimensional stochastic process and  $\mathbf{N}(s)$  is the Ornstein-Uhlenbeck process defined by  $\mathbf{N}(s) = \mathbf{B}(s) + ab' \int_0^s \mathbf{N}(u) du$ , where  $\mathbf{B}(s)$  is a standard Brownian motion,  $a = (\alpha'_\perp \Omega \alpha_\perp)^{-1/2} \alpha'_\perp \alpha_1$  and  $b = (\alpha'_\perp \Omega \alpha_\perp)^{1/2} (\beta'_\perp \alpha_\perp)^{-1} \beta'_\perp \beta_1$ . The specific form of  $\mathbf{F}(s)$  varies with the assumptions regarding the deterministic term. The different  $\mathbf{F}(s)$  processes are given in Table 2. For example,  $LR_{trace}^0(r_0) \xrightarrow{d} \text{tr}(\mathcal{D}^0)$  and  $LR_{max}^0(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D}^0)$  where

$$\mathcal{D}^0 = \left( \int_0^1 \mathbf{N} d\mathbf{N}' \right)' \left( \int_0^1 \mathbf{N} \mathbf{N}' ds \right)^{-1} \left( \int_0^1 \mathbf{N} d\mathbf{N}' \right)$$

under the local alternatives (2.6). Obviously, the local power depends on the number of common trends,  $n - r_0$ , under the null hypothesis and, via  $a$  and  $b$ , on the ‘distance’ from the null hypothesis. The null distribution results by setting  $a = 0$ .

The asymptotic distributions of the other statistics, except  $LR^{GLS}$ , are obtained in an analogous way. Most of the limiting distributions in Table 2 were derived by Saikkonen & Lütkepohl (1999). The other limiting distributions for trace statistics are also given in

**Table 2.** Limiting Distributions of Tests Based on  $\left(\int_0^1 \mathbf{F}d\mathbf{N}'\right)' \left(\int_0^1 \mathbf{F}\mathbf{F}'ds\right)^{-1} \left(\int_0^1 \mathbf{F}d\mathbf{N}'\right)$

Test statistic	$\mathbf{F}(s)$	$\mathbf{F}_t$
$LR^0$	$\mathbf{N}(s)$	$\mathbf{N}_{t-1}$
$LR^{i0}$	$\mathbf{N}(s) - \int_0^1 \mathbf{N}(u)du$	$\mathbf{N}_{t-1} - T^{-1} \sum_{t=1}^T \mathbf{N}_{t-1}$
$LR^*$	$[\mathbf{N}(s) : 1]'$	$[\mathbf{N}'_{t-1} : 1]'$
$LR^{SL}$	$\mathbf{N}(s)$	$\mathbf{N}_{t-1}$
$LR^+$	$[(\mathbf{N}(s) - \int_0^1 \mathbf{N}(u)du)' : s - \frac{1}{2}]'$	$[(\mathbf{N}_{t-1} - T^{-1} \sum_{t=1}^T \mathbf{N}_{t-1})' : s - \frac{1}{2}]'$
$LR^{PC}$	trend adjusted $\mathbf{N}(s)$	trend adjusted $\mathbf{N}_{t-1}$

the references presented in Table 1. The corresponding results for the maximum eigenvalue statistics have not been derived explicitly to the best of our knowledge. Therefore a proof is given in the Appendix.

For the  $LR^{GLS}$  tests, the asymptotic distributions are based on

$$\left(\int_0^1 \mathbf{N}_*d\mathbf{N}'_*\right)' \left(\int_0^1 \mathbf{N}_*\mathbf{N}'_*ds\right)^{-1} \left(\int_0^1 \mathbf{N}_*d\mathbf{N}'_*\right),$$

where  $\mathbf{N}_*(s) = \mathbf{N}(s) - s\mathbf{N}(1)$  and  $\int_0^1 \mathbf{N}_*d\mathbf{N}'_*$  abbreviates  $\int_0^1 \mathbf{N}d\mathbf{N}' - \int_0^1 \mathbf{N}ds\mathbf{N}(1)' - \mathbf{N}(1) \int_0^1 s d\mathbf{N}' + \int_0^1 s ds\mathbf{N}(1)\mathbf{N}(1)'$ . This limiting distribution is given in Lütkepohl & Saikkonen (2000).

## 4 Local Power Comparison

From the limiting distributions in Table 2 it follows that the local power, that is, the relative rejection frequency if  $H(r_0)$  is tested and  $H_T(r_0)$  is true, depends on  $\alpha$ ,  $\beta$ ,  $\Omega$ ,  $\alpha_1$  and  $\beta_1$  only through  $a$  and  $b$ . This implies, for instance, for the case  $r - r_0 = 1$ , where  $\alpha_1$  and  $\beta_1$  are  $(n \times 1)$  vectors, that the limiting distributions only depend on the two parameters

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

A different parameterization is used in Johansen (1995) and Saikkonen & Lütkepohl (1999). The present one is used in Hubrich et al. (2001) 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 distance from the  $I(2)$



parameter space. We will use the quantities  $l$  and  $d$  in comparing the local power of the tests.

For a local power comparison we consider first the case where  $\alpha_1$  and  $\beta_1$  are  $(n \times 1)$  vectors and simulate the discrete time counterpart of the  $(n - r_0)$ -dimensional Ornstein-Uhlenbeck process  $\mathbf{N}(s)$  as  $\Delta \mathbf{N}_t = \frac{1}{T} a b' \mathbf{N}_{t-1} + \epsilon_t$  ( $t = 1, \dots, T = 1000$ ) with  $\epsilon_t \sim \text{iid} N(0, I_{n-r_0})$ ,  $\mathbf{N}_0 = 0$ ,

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

and

$$a' = \begin{cases} \left( -\sqrt{l^2 d^2}, \sqrt{l^2(1-d^2)} \right) & \text{for } n - r_0 = 2 \\ \left( -\sqrt{l^2 d^2}, \sqrt{l^2(1-d^2)}, 0 \right) & \text{for } n - r_0 = 3 \end{cases}.$$

From the  $\mathbf{N}_t$  we compute  $A_T = T^{-2} \sum_{t=1}^T \mathbf{F}_t \mathbf{F}_t'$  and  $B_T = T^{-1} \sum_{t=1}^T \mathbf{F}_t \Delta \mathbf{N}_t'$ , where the  $\mathbf{F}_t$  are also given in Table 2. The limiting distributions of the trace and maximum eigenvalue test statistics are then simulated as  $\text{tr}(B_T' A_T^{-1} B_T)$  and  $\lambda_{max}(B_T' A_T^{-1} B_T)$ , respectively. Furthermore, using

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

and

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

with  $\overline{\Delta \mathbf{N}} = T^{-1} \sum_{t=1}^T \Delta \mathbf{N}_t$  gives the limiting distributions of the  $LR^{GLS}$  statistics in an analogous fashion. The resulting rejection frequencies for the cases  $n - r_0 = 2$  and  $n - r_0 = 3$  for different values of  $d$  and  $l$  are plotted in Figures 1 and 2. Note that the two types of tests are identical for  $n - r_0 = 1$  and, hence, a comparison is not meaningful for this case. Note also that  $LR^0$  and  $LR^{SL}$  have the same asymptotic distributions. Therefore only the local power curves for the latter tests are depicted in the figures.

Figures 1 and 2 reveal that the results for the trace and the maximum eigenvalue versions of the tests are quite similar. Clearly, this result is not surprising because, if there is just one additional cointegration relation, both tests check against the appropriate alternative hypothesis.

Comparing the different proposals for treating the deterministic terms, it can be seen that the tests assuming no linear trend outperform the tests allowing for a trend. However,

the relative performance of the test variants is not influenced by the inclusion of a trend (see Panels A and B in Figures 1 and 2). The same can be said about the parameter  $d$  measuring the distance from the  $I(2)$  region. These results are in line with those of Saikkonen & Lütkepohl (1999) and Hubrich et al. (2001).

The former setup with just one additional cointegrating relation may favor the maximum eigenvalue tests. These tests consider the alternative hypothesis that there is just one extra cointegration relation whereas the trace variants test against a more general alternative. Therefore, we have also determined the local power for the case  $r - r_0 = 2$ , that is, under the local alternative, the process has two extra cointegration relations. In this case we use

$$b' = \begin{cases} I_2 & \text{for } n - r_0 = 2 \\ [I_2 : 0] & \text{for } n - r_0 = 3 \end{cases}$$

and

$$a' = \begin{cases} \begin{pmatrix} -\sqrt{l^2 d^2} & \sqrt{l^2(1-d^2)} \\ c & 0 \end{pmatrix} & \text{for } n - r_0 = 2 \\ \begin{pmatrix} -\sqrt{l^2 d^2} & \sqrt{l^2(1-d^2)} & 0 \\ c & 0 & 0 \end{pmatrix} & \text{for } n - r_0 = 3. \end{cases}$$

The resulting local power of the tests for  $d = 0.25$  and different values of  $c$  and  $l$  is depicted in Figures 3 and 4.

It can be seen that the trace and the maximum eigenvalue tests perform again quite similarly. On the other hand, the differences between the alternative proposals for treating deterministic terms are more pronounced now.  $LR^+$  and  $LR^{SL}$  have the highest local power among the tests in their respective groups, whereas  $LR^{i0}$  is generally outperformed by its competitors. As in the case of one extra cointegrating relation the parameter  $d$  has no important effect on the relative properties of the test versions. Therefore we do not present the results for other values of  $d$  than 0.25.

In summary, based on local power, no clear recommendations about the preferred use of either the trace or the maximum eigenvalue tests can be given. However, local power properties are informative about the performance of the tests in large samples when alternatives close to the null hypothesis are considered. Therefore a small sample comparison of the tests is performed in the next section in order to get further insight into the relative performance of the two different test versions.

## 5 Small Sample Comparison of Tests

We use the following process  $x_t$  to compare the maximum eigenvalue and trace tests in small samples:

$$x_t = \begin{bmatrix} \Psi & 0 \\ 0 & I_{n-r} \end{bmatrix} x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid } N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_r & \Theta \\ \Theta' & I_{n-r} \end{bmatrix} \right), \quad (5.1)$$

where  $\Psi = \text{diag}(\psi_1, \dots, \psi_r)$  is an  $(r \times r)$  diagonal matrix and  $\Theta$  is  $(r \times (n - r))$ . This process was also used in simulations by Toda (1994, 1995) and a number of other authors. Bivariate, three- and four-dimensional processes will be considered. As mentioned in the introduction, Toda (1994) reports results for bivariate processes only. In the bivariate case, if  $r = 0$ ,  $\Psi$  and  $\Theta$  vanish and the process consists of two nonstationary components. If the cointegrating rank is 1,  $\Psi = \psi_1$  with  $|\psi_1| < 1$  and  $\Theta = \theta$  is a scalar which represents the instantaneous correlation between the two components. For three- and four-dimensional processes the matrices  $\Psi$  and  $\Theta$  have analogous interpretations.

