Bootstrapping Systems Cointegration Tests With a Prior Adjustment for Deterministic Terms

Carsten Trenkler*
University of Mannheim
Department of Economics
Chair of Empirical Economics
L7, 3-5
D-68131 Mannheim
Germany
Tel.: +49-621-181-1852
Fax.: +49-621-181-1931
email: trenkler@uni-mannheim.de

*This research was supported by the Deutsche Forschungsgemeinschaft (DFG) through the SFB 649 "Economic Risk". Moreover, I am grateful to an anonymous referee, the Co-editor Bruce Hansen, Anders Rygh Swensen, Helmut Lütkepohl, Enno Mammen, Kyusang Yu, and participants of the "Unit Root and Cointegration Testing Conference" at the University of Algarve for many helpful comments and suggestions. An earlier version of the paper was written while I was a research assistant at the Humboldt-Universität zu Berlin, Department of Economics.

Abstract

In this paper we analyse bootstrap procedures for systems cointegration tests with a prior adjustment for deterministic terms suggested by Saikkonen, Lütkepohl & Trenkler (2006) and Saikkonen & Lütkepohl (2000b). The asymptotic properties of the bootstrap test procedures are derived and their small sample properties are studied. The simulation study also considers the standard asymptotic test versions and the Johansen cointegration test for comparison.

Keywords: Bootstrap, Systems cointegration tests, VEC models
JEL classification: C12, C13, C15, C32

1 Introduction

Recently, Swensen (2006a) has given a theoretical justification for the use of bootstrap methods to test for the cointegrating rank. He shows that recursive bootstrap versions of the Johansen test have the same limiting distribution as the asymptotic test versions and his simulation findings demonstrate that bootstrap procedures can lead to an improvement in the tests' sizes in small samples. Such improvements could be expected since the cointegrating rank test statistic is asymptotical pivotal (compare e.g. Horowitz 2001). Moreover, Park (2003) shows that a bootstrap algorithm, which can be regarded as a univariate counterpart of the algorithms considered by Swensen (2006a), offers asymptotic refinements in finite samples for the ADF unit root test.

We follow Swensen (2006a) and analyze two variants of recursive bootstrap algorithms for systems cointegration tests which have been suggested by Saikkonen et al. (2006) and Saikkonen & Lütkepohl (2000b). It is assumed that the data generating process can be decomposed into an unobservable zero-mean stochastic part, which follows a vector autoregressive process, and a deterministic component, which consists of a level term and a linear trend. The common idea of the test procedures is to adjust the original time series by the estimated deterministic terms in a first step. Then, a likelihood ratio type test as described e.g. in Johansen (1995) is applied to the adjusted data. This model setup requires to introduce an appropriate recursion for bootstrap data which are adjusted for deterministic terms. Once this is achieved, it can be shown that the bootstrap tests have the same null limiting distributions as the corresponding asymptotic test procedures by Saikkonen et al. (2006) and Saikkonen & Lütkepohl (2000b).

The two bootstrap schemes are based on the vector error correction representation of the cointegrated VAR model. They differ with respect to the residuals and estimators of the short-run dynamics, including the constant, used for generating the bootstrap sample. The first scheme obtains residuals and estimates of the short-run dynamics from a model which does not impose a reduced rank. In contrast, the second one refers to a model imposing the cointegrating rank tested. The latter model is always used to obtain estimates of the parameters associated with the error correction term. Our simulation results show that the first bootstrap scheme can produce clearly inferior small sample properties depending on the specification of the deterministic part in the data generating process. This results from combining estimates from two different models in situations in which the restricted and unrestricted estimates of the constant term clearly differ. Nevertheless, our simulation results indicate that applying bootstrap methods can provide improved approximations of the small sample distributions of the tests with prior adjustment of deterministic terms. Moreover, the bootstrap versions of Saikkonen et al. (2006) produce slightly better small sample properties compared to the bootstrapped Johansen tests in certain situations. However, the bootstrap counterparts of Saikkonen & Lütkepohl (2000b) are often inferior, especially the one based on the first bootstrap scheme.

The rest of the paper is structured as follows. The next section describes the system cointegration tests and Section 3 introduces the bootstrap procedures. Their asymptotic properties are discussed in Section 4 and the simulation results are presented in Section 5. Section 6 summarizes and concludes. Proofs and response surface coefficients to determine p-values for the test of Saikkonen et al. (2006) are deferred to two Appendices.

The following notation is used throughout. If A is an $(n \times m)$ matrix of full column rank $(n > m)$, its orthogonal complement $A_\perp$ is an $(n \times (n-m))$ matrix of full column rank and such that $A'A_\perp = 0$. Furthermore, let $\bar{A} = A(A'A)^{-1}$. A $(n \times n)$ identity matrix is denoted by I_n. LS and GLS are used to abbreviate least squares and generalized least squares respectively, LR relates to likelihood ratio, RR is short for reduced rank, and DGP refers to data generating process. The abbreviations r.h.s. and l.h.s. stand for right- and left-hand-side, respectively.

2 Test Procedures

Let us consider a n-dimensional times series $y_t = (y_{1t}, \dots, y_{nt})'$, $t = 1, \dots, T$, which is generated by

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

where μ_0 and μ_1 are unknown $(n \times 1)$ parameter vectors. Hence, the deterministic part consists of a constant and a linear trend. The term x_t is an unobservable stochastic error process for which we make the following assumption.

Assumption 1. The process x_t is integrated of order at most $I(1)$ with cointegrating rank r and follows a vector autoregressive process of order p, VAR(p),

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t, \quad t = 1, 2, \dots, \tag{2.2}$$

where A_j are $(n \times n)$ coefficient matrices and the initial values are such that $x_t = 0$, $t \leq 0$. For the error terms we assume $\varepsilon_t \sim$ i.i.d.$(0, \Omega)$ with positive definite covariance matrix Ω and $E(\varepsilon_t^4) < \infty$.

The initial value condition $x_t = 0$, $t \leq 0$, is imposed to facilitate the derivation of asymptotic properties. However, the asymptotic results also hold if the initial values have a fixed probability distribution, which does not depend on the sample size. Under Assumption 1, the process x_t has the usual vector error correction model (VECM) form

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

where Π and Γ_j, $j = 1, \ldots, p-1$, are $(n \times n)$ unknown parameter matrices. Because the cointegrating rank is r, the matrix Π can be written as $\Pi = \alpha\beta'$, where α and β are $(n \times r)$ matrices of full column rank. Obviously, the cointegrating rank is equal to the rank of the matrix Π. As is well-known, $\beta' x_t$ and Δx_t are zero mean $I(0)$ processes. Defining $\Gamma = I_n - \Gamma_1 - \cdots - \Gamma_{p-1}$ and $C = \beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}\alpha'_\perp$, we have $x_t = C \sum_{j=1}^{t} \varepsilon_j + \xi_t$, $t = 1, 2, \ldots,$ where ξ_t is a zero mean $I(0)$ process.

Multiplying (2.1) by $A(L) = I_n - A_1 L - \cdots - A_p L^p = I_n \Delta - \Pi L - \Gamma_1 \Delta L - \cdots - \Gamma_{p-1} \Delta L^{p-1}$ and rearranging yields the VECM representation for y_t

$$\Delta y_t = \nu + \alpha(\beta' y_{t-1} - \phi(t-1)) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \quad t = p+1, p+2, \ldots, \tag{2.3}$$

where $\nu = -\Pi\mu_0 + \Gamma\mu_1 = -\alpha\theta + \Gamma\mu_1$, $\theta = \beta'\mu_0$, and $\phi = \beta'\mu_1$.

We consider the so-called trace test version, i.e. we aim to test the pair of hypotheses

$$H_0(r_0) : \text{rk}(\Pi) = r_0 \quad vs. \quad H_1(r_0) : \text{rk}(\Pi) > r_0. \tag{2.4}$$

Saikkonen et al. (2006) suggest to detrend the data before applying a Johansen LR-type cointegration test. To this end, the trend parameter vector μ_1 will be estimated by

$$\tilde{\mu}_1 = \tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\phi} + \tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1}\tilde{\phi}_*, \tag{2.5}$$

where $\tilde{\beta}$, $\tilde{\beta}_\perp$, $\tilde{\phi}$, and $\tilde{\phi}_* = \tilde{\beta}'_\perp \tilde{C}(\tilde{\nu} - \tilde{\Gamma}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\phi})$ are obtained from a RR estimation of (2.3). Note that $\tilde{\phi}_*$ can be regarded as the estimator of $\phi_* = \beta'_\perp \mu_1 = \beta'_\perp C(\nu - \Gamma\beta(\beta'\beta)^{-1}\phi)$.

Using $\tilde{\mu}_1$, we adjust y_t for a linear trend and obtain $\tilde{y}^c_t = y_t - \tilde{\mu}_1 t$. Then, regress $\Delta\tilde{y}^c_t$ and $(\tilde{y}^{c\prime}_{t-1} : 1)'$ on $(\Delta\tilde{y}^{c\prime}_{t-j}, \ldots, \Delta\tilde{y}^{c\prime}_{t-j})'$ to obtain the residuals $\tilde{R}_{0t}$ and $\tilde{R}_{1t}$, respectively. Next, consider $\tilde{S}_{ij} = T^{-1} \sum_{t=1}^{T} \tilde{R}_{it}\tilde{R}'_{jt}$, $i, j = 0, 1$. Solving $det(\lambda\tilde{S}_{11} - \tilde{S}_{10}\tilde{S}^{-1}_{00}\tilde{S}_{01}) = 0$, we obtain the ordered generalized eigenvalues $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$. Hence, the LR test statistic for the pair of hypotheses in (2.4) is[1]

$$SLT(r_0) = -T \sum_{j=r_0+1}^{n} \log(1 - \tilde{\lambda}_j).$$

Saikkonen et al. (2006) have derived the limiting distributions for $SLT(r_0)$ given in Theorem 1.

[1]Saikkonen et al. (2006) allow in addition for a level shift at unknown time. The test statistic considered here can be obtained by deleting all terms associated with the level shift. This has no effect on the limiting distribution.

Theorem 1. Under Assumption 1 and if $H_0(r_0)$ is true, then

$$P(SLT(r_0) \leq x) \longrightarrow P(\text{tr}(D_{SLT}) \leq x)$$

for all x as $T \rightarrow \infty$, where

$$D_{SLT} = \left(\int_0^1 B_+(s) dB_*(s)' \right)' \left(\int_0^1 B_+(s) B_+(s)' ds \right)^{-1} \left(\int_0^1 B_+(s) dB_*(s)' \right),$$

$B(s)$ is a $(n - r_0)$-dimensional standard Brownian motion, $B_*(s) = B(s) - sB(1)$ is a $(n - r_0)$-dimensional Brownian bridge, $B_+(s) = [B_*(s)', 1]'$ and $dB_*(s) = dB(s) - dsB(1)$.

Note that the limiting distribution of $SLT(r_0)$ is formally similar to its counterpart in Theorem 6.3 of Johansen (1995) where a standard Brownian motion appears instead of the Brownian bridge in Theorem 1. The Brownian bridge results from estimating μ_1 since an estimation error regarding the trend component affects the standardized sample moments involving $\beta_\perp' \tilde{y}_t^c$.

