Diskussionsschriften

The Performance of Subspace Algorithm Cointegration Analysis: A Simulation Study

Dietmar Bauer
Martin Wagner

03-08

May 2003

Universität Bern
Volkswirtschaftliches Institut
Gesellschaftstrasse 49
3012 Bern, Switzerland
Tel: 41 (0)31 631 45 06
Web: www.vwi.unibe.ch

The Performance of Subspace Algorithm Cointegration Analysis: A Simulation Study

Dietmar Bauer *
Inst. for Econometrics,
Operations Research and System Theory
TU Wien
Argentinierstr. 8, A-1040 Wien

Martin Wagner†
Department of Economics
University of Bern
Gesellschaftsstrasse 49
CH-3012 Bern

Abstract

This paper presents a simulation study that assesses the finite sample performance of the subspace algorithm cointegration analysis developed in Bauer and Wagner (2002b). The method is formulated in the state space framework, which is equivalent to the VARMA framework, in a sense made precise in the paper. This implies applicability to VARMA processes. The paper proposes and compares six different tests for the cointegrating rank. The simulations investigate four issues: the order estimation, the size performance of the proposed tests, the accuracy of the estimation of the cointegrating space and the forecasting performance. The simulations are performed on a set of trivariate processes with cointegrating ranks ranging from zero to three as well as on processes of output dimension four and cointegrating rank two. We analyze the influence of the sample size on the results as well as the sensitivity of the results with respect to stable poles approaching the unit circle. All results are compared to benchmark results obtained by applying the Johansen procedure on VAR models fitted to the data.
The simulations show advantages of subspace algorithm cointegration analysis for the small sample performance of the tests for the cointegrating rank in many cases. However, we find that the accuracy of the subspace algorithm based estimation of the cointegrating space is unsatisfactory for the four-dimensional simulated systems. The forecasting performance is grosso modo comparable to the results obtained by applying the Johansen methodology on VAR approximations, although for very small sample sizes the forecasts based on VAR approximations outperform the subspace forecasts.
The appendix provides critical values for the test statistics.

JEL Classification: C13, C32
Keywords: State space representation, cointegration, subspace algorithms, simulation study

*Support by the Austrian FWF under the project number P-14438-INF is gratefully acknowledged.
†Corresponding author, Tel.: ++41 31 6314778, Fax: ++41 31 6313992, e-mail: martin.wagner@vwi.unibe.ch.

1 Introduction

This paper complements the theoretical results derived in Bauer and Wagner (2002b) for cointegration analysis in the state space framework based on subspace algorithms with a simulation study. The state space framework is an (in a certain sense made precise below) equivalent alternative to the VARMA framework that has surprisingly not received a lot of attention in cointegration analysis. Early exceptions are given by the work of Aoki and his collaborators, see e.g. Aoki (1990, Chapter 9), Aoki and Havenner (1989) or Aoki and Havenner (1997). All these papers however lack a thorough statistical foundation, including an investigation of the structure of state space systems for integrated processes, the specification of integer parameters like the system order or testing for the cointegrating rank. To a certain extent, these gaps are filled in Bauer and Wagner (2001) and Bauer and Wagner (2002b). The first paper shows that the state space framework is very suitable for the analysis of processes with unit roots, in particular also with respect to the cointegration properties. Based on the structure theory presented in Bauer and Wagner (2001), Bauer and Wagner (2002b) suggests to use the so called subspace algorithms for estimation of and testing in I(1) processes in the state space framework.

The vast majority of cointegration studies is performed using VAR models (see e.g. Johansen, 1995, and the references contained therein) or using static or dynamic regression models (from a long list of contributions see e.g. Engle and Granger, 1987; Stock and Watson, 1988; Phillips, 1995). Saikkonen (1992) shows that the Johansen methodology can be applied also to VARMA processes if the lag length of an autoregressive approximation increases with the sample size at a sufficient rate. Also, the regression based approaches mentioned above can often be applied to VARMA processes by introducing correction factors for serial correlation in a suitable way. These regression based approaches are in a sense focused on the estimation of the cointegrating relationships and neglect all other parameters in a first step. The latter, however, can often be recovered in a second step, because of the usual super-consistency of the estimation of the cointegrating space. Alternatively, Yap and Reinsel (1995) derive both an estimator which is asymptotically equivalent to the maximum likelihood estimator for Gaussian I(1) VARMA processes, and tests for the number of cointegrating relationships. Bauer and Wagner (2002a) derive pseudo ML estimators for processes integrated at any point on the unit circle with the corresponding integration orders all equal

to one for VARMA processes, albeit their analysis is formulated in the state space framework. Developing methods that are applicable for VARMA processes is interesting for at least two reasons. First of all, on a theoretical level applicability to a wider class of processes is a virtue per se. More interestingly for the applied researcher, the increased flexibility gained from a wider process class may lead to a more parsimonious and more accurate description of the underlying data generating process.

Subspace algorithms originated in the engineering literature in the 1980ies and are a computationally cheap alternative to more classical estimation procedures based on the optimization of some criterion function, like e.g. the Gaussian likelihood. In the literature a variety of subspace algorithms have been proposed for stationary processes, see e.g. Larimore (1983), Van Overschee and DeMoor (1994) or Verhaegen (1994). Bauer and Wagner (2002b) adapt the subspace algorithm `CCA` (Larimore, 1983) to result in consistent estimation also for I(1) processes. The estimators for the cointegrating space are shown to be super-consistent as usual. That paper also provides a number of distributional results, which can be used to derive procedures for the specification of the integer parameters determining the model structure, i.e. the order of the system and the cointegrating rank. The present paper aims at shedding light on the finite sample performance of subspace algorithm cointegration analysis. This is done via simulations. The methodological innovation with respect to Bauer and Wagner (2002b) is that here we propose and compare in total six different tests for the cointegrating rank and an additional order estimation criterion.

The performance of the estimation methods and tests is analyzed with respect to several issues: The first is the assessment of the properties of the order estimation. The second is the actual finite-sample size of the various tests. The third is the accuracy of the estimation of the cointegrating space and the fourth is the forecasting performance of the estimated state space models. For all these aspects we investigate the sensitivity of the results with respect to sample size (100,...,500) and with respect to (real or complex conjugate) stable poles of the transfer function approaching the unit circle (see below for definitions). The forecasting performance is investigated for varying forecasting horizons. The results obtained by applying the Johansen methodology on VAR models fitted to the simulated data serve as a benchmark.

The paper is organized as follows: Section 2 discusses state space models and some relationships to VARMA models as well as the assumptions. Section 3 is devoted to a detailed

discussion of the estimation and test procedures. Section 4 presents the simulation design as well as the results obtained from the simulations. Section 5 summarizes and concludes the paper. The appendix provides tables with critical values for the proposed tests for the number of stochastic trends.

Throughout the paper we use I_m to denote the $m \times m$ identity matrix. $\otimes$ is used to denote the Kronecker product. $\lambda_{max}(A)$ denotes an eigenvalue of maximum modulus of a matrix A. The transpose of a matrix will be denoted using $'$.

2 State Space Models: Basics and Assumptions

This section very briefly presents the model class considered in this paper and illustrates the connections to the VARMA framework. A detailed discussion for the case of stationary processes is given in Hannan and Deistler (1988), while Bauer and Wagner (2001) builds the basis for the case of processes with arbitrary unit roots and integer integration orders corresponding to the unit roots.

We consider linear, time invariant, finite dimensional, discrete time systems of the form

$$\begin{array}{rcl} x_{t+1} = \left[\begin{array}{c} x_{t+1,1} \\ x_{t+1,st} \end{array}\right] & = & \left[\begin{array}{cc} I_c & 0 \\ 0 & A_{st} \end{array}\right] x_t + \left[\begin{array}{c} K_1 \\ K_{st} \end{array}\right] \varepsilon_t \\ y_t & = & \left[\begin{array}{cc} C_1 & C_{st} \end{array}\right] x_t + \varepsilon_t \end{array} \tag{1}$$

where $\{y_t\}_{t\in\mathbb{N}}$ denotes the s-dimensional output series observed for $t = 1, \ldots, T$. $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ denotes the s-dimensional white noise innovation sequence. $A_{st} \in \mathbb{R}^{(n-c)\times(n-c)}$, $K = [K_1', K_{st}']' \in \mathbb{R}^{n\times s}$, $C = [C_1, C_{st}] \in \mathbb{R}^{s\times n}$ and $x_t = [x_{t,1}', x_{t,st}']' \in \mathbb{R}^n$ denotes the n-dimensional unobserved state sequence. The integer $c \leq s$ denotes the number of stochastic (or common) trends, as will become clear in the discussion below. $C_1'C_1 = I_c$ and K_1 is assumed to be of full row rank. Furthermore $|\lambda_{max}(A_{st})| < 1$ and $|\lambda_{max}(A - KC)| < 1$ will always be assumed, where $A = \left[\begin{array}{cc} I_c & 0 \\ 0 & A_{st} \end{array}\right]$. The recursion is assumed to be started at some random variable with finite variance x_1, which is independent of the innovation sequence $\{\varepsilon_t\}_{t\in\mathbb{N}}$. Throughout the paper ε_t is assumed to be an ergodic strictly stationary martingale difference sequence fulfilling the following conditions:

$$\begin{array}{rl} \mathbb{E}\{\varepsilon_t|\mathcal{F}_{t-1}\} = 0, & \mathbb{E}\{\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\} = \mathbb{E}\{\varepsilon_t\varepsilon_t'\} = \Omega > 0 \\ \mathbb{E}\{\varepsilon_{t,a}\varepsilon_{t,b}\varepsilon_{t,c}|\mathcal{F}_{t-1}\} = \omega_{a,b,c}, & \mathbb{E}\varepsilon_{t,a}^4 < \infty \end{array} \tag{2}$$

where $\varepsilon_{t,a}$ denotes the a-th component of the vector ε_t and $\mathcal{F}_{t-1}$ denotes the σ-algebra spanned by the past, i.e. by $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_1$ and x_1. $\omega_{a,b,c}$ is a constant not depending on t.

The representation (1) is called *state space representation of the process* $\{y_t\}_{t\in\mathbb{N}}$. Note that any basis change of the (unobserved) state leads to an equivalent representation of the process $\{y_t\}_{t\in\mathbb{N}}$. Using the new state, $z_t = Tx_t$ say, for a nonsingular matrix $T \in \mathbb{R}^{n\times n}$, the system matrices are transformed into the triple (TAT^{-1}, TK, CT^{-1}). Hence state space representations, like VARMA representations, are not unique. A state space system is called *minimal*, if no other state space representation with smaller state dimension exists. In this case, the dimension of the state, i.e. n, is called *order of the system*. If one considers only minimal representations, it can be shown that all state space representations of a given process $\{y_t\}_{t\in\mathbb{N}}$ are related by transformations of the basis of the state.

Note that the representation given in (1) is subject to a number of restrictions and hence is not a general state space system: The matrix A describing the dynamics of the state is block-diagonal with the northwest corner equal to the identity matrix. Furthermore the first block-column in the matrix C, i.e. C_1, is restricted to be orthonormal, i.e. fulfilling $C_1'C_1 = I_c$. These additional restrictions do not identify the system matrices completely, i.e. every basis transformation of the state with a matrix $T = \mathrm{diag}(Q_1, T_{st})$, where $Q_1 \in \mathbb{R}^{c\times c}$ is orthonormal and T_{st} is nonsingular, leads to an equivalent representation of the system obeying the mentioned restrictions. Bauer and Wagner (2001) provide the additional restrictions to define a unique representation and hence achieve identifiability.[1]

Using the recursive nature of the state it follows from (1) that

$$\begin{aligned} y_t &= C_1x_{t,1} + C_{st}x_{t,st} + \varepsilon_t = C_1x_{t-1,1} + C_{st}A_{st}x_{t-1,st} + \varepsilon_t + C_1K_1\varepsilon_{t-1} + C_{st}K_{st}\varepsilon_{t-1} \\ &= \ldots = C_1K_1\textstyle\sum_{j=1}^{t-1}\varepsilon_{t-j} + C_1x_{1,1} + \sum_{j=0}^{t-1}K_{st}(j)\varepsilon_{t-j} + C_{st}A_{st}^{t-1}x_{1,st} \\ &= C_1K_1\textstyle\sum_{j=1}^{t-1}\varepsilon_{t-j} + C_1x_{1,1} + y_{t,st} \end{aligned} \tag{3}$$

where $K_{st}(0) = I_s, K_{st}(j) = C_{st}A_{st}^{j-1}K_{st}, j > 0$ and the last equation defines $y_{t,st}$. This decomposition shows the advantages of the block-diagonality of the A-matrix and the partitioning of the state into $x_{t,1}$ and $x_{t,st}$: It follows immediately that $C_1K_1\sum_{j=1}^{t-1}\varepsilon_{t-j}$ is an integrated process. For the second component of the state $x_{t,st}$ note that, due to the assumption $|\lambda_{max}(A_{st})| < 1$, the coefficients $K_{st}(j)$ converge to zero exponentially fast and thus $\sum_{j=0}^{\infty}K_{st}(j)\varepsilon_{t-j}$ is a stationary process. Hence, using the special choice $x_{1,st} = \sum_{j=0}^{\infty}A_{st}^jK_{st}\varepsilon_{-j}$ for the stable part of the initial state, equation (3) provides a decomposition of the output y_t into an integrated part $(C_1K_1\sum_{j=1}^{t-1}\varepsilon_{t-j})$, a random initial effect $(C_1x_{1,1})$ and a stationary

[1] The details in this respect are of no importance for the discussion here and therefore omitted.