In the simulations the parameter values of the deterministic components are all zero, i.e.,  $\mu_i = 0$  ( $i = 0, 1$ ) throughout. In other words, the deterministic part is actually zero. Because this case is not excluded by the tests allowing, for example, for a linear trend, the same generated time series can be used for comparing all the tests. The sample size used in the simulations is  $T = 100$ . In addition, 50 presample values were generated, starting with an initial value of zero. The number of replications is 1000. The rejection frequencies given in Tables 3 - 5 and Figures 5 - 8 are based on asymptotic critical values for a test level of 5%. The rejection frequencies are not size corrected because a size correction is generally not available in practice.

For a given set of parameter values and a given sample size, the results for the test statistics are based on the same generated time series. Hence, the corresponding entries in the tables and the figures are not independent. Given our 1000 replications, the standard error of an estimator of a true rejection probability  $P$  is  $s_P = \sqrt{P(1 - P)/1000}$  so that, for example,  $s_{0.05} = 0.0069$  and, hence, the two-standard-error confidence interval for  $P = 0.05$  is  $[0.036, 0.064]$ . Note also that the results for testing  $H(1) : \text{rk}(\Pi) = 1$  are not conditioned on the outcome of the test of  $H(0) : \text{rk}(\Pi) = 0$  etc.. In other words, the test results do not refer to sequential test procedures. The small sample properties of  $LR^0$  are not presented here because the test's assumptions regarding the deterministic terms are not very realistic

**Table 3.** Relative Rejection Frequencies of Tests for  $H(0) : r = 0$  Based on Bivariate DGP with Cointegrating Rank  $r = 0$ , VAR Order  $p = 1$ ,  $\theta = 0$ , Sample Size  $T = 100$ , Nominal Significance Level 0.05.

Test Statistic		Test Statistic	
$LR_{max}^{i0}$	0.065	$LR_{max}^+$	0.061
$LR_{trace}^{i0}$	0.060	$LR_{trace}^+$	0.060
$LR_{max}^*$	0.056	$LR_{max}^{PC}$	0.060
$LR_{trace}^*$	0.069	$LR_{trace}^{PC}$	0.060
$LR_{max}^{SL}$	0.066	$LR_{max}^{GLS}$	0.054
$LR_{trace}^{SL}$	0.062	$LR_{trace}^{GLS}$	0.052

for applied work and are therefore not of much interest from that perspective.

The sizes of the tests for the bivariate DGP with  $r = 0$  is shown in Table 3. Since the trace and the maximum eigenvalue tests are identical for testing  $H(1) : r = 1$  we just present the results for  $H(0) : r = 0$ . All tests have roughly a correct size. Obviously, the generated sizes for both the trace and the maximum eigenvalue tests are very similar irrespective of the specific test proposal. The same result was also found for three-dimensional DGPs when  $H(0) : r = 0$  or  $H(1) : r = 1$  are tested (see Table 4). Nevertheless, the sizes of the trace variants of the  $LR^+$  and  $LR^{PC}$  tests exceed the sizes of the corresponding maximum eigenvalue versions a bit for  $\psi_1 = 0.9$  and  $\psi_1 = 0.8$  in case of large innovation correlation ( $\Theta = (0.4, 0.8)$ ). For  $\Theta = (0.4, 0.8)$  we also observe that the trace variants of  $LR^+$  and  $LR^{PC}$  reject slightly too often. Both the trace and the maximum eigenvalue versions of  $LR^{i0}$  and  $LR^*$  are affected by this problem as well. In contrast, in the absence of innovation correlation ( $\Theta = (0, 0)$ ) the tests are quite conservative, especially the maximum eigenvalue variants. However, the  $LR^{SL}$  tests have reasonable size properties for both kinds of innovation correlation and therefore outperform the other tests in this respect.

In the absence of innovation correlation ( $\Theta = (0, 0, 0)$ ) we observe similar size properties of the tests for four-dimensional DGPs with  $r = 1$  as for three-dimensional ones. However, in case of high innovation correlation ( $\Theta = (0.4, 0.4, 0.8)$ ) the problem of an excessive size distortion is much more severe as shown in Table 5. All tests, except  $LR^{GLS}$ , reject far too often. Gonzalo & Pitarakis (1999) have pointed out that this kind of size distortion is

**Table 4.** Relative Rejection Frequencies of Tests for Three-dimensional DGPs with Cointegrating Rank  $r = 0$  or 1, VAR Order  $p = 1$ , Sample Size  $T = 100$ , Nominal Significance Level 0.05.

Test Statistic	$r = 0^\dagger$ ( $H(0) : r = 0$ ) $\psi_1 = 1$	$r = 1$ ( $H(1) : r = 1$ )			$r = 1$ ( $H(1) : r = 1$ )		
		$\psi_1 = 0.9$	$\psi_1 = 0.8$	$\psi_1 = 0.7$	$\psi_1 = 0.9$	$\psi_1 = 0.8$	$\psi_1 = 0.7$
		$\Theta = (0, 0)$			$\Theta = (0.4, 0.8)$		
$LR_{max}^{i0}$	0.057	0.010	0.026	0.043	0.073	0.077	0.068
$LR_{trace}^{i0}$	0.075	0.021	0.036	0.045	0.075	0.077	0.072
$LR_{max}^*$	0.066	0.007	0.019	0.033	0.069	0.071	0.067
$LR_{trace}^*$	0.071	0.015	0.027	0.043	0.076	0.072	0.072
$LR_{max}^{SL}$	0.059	0.015	0.035	0.045	0.057	0.055	0.042
$LR_{trace}^{SL}$	0.066	0.017	0.040	0.050	0.069	0.056	0.055
$LR_{max}^+$	0.059	0.005	0.008	0.024	0.057	0.064	0.067
$LR_{trace}^+$	0.065	0.009	0.018	0.034	0.075	0.080	0.074
$LR_{max}^{PC}$	0.058	0.004	0.006	0.024	0.044	0.055	0.056
$LR_{trace}^{PC}$	0.054	0.010	0.019	0.033	0.064	0.074	0.063
$LR_{max}^{GLS}$	0.060	0.009	0.026	0.033	0.015	0.035	0.045
$LR_{trace}^{GLS}$	0.058	0.007	0.017	0.035	0.017	0.040	0.050

<sup>†</sup> The results for  $r = 0$  are independent of the innovation correlation.

a typical problem of LR type tests emerging in large systems. However, the excessive size distortion is less pronounced for the maximum eigenvalue tests than for the trace tests. So the former have a slight advantage in this respect.

The four-dimensional case also allows us to compare the size of the test variants for processes with two cointegration relations ( $r = 2$ ). As the relative performance of the trace and maximum eigenvalue tests is almost identical, we do not show the results here. Nevertheless, we mention that all tests are very conservative in this case. Generally, the sizes of the tests do not exceed the value of 0.02 and are often below 0.01. Only the  $LR^{SL}$  tests are a bit less conservative.

To sum up, the size properties of the trace and the maximum eigenvalue variants are in general rather similar for all test proposals. However, in specific situations the trace tests suffer more from an excessive size distortion than the maximum eigenvalue tests.

**Table 5.** Relative Rejection Frequencies of Tests for  $H(1) : r = 1$  Based on Four-dimensional DGPs with Cointegrating Rank  $r = 1$ ,  $\Theta = (0.4, 0.4, 0.8)$ , VAR Order  $p = 1$ , Sample Size  $T = 100$ , Nominal Significance Level 0.05.

Test Statistic	$r = 1$			Test Statistic	$r = 1$		
	$\psi_1 = 0.9$	$\psi_1 = 0.8$	$\psi_1 = 0.7$		$\psi_1 = 0.9$	$\psi_1 = 0.8$	$\psi_1 = 0.7$
$LR_{max}^{i0}$	0.108	0.087	0.077	$LR_{max}^+$	0.106	0.088	0.074
$LR_{trace}^{i0}$	0.131	0.101	0.090	$LR_{trace}^+$	0.143	0.118	0.104
$LR_{max}^*$	0.104	0.087	0.080	$LR_{max}^{PC}$	0.100	0.084	0.077
$LR_{trace}^*$	0.130	0.100	0.093	$LR_{trace}^{PC}$	0.121	0.100	0.086
$LR_{max}^{SL}$	0.095	0.078	0.073	$LR_{max}^{GLS}$	0.063	0.062	0.062
$LR_{trace}^{SL}$	0.119	0.084	0.070	$LR_{trace}^{GLS}$	0.065	0.059	0.061

The small sample power of the tests is depicted in the Figures 5 - 8. Figure 5 shows that the trace and maximum eigenvalue tests perform quite similarly in the bivariate case with  $r = 0$  or 1 when  $H(0) : r = 0$  is tested. In accordance with results by Toda (1994) for tests which allow for a time trend, the trace tests are slightly better for alternatives close to the null hypothesis and the maximum eigenvalue tests have an advantage in situations further away from the null hypothesis. Interestingly, this pattern can be observed independently of the innovation correlation and the tests' assumption regarding the deterministic terms. Obviously, the differences are quite small and may be due to sampling variability in our simulations. It is worth noting that the tests assuming no linear trend have higher small sample power than those allowing for a trend. Hence, it pays to specify the deterministic terms properly. On the other hand, comparing the different test versions within the respective groups ( $\mu_1 = 0$  or  $\mu_1 \neq 0$ ), a clear winner for all situations is not found.

Compared to the foregoing bivariate setup all tests have a lower small sample power when the corresponding three- and four-dimensional DGPs with  $r = 0$  or 1 are considered. Nevertheless, the relative characteristics of the test variants remain unchanged, so we do not present the respective graphs here.

We have also simulated the small sample power for three- and four-dimensional processes with two cointegration relations when testing  $H(0)$  and  $H(1)$ . The qualitative results are largely the same for both dimensions. Therefore we just refer to the three-dimensional case. In Figure 6, the power for the case of testing  $H(0)$  is depicted. In this case, for  $|\psi_1| < 1$ , there

are two more cointegrating relations than specified in the null hypothesis (i.e.,  $r - r_0 = 2$ ). Hence, for the maximum eigenvalue tests the null and alternative hypotheses are incorrect and, thus, the trace tests may have an advantage. This is clearly reflected in the power curves. In the absence of innovation correlation ( $\Theta = (0, 0)'$ ) the small sample power of the trace tests exceeds the power of the maximum eigenvalue tests by up to about 20 percentage points. In some cases, the power curves of the trace tests are clearly steeper than those of the maximum eigenvalue tests so that the superior performance of the former is only partly attributable to their larger size.

When testing  $H(1)$ , there is again just one extra cointegration relation under the alternative ( $r - r_0 = 1$ ) and the power of all tests is rather low for  $\psi_2 = 0.9$ , as can be seen in Figure 7. Obviously, in some situations it is not very likely to find a cointegrating rank of two. Setting the second autoregressive eigenvalue to  $\psi_2 = 0.7$  increases the small sample power of the tests for  $\Theta = (0, 0)'$  remarkably (see Figure 8). In the case of testing  $H(0)$  so that  $r - r_0 = 2$  (Panels A and B), the advantage of the trace tests is even more pronounced than for the DGPs with  $\psi_2 = 0.9$ . When the null hypothesis is  $H(1)$  and, hence,  $r - r_0 = 1$ , both test versions are once more rather similar (Panels C and D).