Saikkonen & Lütkepohl (2000b) propose to adjust the data for both the level and trend component using feasible GLS estimators of μ_0 and μ_1. However, the level parameter vector μ_0 cannot be consistently estimated in the direction of $\beta_\perp$ because μ_0 is not identified in that direction in (2.3). Therefore, Saikkonen et al. (2006) have suggested to avoid estimating μ_0. Nevertheless, an estimator of $\beta_\perp' \mu_0$ that is bounded in probability is sufficient to obtain a limiting distribution similar to the one in Theorem 1. However, such estimators are partly determined by initial values, what may result in inferior small sample properties. In fact, our simulation study illustrates that the strong initial value dependence is rather problematic for bootstrapping the test of Saikkonen & Lütkepohl (2000b). That is why the application of the corresponding bootstrap tests is not recommended. We, therefore, focus on the procedure of Saikkonen et al. (2006) when describing the bootstrap framework and deriving the theoretical results in the following. Nevertheless, the bootstrap test versions of Saikkonen & Lütkepohl (2000b) are asymptotically valid. This is explained in detail in Trenkler (2006), the discussion paper version of this article.

3 Bootstrap Algorithms

Swensen (2006a) considers the following recursive bootstrap, also labelled with Algorithm 1, to generate the so-called pseudo or bootstrapped observations $y_1^*, \ldots, y_T^*$.

Algorithm 1.

(1) Estimate an unrestricted version of (2.3) setting $r = n$ in order to obtain the estimators $\hat{\nu}$ and $\hat{\Gamma}_j$, $j = 1, \ldots, p-1$, and the OLS residuals $\hat{\varepsilon}_{p+1}, \ldots, \hat{\varepsilon}_T$.

(2) The remaining parameters are estimated by performing a RR regression of (2.3) under the rank hypothesis $H_0(r_0) : \text{rk}(\Pi) = r_0$. Let $\tilde{\alpha}$, $\tilde{\beta}$, and $\tilde{\phi}$ be the corresponding estimators.

(3) Check whether the roots of the equation $\det[\tilde{A}(z)] = 0$, where

$$\hat{A}(z) = (1 - z)I_n - \tilde{\alpha}\tilde{\beta}'z - \hat{\Gamma}_1(1 - z)z - \cdots - \hat{\Gamma}_{p-1}(1 - z)z^{p-1},$$

are equal to 1 or outside the unit circle and whether $\tilde{\alpha}'_\perp \hat{\Gamma} \tilde{\beta}_\perp$ is nonsingular with $\hat{\Gamma} = I_n - \hat{\Gamma}_1 - \cdots - \hat{\Gamma}_p$.

(4) If so, compute y_t^*, $t = p + 1, \ldots, T$, recursively from

$$\Delta y_t^* = \hat{\nu} + \tilde{\alpha}(\tilde{\beta}' y_{t-1}^* - \tilde{\phi}(t - 1)) + \sum_{j=1}^{p-1} \hat{\Gamma}_j \Delta y_{t-j}^* + \varepsilon_t^*, \tag{3.1}$$

with sampled residuals ε_t^* drawn with replacement from the estimated residuals $\hat{\varepsilon}_{p+1}, \ldots,$ $\hat{\varepsilon}_T$. The starting values of the recursion, $y_1^*, \ldots, y_p^*$, are set equal to $y_1, \ldots, y_p$.

Requirement (3) assures that the generated pseudo observations are indeed $I(1)$. If this condition is not satisfied, one may refer to a more appropriate resampling scheme as pointed out by Swensen (2006a). As usual in the literature, the asterisk refers to quantities related to the bootstrap procedure. Accordingly, the probability measure P^* used later on refers to the conditional distribution of $y_1^*, \ldots, y_T^*$ given the observations $y_1, \ldots, y_T$ and E^* is the corresponding expectation operator given the observations. Note that the pseudo observations $y_1^*, \ldots, y_T^*$ depend on the sample size T.

The bootstrap test version is then computed by running the same test procedures as described above with respect to the pseudo series $y_1^*, \ldots, y_T^*$. Hence, we detrend y_t^* according to $\tilde{y}_t^{c*} = y_t^* - \tilde{\mu}_1^* t$, where $\tilde{\mu}_1^* = \tilde{\beta}^*(\tilde{\beta}^{*\prime}\tilde{\beta}^*)^{-1}\tilde{\phi}^* + \tilde{\beta}_\perp^*(\tilde{\beta}_\perp^{*\prime}\tilde{\beta}_\perp^*)^{-1}\tilde{\phi}_*^*$ and $\tilde{\beta}^*$, $\tilde{\beta}_\perp^*$, $\tilde{\phi}^*$, and $\tilde{\phi}_*^*$ are RR estimators of the VECM (3.1). Then, $\Delta\tilde{y}_t^{c*}$ and $(\tilde{y}_{t-1}^{c*\prime} : 1)'$ are regressed on $(\Delta\tilde{y}_{t-j}^{c*\prime}, \ldots, \Delta\tilde{y}_{t-j}^{c*\prime})'$ to obtain $\tilde{R}_{0t}^*$ and $\tilde{R}_{1t}^*$, respectively. We solve $det(\lambda\tilde{S}_{11}^* - \tilde{S}_{10}^*\tilde{S}_{00}^{*-1}\tilde{S}_{01}^*) = 0$, where $\tilde{S}_{ij}^* = T^{-1}\sum_{t=1}^{T}\tilde{R}_{it}\tilde{R}_{jt}^{*\prime}$, $i, j = 0, 1$, and obtain the generalized eigenvalues $\tilde{\lambda}_1^* \geq \cdots \geq \tilde{\lambda}_n^*$. Hence, the bootstrap test statistic for the pair of hypotheses in (2.4) is

$$SLT_1^*(r_0) = -T \sum_{j=r_0+1}^{n} \log(1 - \tilde{\lambda}_j^*).$$

Let us denote the cumulative distribution function of $SLT_1^*(r_0)$ by $F_{r_0;SLT,1}^*$. The latter is a conditional distribution given the observations $y_1, \ldots, y_T$. Then, we reject the null hypothesis of r_0 cointegrating relations at some chosen significance level δ if

$$1 - F_{r_0;SLT,1}^*(SLT(r_0)) \leq \delta. \tag{3.2}$$

The bootstrap distribution is usually a very complicated function of the observations. Therefore, it is approximated by repeating the bootstrap algorithm a large number of times, say M.[2] Then, we count the number of instances for which $SLT_1^*(r_0) > SLT(r_0)$, say $MR_{SLT,1}$. Finally, we reject $H_0(r_0)$ if

$$MR_{SLT,1}/M \leq \delta. \tag{3.3}$$

For $M \to \infty$, the expression (3.3) converges to (3.2).

As an alternative to Algorithm 1, we consider a bootstrap procedure that exclusively refers to the estimated version of the VECM (2.3) for obtaining estimators and residuals. Thus, in contrast to step (1) of Algorithm 1, the estimators for ν and Γ_j, $j = 1, \ldots, p-1$, as well as the residuals are obtained by performing a RR regression with respect to (2.3) like in step (2). Steps (3) and (4) of Algorithm 1 are adjusted accordingly, i.e. $\hat{\nu}$, and $\hat{\Gamma}_j$, $j = 1, \ldots, p-1$, are replaced by the corresponding RR estimators and the bootstrap error terms are drawn from the residuals of the estimated VECM (2.3). This adjusted algorithm will be labeled as Algorithm 2 and the resulting test statistic will be noted by $SLT_2^*(r_0)$.[3]

Two remarks on the algorithms are in order. First, Swensen (2006a) has shown that a modified version of Algorithm 1 in relation to the Johansen test can be used to consistently estimate the cointegrating rank. Not surprisingly, the modification is based on a sequential application of the bootstrap test. However, Swensen (2006a) was not able to prove that a corresponding adjustment of Algorithm 2 leads to a consistent estimation of the cointegrating rank. The problem to be dealt with is that the bootstrap samples have to be generated for other ranks than the true one if a sequence of tests is performed. Nevertheless, tests based on Algorithm 2 are valid for a specific rank hypothesis and this is the setup we focus on in this paper.

Second, as pointed out by Swensen (2006a), the difference between using unrestricted (Algorithm 1) and restricted (Algorithm 2) residuals and estimators relates to the univariate setups using either OLS residuals or first differences of the data to generate bootstrap observations. Paparoditis & Politis (2005, 2003) have found that bootstrap unit root tests based on unrestricted residuals may have favourable power properties under the alternative. However, restricted and unrestricted residuals are strongly positively correlated in our simulations. Note, further, that restricted and unrestricted estimates are not combined in the univariate setup because of the missing error correction term. But it is this combination within Algorithm 1 that can cause inferior small sample properties in case of nonzero deterministic terms according to our simulation results.

[2] Davidson & MacKinnon (2000) suggest and analyse procedures to determine M in order to assure an appropriate approximation.

[3] Note that our Algorithm 2 is not equal to Algorithm 2 in Swensen (2006a). The latter one describes the sequential bootstrap procedure used to estimate the cointegrating rank.

4 Asymptotic Distributions of Bootstrap Test Statistics

We first consider the derivation of the asymptotic distribution of $SLT_1^*(r_0)$. To this end, we have to rely on a Granger representation of the detrended bootstrap data $y_t^{c*} = y_t^* - \tilde{\mu}_1 t$, $t = 1, \ldots, T$, which is given in the following lemma.

Lemma 1. Under Assumption 1, the bootstrap elements y_t^{c*} have the following representation

$$y_t^{c*} = \tilde{C} \sum_{i=p+1}^{t} \varepsilon_i^* + \tilde{\tau}_0 + \sqrt{T} R_{t,T}^{c*},\ t = p+1, \ldots, T, \tag{4.1}$$

where for all $\epsilon > 0$, $P^*(\max_{p+1 \leq t \leq T} |R_{t,T}^{c*}| > \epsilon) \rightarrow 0$ in probability as $T \rightarrow \infty$, $|\cdot|$ is the Euclidean norm, $\tilde{C} = \tilde{\beta}_\perp (\tilde{\alpha}_\perp' \hat{\Gamma} \tilde{\beta}_\perp)^{-1} \tilde{\alpha}_\perp'$, and $\tilde{\tau}_0 = \bar{\tilde{\beta}} \tilde{\theta}$. Moreover, $E^*[\varepsilon_t^* \varepsilon_t^{*\prime}] = \hat{\Omega} \rightarrow \Omega$ in probability as $T \rightarrow \infty$.

Lemma 1 is proven in Appendix A. The proof makes use of the fact that y_t^{c*} follows the recursive scheme

$$\Delta y_t^{c*} = \tilde{\alpha}(\tilde{\beta}' y_{t-1}^{c*} - \tilde{\theta}) + \sum_{j=1}^{p-1} \hat{\Gamma}_j \Delta y_{t-j}^{c*} + \varepsilon_t^*, \quad t = p+1, \ldots T, \tag{4.2}$$

with $y_t^{c*} = y_t^* - \tilde{\mu}_1 t, t = 1, \ldots, p$, and $\tilde{\alpha}\tilde{\theta} = \hat{\Gamma}\tilde{\mu}_1 - \hat{\nu}$ such that $\tilde{\theta} = (\tilde{\alpha}'\tilde{\alpha})^{-1}\tilde{\alpha}'(\hat{\Gamma}\tilde{\mu}_1 - \hat{\nu})$. This recursion for y_t^{c*} follows from (3.1) and the definition of $\tilde{\mu}_1$ in (2.5). We use $y_t^{c*} = y_t^* - \tilde{\mu}_1 t$, and not $y_t^c = y_t - \mu_1 t$, $t = 1, \ldots, p$, to initialize (4.2) in order to assure that $y_t^* = y_t^{c*} + \tilde{\mu}_1 t$ for all $t = 1, \ldots, T$. Note that $\tilde{\theta}$ can be interpreted as an estimator for $\theta = \beta' \mu_0$ (compare Saikkonen & Lütkepohl 2000b).