part $(y_{t,st})$. Consider the first difference of the process (for $t > 1$):

$$y_t - y_{t-1} = C_1 K_1 \varepsilon_{t-1} + \varepsilon_t + \sum_{j=0}^{t-2} (K_{st}(j+1) - K_{st}(j)) \varepsilon_{t-1-j} + C_{st}(A_{st}^{t-1} - A_{st}^{t-2}) x_{1,st}$$

where for the specific choice of $x_{1,st}$ given above the sum of the last two terms is a stationary process. According to e.g. the definition in Johansen (1995) this shows that $\{y_t\}_{t\in\mathbb{N}}$ is an I(1) process. Bauer and Wagner (2001) prove that also the contrary is true: Consider a process $\{y_t\}_{t\in\mathbb{N}}$, such that there exists an initial value y_0 such that $y_t - y_{t-1} = v_t, t \in \mathbb{N}$, where v_t is a stationary VARMA process, i.e. the stationary solution to the vector difference equation

$$v_t = A_1 v_{t-1} + \cdots + A_p v_{t-p} + \varepsilon_t + B_1 \varepsilon_{t-1} + \cdots + B_q \varepsilon_{t-q}, t \in \mathbb{Z}$$

where (with z denoting a complex variable) the polynomials $a(z) = I_s - A_1 z - \cdots - A_p z^p, b(z) = I_s + B_1 z + \cdots + B_q z^q$ are left coprime, the roots of $\det a(z)$ are outside the unit circle and the roots of $\det b(z)$ are on or outside the unit circle, but $b(1) \neq 0$. Then Lemma 1 in Bauer and Wagner (2001) shows that there exists a linearly deterministic random process $\{d_t\}_{t\in\mathbb{N}}$, such that $y_t - d_t$ has a minimal state space representation. In this sense the VARMA and the state space framework are equivalent. It also follows that $k(z) = I_s + zC(I_n - zA)^{-1}K = ((1-z)a(z))^{-1}b(z)$. Note that the transfer function $k(z)$ is invariant for all state space and VARMA representations. Hence the prime object of interest for estimation is the transfer function $k(z)$ rather than any particular (state space) representation itself. From the above considerations it also follows that the restriction $|\lambda_{max}(A_{st})| < 1$ corresponds to the stability assumption of no root of $\det a(z)$ to lie inside or on the unit circle. The so called strict minimum-phase assumption $|\lambda_{max}(A - KC)| < 1$ ensures that the transfer function $k(z)$ is invertible for $|z| \leq 1$. For details on these correspondences see Hannan and Deistler (1988, Chapter 1).

From representation (3) also the cointegration properties of $\{y_t\}_{t\in\mathbb{N}}$ follow: Let the initial state $x_{1,st}$ be chosen such that $y_{t,st}$ is stationary. Choose a matrix $\beta \in \mathbb{R}^{s\times(s-c)}$ of full column rank, such that $\beta' C_1 = 0$. Then $\beta' y_t = \beta' y_{t,st}$ is stationary and hence the columns of β span the cointegrating space. For other initial conditions $x_{1,st}$ it follows that $\beta' y_t$ is asymptotically stationary. Due to the full rank of both C_1 and K_1 in any minimal representation there cannot be more linearly independent cointegrating relationships and hence C_1 parameterizes the orthogonal complement of the cointegrating space, which motivates the orthonormality constraint $C_1' C_1 = I_c$. The above considerations also directly imply $c \leq s$. Note that also the

well known relationship for I(1) systems, that the sum of the dimension of the cointegrating space $(s-c)$ and the number of stochastic trends (c) equals the dimension of the process s is directly evident in the state space representation. As a further remark note that the number of stochastic trends is identical to the multiplicity of the eigenvalue 1 of A. For a more detailed account on the 1-to-1 relationships between the structure of the eigenvalues of A and the (co-)integration properties of $\{y_t\}_{t\in\mathbb{N}}$ see Bauer and Wagner (2001).

3 Description of the Estimation Method and Tests

This section is intended to give a description of the subspace algorithm developed for estimation of cointegrated processes in Bauer and Wagner (2002b). Both, the estimation procedure and the tests for the number of stochastic trends are explained to such an extent, that the reader is enabled to implement the method herself.[2] Readers interested in the proofs are referred to Bauer and Wagner (2002b). Based on the results provided therein, we propose in this paper six different tests for the number of stochastic trends and also an additional order estimation criterion.

The basic idea underlying subspace algorithms is an appropriate interpretation of the state. Solving the system equations (1) one obtains (analogously to the evaluations in (3))

$$y_{t+j} = CA^j x_t + \sum_{i=0}^{j-1} CA^i K \varepsilon_{t+j-i-1} + \varepsilon_{t+j}. \tag{4}$$

From the system equations (1) it also follows that $x_t = (A-KC)^{t-1}x_1 + \sum_{i=0}^{t-2}(A-KC)^i K y_{t-i-1}$, i.e. the state is *contained* in the space spanned by the past of the output y_t and the initial state x_1. Since the state x_t and the noise $\varepsilon_{t+l}, l \geq 0$ are uncorrelated by assumption, this implies that the best linear predictor of $y_{t+j}, j \geq 0$ given the knowledge of $y_{t-1}, \ldots, y_1$ and x_1, denoted by $y(t+j|t)$, is given by

$$y(t+j|t) = CA^j x_t.$$

Thus, the state x_t is a basis for the predictor space for the whole future of y_t, i.e. for $y_{t+j}, j \geq 0$, and is, as noted above, contained in the past of the time series, $y_{t-j}, 1 \leq j \leq t-1$ and x_1. Next choose two indices f and p, both larger or equal than n, and define

$$Y_{t,f}^{+} = [y_t', y_{t+1}', \ldots, y_{t+f-1}']', \quad Y_{t,p}^{-} = [y_{t-1}', y_{t-2}', \ldots, y_{t-p}']', \quad E_{t,f}^{+} = [\varepsilon_t', \varepsilon_{t+1}', \ldots, \varepsilon_{t+f-1}']'.$$

[2] `GAUSS` and `MATLAB` code is available from the authors upon request.

Furthermore let

$$\mathcal{O}_f = [C', A'C', \ldots, (A^{f-1})'C']', \quad \mathcal{K}_p = [K, (A-KC)K, \ldots, (A-KC)^{p-1}K]$$

and let $\mathcal{E}_f$ denote the matrix with i-th block-row $[CA^{i-2}K, \ldots, CK, I_s, 0]$ for $i \geq 2$ and $[I_s, 0, \ldots, 0]$ as its first block-row. Then it follows from the system equations (1) that (for $t > p$)

$$Y_{t,f}^+ = \mathcal{O}_f \mathcal{K}_p Y_{t,p}^- + \mathcal{O}_f (A-KC)^p x_{t-p} + \mathcal{E}_f E_{t,f}^+. \tag{5}$$

Noting that for $p \to \infty$ due to the strict minimum-phase assumption the term $(A-KC)^p$ vanishes, the above observations motivate the following procedure for given integers f and p:

1) In a first step regress $Y_{t,f}^+$ on $Y_{t,p}^-$ to obtain an estimator $\hat{\beta}_{f,p}$ of $\mathcal{O}_f \mathcal{K}_p$. Due to the construction of the variables $Y_{t,f}^+$ and $Y_{t,p}^-$, the sample range in this regression is $t = p+1, \ldots, T-f+1$. We denote the effective sample size by $T_{f,p} = T - f - p + 1$.

2) Denoting the true system order by n, it follows that $\mathcal{O}_f \mathcal{K}_p$ has rank n for $f, p \geq n$. Typically the estimator $\hat{\beta}_{f,p}$ of $\mathcal{O}_f \mathcal{K}_p$ has full rank equal to $\min(f,p) \times s$. Thus, a rank n approximation of $\hat{\beta}_{f,p}$ with decomposition $\hat{\mathcal{O}_f} \hat{\mathcal{K}_p}$ has to be constructed. In this step also the order n of the system has to be specified. The details of the order specification and the rank n approximation are given below.

3) Use the derived estimator $\hat{\mathcal{K}_p}$ to obtain an estimator of the state $\hat{x}_t = \hat{\mathcal{K}_p} Y_{t,p}^-$, for $t = p+1, \ldots, T+1$.

4) Given the estimated state, the system equations (1) can be used to obtain estimators $(\hat{A}, \hat{K}, \hat{C})$ of the system matrices (A, K, C) as follows:

   i) Regress y_t on $\hat{x}_t, t = p+1, \ldots, T$ to obtain $\hat{C}$ and residuals $\hat{\varepsilon}_t = y_t - \hat{C}\hat{x}_t$.

   ii) Thus, $\hat{\Omega} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$ is an estimator of the innovation variance Ω.

   iii) Regress $\hat{x}_{t+1}$ on $\hat{x}_t$ and $\hat{\varepsilon}_t, t = p+1, \ldots, T$ to obtain $\hat{A}$ and $\hat{K}$.

Let us now turn to the open points, i.e. to the specification of the integers f and p, to the estimation of the system order n and to the rank n approximation $\hat{\mathcal{O}_f} \hat{\mathcal{K}_p}$ of the estimator $\hat{\beta}_{f,p}$.

We commence with the specification of the integers f and p. From the consistency proofs in Bauer and Wagner (2002b) the integers are required to fulfill the following restrictions:

$f \geq n, p \geq \frac{-d \log T}{\log |\lambda_{max}(A-KC)|}$ for some $d > 1$. Both restrictions rely on unknown population quantities, the system order respectively the system matrices. In the literature it has been suggested to use $f = p = 2\hat{p}_{AIC}$, where $\hat{p}_{AIC}$ denotes the estimated order for an autoregression fitted to y_t. This specific choice for f and p is motivated by the properties of `AIC` order estimation for stationary processes, where this choice ensures (almost sure) fulfillment of the above mentioned restrictions for f and p (cf. Hannan and Deistler, 1988, Theorem 6.6.3).

The rank n approximation is not performed directly on $\hat{\beta}_{f,p}$, but on a transformed matrix $\hat{W_f^+}\hat{\beta}_{f,p}\hat{W_p^-}$. Let $\hat{W_f^+}\hat{\beta}_{f,p}\hat{W_p^-} = \hat{U}\hat{\Sigma}\hat{V}'$ be the singular value decomposition, where $\hat{U}$ contains the left singular vectors, $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_{min(f,p)s})$ contains the singular values ordered decreasing in size and $\hat{V}$ contains the right singular vectors. The subspace algorithms proposed in the literature differ i.a. in their choices of these weighting matrices. Let $\hat{\Gamma_f^+} = \frac{1}{T_{f,p}} \sum_{t=p+1}^{T-f+1} Y_{t,f}^+ (Y_{t,f}^+)'$ and $\hat{\Gamma_p^-} = \frac{1}{T_{f,p}} \sum_{t=p+1}^{T-f+1} Y_{t,p}^- (Y_{t,p}^-)'$ denote the (non-centered) sample covariances[3] of $Y_{t,f}^+$ and $Y_{t,p}^-$. `CCA` uses $\hat{W_f^+} = (\hat{\Gamma_f^+})^{-1/2}$ and $\hat{W_p^-} = (\hat{\Gamma_p^-})^{1/2}$ respectively, which results in consistent estimators of the system matrices for stationary processes, see e.g. Bauer *et al.* (1999).[4] With this choice of weighting matrices the algorithm amounts to an estimation of the canonical correlations between $Y_{t,f}^+$ and $Y_{t,p}^-$, which explains the name *Canonical Correlation Analysis* or in short `CCA` algorithm.

Since the rank of $\hat{W_f^+}\mathcal{O}_f\mathcal{K}_p\hat{W_p^-}$ is equal to the system order n, only the first n singular values in the SVD of $\hat{W_f^+}\mathcal{O}_f\mathcal{K}_p\hat{W_p^-}$ are nonzero and the remaining ones are equal to zero. In finite samples the replacement of $\mathcal{O}_f\mathcal{K}_p$ by $\hat{\beta}_{f,p}$ will result in all singular values $\hat{\sigma}_1, \ldots, \hat{\sigma}_{min(f,p)s}$ typically being nonzero. Asymptotically, $\hat{\sigma}_1, \ldots, \hat{\sigma}_n$ converge to their positive limits and $\hat{\sigma}_{n+1}, \ldots, \hat{\sigma}_{min(f,p)s}$ converge to zero. Order estimation is based on these convergence properties. In Bauer and Wagner (2002b) the following criterion was defined:

$$SVC(n) = \hat{\sigma}_{n+1}^2 + 2nsH_T/T. \tag{6}$$

A variety of simulation experiments however led us to in addition consider the following criterion that has preferable, compared to SVC, finite sample performance:[5]

$$BA(n) = -\log(1 - \hat{\sigma}_{n+1}^2) + 2nsH_T/T. \tag{7}$$

[3] Alternatively the summation can be limited to range from $1 \leq t \leq T$, where $y_t = 0$ for $t \leq 0$ and $t > T$ is used. This changes only the computations and does not change the asymptotic results derived for the method.

[4] $X^{1/2}$ here denotes the Cholesky factor of the positive definite matrix X such that $X^{1/2}(X^{1/2})' = X$.

[5] It is straightforward to see that the two criteria $SVC(n)$ and $BA(n)$ have the same asymptotic properties.

