European University Institute

Department of Economics

EUI Working Papers

ECO 2008/24

Testing for the Cointegrating Rank of a Vector Autoregressive Process with Uncertain Deterministic Trend Term

Matei Demetrescu, Helmut Lütkepohl and Pentti Saikkonen

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

Testing for the Cointegrating Rank of a Vector Autoregressive Process with Uncertain Deterministic Trend Term

MATEI DEMETRESCU, HELMUT LÜTKEPOHL

and

PENTTI SAIKKONEN

EUI Working Paper **ECO** 2008/24

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

May 28, 2008

Testing for the Cointegrating Rank of a Vector Autoregressive Process with Uncertain Deterministic Trend Term [1]

Matei Demetrescu
European University Institute, Florence and Goethe University Frankfurt
email: matei.demetrescu@eui.eu

Helmut Lütkepohl
European University Institute
email: helmut.luetkepohl@eui.eu

Pentti Saikkonen
University of Helsinki
email: saikkone@valt.helsinki.fi

Abstract. When applying Johansen's procedure for determining the cointegrating rank to systems of variables with linear deterministic trends, there are two possible tests to choose from. One test allows for a trend in the cointegration relations and the other one restricts the trend to be orthogonal to the cointegration relations. The first test is known to have reduced power relative to the second one if there is in fact no trend in the cointegration relations, whereas the second one is based on a misspecified model if the linear trend is not orthogonal to the cointegration relations. Hence, the treatment of the linear trend term is crucial for the outcome of the rank determination procedure. We compare two alternative testing strategies which are applicable if there is uncertainty regarding the proper trend specification. In the first one a specific cointegrating rank is rejected if one of the two tests rejects and in the second one the trend term is decided upon by a pretest. The first strategy is shown to be preferable in applied work.

Key Words: Cointegration analysis, likelihood ratio test, vector autoregressive model, vector error correction model
JEL classification: C32

[1] The research for this paper was carried out while the first author was a Max Weber Fellow and the third author was a Fernand Braudel Fellow at the European University Institute in Florence.

1 Introduction

In a vector autoregressive (VAR) analysis with integrated variables determining the cointegrating rank is central for setting up a well specified model. The most popular method used for this purpose is the Johansen (1995) sequence of cointegrating rank tests which are based on the likelihood ratio (LR) principle. It is well-known that the asymptotic distributions of these tests depend on the deterministic term which is present in the data generation process (DGP). Moreover, it is also known that the power of the tests depends on the deterministic term allowed for in the model. More precisely, if the deterministic term is over-specified the power may suffer substantially (Doornik, Hendry and Nielsen (1998), Saikkonen and Lütkepohl (1999, 2000)); if a linear trend is allowed for while a constant is sufficient to capture the data properties, a more powerful version of the cointegrating rank test can be obtained by allowing only for a constant and no linear trend. Johansen (1995) also proposes tests that can help in choosing the deterministic term. Hence, given a null hypothesis of a specific cointegrating rank, one may test for the deterministic term first and then use the cointegrating rank test with the deterministic term suggested by the pretest. Pretesting has in fact been reported in the literature, e.g., by Crowder and Hoffman (1996) and Peytrignet and Stahel (1998).

On the other hand, practitioners often proceed in a different way if there is uncertainty regarding the deterministic term. They perform tests based on models with different possible deterministic terms and then decide on the cointegrating rank in some way taking into account all the test results (e.g., Hubrich (2001)). In this study we will formalize this procedure and compare it to the aforementioned pretest procedure.

In this context, the three most popular model versions in applied work are: (i) a model for variables without linear trend, (ii) a model where at least one of the variables has a linear trend but the cointegration relations are trend free and (iii) a model with a general linear trend which may also be part of the cointegration relations. In practice many economic variables are known to have a deterministic time trend. Moreover, it can be checked by univariate tests whether some of the variables are well modelled by including a linear trend. If so, the choice between (ii) and (iii) becomes relevant. For the practitioner the main problem in the multivariate case becomes thus the choice between (ii) and (iii). Hence, we will focus on this case in the following.

Clearly, focussing on a decision between (ii) and (iii) assumes that some pretesting has been done on the univariate series. Such pretesting gives rise to additional questions regarding the properties of the overall procedure. These questions are quite delicate and challenging as is known from Harvey, Leybourne and Taylor (2006), for example. Still the problem of choosing between a model with a trend in the cointegration relations and one with a linear trend which is orthogonal to the cointegration relations is a relevant one and this is the subject of the present paper.

It appears that many applied economists have a preference for (ii) based on a priori grounds. If a cointegration relation is interpreted as an equilibrium relation, a linear trend in that relation may not be very plausible. For example, Lettau and Ludvigson (2001), Coenen and Vega (2001), Ericsson and Sharma (1998), Funke and Rahn (2005), Stephan (2006) and Ribba (2006) apply cointegration tests which allow for a linear trend in the variables but not in the cointegration relations in various contexts. On the other hand, Hubrich (2001) applies both tests with and without allowing for a linear trend in the cointegration relations and she checks the robustness of her results. Crowder and Hoffman (1996) perform a test for the correct trend specification and, based on its outcome, eliminate the trend from the cointegration relations.

It will be shown in this paper that a testing sequence for the cointegrating rank of a VAR process based on the LR test which assumes a trend orthogonal to the cointegration relations is asymptotically likely to end up with a cointegrating rank smaller than the true one if the linear trend is in fact also in the cointegration relations. This result suggests that in applied work, if there is uncertainty with respect to the correct trend specification, one may perform both tests, with and without trend in the cointegration relations, and reject a given cointegrating rank if one of the tests rejects. This procedure will be shown to work well relative to a procedure based on pretesting for the correct trend specification. A corresponding result for unit root tests was obtained by Harvey et al. (2006).

The structure of this study is as follows: In the next section the general model setup is presented. In Section 3 the procedures for determining the cointegrating rank are discussed and in Section 4 the results of a small sample comparison of these procedures are reported. Section 5 concludes. Finally, the Appendix contains the derivation of the limiting distribution of the cointegrating rank test applied to a misspecified model which does not allow for a linear time trend in the cointegration relations although there is one.

Throughout the paper we use the following abbreviations: ML for maximum likelihood, LR for likelihood ratio, DGP for data generation process, VAR for vector autoregressive and VECM for vector error correction model. Moreover, the differencing operator is signified by Δ, that is, for a stochastic process x_t, $\Delta x_t = x_t - x_{t-1}$. A stationary (short memory) or asymptotically stationary process will sometimes be referred to as an $I(0)$ process and a nonstationary process which becomes stationary after differencing once is called $I(1)$ process. A normal (Gaussian) distribution with mean μ and variance (covariance matrix) Σ is denoted by $\mathcal{N}(\mu, \Sigma)$. Furthermore, $\mathbb{R}$ stands for the set of real numbers. For a matrix A, $\text{rk}(A)$ denotes its rank and $A_\perp$ denotes an orthogonal complement.

2 The Model Setup

We consider a K-dimensional system of $I(1)$ variables $y_t = (y_{1t}, \dots, y_{Kt})'$ with deterministic term μ_t such that

$$y_t = \mu_t + x_t, \tag{1}$$

where $\mu_t = \mu_0 + \mu_1 t$ is a K-dimensional linear trend term and x_t is a K-dimensional zero mean VAR(p) process with VECM representation

$$\Delta x_t = \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{p-1} \Delta x_{t-p+1} + u_t. \tag{2}$$

The $(K \times K)$ matrix Π is assumed to have rank r which is the cointegrating rank of x_t and, hence, of y_t. The Γ_j's $(j = 1, \dots, p-1)$ are $(K \times K)$ coefficient matrices and the error term u_t is an independently, identically distributed white noise process with zero mean and nonsingular covariance matrix $E(u_t u_t') = \Sigma_u$. For simplicity we also assume that u_t is Gaussian. Thereby our tests are proper LR tests. For our arguments this assumption is not essential and our results hold under more general assumptions, as usual. In fact, our results are valid whenever the cointegrating rank tests to be discussed in the following have their usual asymptotic properties.

For the deterministic term we consider the following alternative possibilities:

1. $\mu_1 \neq 0$ and $\Pi\mu_1 = 0$, that is, there is a trend in the variables which is, however, orthogonal to the cointegration relations.

2. $\mu_1 \neq 0$ and $\Pi\mu_1 \neq 0$, that is, the trend is fully general and, hence, it is also part of the cointegration relations.

Notice that $\text{rk}(\Pi) = r$ implies that $\Pi = \alpha\beta'$ for suitable $(K \times r)$ matrices α and β of rank r and $\beta' y_t$ represent the cointegration relations. Hence, $\Pi\mu_1 = 0$ is equivalent to $\beta'\mu_1 = 0$ which shows that $\Pi\mu_1 = 0$ is just another way of stating that the linear trend is orthogonal to the cointegration relations. For both linear trend specifications we can write the generation process of the observed variables y_t in VECM form as

$$\Delta y_t = \nu + \Pi^{(i)} y_{t-1}^{(i)} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \tag{3}$$

where ν is an intercept term and the superscript i refers to the two cases of deterministic terms. Hence,

$$\Pi^{(i)} = \begin{cases} \Pi \ (K \times K), & \text{for } i = 1, \\ \Pi^* \ (K \times (K+1)), & \text{for } i = 2. \end{cases} \tag{4}$$

Here the first K columns of Π^* are equal to Π. Accordingly,

$$y_{t-1}^{(i)} = \begin{cases} y_{t-1}, & \text{for } i = 1, \\ (y_{t-1}', t-1)', & \text{for } i = 2 \end{cases} \tag{5}$$

(see, e.g., Lütkepohl (2005, Section 6.4) for details).

For given cointegrating rank r, the relevant model can be estimated by Johansen's reduced rank regression method in both cases. Under our Gaussian assumptions this method delivers ML estimators. Since the cointegrating rank r is usually unknown, testing procedures for determining r will be discussed in the next section.