Lemma 1 shows that y_t^{c*} consists of a stochastic trend, a level term and a stationary remainder term. Hence, following Swensen (2006a, 2006b) and Saikkonen et al. (2006) we derive the asymptotic distribution of $SLT_1^*(r_0)$, which is given below.

Proposition 1. Under Assumption 1, given (4.2), and if $H_0(r_0)$ is true, then

$$P^*(SLT_1^*(r_0) \leq x) \longrightarrow P(\mathrm{tr}(D_{SLT}) \leq x),$$

in probability for all x as $T \rightarrow \infty$, where D_{SLT} is defined as in Theorem 1.

The proof of Proposition 1 in Appendix A relies on two main results. First, the partial sum of the bootstrap error terms, i.e. $S_T^*(s) = T^{-1/2} \sum_{i=k+1}^{[sT]} \varepsilon_i^*$, $0 \leq s \leq 1$, converges weakly in probability toward a n-dimensional Brownian motion W with covariance matrix Ω (compare Swensen 2006a). Second, as shown in Lemma A.1, the asymptotic properties of the difference

$\tilde{\mu}_1^* - \tilde{\mu}$ exactly correspond to the ones of $\tilde{\mu}_1 - \mu$.[4] Thus, the consequence of adjusting the bootstrap data by the estimated trend component $\tilde{\mu}_1^* t$, i.e. using $\tilde{y}_t^{c*} = y_t^* - \tilde{\mu}_1^* t$ instead of $y_t^{c*} = y_t^* - \tilde{\mu}_1 t$, are the same as in the asymptotic test setup: a Brownian bridge instead of a standard Brownian motion appears in the limiting distribution.

The same limiting result is obtained for $SLT_2^*(r_0)$. According to Remark 3 of Swensen (2006a), the relevant asymptotic results of Swensen (2006a) also hold with respect to the restricted bootstrap residuals and estimators used within Algorithm 2. However, we have to replace $\hat{\nu}$ and $\hat{\Gamma}_i$, $j = 1, \ldots, p-1$, by the corresponding RR estimators based on (2.3).

5 Monte Carlo Study

To analyse and compare the small sample properties of the bootstrap and asymptotic test versions, we first use bivariate versions of a VAR(1) process suggested by Toda (1994, 1995). The process has the following structure

$$x_t = \begin{bmatrix} a_1 & 0 \\ 0 & 1 \end{bmatrix} x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}\right). \tag{5.1}$$

Obviously, the parameter a_1 determines the cointegrating rank. If $|a_1| < 1$, $r = 1$ and θ describes the instantaneous correlation between the stationary and nonstationary components. In contrast, if $a_1 = 1$ the cointegrating rank is zero and we set $\theta = 0$ since the test results do not depend on θ in this case (see Toda 1994, 1995). The starting values are set to zero.

The computations are performed by using programs written in GAUSS V6 for Windows. The RNDNS function with a fixed seed has been used to generate standard normally distributed random numbers. The number of replications is $R = 10000$. For determining the quantiles of the empirical bootstrap distributions we use $M = 1000$ bootstrap replications. We believe that these numbers of replications are large enough to obtain sufficiently precise estimations of the tests' true rejection frequencies. Since the overall replications are Bernoulli trials the standard deviation of a rejection frequency is limited by $\sqrt{p(1-p)/R} \leq \sqrt{1/4R} = 0.005$. Note, however, that this limit ignores the uncertainty involved in the bootstrap simulations and the simulation of the critical values. The sample sizes are $T = 50$, $T = 100$, and $T = 200$.

The simulations are performed with respect to 1%, 5%, and 10% significance levels but the presentation focusses on the 5% level, since the relative results are very similar for the other nominal levels. The rejection frequencies of the asymptotic GLS and JOH tests are based on asymptotic

[4]Note that determining the properties of $\tilde{\mu}_1^*$ relative to μ_1 would require to additionally take account of the term $(\mu_1 - \tilde{\mu}_1)t$ since $\tilde{y}_t^{c*} \neq y_t^{c*} - (\tilde{\mu}_1^* - \mu_1)t$ so that $\tilde{y}_t^{c*} = y_t^{c*} - (\tilde{\mu}_1^* - \mu_1)t - (\mu_1 - \tilde{\mu}_1)t$ had to be used.

Table 1. Rejection Frequencies of Tests for Bivariate Toda-DGP (5.1), VAR Order $p = 1$, Nominal Significance Level 0.05, $R = 10000$

	Panel A: $a_1 = 1$ ($r = 0$), $H_0 : r_0 = 0$			Panel B: $a_1 = 0.9$ ($r = 1$), $H_0 : r_0 = 1$			Panel C: $a_1 = 0.7$ ($r = 1$), $H_0 : r_0 = 1$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
SLT	0.0588	0.0575	0.0540	0.0281	0.0578	0.0654	0.0598	0.0581	0.0626
SLT_1^*	0.0524	0.0540	0.0502	0.0276	0.0491	0.0552	0.0506	0.0526	0.0586
SLT_2^*	0.0531	0.0535	0.0514	0.0267	0.0477	0.0568	0.0507	0.0534	0.0579
GLS	0.0535	0.0499	0.0519	0.0149	0.0240	0.0302	0.0360	0.0397	0.0486
GLS_1^*	0.0554	0.0493	0.0499	0.0360	0.0612	0.0798	0.0563	0.0546	0.0576
GLS_2^*	0.0535	0.0488	0.0518	0.0282	0.0400	0.0462	0.0444	0.0502	0.0557
JOH	0.0602	0.0544	0.0548	0.0131	0.0369	0.0635	0.0501	0.0589	0.0607
JOH_1^*	0.0513	0.0516	0.0527	0.0134	0.0345	0.0561	0.0445	0.0531	0.0558
JOH_2^*	0.0503	0.0520	0.0520	0.0130	0.0359	0.0578	0.0442	0.0528	0.0570

critical values computed from response surfaces given in Trenkler (2008) and Doornik (1998) respectively. The coefficients of a corresponding response surface for the SLT test are presented in Appendix B. We use the response surface critical values since they deliver a more accurate approximation of the tests' asymptotic properties than the standard tabulated critical values (see Doornik 1998).

The results for different versions of the bivariate Toda-DGP (5.1) are collected in Tables 1 and 2. Table 1 contains the rejection frequencies of the tests for a correct null hypothesis. We will use the term *size* with respect to these frequencies, although this use does not coincide with the exact definition of the size of a test. The definition would require to maximize the power function over the whole parameter space associated with the null hypothesis. However, such a maximization is not done in our simulation framework.

Table 1 shows that clear size improvements can be observed if the bootstrap test versions are applied. This is the case both for processes with a true cointegrating rank of $r = 0$ and of $r = 1$. The GLS_1^* test, however, rejects sometimes too often such that significant deteriorations may occur (see Panels B and C in Table 1).

We have also performed simulations with a_1 equal to 0.8, 0.6 and 0.5 in order to evaluate whether the parameter a_1 has a systematic impact on whether and how the application of bootstrap tests may improve the tests' sizes. Our overall conclusion is that the values of a_1 seem to be less important in these respects than e.g. the sample size T. Interestingly, the most pronounced improvements occur for $T = 50$, although we also observe a number of significant improvements in case of $T = 200$.

Table 2 displays the small sample power for the case of a true cointegrating rank of $r = 1$ ($a_1 = 0.9$ and $a_1 = 0.7$) if the null hypothesis $H_0(r_0) : r_0 = 0$ is tested. As can be seen, the powers of corresponding bootstrap and asymptotic tests are close to each other. The slightly smaller power levels of the bootstrap tests compared to the asymptotic procedures are in line with the tests' size values given in Table 1.

In addition to the simple process (5.1) we have also used a more complex data-based DGP by referring to an empirical study by King, Plosser, Stock & Watson (1991). This DGP has also been chosen in order to analyse the dependence of the tests' performance on the initial values and the deterministic component based on a real data setup. To the best of our knowledge, this issue has not been yet studied in relation to bootstrap cointegration tests. King et al. (1991) analyse a small macroeconomic model for the U.S. which consists of the logarithms of per-capita private real GNP, per-capita real consumption, and per-capita gross private domestic fixed investment. Based on their recommendations we estimate a VECM with one lag and two restricted cointegrating relationships using quarterly data in logarithms for the period 1949:1-1988:4. The two cointegrating relations

Table 2. Rejection Frequencies of Tests for Bivariate Toda-DGP (5.1) with True Cointegrating Rank $r = 1$, Rank under H_0 is $r_0 = 0$, VAR Order $p = 1$, $\theta = 0.8$, Nominal Significance Level 0.05,$R = 10000$

	Panel A: $a_1 = 0.9$			Panel B: $a_1 = 0.7$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
SLT	0.1817	0.5060	0.9551	0.7910	0.9938	1.0000
SLT_1^*	0.1644	0.4908	0.9517	0.7718	0.9933	1.0000
SLT_2^*	0.1640	0.4919	0.9519	0.7721	0.9935	0.9999
GLS	0.1557	0.4755	0.9237	0.7152	0.9729	0.9996
GLS_1^*	0.1571	0.4693	0.9221	0.7161	0.9739	0.9998
GLS_2^*	0.1548	0.4734	0.9222	0.7139	0.9724	0.9996
JOH	0.1354	0.4038	0.9283	0.7299	0.9990	1.0000
JOH_1^*	0.1219	0.3876	0.9250	0.7019	0.9988	1.0000
JOH_2^*	0.1205	0.3902	0.9239	0.7017	0.9988	1.0000

describe the differences between consumption and GNP and between investment and GNP. In order to reduce the impact of estimates of possibly insignificant parameters, a subset VECM with zero restrictions has been estimated. The subset restrictions have been searched for by a *Top-Down* strategy employing the Akaike Information Criterion (AIC) (see e.g. Brüggemann 2004, Chapter 3). We obtain the following estimated process, which is used for the simulations[5]

$$\begin{aligned}\Delta y_t =& \nu + \begin{pmatrix} 0 & -0.026 \\ 0.217 & -0.150 \\ 0.126 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} y_{t-1} + \begin{pmatrix} 0 & 0 & 0.154 \\ 0 & 0.282 & 0.660 \\ 0.272 & 0.162 & 0 \end{pmatrix} \Delta y_{t-1} \\ &+ \varepsilon_t, \quad \varepsilon_t \sim \text{ iid } N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, 10^{-4} \begin{bmatrix} 0.588 & 0.821 & 0.465 \\ 0.821 & 4.870 & 1.688 \\ 0.465 & 1.688 & 1.376 \end{bmatrix} \right).\end{aligned} \tag{5.2}$$

Hence, we consider the case of a linear trend which is orthogonal to the cointegration space. To analyse the impact of the deterministic component and the initial values we will use various specifications for the latter ones and the parameter vector ν given in Table 3. The simulation details mentioned above in relation to the Toda-DGP (5.1) also apply to the current setup (5.2).

[5]For the simulation program the parameters are stored as 8-bit, double precision. This assures that the characteristic roots obey the restrictions imposed on the DGP.