Here $H_T > 0, H_T/T \to 0$ denotes a penalty term, which determines the asymptotic properties of the estimated order. Note that $2ns$ is the number of parameters in a model with state dimension n, (see e.g. Hannan and Deistler, 1988, Theorem 2.6.3). In this paper we only report the results based on the order estimated by using $BA(n)$. The estimated order, $\hat{n}$ say, is then given by the minimizing argument of the criterion function $BA(n)$. Thus, for the specified rank n, where e.g. $n = \hat{n}$, decompose the SVD in two parts:

$$\hat{W_f^+}\hat{\beta}_{f,p}\hat{W_p^-} = \hat{U}\hat{\Sigma}\hat{V}' = \hat{U}_n\hat{\Sigma}_n\hat{V}_n' + \hat{R}_n$$

where $\hat{U}_n \in \mathbb{R}^{fs\times n}, \hat{V}_n \in \mathbb{R}^{ps\times n}$ and $\hat{\Sigma}_n \in \mathbb{R}^{n\times n}$. Here $\hat{\Sigma}_n = \mathrm{diag}(\hat{\sigma}_1, \ldots, \hat{\sigma}_n)$ contains the n dominant singular values ordered decreasing in size, i.e. $1 \geq \hat{\sigma}_1 \geq \ldots \geq \hat{\sigma}_n > 0$. The matrices $\hat{U}_n$ and $\hat{V}_n$ contain the corresponding left and right singular vectors. The remaining singular values and vectors are attributed to $\hat{R}_n$ and are neglected. The rank n approximation of $\hat{\beta}_{f,p}$ is given by $\hat{\mathcal{O}}_f\hat{\mathcal{K}}_p = [(\hat{W_f^+})^{-1}\hat{U}_n][\hat{\Sigma}_n\hat{V}_n'(\hat{W_p^-})^{-1}]$ and thus $\hat{\mathcal{K}}_p = \hat{\Sigma}_n\hat{V}_n'(\hat{W_p^-})^{-1}$. This completes the description of the standard algorithm.

For integrated processes of the form (1), the procedure described above has to be *adapted* in order to achieve consistent estimation of the stationary subsystem (A_{st}, K_{st}, C_{st}). For correctly specified c, a consistent estimator $\hat{C}_1$ of C_1 can be derived by applying the `CCA` algorithm in its standard form as described above (see Bauer and Wagner, 2002b, Theorem 2 or Theorem 1 below), i.e. $T^\gamma\|\hat{C}_1 - C_1\| \to 0$, for $0 < \gamma < 1$ holds. Let $r = s - c$ denote the cointegrating rank. Next denote with $\hat{\tilde{C}} = [\hat{C}_1, \hat{C}_1^\perp]'$, where $\hat{C}_1^\perp \in \mathbb{R}^{s\times r}, \hat{C}_1'\hat{C}_1^\perp = 0$ and $(\hat{C}_1^\perp)'\hat{C}_1^\perp = I_r$. Define a new weighting matrix $\widehat{W_{f,C_1}^+} = [(I \otimes \hat{\tilde{C}})\frac{1}{T_{f,p}}\sum_{t=p+1}^{T-f+1} Y_{t,f}^+(Y_{t,f}^+)'(I \otimes \hat{\tilde{C}})']^{-1/2}(I \otimes \hat{\tilde{C}})$, using again the Cholesky decomposition as the square root of a matrix. In combination with the modified weighting matrix also the estimator of $\hat{\mathcal{K}}_p$ is modified: For any choice of weighting matrices, the estimated matrix $\hat{\mathcal{K}}_p = \hat{\Sigma}_n\hat{V}_n'(\hat{W_p^-})^{-1}$ can alternatively be written as $\hat{\mathcal{K}}_p = \hat{U}_n'\hat{W_f^+}\hat{\beta}_{f,p}$. If the modified weighting matrix $\widehat{W_{f,C_1}^+}$ is used, the corresponding matrix of left singular vectors $\hat{U}_n$ is changed to $\hat{U}_{n,c}$, where

$$\hat{U}_{n,c} = \begin{bmatrix} I_c & 0^{c\times(n-c)} \\ 0^{(fs-c)\times c} & \hat{U}(2,2) \end{bmatrix}.$$

$\hat{U}(2,2)$ here denotes the $(2,2)$-block of the matrix $\hat{U}_n$ of appropriate dimensions.[6] Note that $\hat{U}_n$ converges for fixed f to a matrix of this structure, i.e. to a matrix that has the $(1,2)$ and

[6]From a theoretical point of view, to achieve consistency only the (2,1)-block of the matrix $\hat{U}_n$ has to be replaced by a null-block.

the $(2,1)$ block equal to zero, while the $(1,1)$ block is equal to I_c. Thus, under the assumption of c stochastic trends the adapted subspace procedure can be described as follows:

1) Perform steps 1) to 4)i) of the standard `CCA` subspace algorithm as described above. This also includes the order estimation using e.g. $BA(n)$.

2) Use the estimator $\hat{C}_1$ (the heading $s \times c$ sub-block of $\hat{C}$) to construct the modified weighting matrix $\widehat{W^+_{f,C_1}}$ and recalculate the SVD as $\hat{U}\hat{\Sigma}\hat{V}' = \widehat{W^+_{f,C_1}}\hat{\beta}_{f,p}\hat{W}^-_p$ (using identical notation for the SVD as above).

3) Compute $\hat{U}_{n,c}$ and generate the adapted estimator of $\hat{\mathcal{K}}_{p,C_1} = \hat{U}'_{n,c}\widehat{W^+_{f,C_1}}\hat{\beta}_{f,p}$

4) Use the adapted estimator $\hat{\mathcal{K}}_{p,C_1}$ to obtain the adapted estimator of the state vector $\hat{x}_{t,c} = \hat{\mathcal{K}}_{p,C_1}Y^-_{t,p}, t = p+1,\ldots,T+1$.

5) Use, as in item 4) of the standard `CCA` algorithm, the system equations to obtain estimators $(\hat{A}_c, \hat{K}_c, \hat{C}_c)$ of the system matrices.

For stationary processes, i.e. when $r = s$ and thus $c = 0$, the adapted procedure coincides with the standard `CCA` procedure.

In the algorithms described above, the matrix $\hat{A}_c$ is obtained by regressing $\hat{x}_{t+1,c}$ on $\hat{x}_{t,c}$. This being an unrestricted regression, there is no guarantee that $\hat{A}_c$ is similar to a matrix of the form given in (1). Typically due to the consistency property stated below, the c largest eigenvalues of the estimated $\hat{A}_c$ will be close to 1 but not identically equal to 1. The estimation step can however be easily modified to deliver a system that is exactly cointegrated. A reduced rank regression approach delivers the required result: Assume again that there are c stochastic trends in y_t, then the rank of the matrix $(A - I_n)$ is given by $n - c$. Thus, alternative estimators $\tilde{A}_c$ and $\tilde{K}_c$ can be obtained from a reduced rank regression

$$\hat{x}_{t+1,c} - \hat{x}_{t,c} = (\tilde{A}_c - I_n)\hat{x}_{t,c} + \tilde{K}_c\hat{\varepsilon}_{t,c} + r_t \tag{8}$$

under the constraint that $\text{rank}(\tilde{A}_c - I_n) = n - c$. This approach results by construction in an estimated system that is exactly cointegrated with the specified cointegrating rank.[7] In order to distinguish the two estimation approaches for the system matrices, the latter approach using (8) is referred to as *reduced rank regression* approach and the least squares method for

[7]Note the similarity of the above reduced rank regression problem to the regression problem considered by Johansen (1995).

obtaining estimators of A and K is called *unrestricted regression approach.*

The asymptotic properties of these algorithms are cited from Theorems 2 and 3 in Bauer and Wagner (2002b):

Theorem 1 *Let the s-dimensional output y_t be generated according to a system of the form (1) with the ergodic noise ε_t fulfilling assumptions (2). Assume that the true order n of the transfer function $k(z)$ is known. Concerning the indices f and p the following assumptions are made: $f \geq n$ is fixed and $p = p(T) = o((\log T)^a)$ for some $0 < a < \infty, p(T) \geq -d \log T / \log |\lambda_{max}(A - KC)|, d > 1$. Given the true number of stochastic trends c, the standard subspace algorithm results in consistent estimation of order T of the cointegrating space as follows: Denote by $\hat{C}_1$ the matrix of the first c columns of $\hat{C}$. Then $T^\gamma (C_1^\perp)'(\hat{C}_1 - C_1) \to 0$ in probability for $0 < \gamma < 1$.*

Assuming that c is correctly specified and that the adapted subspace procedure is used with an estimate $\hat{C}_1$ consistent of order T, the estimate $\hat{k}(z) = I_s + z\hat{C}_c(I_n - z\hat{A}_c)^{-1}\hat{K}_c$ obtained using the adapted subspace algorithm, converges in probability to the true transfer function $k(z)$. The consistency result also applies to the reduced rank regression approach.

Minimization of either order estimation criterion $SVC(n)$ or $BA(n)$ leads to weakly consistent order estimation for $H_T/(p(T) \log \log T) \to \infty$ and $H_T/T \to 0, H_T > 0$.

Note again that in (Theorem 3 of) Bauer and Wagner (2002b) only the order estimation criterion $SVC(n)$ has been discussed.

In the above discussion the dimension of the cointegrating space, or equivalently the number of stochastic trends, has been assumed to be known or specified by the user. In order to make the approach useful for practical purposes, tests for the number of stochastic trends have to be discussed next. Such tests can be based both on the estimated singular values of $\hat{W_f^+}\hat{\beta}_{f,p}\hat{W_p^-}$ and on the eigenvalues of the matrix $\hat{A}_c$. Let us start with the singular value based test: It can be shown that for a system of order n and with c stochastic trends, the largest c estimated singular values $\hat{\sigma}_1 \geq \ldots \geq \hat{\sigma}_c$ converge to 1 at rate T, whereas the following $n - c$ singular values $\hat{\sigma}_{c+1} \geq \ldots \geq \hat{\sigma}_n$ converge to their positive limits smaller than 1 at rate $T^{1/2}$. In Bauer and Wagner (2002b, Theorem 4) the limiting distribution of $T(1 - \frac{1}{c}\sum_{j=1}^{c} \hat{\sigma}_j^2)$ is derived. It turns out that this limiting distribution depends upon C_1, K_1 and Ω. For any given value of c, the system can be estimated with the adapted `CCA` algorithm as described above, to obtain consistent estimators of C_1, K_1 and Ω, which can then be inserted in the test statistic in order to derive the critical values. The dependence of this test on nuisance parameters is

also its main drawback. This drawback could in principle be overcome, or at least mitigated, by bootstrapping the test statistic (see e.g. Bauer and Wagner, 2000). In the present paper, this nuisance parameter dependent test is not investigated further. The asymptotic behavior of the estimated singular values can however also be used in a simple fashion for obtaining a (possibly rough) *estimator* of the number of stochastic trends. This proceeds analogously to the order estimation, where a threshold is applied to identify significantly non-zero estimated singular values. We exploit the above mentioned fact that the first c estimated singular values converge to 1 at rate T. Take as an *estimator* of the number of stochastic trends the largest integer, $\hat{c}$ say, such that the $\hat{c}$-th singular value $\hat{\sigma}_{\hat{c}}$ is the smallest one for which $\hat{\sigma}_{\hat{c}}^2 > 1 - h_T/T$ holds, with $h_T \to \infty$ and $h_T/T^{1/4} \to 0$ as $T \to \infty$.[8] This threshold based estimator is not advocated to be used without either the singular value based test or some of the eigenvalue based tests described below. It is simply intended to give a first guess concerning an upper bound for the number of stochastic trends.[9]

Nuisance parameter free tests can be based on the eigenvalues of the estimated matrix $\hat{A}_c$ cf. Theorem 5 in Bauer and Wagner (2002b):

Theorem 2 *Let the assumptions of Theorem 1 hold and let the number of stochastic trends be denoted by c. Assume that the adapted subspace procedure under the hypothesis of a correctly specified number of stochastic trends is used.*
Then the asymptotic distribution of the largest c eigenvalues of $T(\hat{A}_c - I_n)$ *is equal to the distribution of the c eigenvalues of* $\int_0^1 W(u)dW(u)'(\int_0^1 W(u)W(u)'du)^{-1}$*, where* $W(u)$ *denotes the standard c-dimensional Brownian motion.*

The idea of these tests is inspired by Stock and Watson (1988) and as in that paper a variety of possibilities to construct tests arises. The tests for the null hypothesis of c stochastic trends can be based on either the c-th largest eigenvalue alone, or on all the c largest eigenvalues, e.g. on their sum. Also, it can be the case that some estimated eigenvalues are complex, thus the issue of whether to take the *real parts* or the *absolute values* in order to define the precise meaning of 'close to 1' arises. This leaves us with four possibilities to construct tests, presented in Table 1. The tests are in fact based on the eigenvalues of $(\hat{A}_c - I_n)$, denoted by

[8] The specific choice of h_T influences the finite sample behavior of this estimator of the number of stochastic trends.

[9] As the singular values are computed within the procedure anyway, essentially no additional computational costs are involved when performing this threshold estimation of the number of stochastic trends. Note that only the estimated singular values from the standard `CCA` algorithm are used.

Test Name, Nr.	$\mu_{c,re}$, I	$\mu_{\sum_c,re}$, II	$\mu_{c,abs}$, III	$\mu_{\sum_c,abs}$, IV
Test Stat.	$T\text{re}(\hat{\mu}_c)$	$T\sum_{i=1}^{c}\text{re}(\hat{\mu}_i)$	$T\text{abs}(\hat{\mu}_c)$	$T\sum_{i=1}^{c}\text{abs}(\hat{\mu}_i)$

Table 1: The four tests based on the eigenvalues of the matrix $\hat{A}_c$. Under the null hypothesis of c stochastic trends, the first c columns of the `CCA` estimator of C are chosen as C_1 and used for the construction of the modified weighting matrix $\widehat{W^+_{f,C_1}}$. *re* denotes the real part of a (possibly) complex number and *abs* denotes the absolute value. In the text the tests are mainly referred to by their numbers I to IV.