3 Testing for the Cointegrating Rank

In the context of the model setup presented in the previous section, we are interested in finding the cointegrating rank r. This quantity is typically chosen by testing a sequence of hypotheses

$$H_0(r_0) : \mathrm{rk}(\Pi) = r_0 \quad \text{versus} \quad H_1(r_0) : \mathrm{rk}(\Pi) > r_0 \tag{6}$$

for $r_0 = 0, 1, \ldots, K-1$. The first rank r_0 for which the null hypothesis cannot be rejected is then chosen as an estimate for r. Alternatively one may consider tests of $H_0 : \mathrm{rk}(\Pi) = r_0$ versus $H_1 : \mathrm{rk}(\Pi) = r_0 + 1$. This choice would result in a completely analogous discussion and is therefore not treated here in order to save space.

Because Gaussian ML estimation is straightforward, LR tests can readily be used for testing (6) (Johansen (1995)). In the following we will denote by $\mathrm{LR}(r_0)$ the LR statistic based on a model with intercept only,

$$\Delta y_t = \nu + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \tag{7}$$

and we use $\mathrm{LR}^*(r_0)$ for the LR statistic based on the model with linear trend term in the cointegration relations. Using this notation, the asymptotic null distributions and the asymptotic distributions under local alternatives are known for both test statistics if the deterministic term is specified properly (see Johansen (1995) and Saikkonen and Lütkepohl (1999, 2000)).

Since we are interested in analyzing the properties of $\mathrm{LR}(r_0)$ more closely, it is useful to provide a more explicit expression of this statistic. Let R_0 and R_1 be the residuals of a regression of Δy_t and y_{t-1}, respectively, on $1, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$ and define $\mathcal{S}_{ij} = T^{-1} R_i R_j$, $i = 0, 1$. Moreover, let $\lambda_1 \geq \cdots \geq \lambda_K \geq 0$ be the ordered eigenvalues of the matrix $\mathcal{S}_{01}\mathcal{S}_{11}^{-1}\mathcal{S}_{10}\mathcal{S}_{00}^{-1}$.

Then

$$\text{LR}(r_0) = -T \sum_{k=r_0+1}^{K} \log(1-\lambda_k). \tag{8}$$

In other words, the test statistic is made up of the $K-r_0$ smallest eigenvalues of $\mathcal{S}_{01}\mathcal{S}_{11}^{-1}\mathcal{S}_{10}\mathcal{S}_{00}^{-1}$. If the true cointegrating rank $r = r_0$, the matrix converges in probability to a matrix with rank r_0 which has $K - r_0$ zero eigenvalues, as the sample size $T \to \infty$. Hence, the limiting values of the $K - r_0$ smallest eigenvalues are zero. If the true cointegrating rank is greater than r_0, at least one of the eigenvalues (λ_{r_0+1}) in the test statistic in (8) will be nonzero asymptotically and, hence, $-T\log(1-\lambda_{r_0+1})$ as well as $\text{LR}(r_0)$ diverge to infinity as the sample size gets large. Thereby the test is consistent. The following proposition shows that the number of zero eigenvalues increases by one if the true DGP contains a linear trend in the cointegration relations which is not accounted for in $\text{LR}(r_0)$. This result will be useful in motivating one of the test procedures for the cointegrating rank when the actual trending properties are unknown.

Proposition 1.
If $r = \text{rk}(\Pi) > 0$ and $\Pi\mu_1 \neq 0$, then $\mathcal{S}_{01}\mathcal{S}_{11}^{-1}\mathcal{S}_{10}\mathcal{S}_{00}^{-1}$ converges in probability to a matrix with exactly $K - r + 1$ zero eigenvalues, as $T \to \infty$. □

In the Appendix we will derive the limiting distribution of $\text{LR}(r)$ under the conditions of Proposition 1, that is, for the case where the rank test is applied to a model with misspecified trend term. As a byproduct we will also prove Proposition 1. Unfortunately, under the conditions of the proposition, the limiting distribution of $\text{LR}(r)$ depends in a complicated way on nuisance parameters and is therefore not directly useful for devising a rank test. The derivation of the limiting distribution of $\text{LR}(r)$ is based on writing the DGP as

$$\begin{aligned}
\Delta y_t &= \nu + \alpha\beta'(y_{t-1} - \mu_1(t-1)) + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t \\
&= \nu + \alpha_2\beta_2'\mu_0 + \alpha_1\beta_1' y_{t-1} + \alpha_2\beta_2'(y_{t-1} - \mu_0 - \mu_1(t-1)) \\
&\quad + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t \\
&= \nu_* + \alpha_1\beta_1' y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + e_t,
\end{aligned} \tag{9}$$

where the cointegration matrix β $(K \times r)$ is chosen to have orthogonal columns and such that $\beta = [\beta_1 : \beta_2]$, where β_1 $(K \times (r-1))$ and β_2 $(K \times 1)$ have the properties $\beta_1'\mu_1 = 0$ and $\beta_2'\mu_1 \neq 0$, $\nu_* = \nu + \alpha_2\beta_2'\mu_0$, and $e_t = u_t + \alpha_2\beta_2' x_{t-1}$. The representation in (9) suggests that the test procedure tries to test the null hypothesis that there are $r-1$ stationary linear combinations of y_t given by $\beta_1' y_{t-1}$ and $K-r+1$ nonstationary linear combinations of which one, $\beta_2' y_t$, is trend stationary and the others, $\beta_\perp' y_t$, are $I(1)$. The main reason why the limiting distribution of the test becomes complicated is that the error term of the relevant model, e_t, is autocorrelated (although stationary). Consequently, the resulting limiting distribution suffers from problems similar to those previously encountered in unit root tests with autocorrelated errors (see, e.g., Phillips (1987) or Phillips and Perron (1988)). In particular, the limiting distribution involves 'second order bias' terms and complications resulting from the fact that the covariance matrix of the error term differs from the long run covariance matrix. Although the residual autocorrelation may be taken care of if data dependent lag order selection procedures are used, as is often the case in applied work, this will not fully eliminate the dependence of the limiting distribution on nuisance parameters because the lagged differences of y_t cannot fully capture the autocorrelation in $\alpha_2\beta_2' x_{t-1}$.

Proposition 1 implies, however, that a test based on a model with misspecified (or better under-specified) deterministic term is likely to terminate a testing sequence for the cointegrating rank too early and, hence, chooses the rank too small because, even for large T there is a positive probability for not rejecting a rank $r_0 = r-1$. Given this result, the procedure used by many practitioners may not be implausible when they do not know the precise deterministic term. They perform tests for both alternative trend specifications and reject a cointegrating rank if one of the tests rejects. If $\mathrm{LR}(r_0)$ is applied although there is a trend in the cointegration relations, then the test tends to terminate too early whereas in this case $\mathrm{LR}^*(r_0)$ will find the true cointegrating rank, r, or even overestimate r at least asymptotically because a test based on $\mathrm{LR}^*(r_0)$ is also consistent. On the other hand, if there is no trend in the cointegration relations, $\mathrm{LR}^*(r_0)$ will have reduced power and will hence have a tendency to choose too small a cointegrating rank while in this case $\mathrm{LR}(r_0)$ has its usual properties and, in particular, the associated test is consistent so that it will reject all cointegrating ranks below the true one at least asymptotically.

The procedure which decides on the basis of the outcome of both tests can be compared formally to a pretest procedure which also tests the deterministic term. As mentioned earlier, pretesting is, for instance, reported by Crowder and Hoffman (1996) and Peytrignet and Stahel (1998). Thus, the following two procedures for choosing an estimate $\hat{r}$ of the true cointegrating rank r will be considered in the following.

Procedure 1: For a given r_0, starting with $r_0 = 0$, use both $\mathrm{LR}(r_0)$ and $\mathrm{LR}^*(r_0)$ to test $H_0(r_0)$. Choose $\hat{r} = r_0$ if none of the tests rejects. Otherwise proceed to testing r_0+1 etc. until a given rank is not rejected by both tests. □

Procedure 2: Choose $\hat{r} = 0$ if none of the tests rejects $H_0(0)$. Otherwise proceed with $r_0 = 1$. For a given $r_0 > 0$, test $H_0 : \Pi\mu_1 = 0$ versus $H_1 : \Pi\mu_1 \neq 0$. If H_0 is not rejected, use $\mathrm{LR}(r_0)$ to test $H_0(r_0)$. If H_0 is rejected, use $\mathrm{LR}^*(r_0)$. Choose $\hat{r} = r_0$ if the appropriate test does not reject $H_0(r_0)$. Otherwise proceed to rank $r_0 + 1$ etc. until a given rank is not rejected. □

If $r_0 = 0$, a pretest is not possible in Procedure 2 because there are no cointegration relations under the null hypothesis. Still $\mathrm{LR}(0)$ and $\mathrm{LR}^*(0)$ differ because they are based on different models. The null hypothesis $r_0 = 0$ is rejected if one of the tests rejects, as in Procedure 1. Thus, the two procedures differ only for $r_0 > 0$.

In the pretest procedure the null hypothesis $H_0 : \Pi\mu_1 = 0$ can be checked by an LR test (e.g., Johansen (1995)). Proposition 1 suggests that this procedure may have reduced power because a pretest may not reject $H_0 : \Pi\mu_1 = 0$ even if the trend is not orthogonal to the cointegration relations. In that case $\mathrm{LR}(r_0)$ is used which may then have low power. In the next section we will report the results of a Monte Carlo study to explore the small sample properties of the two aforementioned procedures for choosing the cointegrating rank.

4 Simulation Study

In this section we investigate the empirical small sample properties of the two tests and the procedures for choosing the cointegrating rank if there is uncertainty about the correct trend specification. We will consider both types of DGPs with and without trend in the cointegration relations. All simulations are done with R programs.