Table 3. Initial Values and Specifications of ν for DGP (5.2)

Specification	ν	Starting Values $\begin{pmatrix} y_1' \\ y_2' \end{pmatrix}$
(1): Table 4	$\begin{pmatrix} -0.038 \\ -0.186 \\ 0.032 \end{pmatrix}$	$\begin{pmatrix} -5.019 & -6.268 & -4.730 \\ -5.012 & -6.277 & -4.733 \end{pmatrix}$
(2): Table 5 (a)	$\begin{pmatrix} -0.038 \\ -0.186 \\ 0.032 \end{pmatrix}$	$\begin{pmatrix} -5.019 & -6.361 & -4.786 \\ -5.014 & -6.357 & -4.782 \end{pmatrix}$
(3): Table 5 (b)	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

We first set $\nu = (-0.038 - 0.186\ 0.032)'$ according to the estimates obtained from the data and use the original observations at the first two data points as starting values (compare specification (1) in Table 3). Table 4 displays the resulting tests' rejection frequencies. We have obtained three main findings. First, in this setting, the bootstrap tests clearly perform worse than the asymptotic tests. This especially applies to the ones based on Algorithm 1, which deteriorate not only in terms of power but also with respect to the size. Thus, secondly, the two bootstrap algorithms can lead to rather different test decisions. Thirdly, the GLS tests do not perform very well in general and GLS_1^* is by far the most inferior procedure.

A closer inspection of the simulation results and further simulations have revealed that the different precision of the unrestricted and restricted estimates of ν is the main key to shed light on these results. This is explained in the following.

As an example, specification (1) in Table 6 displays the average of the 10000 estimates of ν and the corresponding standard deviations for $T = 100$. Obviously, the unrestricted and restricted estimates of the constant term vector clearly deviate from each other and from the true values. The deviation from the true values is especially pronounced for the unrestricted estimates. Given the rather large negative values of the unrestricted estimates, we often observe negative drifts in the data obtained from Algorithm 1 due to the nonstationary autoregressive dynamics. As a result, the pseudo data generated by the algorithms are clearly negatively correlated, although the average correlation of the unrestricted and restricted residuals is usually close to one.

Next, we use initial values in correspondence to the unconditional expectation of y_t that is

Table 4. Rejection Frequencies of Tests for KPSW-DGP (5.2) with $\nu = (-0.038\ -0.186\ 0.032)'$ and empirical initial values, True Cointegrating Rank $r = 2$, VAR Order $p = 2$, Nominal Significance Level 0.05, $R = 10000$

	Panel A: $H_0 : r_0 = 2$			Panel B: $H_0 : r_0 = 1$			Panel C: $H_0 : r_0 = 0$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
SLT	0.0138	0.0260	0.0506	0.0597	0.1238	0.3645	0.2790	0.6163	0.9905
SLT_1^*	0.0014	0.0077	0.0271	0.0108	0.0643	0.3008	0.1117	0.4903	0.9847
SLT_2^*	0.0101	0.0236	0.0468	0.0276	0.0968	0.3307	0.1290	0.5055	0.9857
GLS	0.0043	0.0079	0.0153	0.0237	0.0409	0.0836	0.1801	0.4638	0.9404
GLS_1^*	0.0652	0.1058	0.0827	0.0000	0.0005	0.0088	0.0000	0.0005	0.3464
GLS_2^*	0.0309	0.0325	0.0373	0.0329	0.0677	0.1259	0.1131	0.4038	0.9288
JOH	0.0047	0.0102	0.0301	0.0413	0.0980	0.3571	0.2905	0.5995	0.9943
JOH_1^*	0.0005	0.0033	0.0167	0.0079	0.0448	0.2947	0.1060	0.4582	0.9908
JOH_2^*	0.0037	0.0108	0.0319	0.0177	0.0774	0.3311	0.1220	0.4747	0.9909

obtained for $\nu = (-0.038 - 0.186\ 0.032)'$. To achieve this, we recover μ_0 and μ_1 in (2.1) from the parameters of the DGP (5.2). We directly have $\mu_1 = (0.0047\ 0.0047\ 0.0047)'$ according to (2.5). Since μ_0 is only identified in the direction of β by (5.2), we are only able to derive the difference between the components of μ_0 and obtain $\beta'\mu_0 = (\mu_{0,1} - \mu_{0,3}\ \mu_{0,2} - \mu_{0,3})' = (-0.2323\ -1.5751)'$. To determine the initial values we set $y_{1,1}$ equal to the first observation of the empirical consumption data representing the first component of (5.2). Thereby, we assure that the new data have a level comparable to the data simulated before. Hence, we have $y_{1,3} = y_{1,1}+0.2323$, $y_{1,2} = y_{1,3} - 1.5751$, $y_1 = (y_{1,1}\ y_{1,2}\ y_{1,3})'$ and $y_2 = y_1 + \mu_1$ and obtain the values specification (2) in Table 3. The lines in Table 5 with the label (a) behind the test abbreviations display the resulting simulation outcomes.

We see that initializing the DGP according to the unconditional expectation leads to slightly higher rejection frequencies for the asymptotic and bootstrap SLT tests and to slightly smaller ones for the JOH tests. For the GLS tests, however, we observe a very strong increase in the rejection frequencies. Now, GLS and GLS_2^* perform comparable to the corresponding SLT and JOH procedures. Of course, we only compare two sets of initial values but they illustrate the pronounced sensitivity of the GLS procedures to deviations from the unconditional expectation. This is in line with similar results of Elliott & Müller (2003) for the GLS unit root test. Moreover, we cannot attribute our results to changes in the unrestricted and restricted estimates of ν since they are almost unaffected by the variation in the starting values (compare specification (2) in Table 6).

Then, we have analysed quite a few specifications of ν and $y_{1,1}$ while always initializing the DGP at the unconditional expectation. Given the latter, varying ν or $y_{1,1}$ leaves the asymptotic tests unchanged since we only alter trend and level of the data. The RR estimators of the constant and linear trend perfectly capture such changes so that the autoregressive parameter estimates do not respond to variations in ν or $y_{1,1}$. Thus, the pseudo data y_t^*, $t = 1, \ldots, T$, generated by Algorithm 2 are shifted in the same way as the simulated data y_t, $t = 1, \ldots, T$. Accordingly, the bootstrap tests based on Algorithm 2 do not change either. By contrast, the bootstrap procedures related to Algorithm 1 are not invariant to variations in ν or $y_{1,1}$. Of course, also the unrestricted estimators of the constant and trend adjust appropriately. However, the unrestricted estimator of ν is combined with the estimators of the parameters of the error correction term obtained from a different model. Hence, the bootstrap data produced by Algorithm 1 do not simply shift since the constant is incorporated within the autoregressive dynamics in (3.1).

We have found that the unrestricted and restricted estimates of ν are closer to each other and to the true value the smaller the deterministic component of the DGP is. In the extreme case when the deterministic part is zero, i.e. initial values and ν are set to zero, unrestricted and restricted estimates are practically equal (compare specification (3) in Table 6). Accordingly, we observe

Table 5. Rejection Frequencies of Tests for KPSW-DGP (5.2) with (a) $\nu = (-0.038 \ -0.186 \ 0.032)'$ and initial values according to unconditional expectation and (b) $\nu = (0 \ 0 \ 0)'$ and zero initial values, True Cointegrating Rank $r = 2$, VAR Order $p = 2$, Nominal Significance Level 0.05, $R = 10000$

	Panel A: $H_0 : r_0 = 2$			Panel B: $H_0 : r_0 = 1$			Panel C: $H_0 : r_0 = 0$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
SLT (a), (b)	0.0129	0.0285	0.0525	0.0557	0.1323	0.4398	0.2710	0.6235	0.9908
SLT_1^* (a)	0.0008	0.0086	0.0277	0.0100	0.0710	0.3675	0.1074	0.4934	0.9867
SLT_1^* (b)	0.0103	0.0264	0.0462	0.0243	0.1066	0.4020	0.1137	0.5022	0.9872
SLT_2^* (a), (b)	0.0095	0.0252	0.0456	0.0263	0.1066	0.4011	0.1248	0.5123	0.9873
GLS (a), (b)	0.0089	0.0203	0.0374	0.0415	0.1009	0.3332	0.2282	0.6162	0.9814
GLS_1^* (a)	0.0909	0.1754	0.1644	0.0001	0.0002	0.0486	0.0001	0.0077	0.4501
GLS_1^* (b)	0.0458	0.0495	0.0579	0.0533	0.1311	0.3765	0.1420	0.5524	0.9778
GLS_2^* (a), (b)	0.0374	0.0461	0.0555	0.0499	0.1209	0.3663	0.1463	0.5568	0.9768
JOH (a), (b)	0.0019	0.0082	0.0284	0.0328	0.0868	0.3222	0.2628	0.5611	0.9937
JOH_1^* (a)	0.0003	0.0030	0.0137	0.0054	0.0389	0.2622	0.0955	0.4236	0.9887
JOH_1^* (b)	0.0015	0.0093	0.0305	0.0118	0.0669	0.2971	0.0997	0.4300	0.9885
JOH_2^* (a), (b)	0.0016	0.0097	0.0315	0.0134	0.0663	0.2982	0.1126	0.4416	0.9884

Note: The labels behind the test abbreviations refer to the two simulation setups (a) and (b) described in the header of the Table. For the tests that are invariant to the specifications (a) and (b) the rejection frequencies are displayed only once.

correlations of the pseudo-data that are similar to the Toda-DPGs, e.g. the correlations are between 0.3 and 0.7 if $T = 100$. Not surprisingly, both bootstrap algorithms now produce very similar test outcomes. This can be seen when comparing the bootstrap tests' rejection frequencies for the setup (b) shown in the last two lines of each of three partitions of Table 5. Hence, it is obviously the combination of parameter estimates from different models that can lead to stark differences in the bootstrap procedures, namely if the estimates of ν vary strongly. Why is GLS_1^* most negatively affected by that? The combination of estimates from different models together with the pronounced negative drifts in case of nonzero deterministics may lead to a strong mismatch between initial values and the actual properties of the pseudo data of Algorithm 1. This, in turn, may result in poor properties of the level bootstrap estimator, say $\tilde{\mu}_0^*$. In fact, the mean of the adjusted bootstrap data often clearly deviates from zero in contrast to the case of Algorithm 2. This line of argument may explain the additional deterioration of the GLS_1^* test. Given the poor performance of the GLS bootstrap procedures observed in some situations and the strong sensitivity to the specification of the initial values and the deterministic component, we do not recommend to apply the GLS bootstrap tests.

When analysing the relative performance of the bootstrap and asymptotic tests we should remember that the Algorithm 2-bootstrap tests are invariant to changes in ν and $y_{1,1}$ as long as we initialize the data at the unconditional expectation. In contrast, the precision of the restricted estimates of ν strongly varies when changing ν and $y_{1,1}$. Thus, poor estimates of ν cannot generally explain differences in the performance of the asymptotic and Algorithm 2-bootstrap tests. Of course, this also applies to the bootstrap tests based on Algorithm 1. Their worse performance relative to both the asymptotic and Algorithm 2-bootstrap tests, which we often observe, is not primarily due to even worse estimates of ν but follows from a resulting mismatch of parameter estimates stemming from two different models.

Although, the bootstrap procedures perform worse than the asymptotic tests for many versions of the DGP (5.2) applying bootstrap methods can still result in mild improvements in terms of the size in some cases as shown in Table 5, e.g. for the JOH test.