$\hat{\mu}$.[10] Given that explosive systems are not of great concern, the alternative hypothesis is, e.g. for test $\mu_{c,re}$ (I), that the real part of the eigenvalue $\hat{\mu}_c$ is *smaller* than 0. This corresponds to $re(\hat{\lambda}_c) < 1$, with $\hat{\lambda}$ denoting the eigenvalues of $\hat{A}_c$. Thus, the null hypothesis is rejected, if the test statistic is smaller than the corresponding critical value from the simulated asymptotic distribution. Test $\mu_{\sum_c,re}$ (II) is performed exactly analogously.

For the tests based on the absolute values, the null hypotheses that the eigenvalue λ_c is equal to 1 respectively that the eigenvalues $\lambda_i, i = 1, \ldots, c$ are all equal to 1, are rejected if the test statistics $\mu_{c,abs}$ (III) or $\mu_{\sum_c,abs}$ (IV), respectively, are *larger* than the corresponding critical values. In the construction of the test statistic $\mu_{c,abs}$ the relevant eigenvalue to compare with is the absolute value of the largest eigenvalue of the c-dimensional functional of Brownian motions given above. Critical values for the 4 proposed tests are given in the appendix.

For all proposed tests the recursive testing sequence is carried out as follows: Choose or determine an upper bound for the number of stochastic trends. One possible upper bound is to take the maximum possible number of stochastic trends, which is given by the minimum of the system order and the dimension of y_t. Alternatively also the threshold estimator for the number of stochastic trends based on the singular values can be chosen as an upper bound. To make the latter approach feasible, it is necessary to specify the threshold function h_T in such a way that the probability of underestimating the correct number of stochastic trends is small. This is achieved by taking large values for h_T. When using this idea to arrive at the initial upper bound for the number of stochastic trends, in the simulations $h_T = (\log T)^2$ is chosen. For the given initial upper bound value of c, estimate the system using the adapted subspace algorithm. If the null hypothesis of c stochastic trends is rejected, then the test sequence is continued with the null hypothesis of $(c-1)$ stochastic trends and so on. The

[10] The various tests described in the table are computed on the ordering based on either the real parts or the absolute values.

sequence is stopped, when the null hypothesis cannot be rejected anymore, and of course also after rejecting the null hypothesis $c = 1$.

It seems tempting to investigate also different approaches to test for the number of cointegrating relationships. The mentioned Theorem 5 of (Bauer and Wagner, 2002b) basically shows that the replacement of the state x_t by a suitably defined estimator $\hat{x}_{t,c}$ does not change some of the usual asymptotics. In this respect a very natural idea is to replicate the Johansen procedure on the state equation for the estimated state.

In the present context the Johansen procedure is very simple. As the state equation is an autoregression of order one, it simply amounts to a computation of the canonical correlations between $\Delta\hat{x}_{t,c}$ and $\hat{x}_{t-1,c}$. Thus, for the n-dimensional state, the null hypothesis of c stochastic trends can also be tested by performing a Johansen type cointegration test on the state equation with the null hypothesis of $(n-c)$ linearly independent cointegrating relationships. This observation gives rise to two additional tests, replicating the Johansen trace test (test number V) and the Johansen max test (test number VI). The test sequence is performed in a similar manner to the test sequences described above. Start with an initial null hypothesis of c stochastic trends and compute the adapted estimator of the state under this null hypothesis. Then use the Johansen approach for testing the null of $(n-c)$ cointegrating vectors. If the null hypothesis is rejected, re-estimate the system under the null of $(c-1)$ stochastic trends, and perform the Johansen test for the presence of $(n-c+1)$ cointegrating relationships. Iterate until the null hypothesis cannot be rejected anymore, or the null hypothesis $c = 1$ is rejected. Note that a difference to a standard Johansen application in a VAR is that after each step of the testing sequence the system has to be re-estimated, since in the course of the test procedure the adapted algorithm has to be applied for the decreasing number of stochastic trends under the sequence of null hypotheses until rejection occurs.

Now all the ingredients for estimation and testing are collected, we can thus proceed to an investigation of the finite sample properties of the proposed estimation and test procedures.

4 Simulations

In this section we investigate the properties of the proposed estimators and tests. All simulation results are compared with the results obtained by applying the Gaussian ML cointegration analysis for VAR models summarized in Johansen (1995). As the Johansen approach is the workhorse in the literature, it is natural to compare our results to the results obtained

with this method. Although we are simulating VARMA processes, the Johansen approach still leads, as already mentioned in the introduction, to consistent estimators of the transfer function $k(z)$ and asymptotically correct tests, when the lag length of an autoregressive approximation is growing at a sufficient rate with the sample size (see Saikkonen, 1992). The Saikkonen result requires, strictly speaking, the lag length to increase with an exogenous function of the sample. Nevertheless we base our VAR results on lag lengths chosen according to `AIC`, denoted by $\hat{p}_{AIC}$.

We are interested in four aspects: The properties of the order estimation criterion, the size properties of the six proposed tests, the estimation quality of the cointegrating spaces and the forecasting performance of the estimated state space models. With respect to the tests we are interested in the actual size as a function of the sample size and in the sensitivity of the actual size with respect to some stable eigenvalues approaching the unit circle. Given that the tests are based on different quantities – only one eigenvalue, sums of eigenvalues, real parts or absolute values – it may be expected that the performance of the various tests differs for instance when some stable eigenvalues are close to 1. To assess the *quality* of the tests, the results are compared with the Johansen trace and max test results. In the graphs below we will report the percentage of replications in which the test sequences lead to a correct decision concerning the cointegrating rank, this percentage will be called *hit rate*. All individual test steps are performed at nominal size 5%.

The next aspect investigated is the approximation quality of the estimated cointegrating space to the true cointegrating space. For any method of cointegration analysis this is obviously a prime issue. The measure of quality employed is the *gap* between the true and the estimated cointegrating space. The gap is defined as follows: Let M and N denote two linear subspaces of $\mathbb{R}^s$, then the gap $d_H(M,N)$ is given by

$$d_H(M,N) = \max\left(\sup_{x\in M, \|x\|=1} \| (I-Q)x \|, \sup_{x\in N, \|x\|=1} \| (I-P)x \| \right)$$

where Q denotes the orthogonal projection onto N, P the orthogonal projection onto M and $\| x \|$ denotes the Euclidean norm on $\mathbb{R}^s$. The gap is between zero and one and it is e.g. equal to one for spaces of different dimensions. In our simulations we compute the gap between the estimated cointegrating space and the true cointegrating space exclusively under the assumption of the correct specification of the cointegrating rank. This separates the properties of the tests from the estimation properties for the cointegrating space. Three different gaps are

compared, the gap between the initial subspace estimator and the true cointegrating space, the gap between the adapted subspace estimator and the true cointegrating space and the gap between the VAR Johansen estimator and the true cointegrating space.

Finally we investigate the forecasting performance. The forecasting performance is not only interesting in itself, but also for the following reason: The asymptotic distribution of the subspace estimators is unknown, therefore the relative accuracy – as compared to the ML estimators in a state space framework – of the subspace estimators is unclear. If the forecasting performance were worse than for a VAR approximation, this would be an indication for relative (finite sample) inaccuracy of the estimators. When the estimators are relatively accurate, then – as we are estimating state space models – the forecast performance of the state space models should be (marginally) better than the corresponding accuracy for a VAR approximation. The comparison depends crucially upon the approximation accuracy of a low order VAR with respect to the underlying transfer function. This in turn depends upon the zeros of the transfer function, i.e. upon the eigenvalues of $(A - KC)$. Let ρ_0 denote an eigenvalue of maximum modulus of $(A - KC)$, which is due to the strict minimum-phase assumption bounded in absolute value to be smaller than one. Then it can be shown that the order of an autoregressive approximation of the underlying transfer function needed in order to achieve a certain rate of convergence to the true transfer function is an increasing function of $|\rho_0|$. This fact is contained e.g. in the discussion on the properties of `BIC` order estimators in a stationary context in Theorem 6.6.3 of Hannan and Deistler (1988).

Another interesting issue with respect to the forecasting performance is to compare the forecasts derived from the unrestricted regression approach with the forecasts derived from the reduced rank regression approach delivering an exactly cointegrated system. This issue is similar to comparing forecasts from VARs estimated in levels, differences or as error correction models. Intuitively we expect a superior forecasting performance from the models that impose the correct cointegrating structure, at least for our simulated data.

At this point it may be worthwhile to clarify a further computational detail. Given estimated system matrices $(\hat{A}, \hat{K}, \hat{C})$ and an estimator $\hat{x}_{t_0}$ of the initial state x_{t_0} at an initial time point $t_0 \leq T + 1$, the estimated state at time $t = t_0 + 1, \ldots, T + 1$ is given as

$$\hat{x}_t = (\hat{A} - \hat{K}\hat{C})^{t-t_0}\hat{x}_{t_0} + \sum_{j=0}^{t-t_0-1} (\hat{A} - \hat{K}\hat{C})^{t-t_0-j-1}\hat{K}y_{t_0+j}, \quad t = t_0 + 1, \ldots, T + 1.$$

Forecasting itself is then simple, as the best linear prediction for y_{T+h} is, as discussed in Section 3, given by

$$y(T+h|T+1) = \hat{C}\hat{A}^{h-1}\hat{x}_{T+1}, \quad h = 1, 2, \ldots$$

using again the notation and timing convention of Section 3. The choice of t_0 and $\hat{x}_{t_0}$ influences the properties of the forecasts. Three possibilities are $t_0 = 1, \hat{x}_1 = 0$, $t_0 = T+1, \hat{x}_{T+1}$ obtained from the subspace approach, and $t_0 = T-f+1, \hat{x}_{T-f+1}$ obtained from the subspace approach. Somewhat surprisingly, the latter choice led to the best forecasting accuracy throughout a variety of simulations. Thus, in this paper only the results based on this choice are reported. The forecast accuracy is compared to the results obtained by fitting VAR models to the data with the lag length chosen according to `AIC`. We compute forecasts for forecasting horizons $h = 1, \ldots, 8$ and compare them with actual observations. We compute the forecast errors and the root mean squared errors. The latter are defined for each coordinate i and for each forecasting horizon h as:

$$RMSE_i(h) = \sqrt{\frac{1}{h}\sum_{j=1}^{h}(y_i(T+j|T+1) - y_{T+j,i})^2}, \quad i = 1, \ldots, s, \quad h = 1, \ldots, 8,$$

where y_{T+h} are the given out of estimation sample values simulated from the underlying process.

In total six different forecasts are computed and compared, three state space model- and three VAR forecasts: For the subspace approach we compute forecasts based on the adapted estimator derived from the unrestricted regression approach (SUB), from the reduced rank regression subspace estimator (RRR) and from a subspace estimator based on the first difference of the data (SUB-D). For the VAR approximation the three forecasts are based on a VAR estimated in levels (LEV), on a VAR estimated in differences (DIFF) and on an error correction model (ECM). To separate the forecasting properties from the properties of the testing procedures, the forecasts SUB, RRR and ECM are generated from state space respectively VAR models with the correct number of cointegrating relationships.[11]

The listed issues lead to a large number of results. Thus, in the paper only some illustrative results and the main messages can be reported. More detailed results are available upon request. The simulations have been performed using `MATLAB`, the sample sizes are $T = 100, 200, \ldots, 500$ and the number of replications per system and sample size is 1000. In all simulations the

[11] The results combining the testing problem with the forecasting problem are available upon request.

indices f and p are chosen as $2\hat{p}_{AIC}$, where $\hat{p}_{AIC}$ denotes the lag length chosen to minimize the AIC of an autoregressive approximation. The lag lengths for the estimated VAR models are chosen equal to $\hat{p}_{AIC}$. The sequential testing procedures for the cointegrating rank in the VAR are performed as described in detail e.g. in Johansen (1995).

In the following two subsections first the simulation results for some three-dimensional systems are discussed and then the results for a number of four-dimensional systems are reported. The set of three-dimensional systems is designed to assess the effect of stable real valued eigenvalues approaching 1 and to assess whether the performance of the methods is affected by the number of stochastic trends. The four-dimensional systems all have a 2-dimensional cointegrating space, and we use them to assess the sensitivity of subspace algorithm cointegration analysis with respect to complex conjugate eigenvalues approaching the unit circle. This situation is expected to be more problematic, as complex eigenvalues with absolute value one also introduce singular values equal to one.