4.1 Monte Carlo Setup

Time series from DGPs with linear trend in the cointegration relations are generated as

$$y_t = \mu_0 + \mu_1 t + x_t \qquad t = 1, \dots, T, \tag{10}$$

with

$$\mu_0 = 0 \quad \text{and} \quad \mu_1 = c\boldsymbol{\iota}_K, \quad c = 0.1, 0.5,$$

where $\boldsymbol{\iota}_K$ is a $(K \times 1)$ vector of ones, and

$$x_t = \begin{bmatrix} \psi I_r & 0 \\ 0 & I_{K-r} \end{bmatrix} x_{t-1} + u_t, \quad x_0 = 0, \quad u_t = \varphi u_{t-1} + \varepsilon_t. \tag{11}$$

Moreover,

$$\varepsilon_t \sim \mathcal{N}\left(0, \Sigma_u = \begin{bmatrix} I_r & \Theta \\ \Theta' & I_{K-r} \end{bmatrix}\right) \tag{12}$$

is Gaussian white noise. Here the parameter $|\psi| < 1$ and Θ is an $(r \times (K-r))$ matrix. The error process u_t is a VAR(1) with scalar parameter φ, $|\varphi| < 1$, for which we have used different values. Equivalently, we could have written x_t as a VAR(2) process. In (11) this process is expressed such that the unit root and short-term properties are easy to disentangle. For $\varphi = 0$, x_t is a VAR(1), of course. This type of VAR(1) process was also used by Toda (1994) and subsequently in a number of other simulation studies where properties of cointegrating rank tests were explored (see, e.g., Hubrich, Lütkepohl and Saikkonen (2001)). Toda argues that this process is useful for investigating the properties of LR tests for the cointegrating rank because other VAR(1) processes can be obtained from it by linear transformations which leave the tests invariant. Thus, this process allows us to explore the properties of the tests for a wide range of DGPs. We have also used VAR(2) processes because the short-term dynamics play a role in the asymptotic distribution of the LR(r_0) tests if the trend is under-specified. In the following results for three- and five-dimensional DGPs will be presented.

Time series from DGPs with trend orthogonal to the cointegration relations will be generated as

$$y_t = c \begin{bmatrix} 0 \\ \boldsymbol{\iota}_{K-r} \end{bmatrix} + \begin{bmatrix} \psi I_r & 0 \\ 0 & I_{K-r} \end{bmatrix} y_{t-1} + u_t, \quad y_0 = 0, \tag{13}$$

with u_t as in (11). For both types of DGPs we generated 50 presample values to reduce the effects of initial values.

For cointegrating rank r_0, the LR statistic for testing $H_0 : \Pi\mu_1 = 0$ is easy to compute as

$$\mathrm{LR} = T \sum_{k=1}^{r_0} \log[(1-\lambda_k)/(1-\lambda_k^*)],$$

where the λ_k are the eigenvalues based on the model (7) without linear trend term in the cointegration relations and the λ_k^* are the corresponding eigenvalues from a reduced rank regression of a model with such a term. The test statistic has a standard χ^2 limiting distribution with r_0 degrees of freedom if the null hypothesis holds.

4.2 Monte Carlo Results

We have generated three- and five-dimensional time series from the two types of DGPs specified in (10)-(12) and (13) with a range of different values for the parameters ψ, c, φ, Θ and cointegrating rank r. We have also used different sample sizes and data-driven VAR order selection based on AIC in some of our simulations. A small selection of results is presented in Tables 1 - 8. The nominal significance level for all tests is 5% because this is the leading case considered in practice. In a range of simulations we have used $\varphi = 0$ and in that case the VAR order $p = 1$ is assumed to be known. Although in practice the VAR order is typically unknown we have also generated results for known VAR order to separate the two problems of choosing the cointegrating rank and the VAR order. It is not obvious, however, that the VAR order should be fixed at the true order of the DGP if the deterministic trend is under-specified because the misspecification may result in autocorrelated errors, as we have shown in Section 3. Therefore it is of interest to compare the known VAR order case with results based on a data-driven VAR order selection.

In Table 1 results for three-dimensional VAR(1) processes ($\varphi = 0$) with linear trend in the cointegration relations are presented. Here the VAR order is fixed at $p = 1$. In this case the trend is under-specified in the LR(r_0) tests. Clearly this case requires that the true cointegrating rank r is at least one because otherwise there cannot be a trend in the cointegration relations. In Table 1 rejection frequencies for the two tests and the two selection procedures (under the headings Proc 1 and Proc 2) are shown. Notice that the test results are not conditioned on the outcome of the tests for a smaller cointegrating rank. In other words, in Table 1 we are not considering the properties of the testing sequence but those of individual tests.

Testing $H_0(0)$, i.e., $r_0 = 0$, gives an idea about the power of the different tests and procedures. As mentioned in Section 3, for this null hypothesis the two selection procedures are identical. If the true cointegrating rank is $r = 1$ the LR(0) test has very little power relative to LR*(0) and a similar result is obtained for LR(1) and LR*(1) if $r = 2$. This shows the effects of under-specifying the trend on the performance of a test based on LR(r_0).

In Table 2 the relative frequencies of the different cointegrating ranks obtained with the two procedures are presented. In that table it is clearly seen that following Procedure 1 leads to a substantially higher success rate than Procedure 2 if the true cointegrating rank is $r = 2$ and $\Theta = 0$. As expected, both procedures lead to similar results if the true cointegrating rank $r = 1$. Note that in this case there is no pretest involved in the only test for which power is needed, i.e., the test of $r_0 = 0$. Thus, for a three-dimensional process the only case where a clear advantage of any of the procedures can be expected is the $r = 2$ case and here Procedure 1 has a clear lead if $\Theta = 0$. Its relative advantage is reduced, however, if our choice of nonzero Θ matrix is used. In other words, there are processes where the performance of both procedures is similar or Procedure 2 even has a small lead. On the other hand, we have not found a single case where the advantage of Procedure 2 over Procedure 1 was of the same magnitude as the lead of Procedure 1 over Procedure 2 for, e.g., the case $r = 2$, $\Theta = 0$. Notice also that Procedure 1 does not suggest a cointegrating rank in excess of the true one much more often than Procedure 2 in any of the cases and, hence, the actual levels of both procedures are in fact comparable.

In fact, the rejection frequencies presented in Table 1 for Procedure 1 are not substantially greater than the nominal significance level of 5%, if $\Theta = 0$ and the true cointegrating rank is tested. This is especially true when the trend slope is large ($c = 0.5$). In that case the level of Procedure 1 is not much higher than that of Procedure 2. Of course, for true cointegrating rank $r = 1$, one would not expect great differences between the two procedures because for the null hypothesis $H_0(0)$ they are identical by definition and for $H_0(1)$ we can only learn about the actual size properties of the two procedures. Under the rather ideal conditions of our Monte Carlo setup, one would hope for a rejection frequency close to 5% if the null hypothesis is true. In any case, looking at the actual ranks selected in Table 2 it is clear that Procedure 1 selects the true cointegrating rank much more often than Procedure 2 in some cases.

There is one slight problem, however, when a three-dimensional process is considered. In applied work a practitioner may not consider testing $H_0(2)$ in this situation if s/he believes that the trend is orthogonal to the cointegration relations because this would be incompatible with the alternative hypothesis $\text{rk}(\Pi) = K$, that is, the process is stationary under the alternative hypothesis and cannot have a linear trend in model (7) with intercept only which would be in contradiction to the assumption that there is a linear trend in the variables. Therefore it makes sense to check how well our procedures do for higher-dimensional processes for which they may be even more relevant and closer to what practitioners really do in empirical studies.

In Table 3 the relative frequencies of ranks selected by the two procedures for five-dimensional VAR(1) processes are presented. The VAR order is again assumed to be known and fixed at $p = 1$. Again we show results for sample size $T = 100$ although that may be regarded as fairly low for a five-dimensional system. We use this sample size because it is not uncommon in applied work and the general results did not change much for substantially larger sample sizes (e.g., $T = 250$). It can be seen in Table 3 that the differences between the two procedures are not great if the true cointegrating rank is $r = 1$ or if processes with $\Theta \neq 0$ are considered. On the other hand, substantially better results are obtained with Procedure 1 if $\Theta = 0$ and the true cointegrating rank is greater than one. In fact, the true rank may be chosen more than 50% more often by Procedure 1 than by Procedure 2 (see, e.g., the case where $r = 2$ and $c = 0.5$).

Thus, from the results presented so far it is clear that if there is indeed a trend in the cointegrating relations, a researcher who uses Procedure 1 is on the safe side. S/he never looses much relative to the pretest procedure (Procedure 2) and may choose the true cointegrating rank substantially more often for some of the DGPs considered. This general impression was reinforced by experiments with other DGPs and sample sizes even when the VAR order was chosen by AIC. This result was also obtained with ψ's closer to one. In that case, depending on the actual value of ψ, the true cointegrating rank may be underestimated considerably by both procedures, however, in particular for relatively high dimensional processes and sample size $T = 100$. Only for substantially larger sample sizes, e.g., $T = 250$, can the true cointegrating rank be expected to be found with high probability.

To illustrate this point we present results for a more difficult case in Table 4. The DGP underlying that table is a five-dimensional VAR(2) of the type (10)-(12) with $\varphi = -0.8$ and $\psi = 0.8$. The sample size is again $T = 100$ and the VAR order is now chosen by AIC using a maximum order of four.[2] Order selection is based on a VAR model in levels with an intercept term because this appears to be a common approach in practice. We have also used VAR order selection based on trend adjusted data in other experiments and found

[2]Generally we have chosen the maximum VAR order as the integer part of $4(T/100)^{1/4}$ as recommended in some of the related literature (e.g., Schwert (1989), Demetrescu, Kuzin and Hassler (2008)). This choice leads, for example, to $p_{\max} = 4$ for $T = 100$ and $p_{\max} = 5$ for $T = 250$.

qualitatively similar results. In Table 4, if $r > 1$, Procedure 1 still finds the true rank much more often than Procedure 2. Both procedures are not very successful in this respect, however. For example, for $r = 3$ and $\Theta = 0$, Procedure 1 finds the correct rank only in about 2% of the replications while Procedure 2 performs even worse and ends up with only about 1% correct choices. The message of these results is clearly that the probability of finding the correct order with so little sample information in such a difficult situation is very small. Procedure 1 at least tends to get closer to the true rank.