Focussing on the SLT and JOH bootstrap tests we find that the SLT procedures tend to produce higher rejection frequencies in line with the relative properties of the asymptotic tests. Accordingly, we observe slightly better size properties for the bootstrap SLT test in case of the KPSW-DGPs and the Toda-DGPs for $a_1 = 0.9$ in connection with $T = 50$ and $T = 100$ since the tests are somewhat conservative in these situations. Similarly, the powers of the SLT bootstrap tests are sometimes visibly higher than of the JOH procedures, e.g. in case of the Toda-DGP with $a_1 = 0.9$ and $T = 100$. Hence, applying the SLT bootstrap procedure can be beneficial. Note, however, that we only consider two different sets of processes.

Table 6. Average Unrestricted and Restricted Estimates of ν for DGP (5.2) with $T = 100$

Specification	True Values	Unrestricted Estimates	Restricted Estimates $r=0$	$r=1$	$r=2$
(1): Table 4	$\begin{pmatrix} -0.038 \\ -0.186 \\ 0.032 \end{pmatrix}$	$\begin{pmatrix} -0.557 \\ [0.323] \\ -0.560 \\ [0.842] \\ -0.356 \\ [0.480] \end{pmatrix}$	$\begin{pmatrix} 0.004 \\ [0.001] \\ 0.000 \\ [0.002] \\ 0.003 \\ [0.001] \end{pmatrix}$	$\begin{pmatrix} -0.111 \\ [0.232] \\ -0.263 \\ [1.042] \\ -0.086 \\ [0.368] \end{pmatrix}$	$\begin{pmatrix} -0.322 \\ [0.359] \\ -0.398 \\ [0.933] \\ -0.163 \\ [0.554] \end{pmatrix}$
(2): Table 5 (a)	$\begin{pmatrix} -0.038 \\ -0.186 \\ 0.032 \end{pmatrix}$	$\begin{pmatrix} -0.554 \\ [0.321] \\ -0.554 \\ [0.842] \\ -0.347 \\ [0.480] \end{pmatrix}$	$\begin{pmatrix} 0.004 \\ [0.001] \\ 0.000 \\ [0.002] \\ 0.003 \\ [0.001] \end{pmatrix}$	$\begin{pmatrix} -0.113 \\ [0.234] \\ -0.268 \\ [1.043] \\ -0.087 \\ [0.369] \end{pmatrix}$	$\begin{pmatrix} -0.331 \\ [0.361] \\ -0.404 \\ [0.925] \\ -0.167 \\ [0.554] \end{pmatrix}$
(3): Table 5 (b)	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ [0.004] \\ 0.000 \\ [0.008] \\ 0.000 \\ [0.006] \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ [0.001] \\ 0.000 \\ [0.001] \\ 0.000 \\ [0.001] \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ [0.002] \\ 0.000 \\ [0.009] \\ 0.000 \\ [0.003] \end{pmatrix}$	$\begin{pmatrix} 0.000 \\ [0.003] \\ 0.000 \\ [0.008] \\ 0.000 \\ [0.005] \end{pmatrix}$

Note: Standard deviations are given in brackets.

Some further remarks are in order. The discussed effects of the deterministic component becomes less severe the larger sample size is. We should note that our results on the KPSW-DGP are also driven by the small entries in the error term covariance matrix. After dividing by the standard errors, the magnitude of the deterministic components is rather large. We obtain similar results for Toda-DGPs within the framework of (2.1) when using values of μ_0 and μ_1 that are comparable to the scaled magnitudes of ν and $y_{1,1}$.

We have also analysed the correlations between the decisions of the asymptotic and bootstrap tests. The correlations with respect to the corresponding asymptotic and bootstrap tests as well as regarding the two bootstrap procedures are usually above 0.8 and often larger than 0.9 for the Toda-DGPs and the KPSW-DGP with zero deterministics. In contrast, for the KPSW-DGPs with nonzero deterministic components, the correlations involving the Algorithm 1-bootstrap tests can drop below 0.5, in particular for the GLS tests.

Generally, our findings are in line with the simulation results Swensen (2006a, 2005) has obtained for the Johansen test. Swensen (2006a) considers a five-dimensional VAR(1) process with one cointegrating relation, which is similar to a five-dimensional version of the Toda process (5.1). Swensen (2005) additionally employs the Toda-DGP and another DGP that has been used in the literature before. Nevertheless, the difference in power between asymptotic and bootstrap tests seem to be a bit smaller in the setups analysed by Swensen (2006a, 2005). Moreover, Swensen (2005) has not observed pronounced differences between the two bootstrap algorithms. However, the parameters of the deterministic components were set to zero and the initial values were drawn from normal distributions with zero mean. Therefore, the unrestricted and restricted estimates of the constant term vector and, thus, the properties of the two bootstrap test versions should be similar.

6 Summary

We have shown the asymptotic validity of two recursive bootstrap procedures for systems cointegration tests with a prior adjustment for deterministic terms suggested by Saikkonen et al. (2006) and Saikkonen & Lütkepohl (2000b). Our simulation results have confirmed that bootstrap methods can help to improve the size properties of systems cointegration test. The bootstrap versions of Saikkonen et al. (2006) perform better than bootstrapped Johansen tests in certain situations whereas the small sample properties of bootstrap counterparts of Saikkonen & Lütkepohl (2000b) are often inferior. Thus, we do not recommend to apply the latter ones in empirical work. We also found that combining parameter estimates from a VAR, which does not impose the cointegrating rank tested and from a VECM, which employs this rank, can cause strong deteriorations. Such a combination of parameters is done within one of the recursive bootstrap schemes. Accordingly, one should be careful when applying this bootstrap scheme. However, the deteriorations occur if the estimates of the constant term obtained from the two models strongly differ. Thus, if the two bootstrap versions deliver different test outcomes, applied researcher can easily check one likely source of the contradicting results.

Although size improvements are possible, applied researcher also have to be aware of stronger losses in small sample power and size deteriorations. Hence, some of our findings seem to support a remark of Johansen (2002, p. 1930). He does not necessarily expect an improved approximation of the tests' distributions by the use of bootstrap methods. The reason is the nonuniformity in the convergence with respect to the nuisance parameters in the null distributions of the test statistics.

By contrast, Johansen (2002) proposes a small sample correction following Bartlett (1937). However, Swensen (2006a) has shown that bootstrap approximations work better than the correction suggested by Johansen (2002) in a number of situations.

It should be rather straightforward to adjust our asymptotic results such that different assumptions about the presence of deterministic terms can be made. Hence, one may apply valid bootstrap versions of cointegration tests considered by Saikkonen & Lütkepohl (2000a), which allow for level shifts at known time, or of a test procedure excluding a linear trend as e.g. suggested by Saikkonen et al. (2006). Finally, an extension of the bootstrap algorithms in order to consistently determine the cointegration rank may be conducted in future research. The asymptotic results of Swensen (2006a) regarding the Johansen test indicate that a corresponding consistent sequential bootstrap procedure can also be found for tests with a prior adjustment for deterministic terms.

Appendix A: Proofs

Proof of Lemma 1

The Granger representation is obtained from the recursion (4.2), which is a VECM with a restricted constant. Accordingly, we can proceed as in the proof of Lemma 1 of Swensen (2006a) and only need to take account of the differing linear terms. From arguments in Swensen (2006a, 2006b) and Johansen (1995) we obtain the representation

$$
\begin{aligned}
Y_t^{c*} = \begin{pmatrix} \tilde{\beta}' y_t^{c*} \\ \tilde{\beta}_\perp' \Delta y_t^{c*} \end{pmatrix} &= \tilde{\bar{A}}(L)^{-1}(\bar{\tilde{\alpha}}, \bar{\tilde{\alpha}}_\perp)'(\varepsilon_t^* - \tilde{\alpha}\tilde{\theta}) \\
&= \tilde{\bar{C}}(L)(\bar{\tilde{\alpha}}, \bar{\tilde{\alpha}}_\perp)'(\varepsilon_t^* - \tilde{\alpha}\tilde{\theta}) \\
&= \tilde{\bar{C}}(1)(\bar{\tilde{\alpha}}, \bar{\tilde{\alpha}}_\perp)'(\varepsilon_t^* - \tilde{\alpha}\tilde{\theta}) + \tilde{\bar{C}}^{\#}(L)\Delta(\bar{\tilde{\alpha}}, \bar{\tilde{\alpha}}_\perp)'(\varepsilon_t^* - \tilde{\alpha}\tilde{\theta})
\end{aligned}
$$

for some suitable polynomial $\tilde{\bar{A}}(L)$. Since the bootstrap error terms and involved polynomials are the same as in Swensen (2006a) we can focus on the linear part. Consider the decomposition

$$
y_t^{c*} = \bar{\tilde{\beta}}_\perp \tilde{\beta}_\perp' \sum_{i=p+1}^{t} \Delta y_t^{c*} + \bar{\tilde{\beta}} \tilde{\beta}' y_t^{c*} + \bar{\tilde{\beta}}_\perp \tilde{\beta}_\perp' y_p^{c*}. \tag{A.1}
$$

From Swensen (2006b, Proof of Equation (A.2)) we see that only the second term of (A.1) contributes to the linear part in the current setup of a restricted constant. The contribution from $\bar{\tilde{\beta}}\tilde{\beta}' y_t^{c*}$ is given by $\bar{\tilde{\beta}}\tilde{\theta}$ which is equal to $\tilde{\tau}_0$ as defined in Lemma 1.

Finally, the terms involving deterministic components in the definition of the remainder term

$\sqrt{T}R^*_{t,T}$ in Swensen (2006b) are zero in our framework. Hence, we obtain as remainder term

$$\begin{aligned}\sqrt{T}R^{c*}_{t,T} &= (0, \bar{\bar{\beta}}_\perp)\tilde{\tilde{C}}^\#(L)(\bar{\bar{\alpha}}, \bar{\bar{\alpha}}_\perp)'(\varepsilon^*_t - \varepsilon_p) + (\bar{\bar{\beta}}, 0)\tilde{\tilde{C}}(1)(\bar{\bar{\alpha}}, \bar{\bar{\alpha}}_\perp)'\varepsilon^*_t \\ &\quad + (\bar{\bar{\beta}}, 0)\tilde{\tilde{C}}^\#(L)(\bar{\bar{\alpha}}, \bar{\bar{\alpha}}_\perp)'\Delta\varepsilon^*_t + \bar{\bar{\beta}}_\perp\tilde{\beta}'_\perp y^{c*}_p.\end{aligned} \tag{A.2}$$

Note that $\bar{\bar{\beta}}_\perp\tilde{\beta}'_\perp y^{c*}_p = \bar{\bar{\beta}}_\perp\tilde{\beta}'_\perp(y^c_p - (\tilde{\mu}_1 - \mu_1)p) = \bar{\bar{\beta}}_\perp\tilde{\beta}'_\perp y^c_p + O_p(T^{-1/2})$ given the properties of $\tilde{\mu}_1$ such that $\bar{\bar{\beta}}_\perp\tilde{\beta}'_\perp y^{c*}_p$ is bounded in probability. Thus, the result $P^*(\max_{p+1\le t\le T}|R^{c*}_{t,T}| > \epsilon) \to 0$ for all $\epsilon > 0$ in probability as $T \to \infty$ follows from the proof of Lemma 1 in Swensen (2006a). Finally, $E^*[\varepsilon^*_t\varepsilon^{*\prime}_t] = \hat{\Omega} = T^{-1}\sum_{t=p+1}^T \hat{\varepsilon}_t\hat{\varepsilon}'_t \to \Omega$ in probability as $T \to \infty$ is already given in Lemma 1 of Swensen (2006a). This completes the proof of Lemma 1. ■