4.1 Three-Dimensional Systems

The set of three-dimensional systems simulated is based on Saikkonen and Luukkonen (1997) and on Bauer and Wagner (2002b). Compared to the previous investigations we extend the set of systems, in order to assess the sensitivity of the results concerning some eigenvalues being close to 1 but not equal to 1. The systems are given by eleven VARMA(1,1) processes:

$$\Delta y_t = \Psi y_{t-1} + \varepsilon_t - \Gamma_1 \varepsilon_{t-1} \tag{9}$$

with $y_0 = \varepsilon_0 = 0$ and ε_t normally and independently distributed $N(0, \Omega)$. This corresponds to the assumption of a zero initial state. The parameter matrices are given by $\Gamma_1 = C_\gamma \mathrm{diag}(0.297, -0.202, 0) C_\gamma^{-1}$, where

$$C_\gamma = \begin{pmatrix} -0.816 & -0.657 & -0.822 \\ -0.624 & -0.785 & 0.566 \\ -0.488 & 0.475 & 0.174 \end{pmatrix} \tag{10}$$

$$\Omega = \begin{pmatrix} 0.47 & 0.20 & 0.18 \\ 0.20 & 0.32 & 0.27 \\ 0.18 & 0.27 & 0.30 \end{pmatrix} \tag{11}$$

and $\Psi = N \mathrm{diag}(\phi_1, \phi_2, \phi_3) N^{-1} - I_3$ with

$$N^{-1} = \begin{pmatrix} -0.29 & -0.47 & -0.57 \\ -0.01 & -0.85 & 1.00 \\ -0.75 & 1.39 & -0.55 \end{pmatrix} \tag{12}$$

System	1	2	3	4	5	6	7	8	9	10	11
ϕ_1	0.9	0.95	1	1	1	1	1	1	1	1	1
ϕ_2	0.8	0.9	0.8	0.85	0.9	0.95	1	1	1	1	1
ϕ_3	0.7	0.85	0.7	0.75	0.8	0.85	0.7	0.8	0.9	0.95	1

Table 2: Parameter values ϕ_i for the three-dimensional systems 1 to 11.

	System 1				System 3				System 6			
Order	1	2	3	4	1	2	3	4	1	2	3	4
T = 100	0.07	0.22	0.69	0.01	0.01	0.22	0.74	0.03	0	0.02	0.93	0.05
T = 200	0.00	0.01	0.96	0	0	0.01	0.96	0.03	0	0	0.97	0.03
T = 300	0.00	0.00	0.99	0.01	0	0	0.99	0.01	0	0	0.98	0.02
T = 400	0	0	0.99	0.01	0	0.00	0.99	0.00	0	0	0.99	0.01
T = 500	0	0	1	0	0	0	0.99	0.01	0	0	1	0.00
	System 7				System 10				System 11			
Order	1	2	3	4	1	2	3	4	1	2	3	4
T = 100	0.00	0.21	0.76	0.03	0.00	0.02	0.91	0.08	0	0.02	0.93	0.05
T = 200	0	0.01	0.96	0.03	0	0	0.95	0.05	0	0	0.96	0.04
T = 300	0	0	0.98	0.02	0	0	0.98	0.02	0	0	0.98	0.02
T = 400	0	0	0.99	0.01	0	0	0.98	0.02	0	0	0.98	0.02
T = 500	0	0	0.99	0.01	0	0	0.99	0.01	0	0	1	0.00

Table 3: Distribution of the results of the order estimation using the criterion $BA(n)$ with threshold function $H_T = \log T$ for systems 1, 3, 6, 7, 10 and 11.

The parameters ϕ_i are presented in Table 2. The number of parameters $\phi_i = 1$ corresponds to the number of stochastic trends. Thus, we simulate two stationary processes, systems 1 and 2, four systems with one stochastic trend (systems 3 to 6), four systems with two stochastic trends (systems 7 to 10) and one system with three stochastic trends, i.e. an integrated system without cointegration (system 11). All systems are strictly minimum-phase and for all systems $\rho_0 = |\lambda_{max}(A - KC)| = 0.297$. As the systems are only varying with respect to the parameters ϕ_i, which correspond to the poles of the transfer function, it is clear that the zeros of the transfer functions (corresponding to the roots of the moving average polynomial of the ARMA system) are identical for all eleven systems.

In subspace cointegration analysis the first step in the analysis is the estimation of the system order, n. This is crucial, because it imposes also bounds on the possible number of stochastic trends and therefore also on the number of cointegrating relationships in the system. In Table 3 the results concerning the estimated orders of the systems using $BA(n)$ are presented for the various sample sizes for a selection of systems that includes the systems with the worst performance (systems 1 and 3) of the order estimation. It becomes clear from Table 3 that

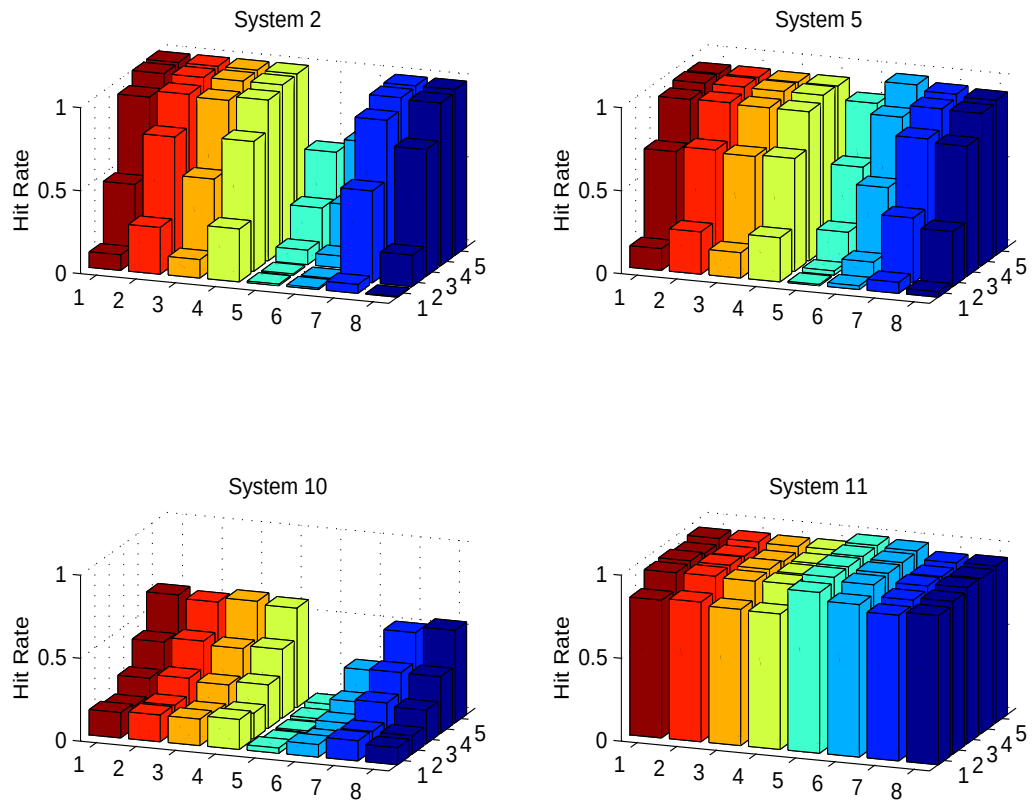

Figure 1: Hit rates of all six proposed tests for the number of stochastic trends (1 to 6) and the Johansen trace and max tests (7 and 8) for all sample sizes $T = 100, \ldots, 500$ from the front to the back for systems 2, 5, 10 and 11.

only for $T = 100$ the true orders are somewhat under-estimated. For all other sample sizes the order estimation turns out to be very accurate, providing in more than 95% of the cases the true order. Only for $T = 100$ the estimated order is smaller than the number of common trends in some replications, and even then only in 2% of the replications. Hence the order estimation is in no conflict with the estimation of the cointegrating rank.

Let us next turn to a comparison of the different tests. For a subset of the simulated systems the results are displayed in Figure 1. The recursive tests are started at the maximum possible number of 3 stochastic trends. A few observations, which hold also true for the systems not displayed in the figure, can be made. First of all, at least for the smaller sample sizes, the two tests based on the sum of the eigenvalues, either on the real parts (test II) or on the absolute values (test IV), have higher hit rates than tests I and III based on only one eigenvalue. This is especially pronounced for systems 1 to 6, having zero or one stochastic trend. This

difference is due to the smaller probability of overestimating the number of stochastic trends for the tests based on the sums. For the larger sample sizes, the performance of these four tests is more or less identical. Tests V and VI prove clearly inferior in these simulations: The hit rates of tests V and VI for sample size $T = 100$ are clearly smaller than for the tests I to IV in most cases. Only for system 11 is the hit rates are almost 100 % for all sample sizes for both test V and VI. This has to be contrasted with the performance on system 10, where even for sample size $T = 500$ the better of the two tests achieves only a hit rate of 27%, which is to be compared to 57% for the better Johansen test on the approximating VAR models and approximately 60 % for test IV. Thus, the tests constructed to replicate the Johansen approach on the state equation with the estimated state are not recommendable based on these simulations.

System 11 and sample size $T = 100$ is the only case, where the results obtained by applying Johansen's tests on a VAR approximation are superior to the results obtained with an application of the subspace tests. The difference is not pronounced, however, since the best Johansen test results in approximately 90% correct decisions, whereas the worst of the tests I to IV achieves 82 %. This difference is small compared to the gain achieved using the subspace tests for the remaining systems. Figure 2 compares the hit rates for sample size $T = 100$ for the Johansen trace test, test IV started at the output dimension as the first null hypothesis and test IV started at the singular value based estimate of the number of stochastic trends (IV, thresh). The picture clearly shows the superiority of test IV for all but the eleventh system. It can also be seen that the performance of all procedures and tests degrades, as expected, when stable eigenvalues of A tend to the unit circle. The starting point of the test sequence, 3 or the threshold estimator of the number of stochastic trends results in almost no difference for the systems with zero or one stochastic trend. For systems 8, 9 and 10 however, starting the testing sequence from the threshold estimator of the number of stochastic trends is beneficial. This comes at the price of smaller performance for system 11 having three stochastic trends. Looking at Figure 2 one notices that the sum of the hit rates for system 10 and system 11 is approximately one for all tests. This is due to the fact that in almost all replications for these two systems the tests result in either one cointegrating vector or none. Since these two systems are very close, a test cannot have a high power for systems 10 and 11 simultaneously. The observed results therefore indicate the underlying size/power tradeoff. None of the tests I to IV has a uniformly better size performance. In most cases the tests based on the sums

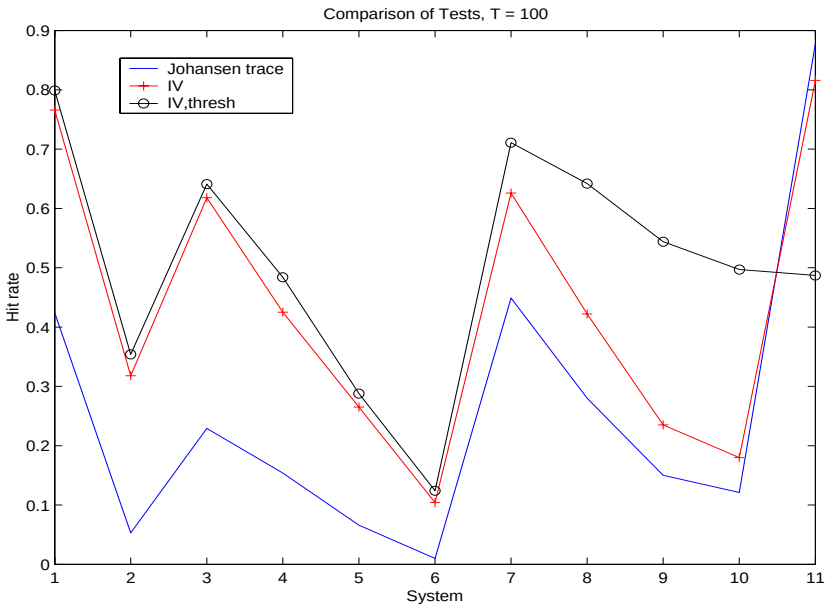

Figure 2: Hit rates for $T = 100$ and all 11 systems. The three tests shown are the Johansen trace test, IV started at the output dimension (i.e. at 3) and the test IV started at the singular value based estimator of the number of stochastic trends (IV, thresh).

of the eigenvalues have a higher hit rate for small sample sizes like $T = 100$. Note again that tests V and VI seem to be, compared to tests I to IV, too sensitive with respect to some stable eigenvalues close to the unit circle. Note, that all tests except for tests V and VI approach the theoretical size for the sample size growing to $T = 500$ very accurately.

The next point on the list is the gap between the true and the estimated cointegrating spaces. In Figure 3 we display density estimators of the logarithms of the gaps between the true and the estimated cointegrating spaces. Three gaps are plotted: The gap between the initial subspace estimator and the true cointegrating space (solid), the gap between the adapted subspace estimator and the true cointegrating space (dashed) and the gap between the Johansen estimator and the true cointegrating space (dash-dotted). The results are displayed from left to right for systems 3 and 6 with 2-dimensional cointegrating spaces and systems 7 and 10 with 1-dimensional cointegrating spaces. These systems are selected because they represent the boundary cases in terms of the magnitude of the stable eigenvalue(s), see Table 2. Qualitatively the same conclusions emerge also for the 'intermediate' systems. The figure is constructed such that the modulus of the stable eigenvalues, respectively eigenvalue, increases from left to right and the sample size increases from top to bottom. These and

all other density estimators presented in this paper are based on a Gaussian kernel with the bandwidth chosen according to Silverman's rule of thumb. The estimated mean of the logarithm of the gap for systems 5, 8 and 9 and all sample sizes are provided in Table 4. A couple of features are clearly visible: The closer the stable eigenvalue(s) tend(s) to 1, the less precise is the estimation of the cointegrating space. This holds true for all methods and is quite pronounced for the smaller sample sizes. The initial subspace estimator shows the largest variances and is most affected by stable eigenvalues close to 1.[12]

The second observation is that the relative performance of the initial and the adapted subspace estimators depends upon the dimension of the cointegrating space. Hence, let us discuss the two cases separately and start with the systems with 2-dimensional cointegrating spaces. For these systems the results for the adapted subspace estimator and the Johansen estimator are quite close to each other. For the smaller sample sizes, the adapted subspace estimators have a smaller mean log gap than the Johansen estimators and for the larger sample sizes the estimators result in essentially identical results, also with respect to robustness when stable eigenvalues tend to 1. Again, for all systems with 2-dimensional cointegrating spaces and for all sample sizes the initial subspace estimator exhibits the largest mean log gap (cf. Table 4). For the systems with the 1-dimensional cointegrating space, the main messages are the same as above: There is substantial sensitivity when the single stable eigenvalue tends to 1, especially for the smaller sample sizes. The adapted subspace estimator is least affected and again the initial subspace estimator is the most sensitive one, at least for small samples. However, and this is surprising, for these systems it turns out that the initial subspace estimator performs comparable to the Johansen estimator, e.g. in terms of the mean log gap. These two, the initial subspace estimator and the Johansen estimator, show better properties than the adapted subspace estimator. Only in the last column, corresponding to system 10, has the adapted subspace estimator the smallest mean log gap to the true cointegrating space.