In Table 5 results for the same DGPs but with sample size $T = 250$ are reported and a substantial improvement regarding the correct choice of the cointegrating rank can be noticed. Still, one may regard success rates of 50% as low. These rates are, e.g., obtained by Procedure 1 if $r = 3$ and $\Theta = 0$. Obviously, Procedure 2 is considerably less successful in this respect for $r > 1$. Thus, our results suggest that Procedure 1 has a particularly great advantage in difficult situations where sample information is scarce.

Of course, the question arises how the tests behave for processes which in fact have no trend in the cointegration relations. This question is considered next by analyzing results obtained for DGP (13). Some results based on three-dimensional versions of DGP (13) are presented in Tables 6 and 7. Now both tests are in principle applicable and should have their usual asymptotic null distributions because both of them are based on properly specified models under the present conditions. Despite this fact, the LR(r_0) test rejects far too often in some cases, although it is designed especially for this situation. For example, if the trend is not very pronounced ($c = 0.1$), it rejects the true null hypothesis $H_0(2)$ in more than 25% of the cases, that is, its actual level can be more than 25% when the nominal level is 5%. Of course, one may argue that in practice, for a three-dimensional process, one would not test $H_0(2)$ with the LR test because it leads to a contradiction under the alternative, as argued earlier. On the other hand, even for $r = 1$ the test rejects a true null hypothesis in about 10% of the cases when the nominal significance level is 5%. Thus, even under ideal conditions it overrejects considerably. This property is also reflected in the rejection frequencies of Procedures 1 and 2 in Table 6.

Looking at the frequencies of ranks chosen in Table 7, it is seen that there is not much difference between the two procedures in any of the cases although Procedure 2 finds the true rank slightly more often in most situations. We have also considered five-dimensional DGPs and show some results in Table 8. Here again none of the two procedures has a substantial lead over the other in any of the cases shown. In fact, now there are cases were Procedure 1 is more successful in finding the true cointegrating rank than Procedure 2 and there are also cases where the reverse is true. But in any case the differences are not large. Given the substantial overrejection of LR(r_0) in some cases, it is also not surprising that there is some chance to overestimate the cointegrating rank if it is in fact small. Thus, the overall conclusion from looking at processes with trend in the variables but not in the cointegration relations is that using either one of the procedures does not result in substantial gains or losses relative to the other one.

This conclusion was also confirmed with other DGPs in difficult situations where VAR(2) processes ($\varphi \neq 0$) and data-dependent order selection were used. In these situations both procedures have a tendency in small samples to overestimate small cointegrating ranks and underestimate large ones. Clearly this reflects the tendency of LR(r_0) to reject the null hypothesis too often in some situations on the one hand, while on the other hand, the power may be quite low in difficult situations when the sample size is small. Overall the results for these cases are qualitatively similar to those shown in Figures 6-8 and are therefore not presented here to save space.

Summarizing the results from all of the experiments, the overall conclusion is that there are DGPs for which Procedure 1, which is based on the outcome of both tests, finds the true cointegrating rank much more often than the pretest procedure (Procedure 2) whereas in other cases both procedures perform in a very similar way. Thus, a practitioner who has based the decision on the cointegrating rank on the outcome of both tests may in fact have done the right thing. In some cases a better decision might have been possible by applying a pretest procedure, however.

5 Conclusions

In this study we have compared two procedures for choosing the cointegrating rank of a VECM when the variables have a deterministic linear time trend of unknown form. In that case there is a choice of two LR tests for the cointegrating rank, the first one allows for a trend not only in the variables but also in the cointegration relations, whereas the second one assumes that the linear trend is orthogonal to the cointegration relations. If there is no linear trend in the cointegration relations (i.e., the linear trend is orthogonal to the cointegration relations), then the second test is preferable because it may be substantially more powerful than the first one. We have derived the asymptotic distribution of the second test if there is actually a linear trend in the cointegration relations and, hence, the test is based on a misspecified model. Unfortunately, in this case the limiting distribution depends in a complicated way on nuisance parameters. It turns out, however, that if the deterministic trend term is under-specified the test tends to be conservative.

Taking into account the asymptotic properties of the tests, two promising procedures for choosing the cointegrating rank of a VAR process are (1) to apply both tests and reject any rank for which one of the tests rejects the null hypothesis and (2) to perform a pretest for the deterministic trend and choose the test for the cointegrating rank on the basis of the outcome of the pretest. Although it is not always fully clear how practitioners actually choose their tests, both possibilities appear to have been used in the literature. Given our theoretical results regarding the properties of the test which ignores a trend in the cointegration relations and, hence, may be applied to a misspecified model, both procedures have an asymptotic justification.

We have performed a Monte Carlo study to investigate the small sample properties of the two procedures. In our simulations the first procedure which is based on the outcome of both tests is overall preferable. It tends to find the true cointegrating rank much more often than the pretest procedure for some of the processes we have considered. Moreover, in those cases where the pretest procedure dominates, it usually has only a small lead over the first procedure. Therefore, based on our simulation results, the first procedure can be recommended. Unfortunately, the LR tests for the cointegrating rank are known to have poor power for processes with large dimension and/or order. Therefore both procedures may not find the true cointegrating rank very often in extreme situations, which do arise in practice, however. Without a reasonably large sample size, finding the true cointegrating rank of a large VAR process cannot be expected. Considering also lower-dimensional subsystems and building up a higher-dimensional model by taking into account the cointegration relations from the lower-dimensional analysis may be worthwhile in this case. This type of specific-to-general specification procedure may in fact be a good strategy more generally when cointegrated variables are considered (e.g., Lütkepohl (2007)).

One possible direction for future research may be to develop a procedure and the related theory which allows applied researchers to decide whether or not a linear trend in the variables should be considered. Recall that we have assumed that a linear deterministic trend is known to be present in at least some of the variables. Although the trending properties of individual variables can be explored with univariate methods, knowing the properties of the overall procedure would be of interest.

Appendix: The Limiting Distribution of LR(r)

We use the notation and model setup of Section 2. Moreover, the space of right-continuous functions on the interval $[0, 1]$ which have left limits is denoted by $D[0, 1]$ and weak convergence on $D[0, 1]$ with respect to the uniform topology is denoted by $\xrightarrow{w}$. Convergence in probability is signified by $\xrightarrow{p}$. Furthermore, $o_p(\cdot)$ and $O_p(\cdot)$ are the usual symbols for stochastic sequences which converge to zero or are bounded, respectively.

The test statistic LR(r_0) is made up of the ordered eigenvalues $\lambda_1 \geq \cdots \geq \lambda_K \geq 0$ of the matrix $\mathcal{S}_{01}\mathcal{S}_{11}^{-1}\mathcal{S}_{10}\mathcal{S}_{00}^{-1}$. In what follows, $r_0 = r$ will be assumed. As in Johansen (1995) these eigenvalues can alternatively be computed as solutions to the determinantal equation

$$\det(\mathcal{S}(\lambda)) = \det(\lambda\mathcal{S}_{11} - \mathcal{S}_{10}\mathcal{S}_{00}^{-1}\mathcal{S}_{01}) = 0, \tag{A.1}$$

where $\mathcal{S}(\lambda)$ abbreviates $\lambda\mathcal{S}_{11} - \mathcal{S}_{10}\mathcal{S}_{00}^{-1}\mathcal{S}_{01}$.

To obtain the limiting distribution of test statistic LR(r) under the conditions of Proposition 1, we follow the pattern in Johansen (1995, p. 158-160) with appropriate modifications. First we have to transform the matrix $\mathcal{S}(\lambda)$ in a suitable manner. To this end, recall from equation (1) that $y_t = \mu_0 + \mu_1 t + x_t$, where x_t is a zero mean cointegrated VAR(p) process with cointegrating rank r. Let β $(K \times r)$ be a matrix of cointegrating vectors with orthogonal columns. Proposition 1 assumes that $\beta'\mu_1 \neq 0$ and postmultiplying β by a suitable orthogonal matrix, we can transform this matrix to the form $[\beta_1 : \beta_2]$ where β_1 $(K \times (r-1))$ and β_2 $(K \times 1)$ have the properties $\beta_1'\mu_1 = 0$ and $\beta_2'\mu_1 \neq 0$, which will henceforth be assumed. Define the matrix $\eta = [\beta_\perp : \beta_2]$ $(K \times (K-r+1))$. The columns of the matrix η are orthogonal and we can find a nonsingular matrix ξ such that $\eta\xi = [\gamma : \beta_2]$, where γ $(K \times (K-r))$ satisfies $\gamma'\mu_1 = 0$. The last column of ξ is a vector with last component unity and all other components zero and the first $K-r$ columns of ξ can be taken as $(\eta'\mu_1)_\perp$. Note that by construction the matrix $\gamma'\beta_\perp$ is nonsingular. One way to see this is to premultiply the identity $\eta\xi = [\gamma : \beta_2]$ by the orthogonal matrix $[\eta : \beta_1]' = [\beta_\perp : \beta_2 : \beta_1]'$ which, by the definitions, yields

$$\begin{bmatrix} (\eta'\eta)\xi \\ 0 \end{bmatrix} = \begin{bmatrix} \beta_\perp'\gamma & 0 \\ \beta_2'\gamma & \beta_2'\beta_2 \\ 0 & 0 \end{bmatrix},$$

where there are $r-1$ zero rows on both sides. Because the matrices $\eta'\eta$ and ξ are nonsingular the matrix on the left hand side has rank $K-r+1$. For the matrix on the right hand side to have the same rank, the rows of the matrix $\beta_\perp'\gamma$ must be linearly independent, implying the nonsingularity of $\gamma'\beta_\perp$.