Proof of Proposition 1

We follow the proof of Theorem 4.1 for $SLT(r_0)$ given in Saikkonen et al. (2006). To this end, we repeatedly make use of Lemmas 1, S1, S2 of Swensen (2006a, 2006b), and results from Saikkonen et al. (2006, Appendix A.3: Proof of Theorem 4.1) and Johansen (1995, Chapters 10, 13). Moreover, it is convenient to use the bootstrap stochastic order symbols introduced by Park (2003) and Chang & Park (2003). Let (X^*_T) be a sequence of bootstrap statistics. Then, $X_T = o^*_p(T^\lambda)$, if for every $\epsilon > 0$, $P^*\{|T^{-\lambda}X^*_T| > \epsilon\} \to 0$ in probability as $T \to \infty$. Furthermore, $X^*_T = O^*_p(T^\lambda)$, if for every $\epsilon > 0$ and $\eta > 0$ given, there exists a constant $K > 0$ such that for all T sufficiently large $P\left(P^*\{|T^\lambda X^*_T| > K\} < \epsilon\right) > 1 - \eta$. These definitions can be adopted for sequences of random vectors. Moreover, following Chang & Park (2003) and Swensen (2006a), we say $X^*_T \xrightarrow{d^*} X$ if the conditional distribution of X^*_T weakly converges to that of X in probability. To be precise, $X^*_T \xrightarrow{d^*} X$ if and only if $E^*[f(X^*_T)] \to E[f(X)]$ in probability for all bounded continuous functions f. In order to determine the convergence orders of product moments of bootstrap series we will apply a corresponding version of Theorem B.13 of Johansen (1995). This is possible since the relevant bootstrap series have a MA representation in terms of the bootstrap error terms based on the results of Swensen (2006b, Proof of Lemma S1).

We first proof the following lemma, which is the counterpart of Lemma A.14 in Saikkonen et al. (2006) that states the asymptotic properties of $\tilde{\mu}_1$.[6]

Lemma A1. Under Assumption 1 and given (4.2), we have

$$\tilde{\beta}'(\tilde{\mu}^*_1 - \tilde{\mu}_1) = O^*_p(T^{-3/2}) \tag{A.3}$$

$$T^{1/2}\tilde{\beta}'_\perp(\tilde{\mu}^*_1 - \tilde{\mu}_1) \xrightarrow{d^*} N(0, \beta'_\perp C\Omega C'\beta_\perp). \tag{A.4}$$

[6]Note again that we do not consider a level shift at unknown time in our framework. Therefore, the model setup and the corresponding proofs in Saikkonen et al. (2006) simplify accordingly.

Since the estimator $\tilde{\mu}_1^*$ is a function of the bootstrap RR estimators regarding the VECM (3.1), we have to derive their convergence properties in a first step. As usual, the bootstrap estimators are indicated by an asterisk. The relevant properties are summarized in Lemma A2. For the derivations we have to assume that $\tilde{\beta}^*$ has been made unique by the normalization $\tilde{\beta}^*[(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\beta}'\tilde{\beta}^*]^{-1} = \tilde{\beta}^*(\bar{\tilde{\beta}}'\tilde{\beta}^*)^{-1}$. A corresponding normalization follows for $\tilde{\alpha}^*$. Note that all other relevant quantities are invariant to these normalizations (compare Saikkonen & Lütkepohl 2000b). Therefore, we do not explicitly indicate the normalized version of the estimators in the following.

Lemma A2. Under Assumption 1, we have

$$\tilde{\alpha}^* - \alpha = O_p^*(T^{-1/2}) \tag{A.5}$$

$$\tilde{\beta}^* - \beta = O_p^*(T^{-1}) \tag{A.6}$$

$$\hat{\nu}^* - \nu = O_p^*(T^{-1/2}) \tag{A.7}$$

$$\hat{\Gamma}_i^* - \Gamma_i = O_p^*(T^{-1/2}) \ (i = 1, \ldots, p-1) \tag{A.8}$$

$$\tilde{\phi}^* - \phi = O_p^*(T^{-3/2}) \tag{A.9}$$

$$\tilde{\Omega}^* - \Omega = O_p^*(T^{-1/2}). \tag{A.10}$$

Proof To proof Lemma A2 we proceed as in the proofs of Lemma 2.1 in Saikkonen & Lütkepohl (2000a) and Lemma A.13 in Saikkonen et al. (2006), where the asymptotic properties of the corresponding sample RR estimators are proven. This also applies to (A.7) and (A.8) since $\hat{\nu}^*$ and $\hat{\Gamma}_i^*$, $i = 1, \ldots, p-1$, are RR estimators. Hence, one has to analyse the framework of the generalized eigenvalue problem that underlies the RR estimation of (3.1). This requires to state a counterpart to Lemma 10.3 in Johansen (1995). Taking account of the necessary adjustments due to the linear trend specification in (3.1), 10.13-10.17 of Johansen (1995, Lemma 10.3) are replaced by the results of Lemmas S1 and S2 of Swensen (2006b). The counterpart to 10.18 of Johansen (1995, Lemma 10.3) can be obtained corresponding to Johansen (1995, pp. 146-147) by referring to Theorem B.13 of Johansen (1995) and the proof of Lemma S1 in Swensen (2006b). From this proof we see that the relevant series have conditional zero expectations and MA representations in terms of the bootstrap error terms. Given the foregoing, (A.5) - (A.9) easily follow.

The convergence result for $\tilde{\Omega}^*$ in (A.10) requires some additional remarks. First define $\Omega^* = T^{-1}\sum_{t=p+1}^{T} \varepsilon_t^* \varepsilon_t^{*\prime}$ and $\tilde{\Omega}^* = T^{-1}\sum_{t=p+1}^{T} \tilde{\varepsilon}_t^* \tilde{\varepsilon}_t^{*\prime}$, where $\tilde{\varepsilon}_t^*$ are the residuals obtained from estimating (3.1). Then, following the idea of the proof of Theorem 13.5 in Johansen (1995) and using Theorem B.13 of Johansen (1995) we obtain $\tilde{\Omega}^* - \Omega^* = o_p^*(T^{-1/2})$. Since the bootstrap error terms ε_t^* are randomly drawn with replacement, they can be regarded as an i.i.d. sample from the empirical distribution given by the estimated residuals $\hat{\varepsilon}_t$. In other words, the bootstrap samples ε_t^* become i.i.d. with covariance matrix $\hat{\Omega} = T^{-1}\sum_{t=p+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$ with respect to the probability measure

P^* (compare Park 2002, 2003 for the univariate case). Because we assume that the fourth moments of the error terms exists, a central limit theorem shows that $\Omega^* - \hat{\Omega} = O_p^*(T^{-1/2})$ (compare Hamilton 1994, p. 340-343, for the standard asymptotic situation). Hence, we get $\tilde{\Omega}^* - \hat{\Omega} = O_p^*(T^{-1/2})$ due the triangular inequality. From Saikkonen & Lütkepohl (2000b, Lemma A5) we conclude that $\hat{\Omega} - \Omega = O_p(T^{-1/2})$. Then, (A.10) follows by referring again to the triangular inequality and the proof of Lemma A.2 is complete. ■

Note that the asymptotic results of the bootstrap RR estimators also hold relative to the sample estimators, i.e. we also have $\tilde{\alpha}^* - \tilde{\alpha} = O_p^*(T^{-1/2})$ etc.. This follows from the corresponding convergence properties of the sample estimators and the triangular inequality. We will use this fact when referring to Lemma A1 later on.

Following Saikkonen et al. (2006) we consider a transformed version of the relevant VECM (3.1) by applying the decomposition $y_t^* \rightarrow \tilde{\mu}_0 + \tilde{\mu}_1 t + x_t^*$, where $\tilde{\mu}_0$ is some estimator of μ_0 with $\tilde{\theta} = \tilde{\beta}'\tilde{\mu}_0$ being the estimator of θ used in (4.2). A possible estimator $\tilde{\mu}_0$ is discussed in Trenkler (2006). We obtain

$$\Delta x_t^* = \tilde{\nu}^{(0)} + \tilde{\alpha}(\tilde{\beta}' x_{t-1}^* - \tilde{\phi}^{(0)}(t-1)) + \sum_{j=1}^{p-1} \hat{\Gamma}_j \Delta x_{t-j}^* + \varepsilon_t^*, \quad t = p+1, p+2, \ldots, \tag{A.11}$$

where $\tilde{\nu}^{(0)} = \hat{\nu} - \hat{\Gamma}\tilde{\mu}_1 + \tilde{\alpha}\tilde{\theta}$ and $\tilde{\phi}^{(0)} = \tilde{\phi} - \tilde{\beta}'\tilde{\mu}_1$. Note that $\tilde{\nu}^{(0)}$ and $\tilde{\phi}^{(0)}$ are zero due to the definitions of $\hat{\nu}$ after (4.2) and $\tilde{\mu}_1$ in (2.5). Analogously to Saikkonen et al. (2006), estimators of $\tilde{\nu}^{(0)}$ and $\tilde{\phi}^{(0)}$ are derived by transforming the RR estimators of the VECM (3.1) such that we have $\tilde{\phi}^{(0)*} = \tilde{\phi}^* - \tilde{\beta}^{*\prime}\tilde{\mu}_1$ and $\tilde{\nu}^{(0)*} = \hat{\nu}^* - \hat{\Gamma}^*\tilde{\mu}_1 + \tilde{\alpha}^*\tilde{\theta}^*$. Note, however, that the estimators can also be obtained from the transformed VECM (A.11), which contains a restricted linear trend and unrestricted mean term as algorithm (3.1). Thus, using (A.9) and (A.7) of Lemma A.1 the following lemma can be deduced since $\tilde{\phi}^{(0)}$ and $\tilde{\nu}^{(0)}$ are zero.

Lemma A3. Under Assumption 1, we have

$$\tilde{\phi}^{(0)*} = O_p^*(T^{-3/2}) \tag{A.12}$$

$$\tilde{\nu}^{(0)*} = O_p^*(T^{-1/2}). \tag{A.13}$$

Applying the definitions for $\tilde{\phi}^{(0)*}$ and $\tilde{\nu}^{(0)*}$, we have $\tilde{\beta}_\perp^{*\prime}\tilde{C}^*\tilde{\nu}^{(0)*} = \tilde{\beta}_\perp^{*\prime}\tilde{\mu}_1^* - \tilde{\beta}_\perp^{*\prime}\tilde{C}^*\hat{\Gamma}^*\tilde{\mu}_1 + \tilde{\beta}_\perp^{*\prime}\tilde{C}^*\hat{\Gamma}^*\bar{\tilde{\beta}}^*\tilde{\beta}^{*\prime}\tilde{\mu}_1 + \tilde{\beta}_\perp^{*\prime}\tilde{C}^*\hat{\Gamma}^*\bar{\tilde{\beta}}^*\tilde{\phi}^{(0)*}$ such that

$$\tilde{\beta}_\perp^{*\prime}(\tilde{\mu}_1^* - \tilde{\mu}_1) = \tilde{\beta}_\perp^{*\prime}\tilde{C}^*(\tilde{\nu}^{(0)*} - \hat{\Gamma}\bar{\tilde{\beta}}^*\tilde{\phi}^{(0)*})$$

corresponding to Saikkonen et al. (2006). Then, by Lemma A.3,

$$\tilde{\beta}_\perp^{*\prime}(\tilde{\mu}_1^* - \tilde{\mu}_1) = \tilde{\beta}_\perp^{*\prime}\tilde{C}^*\tilde{\nu}^{*(0)} + o_p^*(T^{-1/2})$$

follows. The estimator $\tilde{\nu}^{(0)*}$ can be viewed as the LS estimator of $\tilde{\nu}^{(0)}$ in the auxiliary model

$$\Delta x_t^* = \tilde{\nu}^{(0)} + \tilde{\alpha}(\tilde{\beta}^{*\prime}x_{t-1}^* - \tilde{\phi}^{(0)*}(t-1)) + \sum_{j=1}^{p-1}\hat{\Gamma}_j\Delta x_{t-j}^* + u_t^*, \tag{A.14}$$

which is obtained from (A.11) by replacing $\tilde{\beta}$ and $\tilde{\phi}^{(0)}$ by their bootstrap analogs $\tilde{\beta}^*$ and $\tilde{\phi}^{(0)*}$ and by noting that $u_t^* = \varepsilon_t^* - \tilde{\alpha}[(\tilde{\beta}^* - \tilde{\beta})'x_{t-1}^* + \tilde{\phi}^{*(0)}(t-1)]$. Keep in mind that $\tilde{\beta}^{*\prime}x_{t-1}^* = \tilde{\beta}^{*\prime}y_{t-1}^{c*} - \tilde{\theta}^*$, $\Delta y_{t-i}^{c*} = \Delta x_{t-i}^*$ $(i = 0, \dots, p-1)$, and that the estimator $\tilde{\theta}^*$ can be recovered from the RR estimators of (3.1).