In summary, we can conclude that with respect to the small sample power both the trace and the maximum eigenvalue tests have similar properties as well, in line with the local power results. However, in some cases, the trace tests are clearly superior to the corresponding maximum eigenvalue versions in terms of power. Not surprisingly, this happens in particular if the actual rank  $r$  exceeds the rank specified in the null hypothesis,  $r_0$ , by more than one,  $r - r_0 > 1$ . In those cases where the maximum eigenvalue tests dominate, their power advantage is only minor. On the other hand, the latter tests seem to have smaller size distortions than the trace tests. Still, our overall recommendation is to use the trace tests if one wants to apply just one test version. Of course, there is nothing wrong with the common practice to use both versions simultaneously.

## 6 Conclusions

In this study we have compared maximum eigenvalue and trace tests for the cointegrating rank of a VAR process. The comparison is performed for test variants suitable for different types of deterministic terms. More precisely, a couple of tests allowing for a nonzero mean

and a group of tests allowing in addition for a deterministic linear trend are considered. The asymptotic distributions under local alternatives are given and a local power comparison is presented. In that comparison no major differences between corresponding maximum eigenvalue and trace tests are detected. In a small sample simulation comparison it is found, however, that in some situations trace tests tend to have more heavily distorted sizes whereas their power performance is superior to that of the maximum eigenvalue competitors. In particular, the trace tests are advantageous if there are at least two more cointegrating relations in the process than specified under the null hypothesis. Based on our simulations we have a preference for the trace tests. This result justifies the common practice in empirical work to use either both types of tests simultaneously or apply the trace tests exclusively.

In accordance with other authors (see, e.g., Hubrich et al. (2001)) we also found that the alternative LR type test versions for a specific deterministic term sometimes differ more substantially than the corresponding maximum eigenvalue and trace tests. Therefore, based on our simulation results, it appears to be more sensible to apply different versions with respect to the treatment of deterministic terms rather than the maximum eigenvalue and the trace variant of one specific test version.

## Appendix. Derivation of Asymptotic Distributions of Test Statistics

To derive the limiting distributions of the maximum eigenvalue tests considered in Section 3, we show that the general framework used by Saikkonen & Lütkepohl (1999) to obtain the limiting distribution of the LR test in a reduced rank regression model can also be used in the case of the maximum eigenvalue test. This result does not follow automatically from Saikkonen & Lütkepohl (1999) because the method of proof used in that paper was based on an idea in which only the limiting distribution of the sum of the relevant eigenvalues and not the limiting distribution of individual eigenvalues was derived. The following extension of Theorem 1 of Saikkonen & Lütkepohl (1999) shows that the joint limiting distribution of the relevant eigenvalues can be obtained without strengthening the previously used assumptions.

Thus, consider the reduced rank regression model

$$Y_t = AB'X_t + Z_t, \quad t = 1, \dots, T, \quad (A.1)$$

where  $Y_t$  and  $Z_t$  are  $(n \times 1)$  vectors,  $X_t$  is an  $(m \times 1)$  vector with  $m \geq n$ , and  $A$  and  $B$  are



$(n \times r_0)$  and  $(m \times r_0)$  coefficient matrices of full column rank, respectively. The error term  $Z_t$  is assumed to be of the form

$$Z_t = T^{-1}A_1B_1'X_t + \mathcal{E}_t, \quad (A.2)$$

where  $A_1$  and  $B_1$  are  $(n \times (r - r_0))$  and  $(m \times (r - r_0))$  matrices, respectively, with  $r - r_0 > 0$  and  $\mathcal{E}_t$  is the error term under the null hypothesis that (A.1) is a correctly specified model. The matrices  $[A : A_1]$  and  $[B : B_1]$  are supposed to be of full column rank unless the null hypothesis holds, in which case  $A_1 = 0$  and  $B_1$  may also be zero.

It is well known that if  $\mathcal{E}_t$  is Gaussian white noise and  $X_t$  is strictly exogenous or pre-determined the statistical analysis of the reduced rank regression model (A.1) is based on a generalized eigenvalue problem in which one solves the determinantal equation

$$\det(M_{XY}M_{YY}^{-1}M_{YX} - lM_{XX}) = 0, \quad (A.3)$$

where  $M_{XX} = T^{-1} \sum_{t=1}^T X_t X_t'$ ,  $M_{XY} = M_{YX}' = T^{-1} \sum_{t=1}^T X_t Y_t'$  and  $M_{YY} = T^{-1} \sum_{t=1}^T Y_t Y_t'$ . In particular, if  $\hat{l}_1 \geq \dots \geq \hat{l}_n$  are the ordered solutions of (A.3) then the LR test for the null hypothesis that the rank specification in (A.1) is correct is based on the trace test statistic analogous to that given in (3.1). The alternative hypothesis of this LR test is that the column rank of  $A$  and  $B$  is larger than  $r_0$ . If the alternative is considered that the parameter matrix has rank  $r_0 + 1$ , the corresponding LR test is based on the maximum eigenvalue test statistic analogous to (3.2).

Saikkonen & Lütkepohl (1999) obtain the limiting distribution of the trace statistics under the following general assumptions.

**Assumption 1.**

- (i)  $T^{-1} \sum_{t=1}^T B'X_tX_t'B \xrightarrow{p} \Sigma_{BB} > 0$
- (ii)  $T^{-1} \sum_{t=1}^T B_{\perp}'X_tX_t'B = O_p(1)$
- (iii)  $T^{-2} \sum_{t=1}^T X_tX_t' \xrightarrow{d} G$  for some (generally) random  $(m \times m)$  matrix  $G$  with  $B_{\perp}'GB_{\perp} > 0$  and  $B'G = 0$  (a.s.)
- (iv)  $T^{-1/2} \sum_{t=1}^T \mathcal{E}_tX_t'B = O_p(1)$
- (v)  $T^{-1} \sum_{t=1}^T \mathcal{E}_tX_t'B_{\perp} \xrightarrow{d} S$  for some random  $(n \times (m - r_0))$  matrix  $S$

(vi)  $T^{-1} \sum_{t=1}^T \mathcal{E}_t \mathcal{E}_t' = \Sigma_{\mathcal{E}\mathcal{E}} + O_p(T^{-1/2})$  for some fixed matrix  $\Sigma_{\mathcal{E}\mathcal{E}} > 0$

Furthermore, the sequences in (iii) and (v) converge jointly in distribution.  $\square$

Using Assumption 1 we can now prove the following theorem.

**Theorem A.1**

If Assumption 1 holds, the random variables  $T\hat{l}_j$ ,  $j = r_0 + 1, \dots, n$ , converge weakly and jointly to the solutions of the equation

$$\det[(FB'_\perp GB_\perp + A'_\perp S)'(A'_\perp \Sigma_{\mathcal{E}\mathcal{E}} A_\perp)^{-1}(FB'_\perp GB_\perp + A_\perp S) - \rho(B'_\perp GB_\perp)] = 0$$

where  $F = A'_\perp A_1 B'_1 B_\perp (B'_\perp B_\perp)^{-1}$ .  $\square$

**Proof:** From the proof of Lemma 1 of Saikkonen & Lütkepohl (1999) we first obtain the following simple consequences of Assumption 1:

$$M_{YX}B = A\Sigma_{BB} + o_p(1) \stackrel{def}{=} \Sigma_{YB} + o_p(1) \tag{A.4}$$

and

$$M_{YY} = A\Sigma_{BB}A' + \Sigma_{\mathcal{E}\mathcal{E}} + o_p(1) \stackrel{def}{=} \Sigma_{YY} + o_p(1). \tag{A.5}$$

Using the notation  $\Sigma_{BY} = \Sigma'_{YB} = \Sigma_{BB}A'$ , we then also have

$$\Sigma_{YY} = A\Sigma_{BY} + \Sigma_{\mathcal{E}\mathcal{E}}. \tag{A.6}$$

A further straightforward consequence of Assumption 1 to be used below is

$$M_{YX} = O_p(1). \tag{A.7}$$

Now consider the solutions of the generalized eigenvalue problem (A.3). In the proof of Lemma 1 of Saikkonen & Lütkepohl (1999) it is shown that the eigenvalues converge weakly to those of the equation

$$\det(\Sigma_{BY}\Sigma_{YY}^{-1}\Sigma_{YB} - l\Sigma_{BB}) \det(lB'_\perp GB_\perp) = 0.$$

This implies that the  $r_0$  largest eigenvalues of (A.3) converge in probability to positive constants and the  $n - r_0$  smallest eigenvalues converge in probability to zero. Since  $B'_\perp M_{XX} B_\perp =$

$O_p(T)$  it is not difficult to check that  $\hat{l}_j = O_p(T^{-1})$  for  $j = r_0 + 1, \dots, n$ . Thus, to analyze the asymptotic behaviour of these eigenvalues, we use the equation

$$\det(S(\rho)) = 0, \quad (\text{A.8})$$

where  $S(\rho) = M_{XY}M_{YY}^{-1}M_{YX} - \rho(T^{-1}M_{XX})$  and, if  $\hat{\rho}_1 \geq \dots \geq \hat{\rho}_n$  are the solutions of this equation, then  $\hat{\rho}_j = O_p(1)$ ,  $j = r_0 + 1, \dots, n$ , can be assumed.

The solutions of (A.8) do not change if  $S(\rho)$  is premultiplied by the matrix  $[B : B_\perp]'$  and postmultiplied by  $[B : B_\perp]$ . Thus, in the same way as in Johansen (1995, p. 159) we consider the decomposition

$$\begin{aligned} \det([B : B_\perp]'S(\rho)[B : B_\perp]) &= \det(B'S(\rho)B) \det(B'_\perp\{S(\rho) - S(\rho)B[B'S(\rho)B]^{-1}B'S(\rho)\}B_\perp) \\ &\stackrel{def}{=} \det(S_1(\rho)) \det(S_2(\rho)) \end{aligned}$$

For  $j = r_0 + 1, \dots, n$ , we have

$$S_1(\hat{\rho}_j) = B'M_{XY}M_{YY}^{-1}M_{YX}B - \hat{\rho}_j(T^{-1}B'M_{XX}B) = \Sigma_{BY}\Sigma_{YY}^{-1}\Sigma_{YB} + o_p(1), \quad (\text{A.9})$$

where the latter equality is based on (A.4), (A.5), Assumption 1(i) and the fact that  $\hat{\rho}_j = O_p(1)$ . This shows that asymptotically  $\hat{\rho}_{r_0+1}, \dots, \hat{\rho}_n$  are not roots of  $\det(S_1(\rho))$ . Hence, we can consider  $S_2(\rho)$ .