Now consider weak convergence of the process $T^{-1/2}y_{[Ts]}$, $s \in [0,1]$, in the directions of the matrices γ and β_2. We use the notation $\bar{\gamma} = \gamma(\gamma'\gamma)^{-1}$ and similarly for any matrix of full column rank. Recall that, by Granger's representation theorem (e.g., Johansen (1995, Theorem 4.2)), the process x_t can be expressed as

$$x_t = C\sum_{j=1}^{t} u_j + \Phi(L)u_t + A, \tag{A.2}$$

where $C = \beta_\perp(\alpha'_\perp \Psi \beta_\perp)^{-1}\alpha'_\perp$ with $\Psi = I_K - \Gamma_1 - \cdots - \Gamma_{p-1}$, $\Phi(L) = \sum_{j=0}^{\infty} \Phi_j L^j$ with L the lag operator and the coefficient matrices Φ_j decaying to zero exponentially fast, and A depends on initial values and satisfies $\beta' A = 0$. Thus, because $\gamma'\mu_1 = 0$, equations (1) and (A.2) yield

$$T^{-1/2}\bar{\gamma}' y_{[Ts]} = T^{-1/2}\bar{\gamma}' C \sum_{j=1}^{[Ts]} u_j + o_p(1), \tag{A.3}$$

where the latter term on the right hand side is $o_p(1)$ in $D[0,1]$. Similarly, denoting $\tau = (\beta_2'\mu_1)^{-1}\beta_2$ and observing that $\tau'\mu_1 = 1$ and $\tau' C = 0$ we can write

$$T^{-1}\tau' y_{[Ts]} = \frac{[Ts]}{T} + o_p(1) \tag{A.4}$$

with the latter term on the right hand side again $o_p(1)$ in $D[0,1]$. These results can be justified in the same way as their counterparts in the proof of Lemma 10.2 of Johansen (1995) (or by using Theorem B.13 of the same reference). Note, however, that in the latter result the $o_p(1)$ term is only due to a stationary process whereas in Johansen's (1995) Lemma 10.2 it also contains a random walk component. In the same way as in that lemma we find that

$$T^{-1/2}\bar{\gamma}' y_{[Ts]} \xrightarrow{w} \bar{\gamma}' C W(s) \qquad \text{and} \qquad T^{-1}\tau' y_{[Ts]} \xrightarrow{w} s,$$

where $W(s)$ is a Brownian motion with covariance matrix Σ_u. Using the matrix $B_T = \left[\bar{\gamma} : T^{-1/2}\tau\right]$ these two results can be expressed as

$$T^{-1/2} B_T' y_{[Ts]} \xrightarrow{w} G_0(s) \stackrel{def}{=} \begin{bmatrix} \bar{\gamma}' C W(s) \\ s \end{bmatrix}$$

and the corresponding demeaned version can also be obtained as in Johansen's (1995) Lemma 10.2, giving

$$T^{-1/2} B_T'(y_{[Ts]} - \bar{y}) \xrightarrow{w} G(s) \stackrel{def}{=} G_0(s) - \overline{G}_0 = \begin{bmatrix} \bar{\gamma}' C(W(s) - \overline{W}) \\ s - \frac{1}{2} \end{bmatrix}, \tag{A.5}$$

where $\overline{G}_0 = \int_0^1 G_0(s)ds$ and analogously for $\overline{W}$.

Similarly to Johansen (1995, p. 158) we now introduce the transformation matrix $A_T = \left[\beta_1 : T^{-1/2}B_T\right]$ and transform the generalized eigenvalue problem (A.1) used to obtain test statistic LR(r). Instead of $\det(\mathcal{S}(\lambda))$ we can consider $\det(A_T'\mathcal{S}(\lambda)A_T)$ and its weak limit. To this end, we need some notation. As in Johansen (1995, p. 141) we define

$$\text{Cov}\left[\begin{array}{c} \Delta x_t \\ \beta' x_{t-1} \end{array}\middle| \Delta x_{t-1}, ..., \Delta x_{t-p+1}\right] = \left[\begin{array}{cc} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{array}\right].$$

By $\Sigma_{0\beta_1}$ $(K \times (r-1))$ we denote the submatrix of $\Sigma_{0\beta}$ $(K \times r)$ obtained by deleting the last column so that $\Sigma_{0\beta_1}$ is the conditional covariance matrix between Δx_t and $\beta_1' x_{t-1}$, given $\Delta x_{t-1}, ..., \Delta x_{t-p+1}$. Similarly, $\Sigma_{\beta\beta_1}$ $(r \times (r-1))$ is used for the matrix obtained by deleting the last column from $\Sigma_{\beta\beta}$ $(r \times r)$ and $\Sigma_{\beta_1\beta_1}$ $((r-1) \times (r-1))$ signifies the matrix obtained by deleting the last row and last column from $\Sigma_{\beta\beta}$. Properties of these matrices are given in the following lemma.

Lemma 1.

$$\Sigma_{0\beta_1} = \alpha\Sigma_{\beta\beta_1}, \tag{A.6}$$

$$\Sigma_{00} = \alpha\Sigma_{\beta\beta}\alpha' + \Sigma_u, \tag{A.7}$$

and

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta_1}(\Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1})^{-1}\Sigma_{\beta_1 0}\Sigma_{00}^{-1} = a(a'\Sigma_{00}a)^{-1}a', \tag{A.8}$$

where $a = (\alpha\Sigma_{\beta\beta_1})_\perp$ $(K \times (K-r+1))$.

Proof: Result (10.3) in Lemma 10.1 of Johansen (1995) shows that $\Sigma_{0\beta} = \alpha\Sigma_{\beta\beta}$ so that deleting the last columns from both sides gives (A.6). From result (10.4) of the same lemma we also get (A.7). Because the matrix $\Sigma_{\beta\beta}$ $(r \times r)$ is positive definite, the matrix $\Sigma_{\beta\beta_1}$ $(r \times (r-1))$ is of full column rank implying that $\Sigma_{0\beta_1} = \alpha\Sigma_{\beta\beta_1}$ is of full column rank (see (A.6)). Noting that $a_\perp = \Sigma_{0\beta_1}$, we can demonstrate the last result of Lemma 1 by multiplying (A.8) from the right by the matrices $a_\perp$ and $\Sigma_{00}a$. Multiplication by $a_\perp =$

$\Sigma_{0\beta_1}$ clearly yields zero on both sides whereas multiplication by $\Sigma_{00}a$ yields a on the right hand side and on the left hand side we get

$$a - \Sigma_{00}^{-1}(\Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1})^{-1}\Sigma_{\beta_1\beta}\alpha' a = a.$$

This proves Lemma 1. □

The following intermediate results are similar to those in Lemma 10.3 of Johansen (1995).

Lemma 2.
Under the conditions of Proposition 1,

$$\mathcal{S}_{00} \xrightarrow{p} \Sigma_{00}, \tag{A.9}$$

$$\beta_1 \mathcal{S}_{11}\beta_1 \xrightarrow{p} \Sigma_{\beta_1\beta_1}, \tag{A.10}$$

$$\beta_1' \mathcal{S}_{10} \xrightarrow{p} \Sigma_{\beta_1 0}, \tag{A.11}$$

$$T^{-1}B_T'\mathcal{S}_{11}B_T \xrightarrow{w} \int_0^1 GG'ds, \tag{A.12}$$

$$B_T'(\mathcal{S}_{10} - \mathcal{S}_{11}\beta\alpha') \xrightarrow{w} \int_0^1 GdW', \tag{A.13}$$

$$B_T'\mathcal{S}_{11}\beta_1 = O_p(1), \tag{A.14}$$

$$B_T'\mathcal{S}_{10} = O_p(1). \tag{A.15}$$

Proof: From (1) and the fact $\beta_1'\mu_1 = 0$ it follows that Δy_t and $\beta_1' y_t$ are jointly stationary and ergodic processes so that the first three results ((A.9) - (A.11)) can be justified by using the definitions and the law of large numbers in the same way as in the proof of Johansen's (1995) Lemma 10.3. The results

(A.12) and (A.14) are also obtained in the same way as their counterparts in Johansen's (1995) Lemma 10.3. Both make use of Johansen's Theorem B.13 and the former also of (A.5) and the continuous mapping theorem. In view of the same theorem and the expansions (A.3) and (A.4), (A.13) and (A.15) are also readily obtained. □

Now consider the determinant

$$\det(A_T'\mathcal{S}(\lambda)A_T) \xrightarrow{w} \det\left(\begin{bmatrix} \lambda\Sigma_{\beta_1\beta_1} & 0 \\ 0 & \lambda\int_0^1 GG'ds \end{bmatrix} - \begin{bmatrix} \Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1} & 0 \\ 0 & 0 \end{bmatrix}\right)$$

$$= \det(\lambda\Sigma_{\beta_1\beta_1} - \Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1})\det\left(\lambda\int_0^1 GG'ds\right).$$

The weak convergence can be justified by using the definitions and (A.9) - (A.12) (cf. (11.16) in Johansen (1995, p. 158)). Setting the limit equal to zero it is seen that there are $K - r + 1$ zero roots and $r - 1$ positive roots given by the solutions of

$$\det(\lambda\Sigma_{\beta_1\beta_1} - \Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1}) = 0.$$

Thus, the $r - 1$ largest roots of (A.1) converge weakly to the roots of this equation and the rest converge weakly to zero. This proves Proposition 1.