Thus, as already mentioned above, it seems to be the case that the dimension of the cointegrating space has some influence on the relative performance of the initial and the adapted subspace estimators of the cointegrating space. Throughout all simulations however the adapted estimator is less sensitive with respect to eigenvalues close to 1 than the initial subspace estimator. The performance is quite comparable to the Johansen results, in some cases one of

[12] We have also computed mean, median, variance, minimum and maximum of the distributions of the Log gap to have a clear quantitative picture of the differences. Especially for the bigger sample sizes density estimation smooths the (very small) differences remaining.

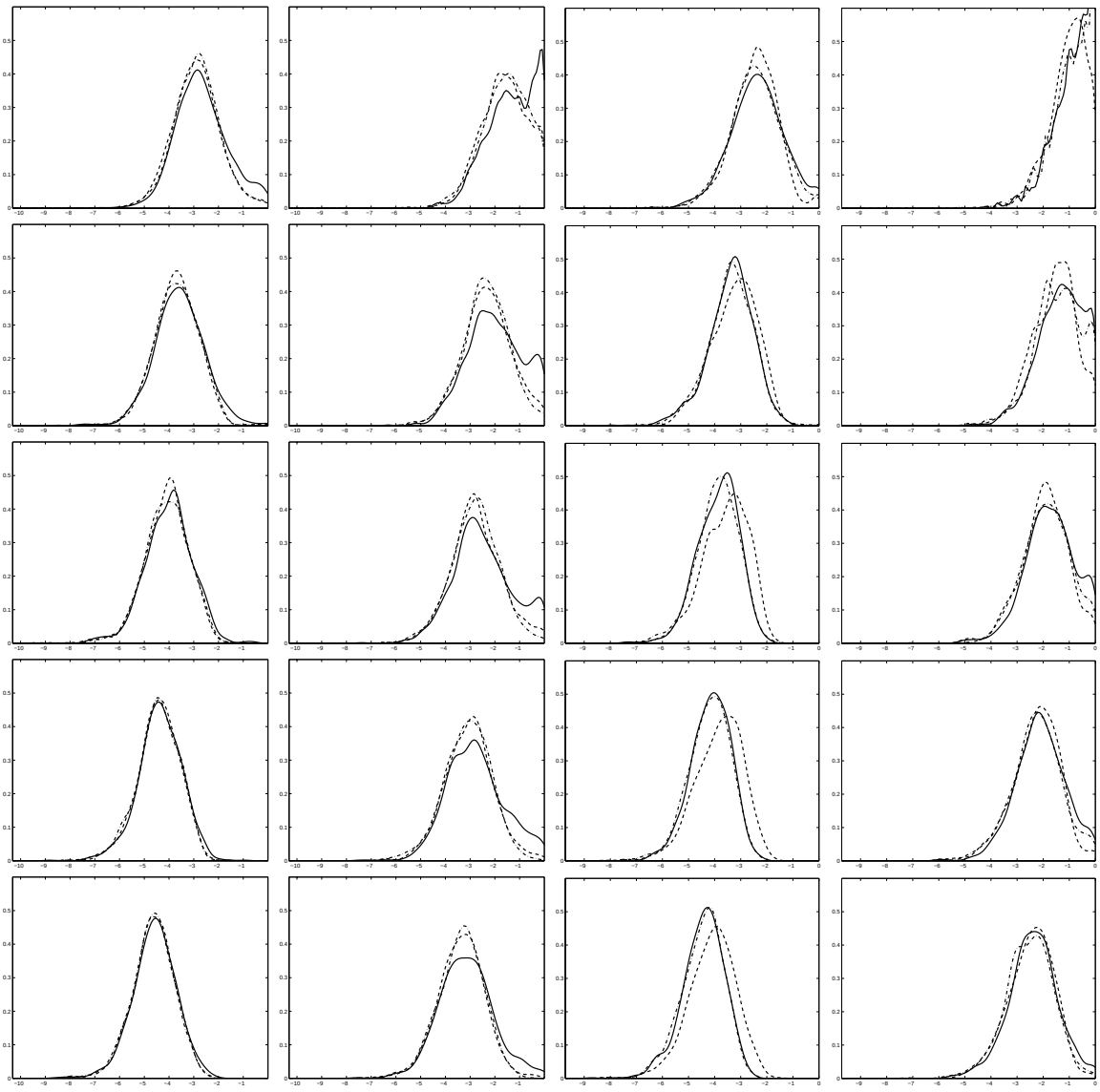

Figure 3: Density plots of Log gaps between true and estimated cointegrating spaces for systems 3 and 6 with 2-dimensional cointegrating spaces, and systems 7 and 10 with 1-dimensional cointegrating spaces. Ordered from system 3 to 10 along columns and along rows for $T = 100, \ldots, 500$. The solid lines correspond to the *initial* subspace estimators, the dashed lines correspond to the *adapted* estimators and the dash-dotted lines correspond to the Johansen estimators derived from a VAR approximation.

	System 5			System 8			System 9		
Estimator	init	adap	VAR	init	adap	VAR	init	adap	VAR
T = 100	-1.9403	-2.2084	-2.1938	-1.9391	-2.1196	-2.0716	-1.2568	-1.4503	-1.3612
T = 200	-2.8033	-3.0641	-3.0860	-2.9064	-2.8035	-2.9614	-2.0821	-2.1749	-2.1238
T = 300	-3.2753	-3.4689	-3.4978	-3.4169	-3.2450	-3.4609	-2.5779	-2.6108	-2.6625
T = 400	-3.6853	-3.8389	-3.8567	-3.7393	-3.4552	-3.7838	-2.9845	-2.8786	-3.0492
T = 500	-3.9417	-4.0494	-4.0722	-3.9589	-3.7663	-4.0232	-3.2359	-3.1275	-3.2982

Table 4: Mean of the logarithms of the gaps for the systems 5, 8 and 9 and all sample sizes. The three estimators are the initial subspace estimator (init), the adapted subspace estimator (adap) and the Johansen estimator (VAR). All numbers are based on 1000 simulated time series. The variances of the estimated mean log gaps are all smaller than 1.35. Therefore an upper bound for the standard deviations of the estimated mean log gaps is 0.037.

the subspace estimators performs marginally better.

The final dimension along which the properties of subspace cointegration analysis is analyzed is the forecasting performance. As mentioned already, six different forecasts are computed and compared. In these experiments the orders of the state space systems are restricted to be bigger or equal than 3. Except for sample size $T = 100$ the best state space model forecasts are derived from the RRR model and the best VAR forecasts stem from the error correction model, see e.g. Table 5 for systems 5 and 8 for $T = 100$ and 200. For the larger sample sizes the differences become even smaller. The results are qualitatively similar for all simulated processes. For $T = 100$ the ECM forecasts are better than the RRR forecasts, measured both by the RMSEs or by looking at the forecast error densities themselves. In many cases for the smallest sample size also forecasts based on first differences show the best performance. For the larger sample sizes the ordering is usually reversed, and the RMSE of the RRR forecasts is slightly smaller than the RMSE of the VAR based ECM forecasts. The results, i.e. the ranking of the different forecasts, is stable over forecasting horizons and also across individual coordinates.[13] It should be noted, however, that the differences are very small and far from being statistically significant and hence the estimates can be considered to be equally good. The fact that the best results are obtained with RRR and ECM is not too surprising. These two forecasts are based on models incorporating the true number of cointegrating relationships. This also is in line with the fact that the better accuracy is most pronounced for the longest forecasting horizon $h = 8$. Density plots of the forecast errors confirm that only for sample size $T = 100$ there seems to be a sizeable difference in the estimation accuracy between the different approaches. Hence we refrain from providing plots.

[13]For this reason we display in the tables only the arithmetic means over the coordinates.

	System 5						System 8					
	$T = 100$											
	SUB	RRR	SUB-D	ECM	LEV	DIFF	SUB	RRR	SUB-D	ECM	LEV	DIFF
$h = 1$	0.576	0.544	0.515	0.506	0.510	0.504	1.153	0.557	0.517	0.506	0.511	0.504
$h = 2$	0.691	0.657	0.629	0.622	0.631	0.617	1.285	0.667	0.629	0.619	0.631	0.616
$h = 3$	0.770	0.730	0.702	0.699	0.711	0.693	1.361	0.736	0.701	0.694	0.712	0.691
$h = 4$	0.835	0.790	0.765	0.760	0.776	0.754	1.422	0.798	0.765	0.760	0.785	0.755
$h = 5$	0.893	0.842	0.818	0.813	0.832	0.809	1.472	0.854	0.822	0.818	0.848	0.813
$h = 6$	0.943	0.885	0.864	0.858	0.881	0.855	1.518	0.902	0.871	0.867	0.906	0.861
$h = 7$	0.990	0.926	0.907	0.901	0.927	0.899	1.559	0.948	0.916	0.914	0.960	0.906
$h = 8$	1.034	0.963	0.947	0.941	0.971	0.940	1.598	0.990	0.957	0.957	1.011	0.948
	System 5						System 8					
	$T = 200$											
	SUB	RRR	SUB-D	ECM	LEV	DIFF	SUB	RRR	SUB-D	ECM	LEV	DIFF
$h = 1$	0.502	0.499	0.507	0.502	0.504	0.505	0.497	0.496	0.500	0.497	0.498	0.498
$h = 2$	0.609	0.606	0.619	0.609	0.612	0.617	0.608	0.604	0.609	0.604	0.608	0.607
$h = 3$	0.676	0.670	0.684	0.673	0.677	0.684	0.691	0.681	0.689	0.684	0.693	0.688
$h = 4$	0.741	0.735	0.749	0.737	0.743	0.748	0.760	0.745	0.755	0.750	0.764	0.757
$h = 5$	0.799	0.791	0.805	0.793	0.801	0.805	0.825	0.803	0.814	0.807	0.829	0.817
$h = 6$	0.851	0.840	0.855	0.841	0.853	0.854	0.882	0.855	0.867	0.858	0.886	0.870
$h = 7$	0.897	0.884	0.900	0.885	0.899	0.899	0.936	0.903	0.916	0.906	0.939	0.919
$h = 8$	0.942	0.925	0.943	0.928	0.944	0.941	0.984	0.946	0.961	0.948	0.987	0.964

Table 5: Comparison of the RMSEs for $h = 1, \ldots, 8$ for $T = 100$ and $T = 200$ for system 5 and system 8 for the six different computed forecasts. The subspace estimators are based on restricting the orders to be greater or equal than 3. For forecasting the RRR and ECM models, the correct number of stochastic trends respectively cointegrating relationships is imposed. The standard deviation of the entries varies between 0.017 and 0.039.

System	1	2	3	4	5	6	7	8
γ	0	0.1	0.2	0.3	0.4	0.5	0.55	0.59
$\lvert 0.8 \pm i\gamma \rvert$	0.800	0.806	0.825	0.854	0.894	0.943	0.971	0.994
ρ_0	0.771	0.735	0.676	0.583	0.428	0.363	0.457	0.522

Table 6: Parameter values γ for the four-dimensional systems 1 to 8 and corresponding absolute value of the pair of complex conjugate eigenvalues $0.8 \pm i\gamma$. ρ_0 denotes the absolute value of the largest eigenvalue of $(A - KC)$.

One motivation for assessing the forecast performance was to assess the estimation efficiency of the subspace estimators. It seems to be the case that for small samples like $T = 100$ the method results in relatively imprecise estimators. Only for $T \geq 200$ start the results to be satisfactory, compared to the Johansen benchmark. It has to be noted again that the state space models estimated with the adapted subspace algorithm do not significantly outperform the approximating VARs in terms of the forecasting performance for any of the sample sizes considered.

4.2 Four-Dimensional Systems

The second set of simulated systems consists of eight state space systems with output and state dimension equal to 4, where a pair of stable complex conjugate eigenvalues is approaching the unit circle. This set-up is chosen to assess the robustness of the methods, especially the tests, with respect to complex conjugate stable eigenvalues tending to the unit circle. The system matrices are given by:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.8 & \gamma \\ 0 & 0 & -\gamma & 0.8 \end{bmatrix}, K = \begin{bmatrix} -0.8777 & -0.7735 & -1.4522 & -1.1218 \\ 0.2587 & 0.1307 & -0.1186 & -0.1913 \\ -0.0700 & 0.1279 & 1.0062 & -0.8179 \\ 0.3500 & -0.0341 & 0.0019 & 0.9628 \end{bmatrix}$$

$$C = \begin{bmatrix} -0.2279 & 1.7240 & -0.1800 & 0.3333 \\ -0.2332 & 0.3520 & 0.2100 & -0.4889 \\ -0.2570 & -0.6560 & 0.3000 & -0.2667 \\ -0.2401 & -1.3160 & -0.3200 & 0.4222 \end{bmatrix}$$

The parameter values chosen for γ and the resulting absolute value of the corresponding pair of complex conjugate stable eigenvalues are given in Table 6. The parameter γ is varied from 0 to 0.59 resulting in absolute values of the stable pair of eigenvalues ranging from 0.8 to 0.994. In the lower row of Table 6 the maximal absolute value of the eigenvalues of $(A - KC)$

is displayed. The innovations ε_t are independently standard normally distributed. For all systems the dimension of the true cointegrating space is given by 2. The following discussion parallels the discussion of the previous subsection, i.e. order estimation, test results, gaps between true and estimated cointegrating spaces and forecasting performance are discussed in turn. The discussion is kept more brief than before as the set-up is identical to the previous subsection and also the results are to a certain extent comparable.