For  $j = r_0 + 1, \dots, n$ ,

$$\begin{aligned} B'_\perp S(\rho)B &= B'_\perp M_{XY}M_{YY}^{-1}M_{YX}B - \hat{\rho}_j(T^{-1}B'_\perp M_{XX}B) \\ &= B'_\perp M_{XY}\Sigma_{YY}^{-1}\Sigma_{YB} + o_p(1). \end{aligned} \quad (\text{A.10})$$

Since here  $\hat{\rho}_j = O_p(1)$ , the latter equality is obtained by using (A.4), (A.5), (A.7) and Assumption 1(ii). By the same arguments we also find that, for  $j = r_0 + 1, \dots, n$ ,

$$\begin{aligned} B'_\perp S(\rho)B_\perp &= B'_\perp M_{XY}M_{YY}^{-1}M_{YX}B_\perp - \hat{\rho}_j(T^{-1}B'_\perp M_{XX}B_\perp) \\ &= B'_\perp M_{XY}\Sigma_{YY}^{-1}M_{YX}B_\perp - \hat{\rho}_j(T^{-1}B'_\perp M_{XX}B_\perp) + o_p(1). \end{aligned} \quad (\text{A.11})$$

From (A.9) - (A.11) it follows that

$$S_2(\hat{\rho}_j) = B'_\perp M_{XY}PM_{YX}B_\perp - \hat{\rho}_j(T^{-1}B'_\perp M_{XX}B_\perp) + o_p(1), \quad j = r_0 + 1, \dots, n, \quad (\text{A.12})$$

where

$$\begin{aligned} P &= \Sigma_{YY}^{-1} - \Sigma_{YY}^{-1}\Sigma_{YB}(\Sigma_{BY}\Sigma_{YY}^{-1}\Sigma_{YB})^{-1}\Sigma_{BY}\Sigma_{YY}^{-1} \\ &= \Sigma_{YY}^{-1} - \Sigma_{YY}^{-1}A(A'\Sigma_{YY}^{-1}A)^{-1}A'\Sigma_{YY}^{-1} \\ &= A_\perp(A'_\perp\Sigma_{\mathcal{E}\mathcal{E}}A_\perp)^{-1}A'_\perp. \end{aligned}$$

Here the second equality is obtained from the definition of  $\Sigma_{YB}$  in (A.4) and the third one can be justified by using the identity  $A'_\perp \Sigma_{YY} = A'_\perp \Sigma_{\mathcal{E}\mathcal{E}}$  obtained from (A.6) in conjunction with the argument used by Johansen (1995, p. 142) to prove his result (10.6). Hence, from equation (A.1) and the definitions it follows that

$$\begin{aligned} B'_\perp M_{XY} P M_{YX} B_\perp &= B'_\perp T^{-1} \sum_{t=1}^T X_t Z'_t P T^{-1} \sum_{t=1}^T Z_t X'_t B_\perp \\ &\xrightarrow{d} (A_1 B'_1 G B_\perp + S)' P (A_1 B'_1 G B_\perp + S) \stackrel{def}{=} H, \end{aligned}$$

where the weak convergence is due to the definition of  $Z_t$  and Assumption 1(iii) and (v). From this result, (A.12), Assumption 1(iii), and the continuity of generalized eigenvalues, we can conclude that  $\hat{\rho}_{r_0+1}, \dots, \hat{\rho}_n$  converge weakly to the solutions of the equation

$$\det[H - \rho(B'_\perp G B_\perp)] = 0. \quad (A.13)$$

Observing that  $B'_1 G B_\perp = B'_1 B_\perp (B'_\perp B_\perp)^{-1} B'_\perp G B_\perp$  (see the proof of Theorem 1 of Saikkonen & Lütkepohl (1999)) and using the definitions it can be readily seen that (A.13) is identical to the equation in the theorem. This completes the proof.  $\square$

Using Theorem A.1 and the results in Saikkonen & Lütkepohl (1999) and Lütkepohl & Saikkonen (2000), it is now a simple matter to demonstrate that the maximum eigenvalue tests considered in this paper have the limiting distributions stated in Section 3. First, note that if  $\rho_1 \geq \dots \geq \rho_n$  are the solutions of the equation in Theorem A.1 or, equivalently, (A.13), then

$$LR_{trace}(r_0) = \sum_{j=r_0+1}^n \hat{\rho}_j + o_p(1) \xrightarrow{d} \sum_{j=r_0+1}^n \rho_j = \text{tr}\{H(B'_\perp G B_\perp)^{-1}\}.$$

Notice that the last expression is identical to that in Theorem 1 of Saikkonen & Lütkepohl (1999) by the definitions and well-known results of matrix algebra. Similarly,

$$LR_{max}(r_0) = \hat{\rho}_{r_0+1} + o_p(1) \xrightarrow{d} \rho_{r_0+1} = \lambda_{max}\{H(B'_\perp G B_\perp)^{-1}\}.$$

Because it has been shown in the aforementioned previous work that the general framework of equations (A.1), (A.2) and Assumption 1 can be applied for deriving the limiting distributions of the LR trace tests considered in Section 3, it follows that the same is true for the corresponding maximum eigenvalue tests. Moreover, it follows that if we have  $LR_{trace}(r_0) \xrightarrow{d} \text{tr}(\mathcal{D})$  then  $LR_{max}(r_0) \xrightarrow{d} \lambda_{max}(\mathcal{D})$  obtains.

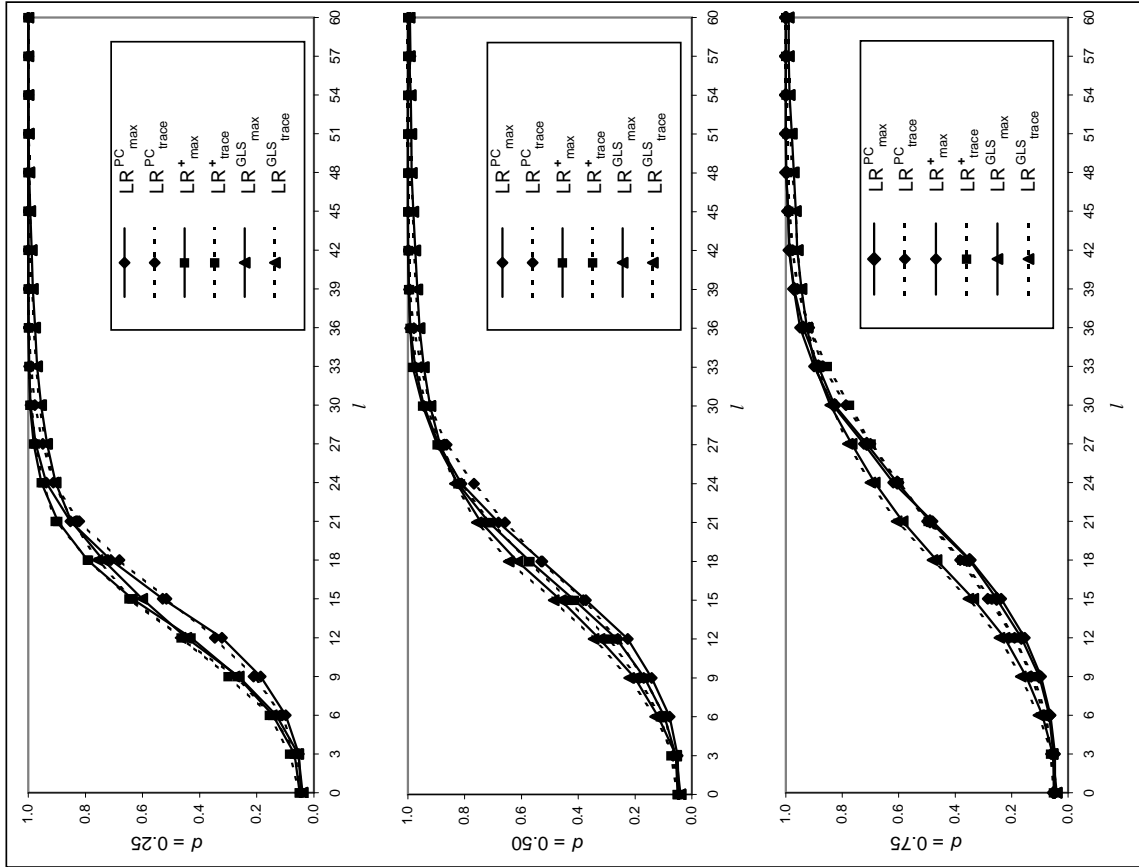
To illustrate this result, consider the test statistics  $LR_{trace}^{SL}(r_0)$  and  $LR_{max}^{SL}(r_0)$  and define  $K(s) = (\alpha'_\perp \Omega \alpha_\perp)^{1/2} N(s)$ . From Section A.4.4 of Saikkonen & Lütkepohl (1999) we can conclude that Theorem A.1 applies with the counterparts of the matrices  $B'_\perp GB$  and  $(FB'_\perp GB_\perp + A'_\perp S)'$  given by  $\int_0^1 K(s)K(s)'ds$  and  $\int_0^1 K(s)dK(s)'$ , respectively. Because the counterpart of  $A'_\perp \Sigma_{\varepsilon\varepsilon} A_\perp$  is  $\alpha'_\perp \Omega \alpha_\perp$ , it follows from Theorem A.1 that the limiting distributions of the test statistics  $LR_{trace}^{SL}(r_0)$  and  $LR_{max}^{SL}(r_0)$  are given by  $\text{tr}(\mathcal{D}^0)$  and  $\lambda_{max}(\mathcal{D}^0)$ , respectively, where  $\mathcal{D}^0$  is as defined in Section 3. Details of the other maximum eigenvalue tests can be worked out similarly.

## References

- Gonzalo, J. & J.-Y. Pitarakis (1999), Dimensionality effect in cointegration analysis, in R. Engle & H. White (eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford: Oxford University Press, pp. 212-229.
- Hubrich, K., H. Lütkepohl & P. Saikkonen (2001), A review of systems cointegration tests, *Econometric Reviews*, forthcoming.
- Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.
- Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, 59, 1551-1580.
- Johansen, S. (1992), Determination of cointegration rank in the presence of a linear trend, *Oxford Bulletin of Economics and Statistics*, 54, 383-397.
- Johansen, S. (1994), The role of the constant and linear terms in cointegration analysis of nonstationary time series, *Econometric Reviews*, 13, 205-231.
- Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.
- Johansen, S. & K. Juselius (1990), Maximum likelihood estimation and inference on cointegration - with applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

- Lütkepohl, H. & P. Saikkonen (2000), Testing for the cointegrating rank of a VAR process with a time trend, *Journal of Econometrics*, 95, 177-198.
- Perron, P. & J.Y. Campbell (1993), A note on Johansen's cointegration procedure when trends are present, *Empirical Economics*, 18, 777-789.
- Saikkonen, P. & H. Lütkepohl (1999), Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory*, 15, 50 - 78.
- Saikkonen, P. & H. Lütkepohl (2000a), Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process, *Journal of Time Series Analysis*, 21, 435 - 456.
- Saikkonen, P. & H. Lütkepohl (2000b), Testing for the cointegrating rank of a VAR process with an intercept, *Econometric Theory*, 16, 373 - 406.
- Saikkonen, P. & R. Luukkonen (1997), Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics*, 81, 93 - 126.
- Toda, H.Y. (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.Y. (1995), Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions, *Econometric Theory*, 11, 1015 - 1032.