To derive an explicit expression for the limiting distribution of LR(r) we can now follow arguments entirely similar to those starting at the top of page 159 of Johansen (1995). First consider the decomposition

$$[\beta_1 : B_T]'\det(\mathcal{S}(\lambda))\,[\beta_1 : B_T]$$

$$= \det(\beta_1'\mathcal{S}(\lambda)\beta_1)\det(B_T'\{\mathcal{S}(\lambda) - \mathcal{S}(\lambda)\beta_1\,[\beta_1'\mathcal{S}(\lambda)\beta_1]^{-1}\,\beta_1'\mathcal{S}(\lambda)\}B_T) \quad \text{(A.16)}$$

and let $T \to \infty$ and $\lambda \to 0$ in such a way that $\rho = T\lambda$ is fixed. As in Johansen (1995, p. 159), in order to derive the asymptotic distribution of the ρ's it suffices to consider the second factor of the right hand side of (A.16). This follows from the fact that

$$\begin{aligned} \beta_1'\mathcal{S}(\lambda)\beta_1 &= \rho T^{-1}\beta_1'\mathcal{S}_{11}\beta_1 - \beta_1'\mathcal{S}_{10}\mathcal{S}_{00}^{-1}\mathcal{S}_{01}\beta_1 \\ &= -\Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1} + o_p(1), \end{aligned}$$

which can be concluded from (A.9) - (A.11) in Lemma 2. Clearly, the determinant of this term does not depend on ρ in the limit. Hence, to study the properties of the roots we have to consider the second factor in (A.16). For this factor, (A.9), (A.11), (A.14) and (A.15) yield

$$\begin{aligned} B_T'\mathcal{S}(\lambda)\beta_1 &= \rho T^{-1}B_T'\mathcal{S}_{11}\beta_1 - B_T'\mathcal{S}_{10}\mathcal{S}_{00}^{-1}\mathcal{S}_{01}\beta_1 \\ &= -B_T'\mathcal{S}_{10}\Sigma_{00}^{-1}\Sigma_{0\beta_1} + o_p(1) \end{aligned}$$

and

$$\begin{aligned} B_T\mathcal{S}(\lambda)B_T &= \rho T^{-1}B_T'\mathcal{S}_{11}B_T - B_T'\mathcal{S}_{10}\mathcal{S}_{00}^{-1}\mathcal{S}_{01}B_T \\ &= \rho T^{-1}B_T'\mathcal{S}_{11}B_T - B_T'\mathcal{S}_{10}\Sigma_{00}^{-1}\mathcal{S}_{01}B_T + o_p(1). \end{aligned}$$

Using these results we find that (cf. Johansen (1995, p. 159))

$$B_T'\{\mathcal{S}(\lambda) - \mathcal{S}(\lambda)\beta_1\left[\beta_1'\mathcal{S}(\lambda)\beta_1\right]^{-1}\beta_1'\mathcal{S}(\lambda)\}B_T$$

$$= \rho T^{-1}B_T'\mathcal{S}_{11}B_T - B_T'\mathcal{S}_{10}N_1\mathcal{S}_{01}B_T + o_p(1),$$

where

$$N_1 = \Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta_1}(\Sigma_{\beta_1 0}\Sigma_{00}^{-1}\Sigma_{0\beta_1})^{-1}\Sigma_{\beta_1 0}\Sigma_{00}^{-1}.$$

From (A.8) it is known that this matrix can be expressed as

$$N_1 = a(a'\Sigma_{00}a)^{-1}a',$$

where a $(K\times(K-r+1))$ is an orthogonal complement of $\alpha\Sigma_{\beta\beta_1}$ $(K\times(r-1))$. It is easy to see that $a = [\alpha_\perp : \bar{\alpha}\kappa]$, where κ is an orthogonal complement of $\Sigma_{\beta\beta_1}$ $(K\times 1)$.

Thus, we have reduced the problem to investigating the weak limit of the roots of

$$\det(\rho T^{-1}B_T'\mathcal{S}_{11}B_T - B_T'\mathcal{S}_{10}a(a'\Sigma_{00}a)^{-1}a'\mathcal{S}_{01}B_T) = 0. \tag{A.17}$$

First we use (A.12) to conclude that $T^{-1}B_T'\mathcal{S}_{11}B_T \xrightarrow{w} \int_0^1 GG'ds$. Next we consider the matrix

$$B_T'\mathcal{S}_{10}a = B_T'(\mathcal{S}_{1u} + \mathcal{S}_{11}\beta\alpha')a \overset{def}{=} B_T'\mathcal{S}_{1v}a,$$

where $\mathcal{S}_{1v}$ is the sample moment matrix between y_{t-1} and the stationary process $v_t = u_t + \alpha\beta' x_{t-1}$ corrected for $(1, \Delta y_{t-1}, ..., \Delta y_{t-p+1})$. The reason why we can define v_t by using x_{t-1} instead of y_{t-1} is that v_t can be obtained from the error correction form (3) by writing $\Pi^{(2)} y_{t-1}^{(2)} = \alpha\beta'(y_{t-1} - \mu_0 - \mu_1(t-1)) + \alpha\beta'\mu_0 = \alpha\beta' x_{t-1} + \alpha\beta'\mu_0$ and including $\alpha\beta'\mu_0$ in the intercept term. Because mean corrected series are used, the change in the intercept term has no effect.

To derive the weak limit of $B_T'\mathcal{S}_{10}a = B_T'\mathcal{S}_{1v}a$ we conclude from (A.2) that the process v_t has the linear representation

$$v_t = u_t + \alpha\beta'\Phi(L)u_{t-1} \stackrel{def}{=} u_t + w_t$$

so that, with obvious notation,

$$B_T'\mathcal{S}_{1v}a = B_T'\mathcal{S}_{1u}a + B_T'\mathcal{S}_{1w}a. \tag{A.18}$$

Since $B_T'(\mathcal{S}_{10} - \mathcal{S}_{11}\beta\alpha') = B_T'\mathcal{S}_{1u}$ we find from (A.13) that

$$B_T'\mathcal{S}_{1u}a \xrightarrow{w} \int_0^1 GdW'a. \tag{A.19}$$

For $B_T'\mathcal{S}_{1w}a$, the other component of $B_T'\mathcal{S}_{1v}a$, we denote $z_t = (\Delta y_{t-1}', \ldots, \Delta y_{t-p+1}')'$ and let $S_{a_{-i}b_{-j}}$ stand for the sample covariance matrix of any two time series a_{t-i} and b_{t-j}. Then,

$$B_T'\mathcal{S}_{1w}a = B_T'S_{y_{-1}w}a - B_T'S_{y_{-1}z}S_{zz}^{-1}S_{zw}a. \tag{A.20}$$

For the first term on the right hand side we can use the definition of w_t and Theorem B.13 of Johansen (1995) to obtain

$$\begin{aligned} B_T'S_{y_{-1}w}a &\xrightarrow{w} \int_0^1 GdW'\Phi(1)'\beta\alpha'a + \begin{bmatrix} \sum_{k=1}^{\infty} \mathrm{Cov}(\Delta y_t, w_{t+k})a \\ 0 \end{bmatrix} \\ &\stackrel{def}{=} \int_0^1 GdW'\Phi(1)'\beta\alpha'a + \Lambda_{\Delta yw}a, \end{aligned} \tag{A.21}$$

where the zero is $(1 \times K)$ and $\mathrm{Cov}(\Delta y_t, w_{t+k})$ on the right hand side can be expressed by using the parameters in the linear representations of the processes Δy_t and w_t. A complication in the second term on the right hand side of (A.20) is that the covariance matrix S_{zw} is not of order $O_p(T^{-1/2})$ as it is when we have u_t in place of w_t (cf. Johansen (1995, p. 148)). Therefore, the second term does not vanish. Because z_t and w_t are jointly stationary and ergodic processes, a law of large numbers and the fact $B_T' S_{y_{-1}z} = O_p(1)$ (to be justified shortly) give

$$B_T' S_{y_{-1}z} S_{zz}^{-1} S_{zw} a = B_T' S_{y_{-1}z} \Sigma_{zz}^{-1} \Sigma_{zw} a + o_p(1), \tag{A.22}$$

where $\Sigma_{zz} = \text{Cov}(z_t)$ and $\Sigma_{zw} = \text{Cov}(z_t, w_t)$. Regarding the matrix $B_T' S_{y_{-1}z}$, consider $B_T' S_{y_{-1}\Delta y_{-j}}$ and conclude from (1), (A.2) and Theorem B.13 of Johansen (1995) that

$$\begin{aligned} B_T' S_{y_{-1}\Delta y_{-j}} &\stackrel{w}{\to} \int_0^1 G dW' C' + \begin{bmatrix} \sum_{k=1}^{\infty} \text{Cov}(\Delta y_t, \Delta y_{t-j+k}) \\ 0 \end{bmatrix} \\ &\stackrel{def}{=} \int_0^1 G dW' C' + \Lambda_{\Delta y \Delta y_{-j}}, \end{aligned}$$

where the zero is $(1 \times K)$. Thus, it follows that

$$\begin{aligned} B_T' S_{y_{-1}z} &\stackrel{w}{\to} \left[\int_0^1 G dW' C' + \Lambda_{\Delta y \Delta y_{-1}} : \cdots : \int_0^1 G dW' C' + \Lambda_{\Delta y \Delta y_{-p+1}} \right] \\ &= \int_0^1 G dW' C' J + \Lambda_{\Delta y z}, \end{aligned} \tag{A.23}$$

where $J = [I_K : \cdots : I_K]$ $(K \times K(p-1))$ and $\Lambda_{\Delta y z} = \left[\Lambda_{\Delta y \Delta y_{-1}} : \cdots : \Lambda_{\Delta y \Delta y_{-p+1}} \right]$ $(K \times K(p-1))$. Note that the last row of $\Lambda_{\Delta y z}$ is zero and the autocovariances in $\Lambda_{\Delta y z}$ can again be expressed by using the parameters in the linear representation of the process Δy_t.