Equation (A.14) implies that the estimator $\tilde{\beta}_\perp^{*\prime}\tilde{C}^*\tilde{\nu}^{*(0)}$ can be obtained by LS from

$$\tilde{\beta}_\perp^{*\prime}\tilde{C}^*\Delta x_t^* = \tilde{\Lambda}p_t^* + \tilde{u}_t^*, \tag{A.15}$$

where $p_t^* = [1, \Delta x_{t-1}^*, \dots, \Delta x_{t-p+1}^*]$, $\tilde{\Lambda}$ is a conformable coefficient matrix and $\tilde{u}_t^* = \tilde{\beta}_\perp^{*\prime}\tilde{C}^*\varepsilon_t^* - \tilde{\beta}_\perp^{*\prime}\tilde{C}^*\tilde{\alpha}[(\tilde{\beta}^* - \tilde{\beta})'x_{t-1}^* + \tilde{\phi}^{*(0)}(t-1)]$. Using the definition of $\tilde{C}$, Lemma A.2, and the assumptions, it is straightforward to show that the asymptotic properties of the LS estimator of $\tilde{\Lambda}$ in the auxiliary regression model (A.15) can be obtained by assuming that the error term equals $\beta_\perp' C\varepsilon_t^*$. Since one can show that the estimation of the intercept term in (A.15) is asymptotically independent of the estimation of the other regression coefficients and because $\tilde{\nu}^{(0)}$ is zero, we obtain

$$T^{1/2}\tilde{\beta}_\perp^{*\prime}(\tilde{\mu}_1^* - \tilde{\mu}_1) = \beta_\perp' C T^{-1/2}\sum_{t=p+1}^{T}\varepsilon_t^* + o_p^*(1).$$

This, Lemma 1 in Swensen (2006a), and a standard central limit theorem yield

$$T^{1/2}\tilde{\beta}_\perp^{*\prime}(\tilde{\mu}_1^* - \mu_1^*) \xrightarrow{d^*} N(0, \beta_\perp' C\Omega C'\beta_\perp). \tag{A.16}$$

Finally, (A.4) is obtained since $\tilde{\beta}_\perp^*$ on the l.h.s. of (A.16) can be replaced by $\tilde{\beta}_\perp$. This follows from $(\tilde{\beta}_\perp^* - \tilde{\beta}_\perp)'(\tilde{\mu}_1^* - \mu_1^*) = o_p^*(T^{-1/2})$ due to Lemmas A2 and A3, and (A.16).

Note that $\tilde{\beta}^{*\prime}(\tilde{\mu}_1^* - \tilde{\mu}_1) = \tilde{\phi}^{(0)*} = O_p^*(T^{-3/2})$ given the definition of $\tilde{\phi}^{(0)*}$ and Lemma A.3. Similar as before, one can derive $(\tilde{\beta}^* - \tilde{\beta})'(\tilde{\mu}_1^* - \mu_1^*) = o_p^*(T^{-1/2})$ such that (A.3) is shown. This completes the proof of Lemma A1. ■

Following the explanations in Section 3, the SLT_1^* test is based on the VECM for the adjusted bootstrap series $\tilde{y}_t^{c*} = y_t^* - \tilde{\mu}_1 t$ given by

$$\Delta\tilde{y}_t^{c*} = \tilde{\alpha}(\tilde{\beta}'\tilde{y}_{t-1}^{c*} - \tilde{\theta}) + \sum_{j=1}^{p-1}\hat{\Gamma}_j\Delta\tilde{y}_{t-j}^{c*} + e_t^*, \tag{A.17}$$

where $e_t^* = \varepsilon_t^* - \tilde{\alpha}\tilde{\beta}'(\tilde{y}_{t-1}^{c*} - y_{t-1}^{c*})(t-1) + \Delta\tilde{y}_t^{c*} - \Delta y_t^{c*} - \sum_{j=1}^{p-1}\hat{\Gamma}_j(\Delta\tilde{y}_{t-j}^{c*} - \Delta y_{t-j}^{c*}) = \varepsilon_t^* + \tilde{\alpha}\tilde{\beta}'(\tilde{\mu}_1^* - \tilde{\mu}_1)(t-1) - \tilde{\Gamma}(\tilde{\mu}_1^* - \mu_1^*)$. Define $\tilde{Z}_{0t}^* = \Delta\tilde{y}_t^{c*}$, $\tilde{Z}_{1t}^* = (\tilde{y}_{t-1}^{c*\prime}, 1)'$, and $\tilde{Z}_{2t}^* = (\Delta\tilde{y}_{t-1}^{c*\prime}, \ldots, \Delta\tilde{y}_{t-p+1}^{c*\prime})'$. Moreover, we apply the definitions of $\tilde{R}_{it}^*$ and $\tilde{S}_{ij}^*$ $(i,j = 0,1)$ from Section 3. In the same way we can obtain the moment matrices S_{ij}^* $(i,j = 0,1)$ from Z_{it}^* and R_{it}^* $(i = 0,1)$, which are defined in terms of $y_t^{c*} = y_t^* - \tilde{\mu}_1 t$. Furthermore, let S_{ij} $(i,j = 0,1)$ be the moment matrices based on the original observations adjusted by the trend component, i.e $y_t^c = y_t - \mu_1 t$. Finally, define $\tilde{\beta}^+ = (\tilde{\beta}', -\tilde{\theta})'$.

To apply the framework of Johansen (1995, Ch. 11) we need corresponding results to Lemmas S1 and S2 of Swensen (2006b). To this end, we consider first

Lemma A4. Under Assumption 1, we have

$$\begin{aligned} \tilde{S}_{00}^* &= \Sigma_{00} + o_p^*(1) \\ \tilde{S}_{01}^*\tilde{\beta}^+ &= \Sigma_{0\beta} + o_p^*(1) \\ \tilde{\beta}^{+\prime}\tilde{S}_{11}^*\tilde{\beta}^+ &= \Sigma_{\beta\beta} + o_p^*(1), \end{aligned}$$

where Σ_{00}, $\Sigma_{0\beta}$, and $\Sigma_{\beta\beta}$ are the respective probability limits of S_{00}, $S_{01}\beta^+$, and $\beta^{+\prime}S_{11}\beta^+$ as $T \to \infty$.

Proof We focus on proving the first statement. The last two ones follow by corresponding arguments. Using the definitions stated before Lemma A4, we can write in accordance with Johansen (1995, Ch. 6)

$$\tilde{S}_{00}^* = T^{-1}\left(\sum_{t=p+1}^{T}\tilde{Z}_{0t}^*\tilde{Z}_{0t}^{*\prime} - \sum_{t=p+1}^{T}\tilde{Z}_{0t}^*\tilde{Z}_{2t}^{*\prime}\left(\sum_{t=p+1}^{T}\tilde{Z}_{2t}^*\tilde{Z}_{2t}^{*\prime}\right)^{-1}\sum_{t=p+1}^{T}\tilde{Z}_{2t}^*\tilde{Z}_{0t}^{*\prime}\right).$$

Note that $\tilde{Z}_{it}^* = Z_{it}^* - (\tilde{\mu}_1^* - \tilde{\mu}_1)$, $i = 0,2$, where $Z_{0t}^* = \Delta y_t^{c*}$ and $Z_{2t}^* = (\Delta y_{t-1}^{c*\prime}, \ldots, \Delta y_{t-p+1}^{c*\prime})'$ have zero conditional on the observations and possess MA representations given Lemma 1 and Swensen (2006b, Lemma S1). Using Proposition 1 and applying Johansen (1995, Theorem B.13) analogously, we have $T^{-1}\sum_{t=p+1}^{T} Z_{it}^*(\tilde{\mu}_1^* - \tilde{\mu}_1)' = o_p^*(1)$, $i = 0,2$, and $T^{-1}\sum_{t=p+1}^{T}(\tilde{\mu}_1^* - \tilde{\mu}_1)(\tilde{\mu}_1^* - \tilde{\mu}_1)' = o_p^*(1)$. Combining these results, yields

$$\begin{aligned} \tilde{S}_{00}^* &= T^{-1}\left(\sum_{t=p+1}^{T}Z_{0t}^*Z_{0t}^{*\prime} - \sum_{t=p+1}^{T}Z_{0t}^*Z_{2t}^{*\prime}\left(\sum_{t=p+1}^{T}Z_{2t}^*Z_{2t}^{*\prime}\right)^{-1}\sum_{t=p+1}^{T}Z_{2t}^*Z_{0t}^{*\prime}\right) + o_p^*(1) \\ &= S_{00}^* + o_p^*(1). \end{aligned}$$

Finally, $S_{00}^* = \Sigma_{00} + o_p^*(1)$ follows from Swensen (2006b, Lemma S1) in conjunction with Lemma 1, what completes the proof. ■

Next, we have

Lemma A5. Under Assumption 1,

$$T^{-1}\tilde{B}_T'\tilde{S}_{11}^*\tilde{B}_T \to \int_0^1 G_+(s)G_+(s)'ds \tag{A.18}$$

$$\tilde{B}_T'\tilde{S}_{10}^*\tilde{\alpha}_\perp \to \int_0^1 G_+(s)dG_*(s)'\alpha_\perp, \tag{A.19}$$

where $\tilde{B}_T = \begin{pmatrix} \tilde{\beta}_\perp & 0 \\ 0 & T^{1/2} \end{pmatrix}$ and $G_+(s) = [B_*(s)'\Omega^{1/2}C'\beta_\perp, 1]'$, $dG_*(s) = \Omega^{1/2}dB_*(s)$, and convergence is weakly in probability.

Proof We first show

$$T^{-1}\tilde{B}_T'\tilde{S}_{11}^*\tilde{B}_T = T^{-2}\tilde{B}_T'\sum_{t=p+1}^{T}\tilde{Z}_{1t}^*\tilde{Z}_{1t}^{*\prime}\tilde{B}_T + o_p^*(1) \tag{A.20}$$

$$\tilde{B}_T'\tilde{S}_{10}^*\tilde{\alpha}_\perp = T^{-1}\tilde{B}_T'\sum_{t=p+1}^{T}\tilde{Z}_{1t}^*\bar{\varepsilon}_t^{*\prime}\tilde{\alpha}_\perp + o_p^*(1) \tag{A.21}$$

where $\bar{\varepsilon}_t^* = \varepsilon_t^* - \hat{\Gamma}\tilde{\beta}_\perp\bar{\tilde{\beta}}_\perp'(\tilde{\mu}_1^* - \tilde{\mu}_1)$.