Panel A: Test with a time trend



Panel B: Tests without a time trend

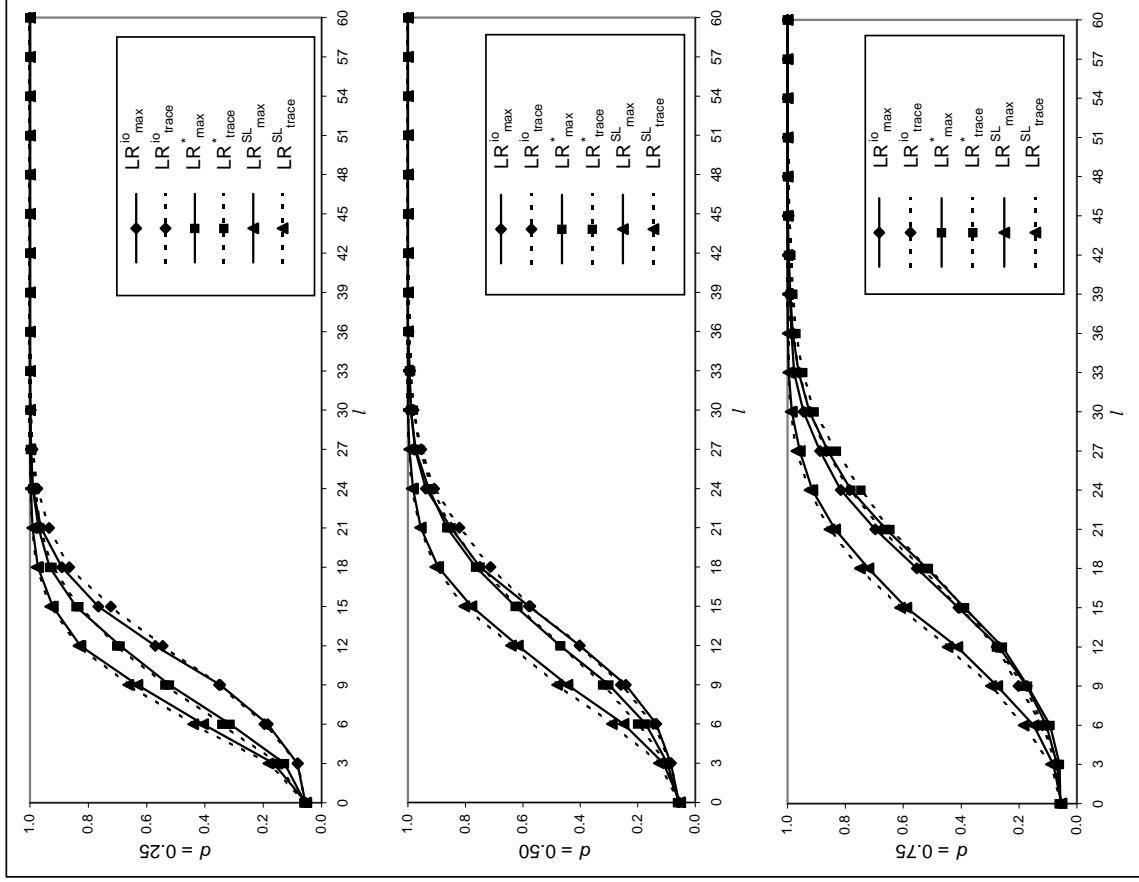
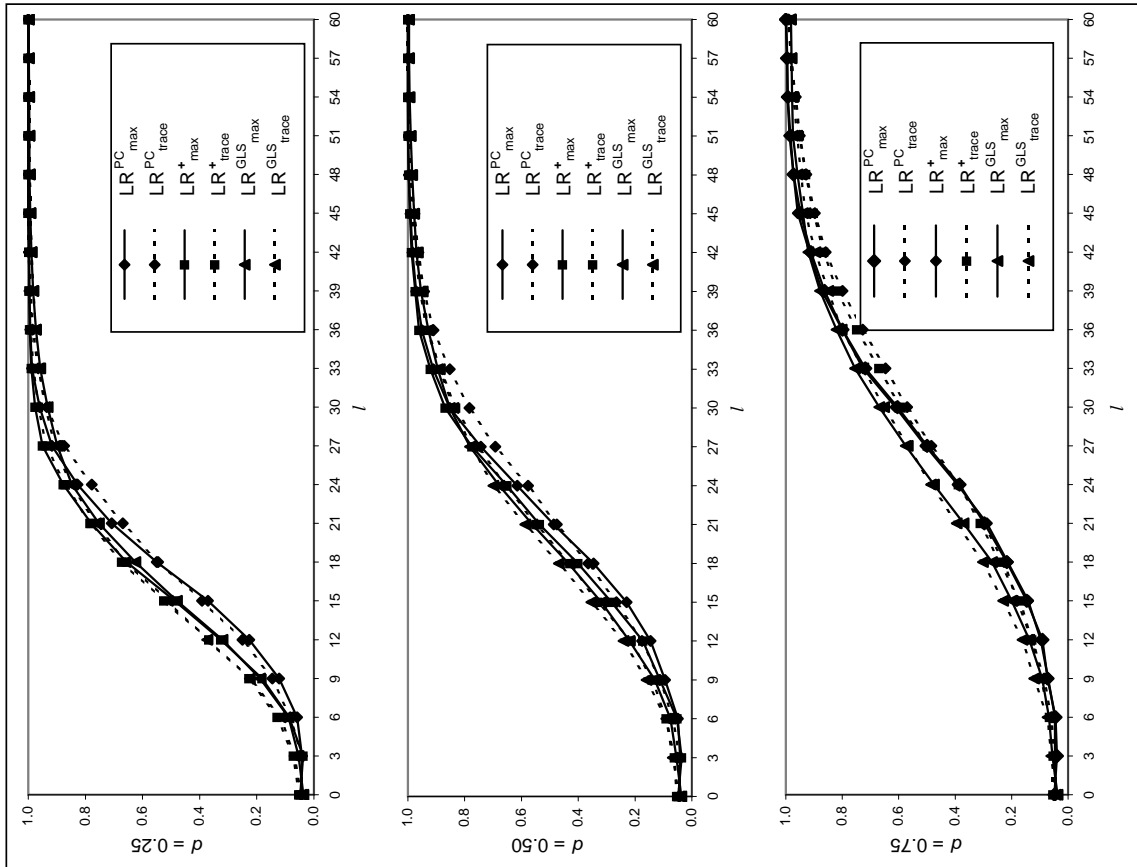


Figure 1. Local power of LR type tests for  $n - r_0 = 2$ .

Panel A: Test with a time trend



Panel B: Tests without a time trend

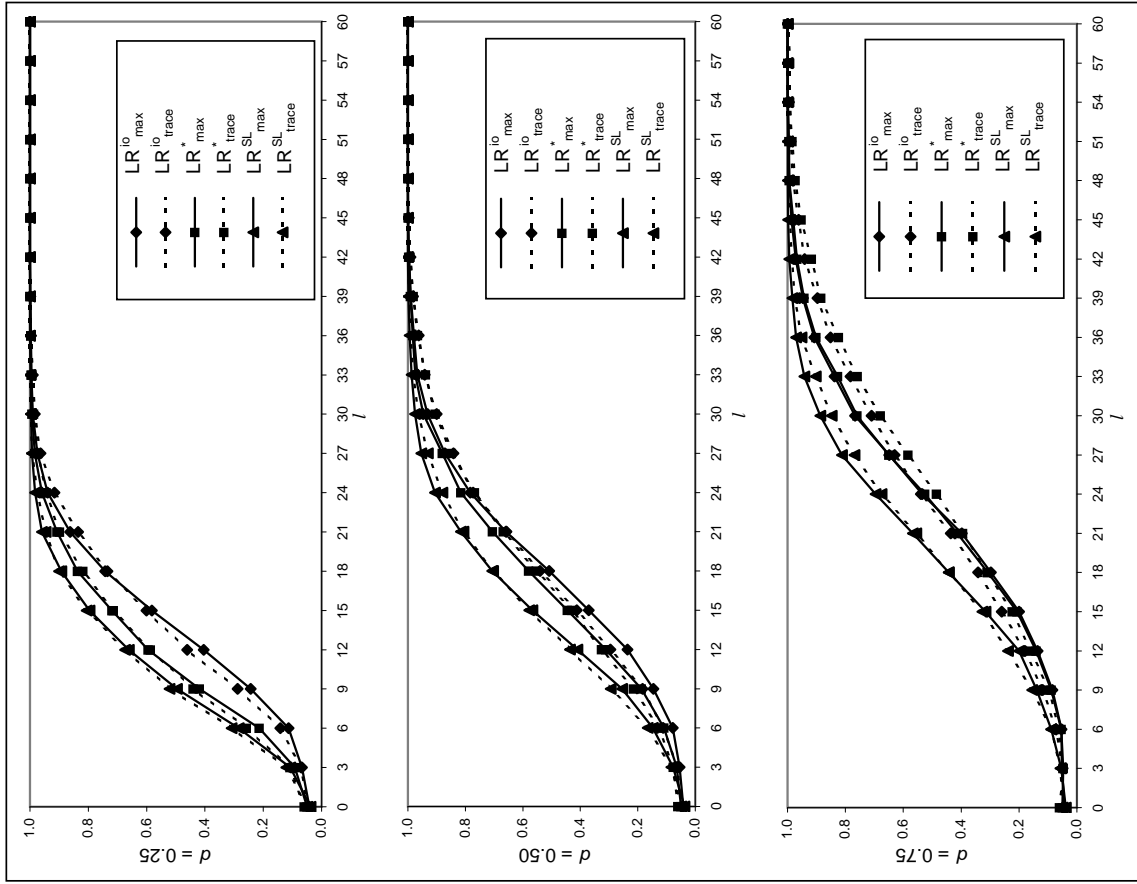
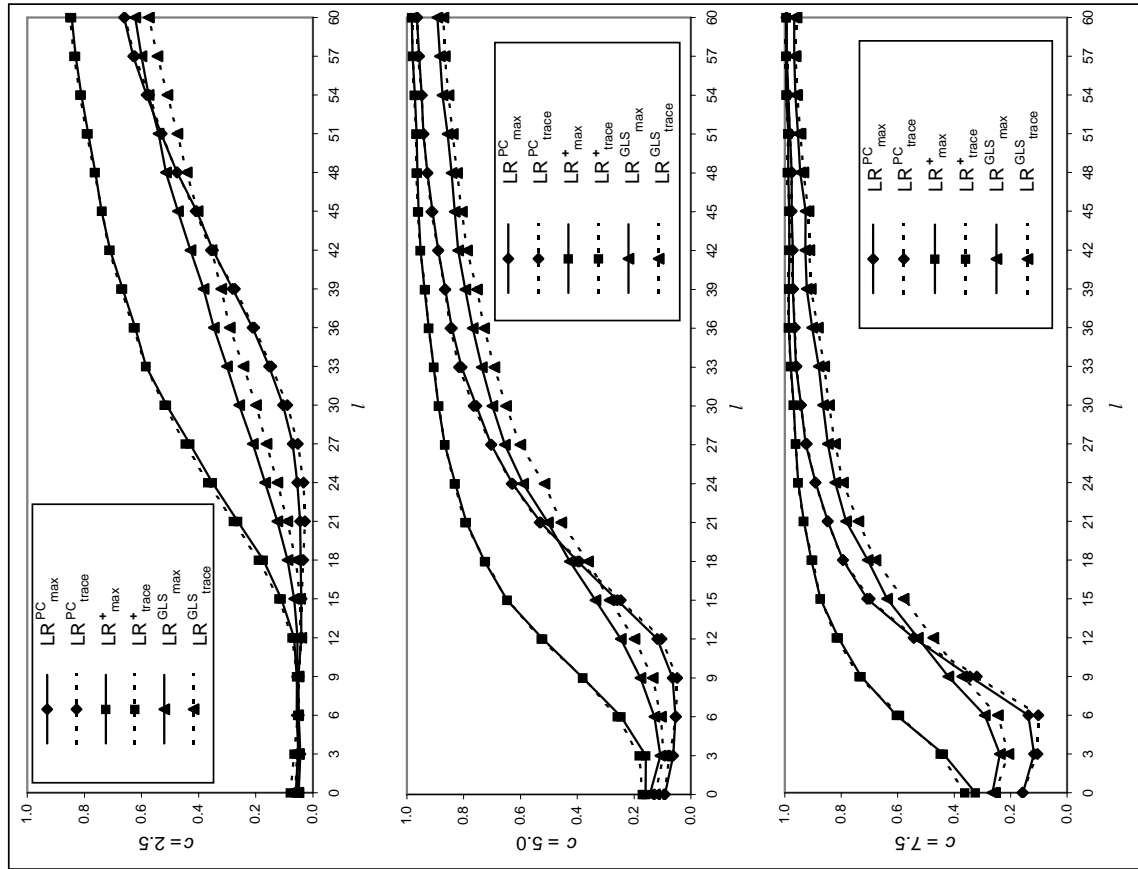