Combining (A.20), (A.21), (A.22) and (A.23) we get

$$B_T' \mathcal{S}_{1w} a \stackrel{w}{\to} \int_0^1 G dW' \Phi(1)' \beta \alpha' a + \Lambda_{\Delta y w} a + \left(\int_0^1 G dW' C' J + \Lambda_{\Delta y z} \right) \Sigma_{zz}^{-1} \Sigma_{zw} a,$$

which in conjunction with (A.18) and (A.19) gives

$$\begin{aligned} B_T' \mathcal{S}_{1v} a &\stackrel{w}{\to} \int_0^1 G dW' a + \int_0^1 G dW' \Phi(1)' \beta \alpha' a + \Lambda_{\Delta y w} a \\ &\quad + \left(\int_0^1 G dW' C' J + \Lambda_{\Delta y z} \right) \Sigma_{zz}^{-1} \Sigma_{zw} a \qquad \text{(A.24)} \\ &\stackrel{def}{=} \int_0^1 G dW' a + \Xi(G, \Lambda_{\Delta y w}, \Lambda_{\Delta y z}). \end{aligned}$$

Now recall that we need to study the weak limit of the roots of (A.17). By (A.12), (A.24) and the identity $B_T'\mathcal{S}_{1v}a = B_T'\mathcal{S}_{01}a$ we get

$$\rho T^{-1}B_T'\mathcal{S}_{11}B_T - B_T'\mathcal{S}_{10}a(a'\Sigma_{00}a)^{-1}a'\mathcal{S}_{01}B_T$$

$$\xrightarrow{w} \rho\left(\int_0^1 GG'ds\right) - \left(\int_0^1 GdW'a + \Xi\right)(a'\Sigma_{00}a)^{-1}\left(\int_0^1 GdW'a + \Xi\right)', \tag{A.25}$$

where we have written Ξ for $\Xi(G, \Lambda_{\Delta yw}, \Lambda_{\Delta yz})$. The limit is a square matrix of order $K - r + 1$. Set the determinant of the limit equal to zero and let $\rho_1 \geq \cdots \geq \rho_{K-r+1} \geq 0$ be the ordered roots. Then we get the following result.

Proposition A.
Under the conditions of Proposition 1,

$$\text{LR}(r) \xrightarrow{w} \sum_{i=2}^{K-r_0+1} \rho_i.$$

□

It is seen that the limiting distribution depends on a number of nuisance parameters and, although some simplifications may be achieved, this dependence appears complicated. For instance, the Brownian motion $\bar{\gamma}'CW(s)$ in the definition of $G(s)$ can be transformed to the standard Brownian motion $B_1(s) = (\bar{\gamma}'C\Sigma_u C'\bar{\gamma})^{-1/2}\bar{\gamma}'CW(s)$ without changing the limiting distribution of $\text{LR}(r_0)$ (cf. Johansen (1995, p. 160)). However, since this transformation changes the matrices $\Lambda_{\Delta yw}$ and $\Lambda_{\Delta yz}$ by premultiplying their first $K-r$ rows by $(\bar{\gamma}'C\Sigma_u C'\bar{\gamma})^{-1/2}$ the resulting simplification (if any) may not be great. Making an analogous transformation from $W(s)$ to $B_2(s)$ to deal with the term dW in (A.25) (see Johansen (1995, p. 160)) does not work in our case because in place of the matrix $a'\Sigma_u a$ we have $a'\Sigma_{00}a$. Also, since $a = [\alpha_\perp : \bar{\alpha}\kappa]$ we can write $\Phi(1)'\beta\alpha' a = \Phi(1)'\beta\,[0 : \kappa]$ and (potentially) achieve a small simplification in the definition of Ξ. However, it seems that any major simplifications are not possible because it is, for instance, unlikely that the effect of the complicated 'second order bias' terms $\Lambda_{\Delta yw}$ and $\Lambda_{\Delta yz}$ could be totally eliminated. Finally, note that the limiting distribution could be derived without using the decomposition of $B_T'\mathcal{S}_{10}a = B_T'\mathcal{S}_{1v}a$ given in (A.18). The given derivation shows better, however, how and why the resulting limiting distribution differs from its counterparts obtained for the corresponding correctly specified models.

References

Coenen, G. and Vega, J. L. (2001). The demand for M3 in the euro area, *Journal of Applied Econometrics* **16**: 727–748.

Crowder, W. J. and Hoffman, D. L. (1996). The long-run relationship between nominal interest rates and inflation: The Fisher equation revisited, *Journal of Money, Credit, and Banking* **28**: 102–118.

Demetrescu, M., Kuzin, V. and Hassler, U. (2008). Long memory testing in the time domain, *Econometric Theory* **24**: 176–215.

Doornik, J. A., Hendry, D. F. and Nielsen, B. (1998). Inference in cointegrating models: UK M1 revisited, *Journal of Economic Surveys* **12**: 533–572.

Ericsson, N. R. and Sharma, S. (1998). Broad money demand and financial liberalization in Greece, *Empirical Economics* **23**: 417–436.

Funke, M. and Rahn, J. (2005). Just how undervalued is the Chinese Renminbi?, *The World Economy* **28**: 465–489.

Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2006). Testing for a unit root when uncertain about the trend, *mimeograph*, University of Nottingham.

Hubrich, K. (2001). *Cointegration Analysis in a German Monetary System*, Physica-Verlag, Heidelberg.

Hubrich, K., Lütkepohl, H. and Saikkonen, P. (2001). A review of systems cointegration tests, *Econometric Reviews* **20**: 247–318.

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

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns, *Journal of Finance* **56**: 815–849.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Lütkepohl, H. (2007). General-to-specific or specific-to-general modelling: An opinion on current econometric terminology, *Journal of Econometrics* **136**: 319–324.

Peytrignet, M. and Stahel, C. (1998). Stability of money demand in Switzerland: A comparison of the M2 and M3 cases, *Empirical Economics* **23**: 437–454.

Phillips, P. C. B. (1987). Time series regression with a unit root, *Econometrica* **55**: 277–301.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika* **75**: 335–346.

Ribba, A. (2006). The joint dynamics of inflation, unemployment and interest rate in the United States since 1980, *Empirical Economics* **31**: 497–511.

Saikkonen, P. and Lütkepohl, H. (1999). Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory* **15**: 50–78.

Saikkonen, P. and Lütkepohl, H. (2000). Testing for the cointegrating rank of a VAR process with an intercept, *Econometric Theory* **16**: 373–406.

Schwert, G. W. (1989). Tests for unit roots: A Monte Carlo investigation, *Journal of Business & Economic Statistics* **7**: 147–159.

Stephan, S. (2006). German exports to the euro area, *Empirical Economics* **31**: 871–882.

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.

Table 1: Relative Rejection Frequencies of $H_0 : \text{rk}(\Pi) = r_0$ vs. $H_1 : \text{rk}(\Pi) > r_0$ in 25 000 Replications Based on Time Series from DGP (10)-(12) with $K = 3$, $\psi = 0.5$, $\varphi = 0$, Varying c, Θ, r and $T = 100$ (the nominal significance level of the tests is 5% and the true VAR order $p = 1$ is used)

		$\Theta = 0$							
		$c = 0.1$				$c = 0.5$			
true rank	r_0	LR(r_0)	LR*(r_0)	Proc 1	Proc 2	LR(r_0)	LR*(r_0)	Proc 1	Proc 2
$r = 1$	0	47.6	90.8	91.5	91.5	26.3	90.9	90.9	90.9
	1	3.3	5.3	7.3	4.4	1.0	5.4	5.7	4.7
	2	4.1	0.3	4.4	0.3	0.6	0.3	1.0	0.3
$r = 2$	0	99.1	100.0	100.0	100.0	98.2	100.0	100.0	100.0
	1	23.4	96.7	96.9	63.4	8.7	97.0	97.0	58.3
	2	3.1	5.1	8.0	4.6	0.1	5.2	5.3	5.1
		$\Theta = (0.8, 0.4)$ or $\Theta' = (0.8, 0.4)$							
$r = 1$	0	87.8	100.0	100.0	100.0	53.0	100.0	100.0	100.0
	1	6.9	6.8	11.1	6.8	1.9	7.0	7.6	7.0
	2	7.1	0.4	7.5	0.4	0.0	0.4	0.5	0.4
$r = 2$	0	99.8	100.0	100.0	100.0	99.4	100.0	100.0	100.0
	1	46.1	99.1	99.3	99.1	6.7	99.2	99.2	99.2
	2	11.6	5.3	16.0	5.3	0.4	5.4	5.8	5.4

Table 2: Relative Frequencies of Cointegrating Ranks Selected in 25 000 Replications Based on Time Series from DGP (10)-(12) with $K = 3$, $\psi = 0.5$, $\varphi = 0$, Varying c, Θ, r and $T = 100$ (the nominal significance level of the tests is 5% and the true VAR order $p = 1$ is used)

		$\Theta = 0$				$\Theta = (0.8, 0.4)$ or $\Theta' = (0.8, 0.4)$			
		$c = 0.1$		$c = 0.5$		$c = 0.1$		$c = 0.5$	
true rank	r_0	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2
$r = 1$	0	8.4	8.4	9.0	9.0	0.0	0.0	0.0	0.0
	1	84.2	87.0	85.2	86.2	88.8	93.1	92.3	92.9
	2	6.0	4.2	5.3	4.4	8.2	6.3	7.1	6.5
$r = 2$	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	3.0	36.5	2.9	41.6	0.6	0.8	0.7	0.7
	2	88.9	59.0	91.6	53.6	83.3	93.7	93.3	93.7

Table 3: Relative Frequencies of Cointegrating Ranks Selected in 25 000 Replications Based on Time Series from DGP (10)-(12) with $K = 5$, $\psi = 0.5$, $\varphi = 0$, Varying c, Θ, r and $T = 100$ (the nominal significance level of the tests is 5% and the true VAR order $p = 1$ is used)

		$\Theta = 0$				$\Theta \neq 0$			
		$c = 0.1$		$c = 0.5$		$c = 0.1$		$c = 0.5$	
true rank	r_0	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2
$r = 1$	0	32.0	32.0	35.2	35.2	5.0	5.0	5.3	5.3
	1	60.4	63.4	58.1	60.1	84.6	87.7	85.8	87.0
	2	6.8	4.1	6.0	4.2	9.2	6.5	8.1	7.0
	3	0.6	0.3	0.5	0.3	0.9	0.5	0.5	0.4
	4	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
$r = 2$	0	1.3	1.3	1.5	1.5	0.0	0.0	0.0	0.0
	1	39.5	58.6	40.5	62.5	16.7	20.8	19.0	21.6
	2	53.5	36.2	53.0	32.1	75.1	72.5	74.4	71.9
	3	5.1	3.4	4.5	3.4	7.2	6.0	6.0	5.9
	4	0.3	0.2	0.3	0.2	0.6	0.4	0.4	0.4
$r = 3$	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	0.8	8.8	0.8	12.7	0.0	0.5	0.0	1.0
	2	32.6	43.3	33.0	41.2	17.9	19.1	18.8	18.7
	3	61.5	44.0	61.6	42.2	75.7	74.8	75.8	74.9
	4	4.4	3.5	4.1	3.4	5.5	5.0	4.8	4.8
$r = 4$	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	0.0	3.4	0.0	5.2	0.0	0.4	0.0	0.9
	3	12.1	27.4	12.5	27.4	7.5	9.1	7.6	9.1
	4	82.4	64.9	82.2	62.7	85.6	85.2	87.3	84.9

For $\Theta \neq 0$ the following matrices were used: $\Theta = (0.4, 0.2, 0.4, 0.2)$, $\Theta = \begin{bmatrix} 0.4 & 0.2 & 0.4 \\ 0.2 & 0.4 & 0.2 \end{bmatrix}$, $\Theta' = \begin{bmatrix} 0.4 & 0.2 & 0.4 \\ 0.2 & 0.4 & 0.2 \end{bmatrix}$ and $\Theta' = (0.4, 0.2, 0.4, 0.2)$ for $r = 1$, $r = 2$, $r = 3$ and $r = 4$, respectively.