Let us start with order estimation again. Basing the order estimation on the criterion $BA(n)$ results for systems 1 to 3 in devastating estimation of the system order. For systems 1 and 2 and even for sample size $T = 500$ the true order is detected in less than 10% of the simulations, for system 3 the corresponding percentage is approximately 23%. For the remaining systems order estimation becomes better, resulting in more than 90% correct decisions for $T = 500$ for systems 4 to 8. For systems 5 to 8 the order is correctly estimated in more than 70 % of the replications already at sample size $T = 100$. However, the estimated orders are always larger or equal than 2, thus the accuracy of the estimation of the number of stochastic trends, 2 for all systems, is not affected by under-estimation of the order.

Let us next turn to the test performance. In the left plot in Figure 4 the hit rates are displayed for all eight systems and all eight different tests for $T = 100$. The test results are based on systems that have been estimated starting the testing sequence with the null hypothesis of $c = 4$ common trends and where the system orders have been restricted to be greater or equal than four. Starting the test sequence alternatively at the singular value based estimator of the number of stochastic trends does not change the picture qualitatively. To a certain extent the same observations that have been made for the three-dimensional systems can also be made for the four-dimensional systems. The main results - for the smaller sample sizes - can be summarized as follows: Test I degrades significantly for stable eigenvalues of A tending to the unit circle. Test II shows the same behavior, however in a less pronounced way. Tests III and IV seem to be unaffected by stable eigenvalues tending to the unit circle. The Johansen results tend to get better with increasing γ. This feature is shared by tests V and VI, which however perform extremely poor for systems 1 to 4. The tests based on the sums of eigenvalues outperform the tests based only on the largest eigenvalue. In the right plot of Figure 4 we compare the two Johansen VAR tests with the two tests based on the sums of eigenvalues, tests II and IV across systems for $T = 100$. The plot shows that the Johansen approach outperforms the subspace based tests on systems 4 to 8 and is outperformed for

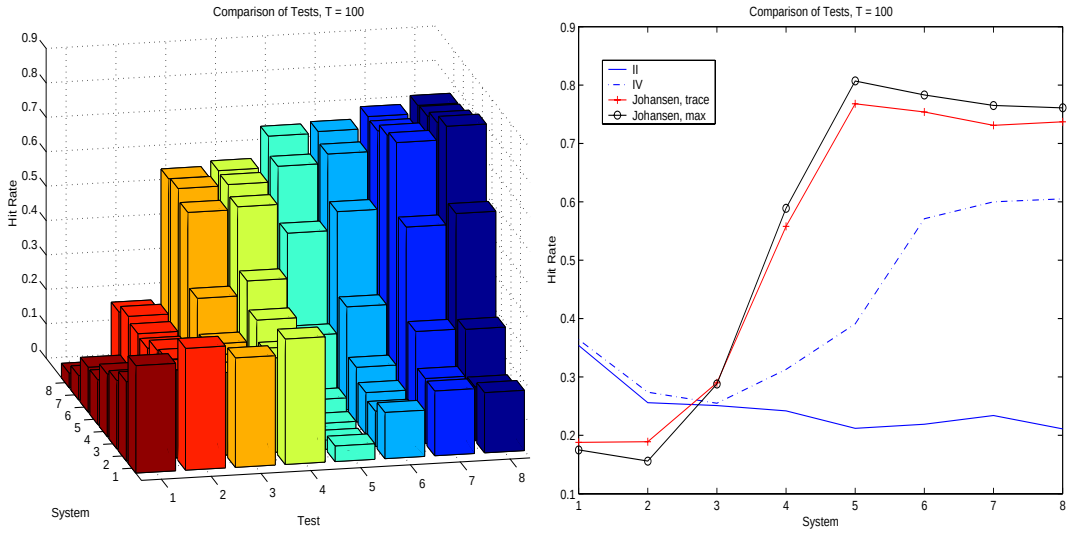

Figure 4: Left plot: Hit rates of all six proposed tests for the number of stochastic trends (1 to 6) and the Johansen trace and max tests (7 and 8) for sample size $T = 100$. From the front to the back the eight systems are arranged. Right plot: For sample size $T = 100$ the hit rates for the two subspace based tests II, IV and the two Johansen tests are shown for all eight systems.

systems 1 and 2. For system 3 all tests perform almost identical

Hence for small samples the results are inconclusive: It may be worthwhile to employ the subspace algorithm cointegration analysis in testing for cointegration, since the tests outperform the Johansen approach in some situations. Roughly speaking, the best performance within tests I to VI is delivered by test IV. Further undocumented investigations lead us to the tentative conclusion that for systems with $|\rho_0|$ close to 1 (in the present context these are the systems with low indices) especially test IV even outperforms the Johansen approach on small samples like $T = 100$. However, more detailed understanding concerning the properties of the various tests has to be gained yet.

For $T = 200$ the first four subspace based tests already are close to the theoretical hit rate achieving a lowest hit rate of 83% (II and IV, system 2), whereas the Johansen approach for system 1 still is quite inaccurate (63 % for the trace test and 67 % for the max test). Thus, for this intermediate sample size the subspace procedure outperforms the Johansen approach.

For $T = 300$ all tests except V and VI hit the correct number of cointegrating relationships in more than 91% of cases, with only small differences. For $T = 400, 500$ the results from all tests except V and VI are not statistically significantly differing from the asymptotic size

0.05 for all systems.

The next issue is again the discussion of the gap between the estimated and the true cointegrating spaces. Figure 5 is set up in the same way as the corresponding picture on gaps in Subsection 4.1. Again the results for four systems are displayed, from left to right systems 1, 3, 5 and 6. The graphs show that the estimation quality of the cointegrating spaces delivered by the subspace methods is quite poor. The results indicate that for larger systems (as the effect is much less pronounced for the three-dimensional systems) and small sample sizes the subspace estimator of the system might probably only be useful as an initial estimator in a pseudo ML estimation procedure (cf e.g. Bauer and Wagner, 2002a).

The final issue investigated is again the forecasting performance. Exactly the same qualitative results as for the three-dimensional systems are found. For $T = 100$ the VAR based forecasts outperform the subspace algorithm based forecasts. For all larger sample sizes the forecasting performance is basically identical for the VAR ECM and the subspace RRR forecasts, see Figure 6 for the RMSE results for system 2 and system 6 as examples. These two estimators that incorporate the correct cointegrating rank deliver again the best forecasts in the RMSE sense among the state space and the VAR forecasts. The forecasts derived from estimation using differenced data are in general worse, for both the state space model estimated with the subspace algorithm and the VAR model. An exception is the smallest sample size $T = 100$, where the subspace estimator on the differenced data delivers the preferable subspace forecasts. Again, as in the previous subsection, the imposition of the correct cointegrating rank improves the forecasting performance at the longer horizons. For one-step forecasting there is virtually no gain. These results for the subspace estimators are to a certain extent surprising, since the estimation accuracy of the cointegrating space is worse for the subspace based approach compared to the VAR estimator. However, this inaccuracy does not show up in terms of degraded forecast accuracy. An explanation for this phenomenon might lie in the fact that the accuracy of the estimated cointegrating space depends on the estimated $\hat{C}_1$, whereas for forecasting the products $\hat{C}\hat{A}^j$ are used. Thus, the product of the matrices $\hat{C}\hat{A}$ seems to be estimated precisely, but the decomposition in $\hat{C}$ and $\hat{A}$ seems to be adversely affected by stable eigenvalues close to the unit circle in $\hat{A}$.

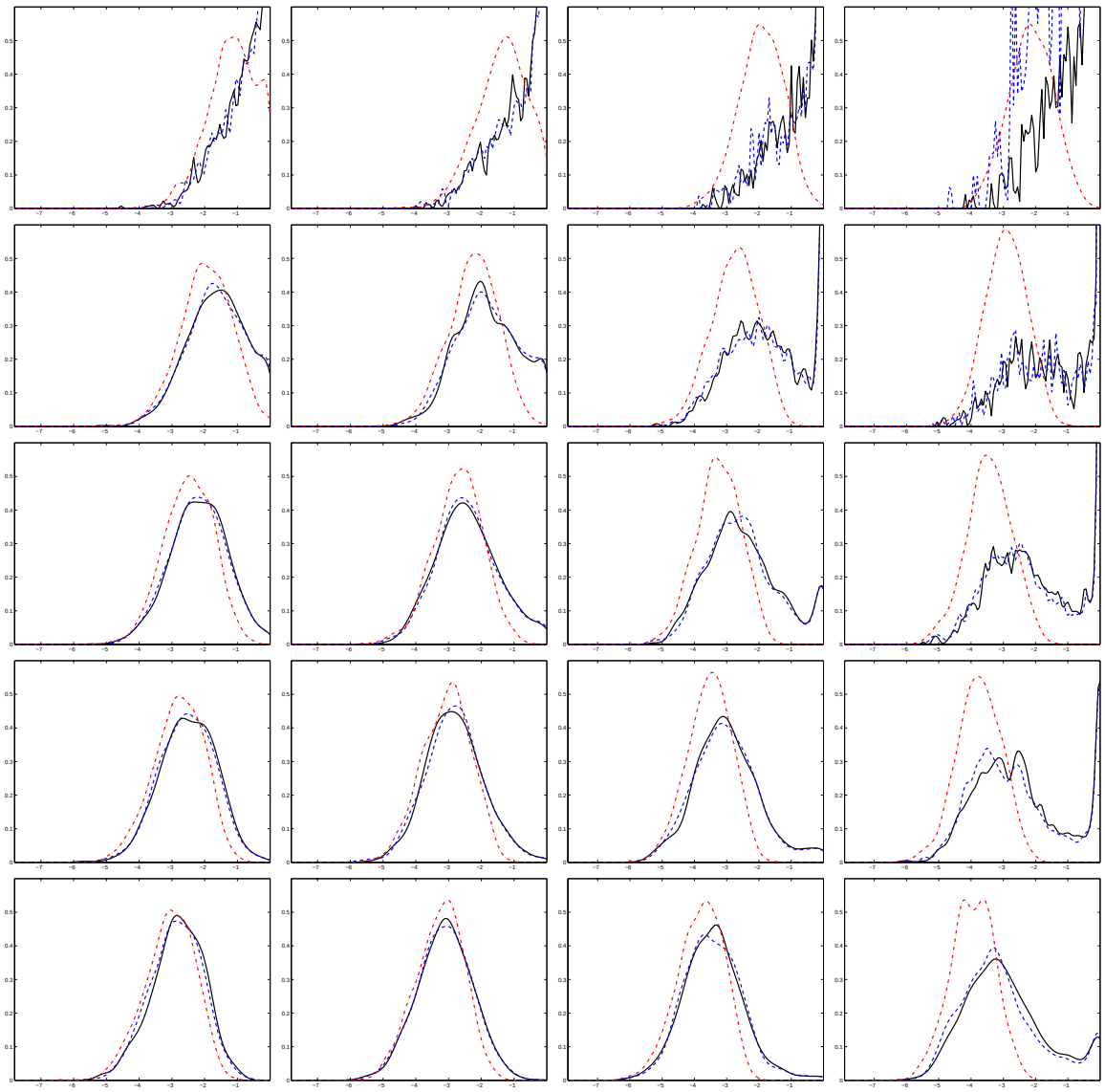

Figure 5: Density plots of Log gaps between true and estimated cointegrating spaces for systems 1, 3, 5 and 6. Ordered from system 1 to 6 along columns and along rows for $T = 100, \ldots, 500$. The solid lines correspond to the *initial* subspace estimators, the dashed lines correspond to the *adapted* estimators and the dash-dotted lines correspond to the Johansen estimators derived from a VAR approximation.

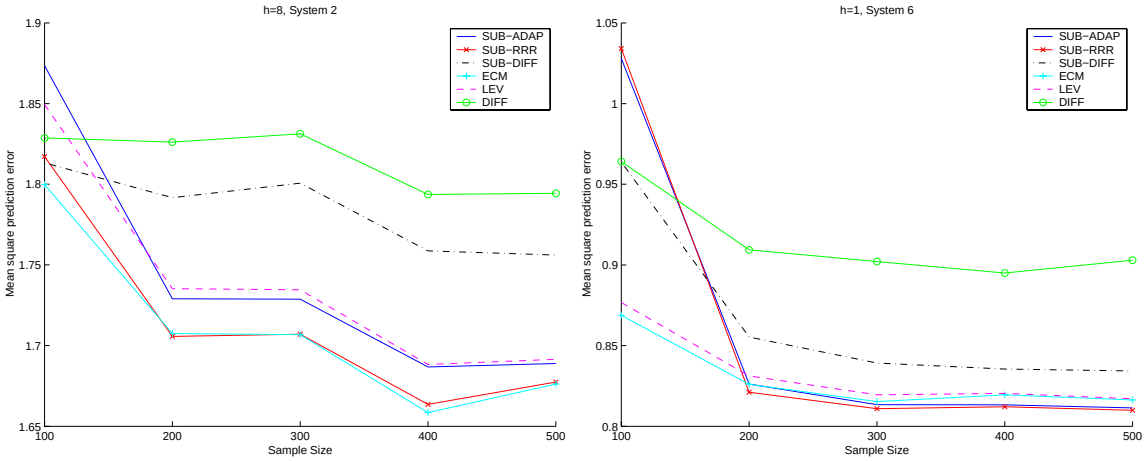

Figure 6: Left plot: Root mean square prediction error for the six forecasts for system 2, forecasting horizon $h = 1$ for sample sizes $T = 100, \ldots, 500$. Right plot: Root mean square prediction error for the six forecasts for system 6, forecasting horizon $h = 8$ for sample sizes $T = 100, \ldots, 500$.