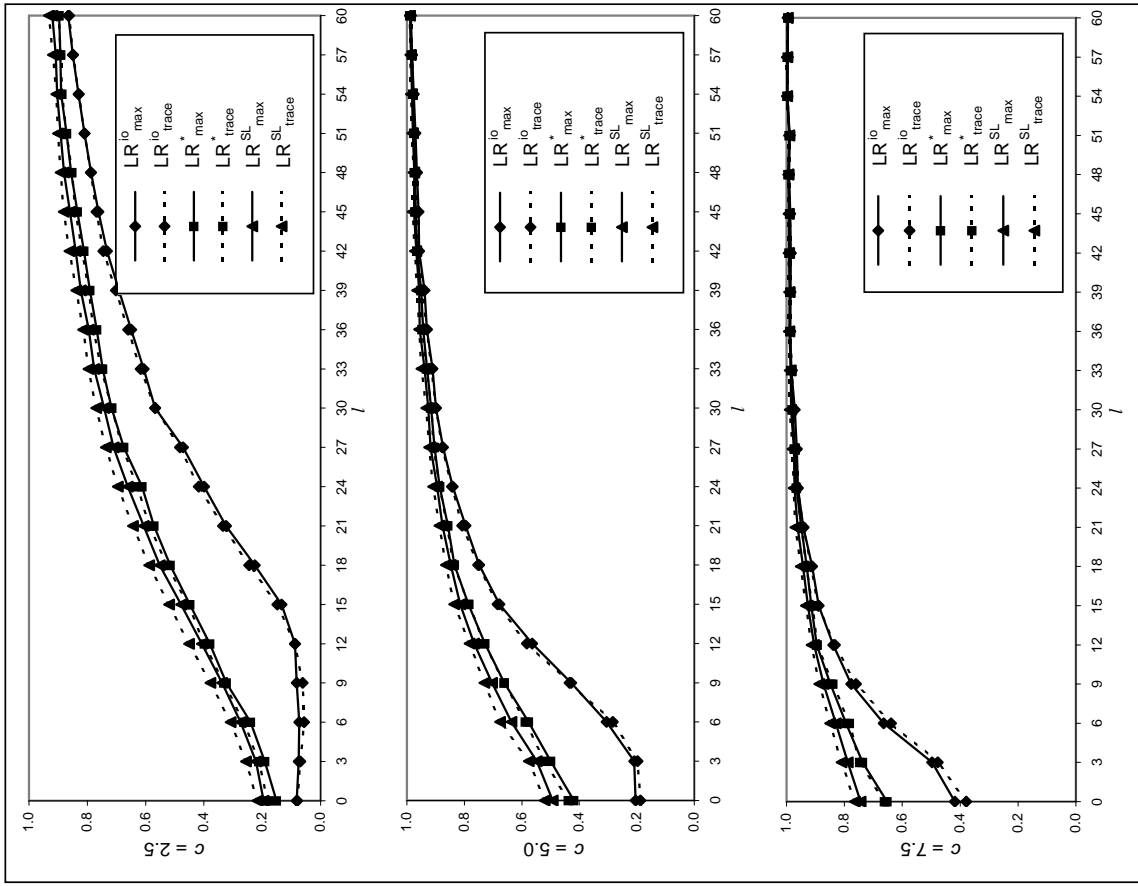
Figure 2. Local power of LR type tests for  $n - r_0 = 3$ .



**Panel A: Test with a time trend**

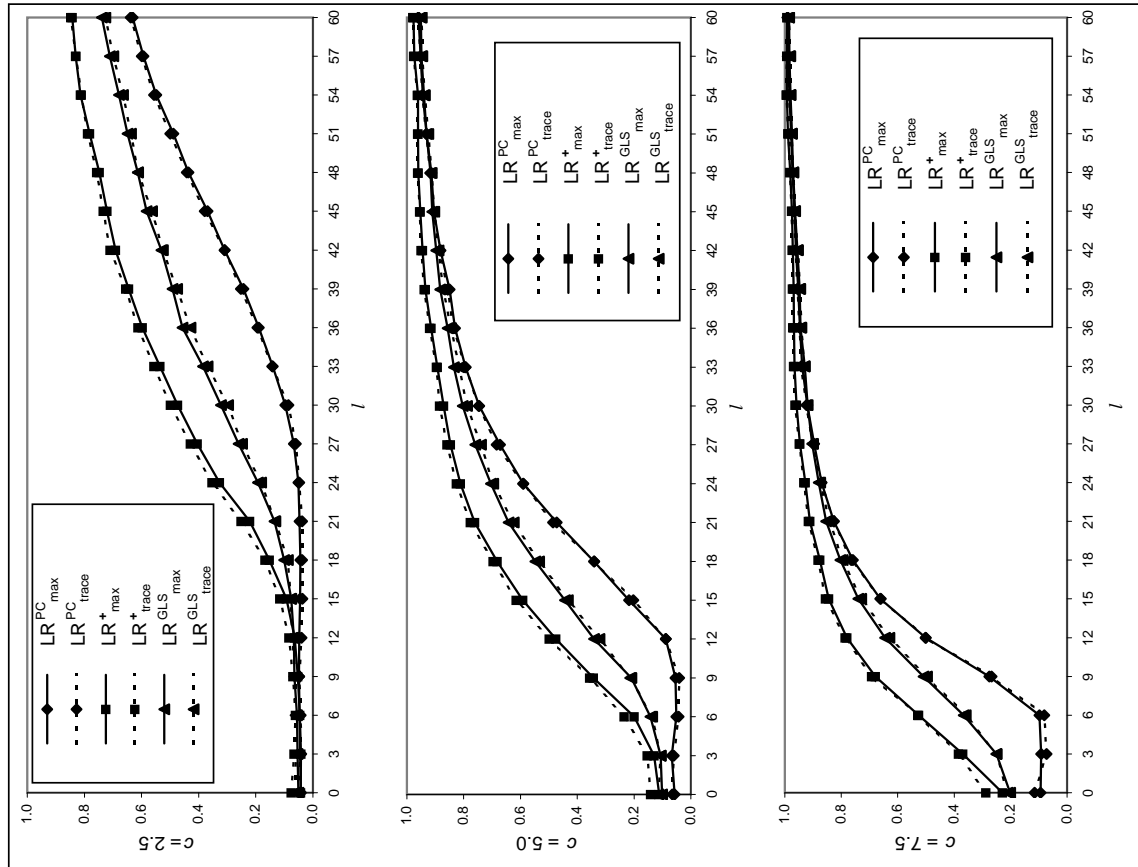


**Panel B: Tests without a time trend**



**Figure 3.** Local power of LR type tests for  $n - r_0 = 2$  and  $d = 0.25$ .

Panel A: Test with a time trend



Panel B: Tests without a time trend

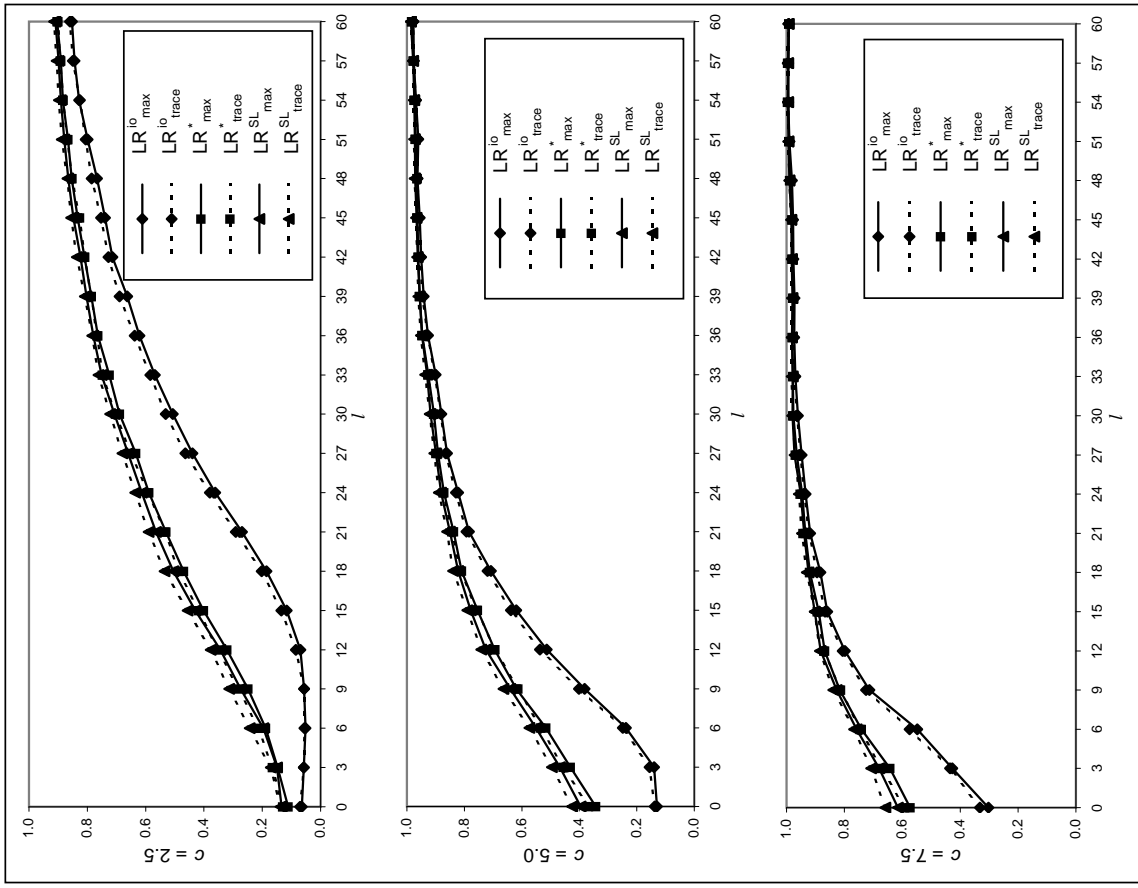


Figure 4. Local power of LR type tests for  $n - r_0 = 3$  and  $d = 0.25$ .