Table 4: Relative Frequencies of Cointegrating Ranks Selected in 25 000 Replications Based on Time Series from DGP (10)-(12) with $K = 5$, $\psi = 0.8$, $\varphi = -0.8$, Varying c, Θ, r and $T = 100$ (the nominal significance level of the tests is 5%, VAR order selected by AIC)

		$\Theta = 0$				$\Theta \neq 0$			
		$c = 0.1$		$c = 0.5$		$c = 0.1$		$c = 0.5$	
true rank	r_0	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2
$r = 1$	0	73.5	73.5	74.2	74.2	58.5	58.5	59.2	59.2
	1	22.1	23.8	21.8	23.2	34.2	36.9	33.9	36.6
	2	3.5	2.2	3.3	2.1	6.0	3.8	5.8	3.5
	3	0.5	0.2	0.5	0.2	0.9	0.5	0.7	0.4
	4	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0
$r = 2$	0	63.4	63.4	64.1	64.1	34.0	34.0	34.8	34.8
	1	29.6	32.2	29.0	31.8	47.7	53.6	47.4	53.6
	2	5.9	3.8	5.8	3.4	15.1	10.2	15.0	9.7
	3	0.9	0.4	0.7	0.3	2.6	1.7	2.3	1.5
	4	0.1	0.0	0.1	0.0	0.3	0.2	0.3	0.2
$r = 3$	0	48.7	48.7	49.0	49.0	23.0	23.0	23.2	23.2
	1	37.7	42.7	37.4	42.8	45.8	55.8	46.0	56.2
	2	11.0	7.1	11.1	6.8	24.3	16.5	24.0	15.9
	3	2.0	1.1	1.9	0.9	5.8	3.9	5.6	3.7
	4	0.3	0.1	0.3	0.1	0.8	0.6	0.8	0.6
$r = 4$	0	31.9	31.9	32.1	32.1	20.3	20.3	20.3	20.3
	1	40.1	49.1	39.8	49.4	39.2	52.1	38.4	52.4
	2	19.9	14.5	20.2	14.1	27.0	19.1	27.3	19.0
	3	6.1	3.5	6.1	3.4	10.3	6.4	10.8	6.3
	4	1.4	0.7	1.3	0.6	2.7	1.6	2.5	1.5

For $\Theta \neq 0$ see footnote of Table 3.

Table 5: Relative Frequencies of Cointegrating Ranks Selected in 25 000 Replications Based on Time Series from DGP (10)-(12) with $K = 5$, $\psi = 0.8$, $\varphi = -0.8$, Varying c, Θ, r and $T = 250$ (the nominal significance level of the tests is 5%, VAR order selected by AIC)

		$\Theta = 0$				$\Theta \neq 0$			
		$c = 0.1$		$c = 0.5$		$c = 0.1$		$c = 0.5$	
true rank	r_0	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2
$r = 1$	0	43.5	43.5	43.9	43.9	8.4	8.4	8.5	8.5
	1	50.7	52.8	50.2	52.3	82.8	84.1	83.1	84.0
	2	5.1	3.2	5.3	3.4	8.0	6.7	7.5	6.7
	3	0.4	0.2	0.4	0.3	0.6	0.5	0.7	0.5
	4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$r = 2$	0	4.0	4.0	3.8	3.8	0.0	0.0	0.0	0.0
	1	48.8	67.1	48.3	68.0	26.0	29.1	26.4	29.5
	2	42.6	25.5	43.1	24.7	67.3	64.3	66.9	63.8
	3	4.0	2.8	4.1	2.9	6.0	5.9	5.9	5.8
	4	0.4	0.2	0.3	0.2	0.4	0.4	0.5	0.5
$r = 3$	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	2.5	20.2	2.7	20.9	0.4	2.5	0.5	2.8
	2	42.7	43.7	42.9	43.2	27.7	26.7	27.9	27.1
	3	50.2	32.6	49.9	32.3	66.4	65.2	66.1	64.6
	4	4.1	3.1	3.9	3.1	5.0	5.0	4.9	4.9
$r = 4$	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.1
	2	0.3	9.4	0.3	9.7	0.1	1.9	0.2	1.9
	3	20.3	31.8	20.5	31.8	13.4	14.6	13.9	15.1
	4	74.3	54.1	74.0	53.9	81.1	78.3	80.7	77.7

For $\Theta \neq 0$ see footnote of Table 3.

Table 6: Relative Rejection Frequencies of $H_0 : \mathrm{rk}(\Pi) = r_0$ vs. $H_1 : \mathrm{rk}(\Pi) > r_0$ in 25 000 Replications Based on Time Series from DGP (13) with $K = 3$, $\psi = 0.5$, $\varphi = 0$, Varying c, Θ, r and $T = 100$ (the nominal significance level of the tests is 5% and the true VAR order $p = 1$ is used)

		$\Theta = 0$							
		$c = 0.1$				$c = 0.5$			
true rank	r_0	LR(r_0)	LR*(r_0)	Proc 1	Proc 2	LR(r_0)	LR*(r_0)	Proc 1	Proc 2
$r = 1$	0	98.5	91.1	98.8	98.8	98.0	91.0	98.4	98.4
	1	9.4	5.3	12.2	9.0	5.5	5.4	9.0	5.5
	2	8.8	0.3	9.1	6.8	2.6	0.3	2.9	1.9
$r = 2$	0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0
	1	99.9	96.6	99.9	99.6	99.9	96.6	99.9	99.5
	2	25.8	4.9	28.6	23.9	6.2	5.1	10.8	6.1
		$\Theta = (0.8, 0.4)$ or $\Theta' = (0.8, 0.4)$							
$r = 1$	0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1	11.1	7.0	14.7	10.8	7.0	6.9	11.3	7.1
	2	9.3	0.4	9.6	7.2	3.2	0.3	3.6	2.5
$r = 2$	0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1	99.9	99.2	100.0	99.9	99.9	99.1	99.9	99.9
	2	25.3	5.5	28.5	23.6	7.3	5.3	11.8	7.1

Table 7: Relative Frequencies of Cointegrating Ranks Selected in 25 000 Replications Based on Time Series from DGP (13) with $K = 3$, $\psi = 0.5$, $\varphi = 0$, Varying c, Θ, r and $T = 100$ (the nominal significance level of the tests is 5% and the true VAR order $p = 1$ is used)

		$\Theta = 0$				$\Theta = (0.8, 0.4)$ or $\Theta' = (0.8, 0.4)$			
		$c = 0.1$		$c = 0.5$		$c = 0.1$		$c = 0.5$	
true rank	r_0	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2
$r = 1$	0	1.2	1.2	1.5	1.5	0.0	0.0	0.0	0.0
	1	86.5	89.7	89.4	92.9	85.2	89.1	88.6	92.8
	2	8.1	6.0	7.5	4.7	10.1	7.3	9.5	5.9
$r = 2$	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	0.0	0.3	0.0	0.4	0.0	0.0	0.0	0.0
	2	71.3	75.7	89.1	93.3	71.4	76.2	88.1	92.7

Table 8: Relative Frequencies of Cointegrating Ranks Selected in 25 000 Replications Based on Time Series from DGP (13) with $K = 5$, $\psi = 0.5$, $\varphi = 0$, Varying c, Θ, r and $T = 100$ (the nominal significance level of the tests is 5% and the true VAR order $p = 1$ is used)

		$\Theta = 0$				$\Theta \neq 0$			
		$c = 0.1$		$c = 0.5$		$c = 0.1$		$c = 0.5$	
true rank	r_0	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2	Proc 1	Proc 2
$r = 1$	0	19.8	19.8	21.8	21.8	1.6	1.6	1.7	1.7
	1	70.1	73.0	68.9	72.1	85.5	89.3	86.4	90.4
	2	8.8	6.3	8.3	5.5	11.3	8.1	10.6	7.0
	3	0.9	0.5	0.7	0.4	1.2	0.7	1.0	0.6
	4	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0
$r = 2$	0	0.1	0.1	0.2	0.2	0.0	0.0	0.0	0.0
	1	20.1	25.8	22.7	28.6	6.6	9.1	7.7	10.4
	2	70.2	67.2	69.4	66.3	81.4	82.5	82.3	82.9
	3	8.0	5.8	6.8	4.4	10.3	7.3	8.9	6.0
	4	0.6	0.4	0.4	0.2	0.7	0.5	0.7	0.4
$r = 3$	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0
	2	8.3	13.3	10.7	15.8	3.1	5.7	4.2	6.9
	3	80.3	78.3	81.5	79.1	84.1	84.9	86.0	87.0
	4	7.2	5.3	6.3	4.1	8.3	6.1	7.9	4.9
$r = 4$	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	3	0.2	2.2	0.3	2.3	0.0	1.1	0.1	1.2
	4	71.8	75.5	89.5	92.0	71.6	76.1	89.1	92.5

For $\Theta \neq 0$ see footnote of Table 3.