Similar as above, we can write

$$\begin{aligned} T^{-1}\tilde{B}_T'\tilde{S}_{11}^*\tilde{B}_T &= T^{-2}\sum_{t=p+1}^{T}\tilde{B}_T'\tilde{Z}_{1t}^*\tilde{Z}_{1t}^{*\prime}\tilde{B}_T \\ &- T^{-2}\sum_{t=p+1}^{T}\tilde{B}_T'\tilde{Z}_{1t}^*\tilde{Z}_{2t}^{*\prime}\left(\sum_{t=p+1}^{T}\tilde{Z}_{2t}^*\tilde{Z}_{2t}^{*\prime}\right)^{-1}\sum_{t=p+1}^{T}\tilde{Z}_{2t}^*\tilde{Z}_{1t}^{*\prime}\tilde{B}_T, \end{aligned} \tag{A.22}$$

where $\tilde{Z}_{1t}^* = Z_{1t}^* - [(\tilde{\mu}_1^* - \tilde{\mu}_1)'(t-1), 0]' = (y_{t-1}^{c*\prime}, 1)' - [(\tilde{\mu}_1^* - \tilde{\mu}_1)'(t-1), 0]'$. Note that $\tilde{\beta}_\perp' y_{t-1}^{c*}$ equals a zero-mean process plus a bounded term conditional on the observations. This follows from the definition of $\tilde{\tau}_0$ in Lemma 1 and $\sqrt{T}R_{t,T}^{c*}$ in (A.2). Then, using Lemma 1 and Proposition 1 in connection with Johansen (1995, Theorem B.13) one can show, corresponding to the proof of Lemma A4, that the second term on the r.h.s. of (A.22) is $o_p^*(1)$, what gives (A.20).

Given the derivations used in Johansen (1995, pp. 90-91) one can show

$$\tilde{S}_{10}^*\tilde{\alpha}_\perp = T^{-1}\left(\sum_{t=p+1}^{T}\tilde{Z}_{1t}^*e_t^{*\prime} - \sum_{t=p+1}^{T}\tilde{Z}_{1t}^*\tilde{Z}_{2t}^{*\prime}\left(\sum_{t=p+1}^{T}\tilde{Z}_{2t}^*\tilde{Z}_{2t}^{*\prime}\right)^{-1}\sum_{t=p+1}^{T}\tilde{Z}_{2t}^*e_t^{*\prime}\right)\tilde{\alpha}_\perp.$$

Note again that $e_t^* = \varepsilon_t^* + \tilde{\alpha}\tilde{\beta}'(\tilde{\mu}_1^* - \tilde{\mu}_1)(t-1) - \hat{\Gamma}(\tilde{\mu}_1^* - \tilde{\mu}_1)$. Hence, we obtain $T^{-1/2}\sum_{t=p+1}^{T}\tilde{Z}_{2t}^*e_t^{*\prime}\tilde{\alpha}_\perp = o_p^*(1)$ such that $\tilde{B}_T'\tilde{S}_{10}^*\tilde{\alpha}_\perp = T^{-1}\sum_{t=p+1}^{T}\tilde{B}_T'\tilde{Z}_{1t}^*e_t^{*\prime}\tilde{\alpha}_\perp + o_p^*(1)$ in connection with previous convergence results of the involved cross-products. Moreover, we can write

$e_t^{*\prime}\tilde{\alpha}_\perp = \{\varepsilon_t^* - \hat{\Gamma}[\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\beta}' + \tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'](\tilde{\mu}_1^* - \tilde{\mu}_1)\}'\tilde{\alpha}_\perp$. Since $\tilde{\beta}'(\tilde{\mu}_1^* - \tilde{\mu}_1) = O_p^*(T^{-3/2})$, $\tilde{B}_T'\tilde{S}_{10}^*\tilde{\alpha}_\perp = T^{-1}\sum_{t=p+1}^T \tilde{B}_T'\tilde{Z}_{1t}^*[\varepsilon_t^* - \hat{\Gamma}\tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'(\tilde{\mu}_1^* - \tilde{\mu}_1)]'\tilde{\alpha}_\perp + o_p^*(1)$ is obtained, what shows (A.21).

Since $\tilde{\beta}_\perp'(\tilde{\mu}_1^* - \mu_1^*) = O_p^*(T^{-1/2})$, $\tilde{Z}_{1t}^*$ cannot be simply replaced by Z_{1t}^* on the r.h.s. of (A.20) and (A.21). In fact, the estimation error with respect to $\tilde{\mu}_1^*$ has to be taken account of when deriving the limit distributional results. This is done next.

We can exactly argue as in Swensen (2006b, Proof of Lemma S2). We refer to the continuous mapping theorem involving the same functionals as in Saikkonen et al. (2006) and related work by Lütkepohl & Saikkonen (2000) and Hansen & Johansen (1998, pp. 118-121). Thus, we only have to show that the process $\{T^{-1/2}y_{[sT]}^{c*}, 0 \leq s \leq 1\}$ converges weakly in probability as element in $D[0,1]^n$, the space of functions that are right continuous and have left limits.

From Lemma 1 we see that the remainder term $R_{t,T}^{c*}$ vanishes asymptotically. The constant term is treated as in Saikkonen et al. (2006). A more detailed description of this treatment in relation to the Johansen test can be found in Hansen & Johansen (1998, pp. 118-121). Therefore, we can focus on $S_T^*(s) = T^{-1/2}\sum_{i=p+1}^{[sT]} \varepsilon_i^*$. This is the same partial sum analysed by Swensen (2006a) who has shown that $S_T^*(s)$ converges weakly in probability to a Brownian motion W with covariance matrix Ω. The latter result is based on the fact that $E^*[f(S_T^*)] \to E[f(W)]$ in probability for all bounded continuous functions f defined on $D[0,1]^n$. We have $\tilde{B}_T'\tilde{Z}_{1t}^* = \{[\tilde{\beta}_\perp' y_{t-1}^{c*} - \tilde{\beta}_\perp'(\tilde{\mu}_1^* - \tilde{\mu}_1)(t-1)]', T^{1/2}\}'$ so that Lemma A5 follows from Proposition 1 and the relationship of the functionals on the r.h.s. of (A.20) and (A.21) with those in equations (A.10)/(A.11) and (A.14)/(A.15) of Lütkepohl & Saikkonen (2000). Note in this respect that the estimator $\tilde{\mu}_1$ used in Lütkepohl & Saikkonen (2000) has the same asymptotic properties as our estimator. ■

Finally, Proposition 2 follows from Lemmas A4 and A5 in the same way as in Johansen (1995, Ch. 11) and Saikkonen et al. (2006) with obvious modifications. ■

Appendix B: Response Surface Coefficients for the *SLT*-Test

The response surface for the SLT test by Saikkonen et al. (2006) is computed exactly in the same way as the one for the GLS test, which is presented in Trenkler (2008). The idea of the response surface is to approximate the limiting distribution given in Theorem 1 by a Gamma distribution with two parameters. These parameters can be related to the mean and variance of the distribution of interest. It turns out, that we only need to derive response surface coefficients with respect to the dimension $d = n - r_0$ in order to estimate the asymptotic mean and variance of the relevant distributions. These coefficients are given in Table 7 and refer to polynomials of d. The calculated moments are then used to fit a Gamma distribution, which can be used to compute p-values or arbitrary quantiles. Further details can be found in Trenkler (2008).

Table 7. Response Surface for Mean and Variance of Asymptotic Distribution of the Trace Test Statistics $SLT(r_0)$

	Mean	Variance
d^2	2.0046	3.0125
d	1.7392	1.9664
$\sqrt{d}$	1.0027	—
1 (constant)	-0.5442	1.4214

References

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests, *Proceedings of the Royal Society of London, Series A* **160**: 268–282.

Brüggemann, R. (2004). *Model Reduction Methods for Vector Autoregressive Processes*, Springer-Verlag, Berlin.

Chang, Y. & Park, J. Y. (2003). A sieve bootstrap for the test of a unit root, *Journal of Time Series Analysis* **24**: 379–400.

Davidson, R. & MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps?, *Econometric Reviews* **19**: 55–69.

Doornik, J. A. (1998). Approximations to the asymptotic distributions of cointegration tests, *Journal of Economic Surveys* **12**: 573–593.

Elliott, G. & Müller, U. K. (2003). Tests for unit roots and the initial condition, *Econometrica* **71**: 1269–1286.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.

Hansen, P. R. & Johansen, S. (1998). *Workbook on Cointegration*, Oxford University Press, Oxford.

Horowitz, J. L. (2001). The bootstrap, *in* J. J. Heckmann & E. E. Leamer (eds), *Handbook of Econometrics, Vol. 5*, North-Holland, Amsterdam.

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

Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model, *Econometrica* **70**: 1929–1961.

King, R. G., Plosser, C. I., Stock, J. H. & Watson, M. W. (1991). Stochastic trends and economic fluctuations, *American Economic Review* **81**: 819–840.

Lütkepohl, H. & Saikkonen, P. (2000). Testing for the cointegration rank of a VAR process with a time trend, *Journal of Econometrics* **95**: 177–198.

Paparoditis, E. & Politis, D. N. (2003). Residual-based block bootstrap for unit root testing, *Econometrica* **71**: 813–855.

Paparoditis, E. & Politis, D. N. (2005). Bootstrapping unit root testing for autoregressive time series, *Journal of the American Statistical Association* **100**: 545–553.

Park, J. Y. (2002). An invariance principle for sieve bootstrap in time series, *Econometric Theory* **18**: 469–490.

Park, J. Y. (2003). Bootstrap unit root tests, *Econometrica* **71**: 1845–1895.

Saikkonen, P. & Lütkepohl, H. (2000a). Testing for the cointegration rank of a VAR process with structural shifts, *Journal of Business & Economic Statistics* **18**: 451–464.

Saikkonen, P. & Lütkepohl, H. (2000b). Trend adjustment prior to testing for the cointegration rank of a VAR process, *Journal of Time Series Analysis* **21**: 435–456.

Saikkonen, P., Lütkepohl, H. & Trenkler, C. (2006). Break date estimation for VAR processes with level shift with an application to cointegration testing, *Econometric Theory,* **22**: 15–68.

Swensen, A. R. (2005). Finite sample properties of a bootstrap procedure for estimation of rank in reduced rank VAR-models, *mimeo*, Department of Mathematics, University of Oslo.

Swensen, A. R. (2006a). Bootstrap algorithms for testing and determining the cointegration rank in VAR models, *Econometrica* **74**: 1699–1714.

Swensen, A. R. (2006b). Suplement to " Bootstrap algorithms for testing and determining the cointegration rank in VAR models", *Econometrica* **74**: 1699–1714.

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

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

Trenkler, C. (2006). Bootstrapping systems cointegration tests with a prior adjustment for deterministic terms, *Discussion Paper 12*, SFB 649, Humboldt-Universität zu Berlin.

Trenkler, C. (2008). Determining p-values for systems cointegration tests with a prior adjustment for deterministic terms, *Computational Statistics,* **23**: 19–39.