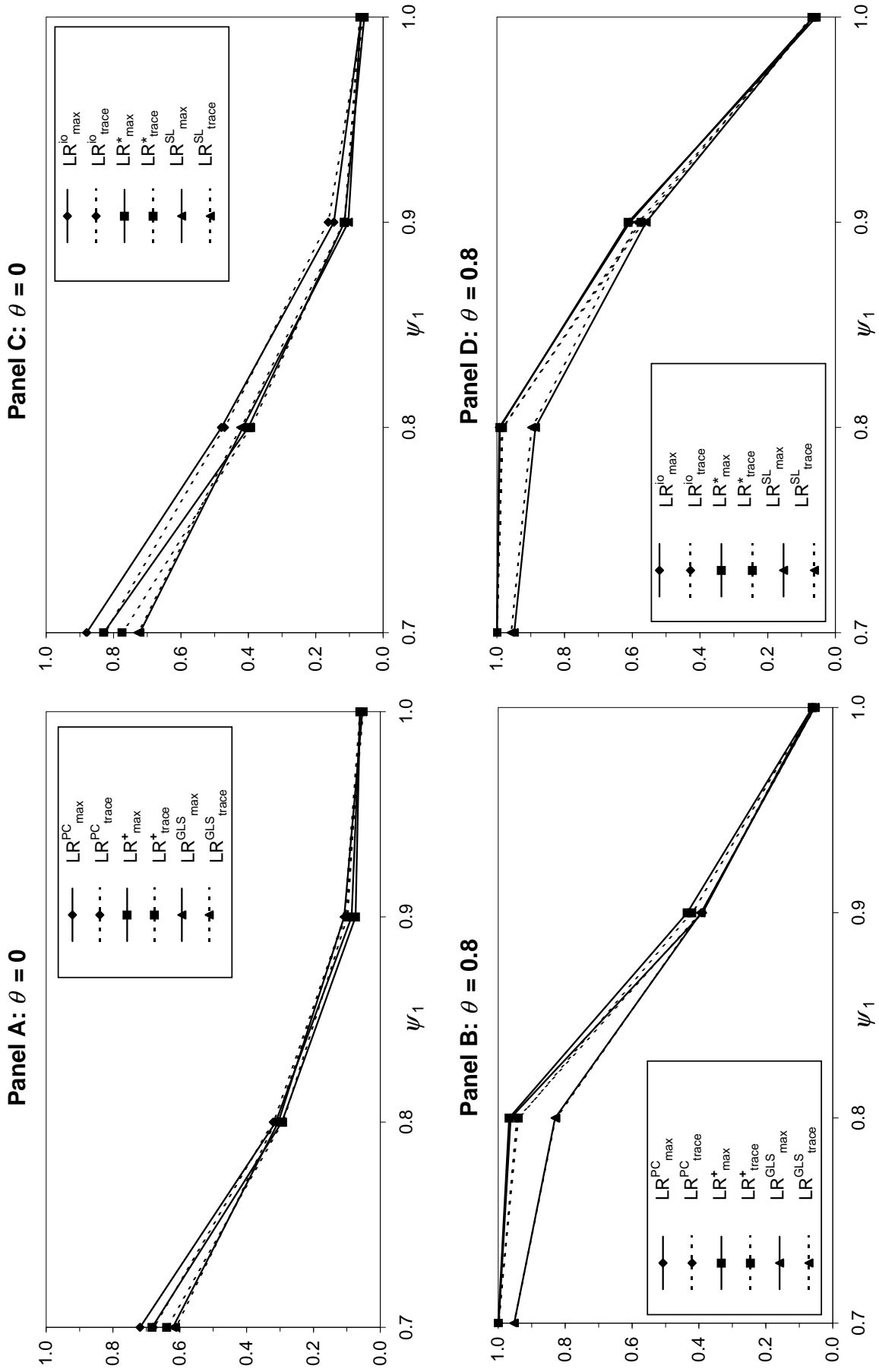
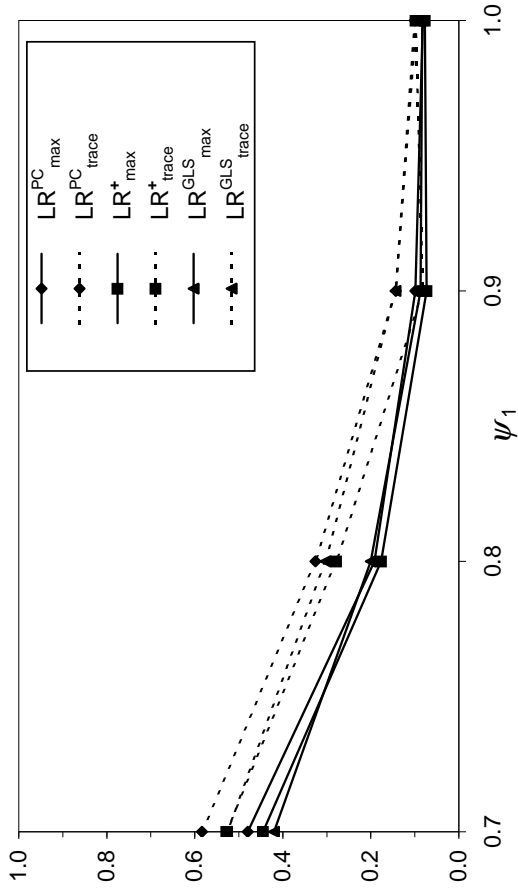
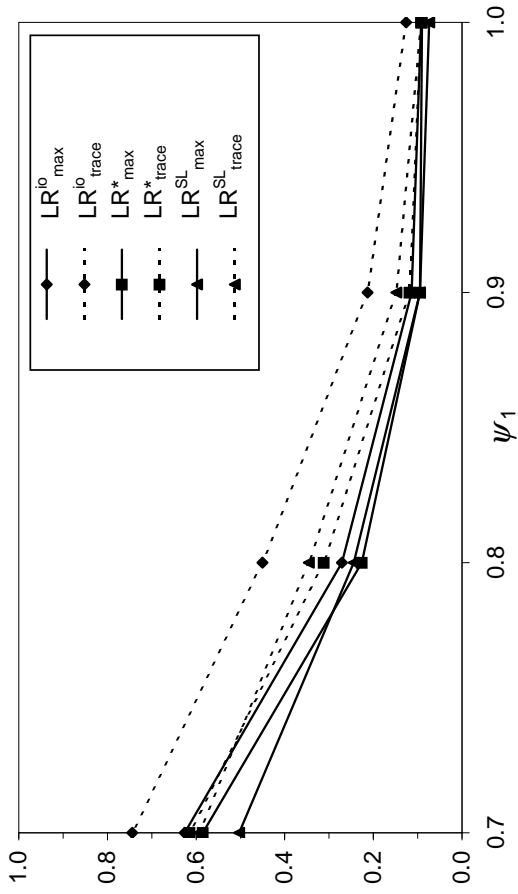


Figure 5. Small sample properties of tests for  $H(0) : r = 0$  based on bivariate DGPs with  $r = 0$  or  $r = 1$ ,  $T = 100$ .