5 Summary and Conclusions

In this paper we have complemented the theoretical results provided in Bauer and Wagner (2002b) by a simulation study. Bauer and Wagner (2002b) derive the asymptotic properties of a particular subspace algorithm adapted to result in consistent system estimation also for cointegrated I(1) VARMA processes, albeit in the equivalent state space representation. The results in Bauer and Wagner (2002b) are applied in this paper to study six different tests. Furthermore let us note that the simulations led to an additional order estimation criterion, with better small sample properties, $BA(n)$.

The finite sample properties examined via simulations are the accuracy of the specification step, i.e. the estimation of the system order and of the tests for the cointegrating rank, the accuracy of the estimated cointegrating space, given the correct specification of the cointegrating rank, and the forecasting performance. The simulations are performed for a set of three-dimensional processes and a set of four-dimensional processes and all results are compared with benchmark results obtained by applying the Johansen methodology on VARs fitted to the data.

With respect to order estimation it turns out that $BA(n)$ leads to poor order estimation for some of the four-dimensional processes, but delivers good results for all three-dimensional processes. It is important to note that in all cases the order is estimated sufficiently large

to allow for recovery of the true number of stochastic trends in the subsequent testing step. Nevertheless, the simulations suggest that better order estimation criteria are required.

Concerning the tests for the number of stochastic trends, the simulations suggest that the subspace algorithm cointegration analysis has some virtues. In a variety of the examples simulated, the small sample improvement compared to Johansen results is substantial. Especially for the three-dimensional, but also for a number of the four-dimensional systems, clear small sample advantages prevail for the subspace method. The results obtained in the simulations of the four-dimensional systems lead us to the suspicion that the magnitude of $|\rho_0|$ might be an indicator for situations where the subspace approach is the preferred method. The final verdict in this respect, however, is still to be provided. Let us note for completeness' sake that for the larger sample sizes the eigenvalue based subspace tests and the Johansen tests performed on VAR approximations all perform well. The only exception are the tests replicating the Johansen procedure on the estimated state equation (tests V and VI). These show inferior accuracy even for the largest sample size and moreover proved to be extremely sensitive with respect to stable poles close to the unit circle.

The estimation of the cointegrating space seems to be a weakness of the subspace method. Especially for the four-dimensional processes the VAR based estimation of the cointegrating space is better than both subspace estimators of the cointegrating space. Thus, it may be worthwhile to use the subspace cointegration analysis in the model specification step (i.e. in determining the system order and the cointegrating rank) and as a provider of initial estimators. These estimators can then be used as the input in a pseudo ML procedure, as laid out in Bauer and Wagner (2002a), to obtain more accurate estimators.

Note finally that the unfavorable estimation of the cointegrating space does not seem to have detrimental effects on the forecasting properties. Across all simulations it turns out that for sample size larger or equal than $T = 200$ the best VAR and the best subspace forecast are of essentially the same quality (measured by RMSE). The best forecasts are for both the VARs and the state space models the forecasts obtained when incorporating the correct cointegrating rank. The superior performance of these forecasts gets more pronounced as the forecasting horizon is increased.

References

Aoki, M., 1990. State Space Modelling of Time Series. *Springer, New York.*

Aoki, M. and A. Havenner, 1989. A method for approximate representation of vector valued time series and its relation to two alternatives. *Journal of Econometrics* **42**, 181–199.

Aoki, M. and A. Havenner, eds. 1997. Applications of Computer Aided Time Series Modeling. Lecture Notes in Statistics. *Springer, New York.*

Bauer, D., Deistler, M., Scherrer, W., 1999. Consistency and Asymptotic Normality of some Subspace Algorithms for Systems without Observed Inputs, Automatica **35**, 1243–1254.

Bauer, D., Wagner, M., 2000. Subspace Algorithm Cointegration Analysis - An Application to Interest Rate Data, in: Nunez-Anton, V., Ferreira, E. eds., Proceedings of the 15th International Workshop in Statistics, 146–151.

Bauer, D., Wagner, M., 2001. A Canonical Form for Unit Root Analysis in the State Space Framework. Mimeo.

Bauer, D., Wagner, M., 2002a. Asymptotic Properties of Pseudo Maximum Likelihood Estimators for Multiple Frequency I(1) Processes. Mimeo.

Bauer, D., Wagner, M., 2002b. Estimating Cointegrated Systems Using Subspace Algorithms, Journal of Econometrics **111**, 47–84.

Engle, R.F., Granger, C.W.J., 1987. Co-Integration and Error Correction: Representation, Estimation and Testing, Econometrica **55**, 251–276.

Hannan, E., Deistler, M., 1988. The Statistical Theory of Linear Systems. *Wiley, New York.*

Johansen, S., 1995. Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models, *Oxford University Press, Oxford.*

Larimore, W.E., 1983. System Identification, Reduced Order Filters and Modelling via Canonical Variate Analysis, in: Rao, H.S., Dorato, P. eds., Proceedings 1983 American Control Conference 2, 445– 451.

Phillips, P.C.B., 1995. Fully Modified Least Squares and Vector Autoregression, Econometrica **63**, 1023–1078.

Saikkonen, P., 1992. Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation, Econometric Theory **8**, 1–27.

Saikkonen, P., Luukkonen, R., 1997. Testing Cointegration in Infinite Order Vector Autoregressive Processes, Journal of Econometrics **81**, 93–126.

Stock, J.H., Watson, M.W., 1988. Testing for Common Trends, Journal of the American Statistical Association **83**, 1097–1107.

Van Overschee, P., DeMoor, B., 1994. N4sid: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems, Automatica **30**, 75–93.

Verhaegen, M., 1994. Identification of the Deterministic Part of Mimo State Space Models given in Innovations Form from Input-Output Data, Automatica **30**, 61–74.

Yap, S.F., Reinsel, G.C., 1995. Estimating and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model, Journal of the American Statistical Association **90**, 253–267.

Appendix

In this appendix the critical values for the simulated asymptotic distribution of the test statistics of the four discussed eigenvalue based tests for the number of stochastic trends, tests I to IV, are presented.

c	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99
1	-13.50	-10.54	-8.11	-5.70	-2.84	-0.89	0.27	0.93	1.28	1.65	2.02
2	-25.08	-20.49	-17.70	-14.17	-9.77	-5.86	-3.01	-1.28	-0.59	-0.16	0.29
3	-35.44	-30.09	-26.16	-22.29	-16.78	-11.62	-7.59	-4.95	-3.81	-3.02	-2.25
4	-43.20	-38.11	-34.48	-29.83	-23.81	-17.64	-12.60	-9.38	-7.80	-6.59	-5.48
5	-51.99	-46.43	-42.04	-37.60	-30.58	-23.60	-18.08	-14.23	-12.27	-10.87	-9.46
6	-60.34	-55.14	-50.57	-45.54	-37.41	-30.02	-23.80	-19.24	-16.94	-15.27	-13.38
7	-69.65	-63.74	-59.13	-53.25	-44.65	-36.32	-29.60	-24.40	-21.74	-19.98	-17.75
8	-78.34	-71.35	-66.30	-60.75	-51.76	-42.83	-35.50	-30.16	-27.35	-25.11	-22.87
9	-85.78	-79.53	-74.13	-67.99	-58.37	-49.12	-41.32	-35.48	-32.45	-29.89	-27.44
10	-94.73	-87.37	-81.66	-75.21	-64.96	-55.22	-47.11	-40.96	-37.75	-34.93	-32.42
11	-102.23	-95.10	-88.34	-82.37	-72.23	-62.03	-53.18	-46.52	-42.82	-40.23	-37.20
12	-108.22	-102.23	-96.38	-89.50	-78.87	-68.61	-59.66	-52.72	-48.95	-45.91	-42.92

Table 7: Percentiles of the asymptotic distribution of the test statistic $\mu_{c,re}$ (test I) for $c = 1, \ldots, 12$.

c	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99
1	-13.50	-10.54	-8.11	-5.70	-2.84	-0.89	0.27	0.93	1.28	1.65	2.02
2	-26.35	-21.73	-18.60	-15.15	-10.50	-6.38	-3.45	-1.56	-0.61	0.17	0.97
3	-42.92	-37.53	-33.13	-28.60	-22.28	-16.39	-11.67	-8.05	-6.36	-5.01	-3.45
4	-61.41	-55.75	-50.53	-45.51	-37.82	-30.31	-23.91	-18.89	-16.39	-14.21	-11.90
5	-85.85	-78.55	-73.19	-66.96	-57.51	-48.20	-40.05	-33.70	-30.18	-27.52	-24.62
6	-113.86	-105.60	-98.71	-92.06	-81.10	-69.96	-60.35	-52.41	-47.96	-44.74	-40.93
7	-144.89	-136.13	-129.68	-121.47	-108.88	-96.15	-84.62	-75.10	-69.94	-65.93	-60.67
8	-179.60	-170.70	-162.73	-154.19	-140.40	-126.21	-113.74	-102.32	-96.34	-91.04	-85.37
9	-218.47	-208.63	-201.16	-191.97	-176.19	-160.09	-145.42	-132.98	-126.53	-120.26	-112.97
10	-265.26	-252.63	-242.93	-232.48	-215.16	-197.79	-181.33	-167.25	-158.99	-152.85	-146.68
11	-313.01	-298.75	-289.21	-277.54	-259.34	-240.00	-221.99	-205.99	-197.61	-189.66	-182.10
12	-361.93	-349.18	-338.80	-326.65	-307.03	-285.78	-266.78	-249.92	-240.10	-231.95	-222.24

Table 8: Percentiles of the asymptotic distribution of the test statistic $\mu_{\sum_c,re}$ (test II) for $c = 1, \ldots, 12$.

c	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99
1	0.02	0.06	0.11	0.21	0.54	1.19	2.80	5.56	7.80	10.06	14.03
2	0.90	1.24	1.61	2.10	3.42	5.98	9.81	14.35	17.44	20.46	24.13
3	3.52	4.18	4.87	5.83	8.08	11.83	16.78	22.34	25.89	29.81	33.97
4	6.87	7.82	8.84	10.26	13.24	17.82	23.61	30.00	34.36	38.28	43.54
5	10.72	11.98	13.32	14.99	18.78	24.00	30.60	37.65	42.65	47.06	52.07
6	15.30	16.80	18.46	20.38	24.51	30.57	37.65	45.35	49.96	54.99	60.59
7	19.95	21.52	23.33	25.78	30.30	36.76	44.68	53.03	58.45	63.02	68.97
8	24.28	26.49	28.67	31.00	35.97	43.16	51.46	60.32	66.34	71.73	78.49
9	29.38	31.89	34.12	36.81	42.11	49.52	58.53	67.81	73.75	80.02	86.34
10	34.80	37.35	39.78	42.82	48.26	56.27	65.57	75.80	82.24	87.80	94.19
11	39.21	42.51	44.99	48.12	54.33	62.67	72.76	82.42	89.37	95.37	102.97
12	45.22	47.86	50.60	54.21	60.68	69.17	79.45	90.06	96.78	102.44	109.72

Table 9: Percentiles of the asymptotic distribution of the test statistic $\mu_{c,abs}$ (test III) for $c = 1, \ldots, 12$.

c	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99
1	0.02	0.06	0.11	0.21	0.54	1.19	2.80	5.56	7.80	10.06	14.03
2	1.55	2.04	2.58	3.37	4.89	7.45	11.27	15.90	19.27	22.31	26.69
3	7.18	8.40	9.55	11.00	13.88	18.31	23.99	29.92	34.16	38.13	42.95
4	17.40	19.35	20.89	23.10	27.68	33.60	40.65	48.06	53.11	57.93	64.23
5	31.29	34.04	36.61	39.73	45.54	53.27	62.17	70.90	77.31	82.28	89.31
6	51.09	54.47	57.34	61.35	68.57	77.54	88.00	98.30	104.97	111.09	118.78
7	73.95	78.30	82.17	87.19	95.84	106.46	118.15	130.00	137.48	144.64	152.79
8	101.08	106.29	111.40	116.72	127.18	139.36	153.21	166.18	174.75	181.95	190.96
9	133.47	139.47	145.38	152.05	163.49	177.15	191.86	206.30	215.71	224.58	233.63
10	172.19	179.03	184.84	191.94	204.58	220.06	236.21	251.32	260.91	269.66	281.73
11	213.83	220.84	227.00	235.41	249.10	265.63	283.93	301.38	312.58	321.53	333.99
12	258.64	267.02	274.93	283.95	299.38	317.68	336.95	355.29	366.90	378.49	391.13

Table 10: Percentiles of the asymptotic distribution of the test statistic $\mu_{\sum_c,abs}$ (test IV) for $c = 1, \ldots, 12$.