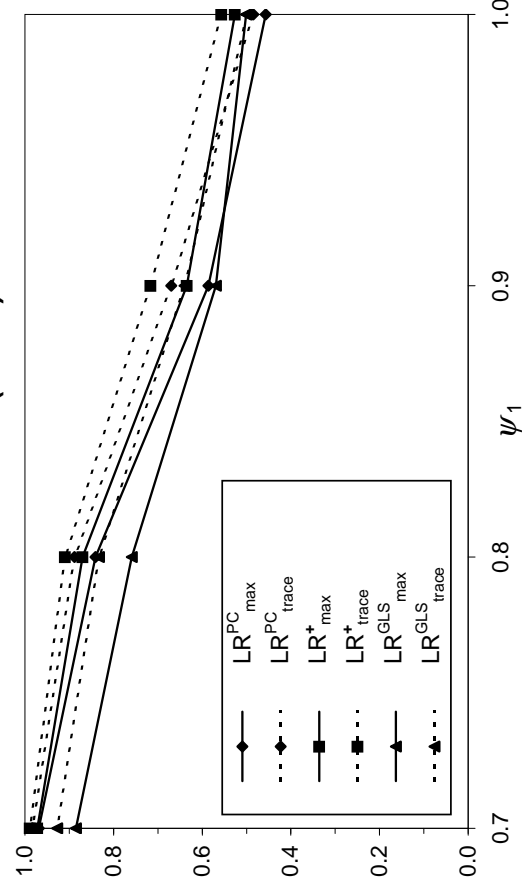
Panel A:  $\Theta = (0\ 0)'$



Panel C:  $\Theta = (0\ 0)'$



Panel B:  $\Theta = (0.4\ 0.8)'$



Panel D:  $\Theta = (0.4\ 0.8)'$

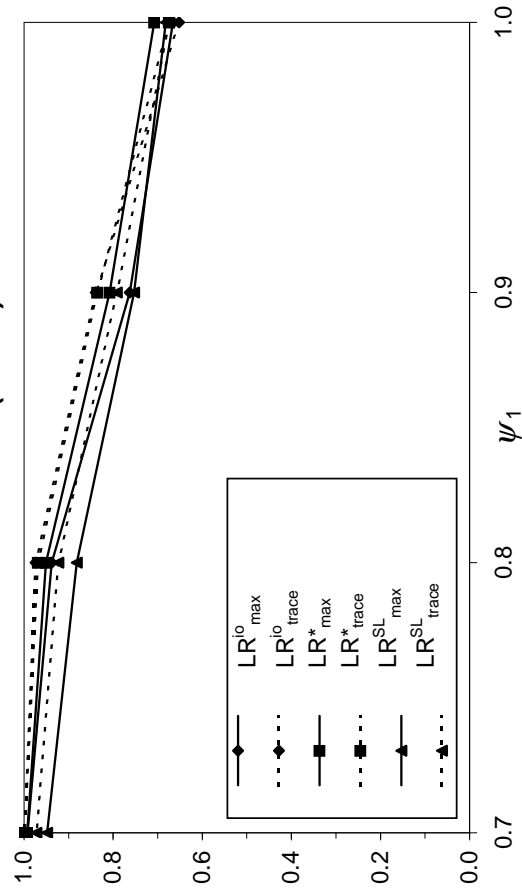
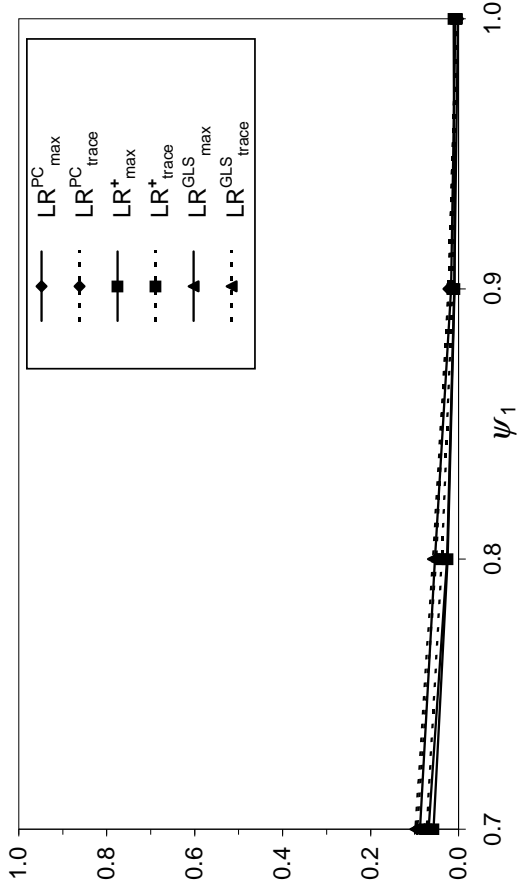
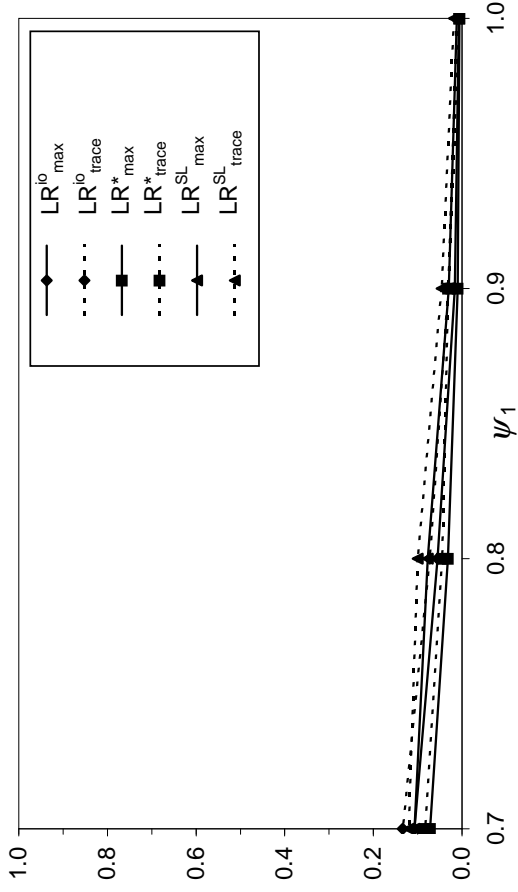


Figure 6. Small sample properties of tests for  $H(0) : r = 0$  based on three-dimensional DGPs with  $r = 1$  or  $r = 2$ ,  $\psi_2 = 0.9$ ,  $T = 100$ .

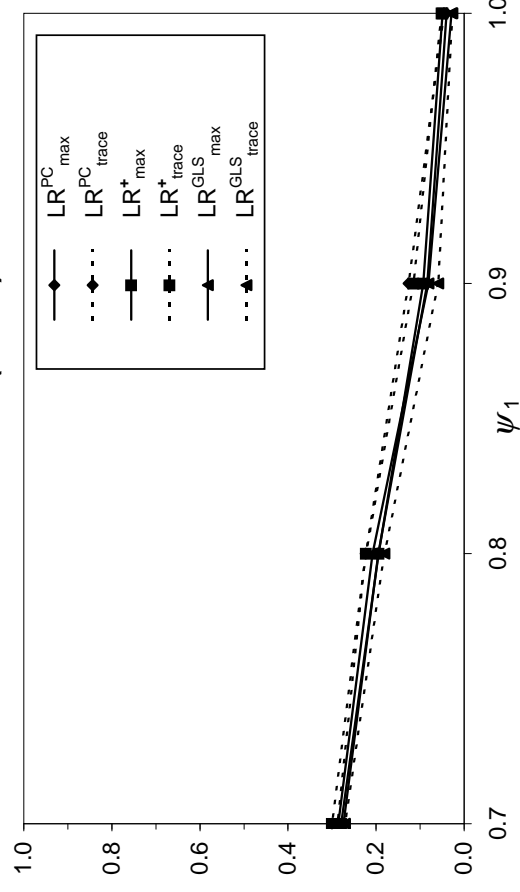
Panel A:  $\Theta = (0 \ 0)'$



Panel C:  $\Theta = (0 \ 0)'$



Panel B:  $\Theta = (0.4 \ 0.8)'$



Panel D:  $\Theta = (0.4 \ 0.8)'$

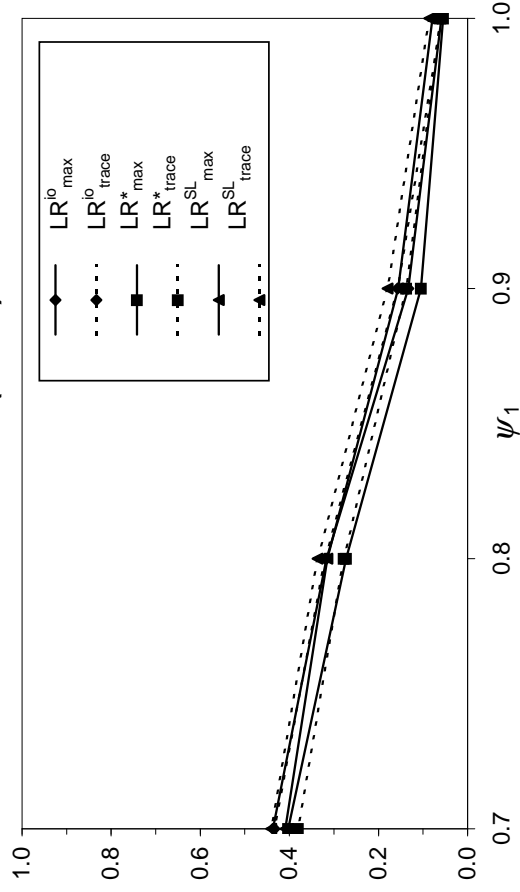


Figure 7. Small sample properties of tests for  $H(1)$ :  $r = 1$  based on three-dimensional DGPs with  $r = 1$  or  $r = 2$ ,  $\psi_2 = 0.9$ ,  $T = 100$ .

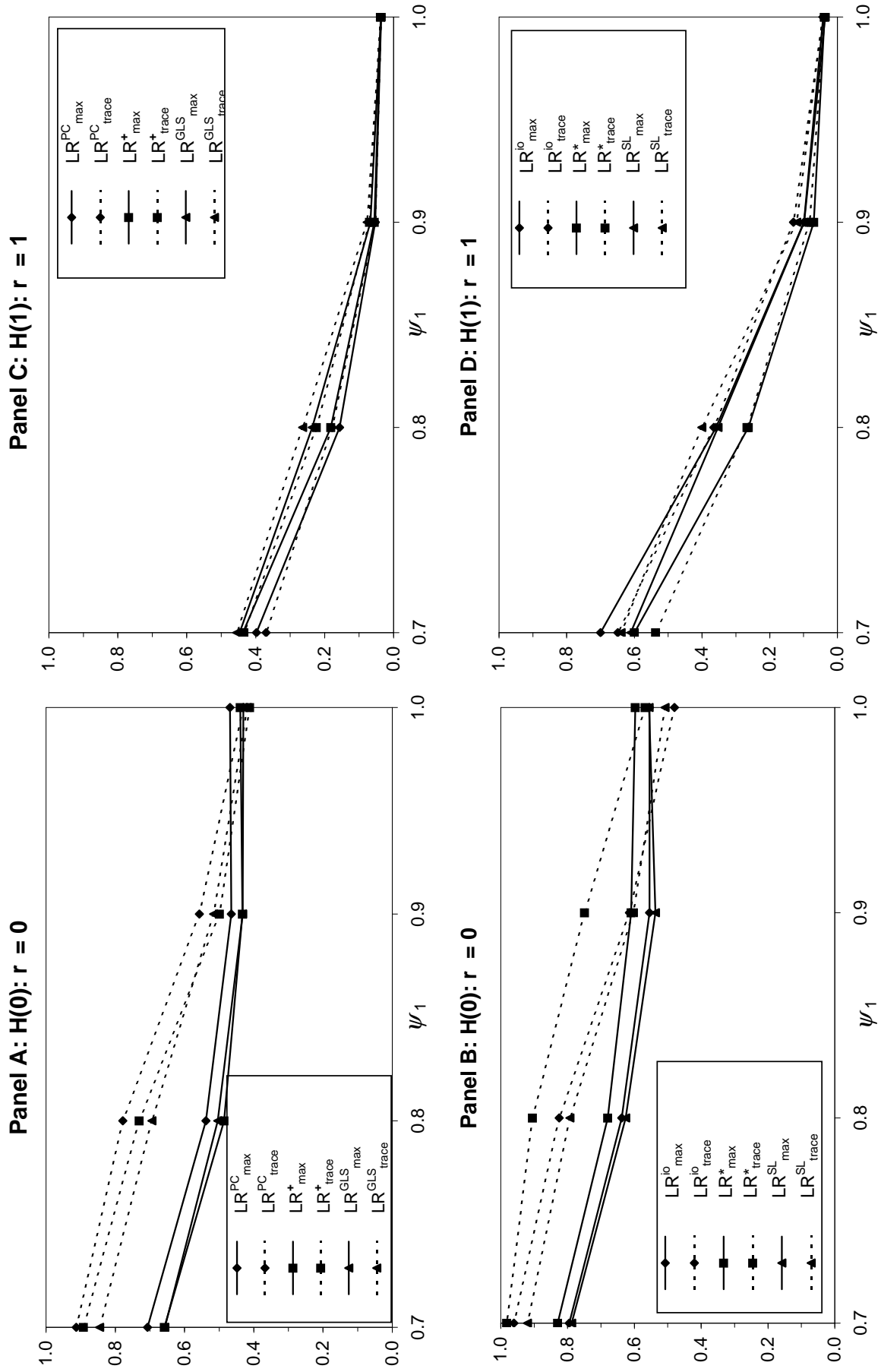


Figure 8. Small sample properties of tests for three-dimensional DGPs with  $r = 1$  or  $r = 2$ ,  $\psi_2 = 0.7$ ,  $\Theta = (0, 0)'$ ,  $T = 100$ .