2009 | 12

Working Paper

Research Department

Bootstrapping the likelihood ratio cointegration test in error correction models with unknown lag order

by Christian Kascha and Carsten Trenkler

NB NORGES BANK

Working papers fra Norges Bank, fra 1992/1 til 2009/2 kan bestilles over e-post.
servicesenter@norges-bank.no
eller ved henvendelse til:
Norges Bank, Abonnementsservice
Postboks 1179 Sentrum
0107 Oslo
Telefon 22 31 63 83, Telefaks 22 41 31 05

Fra 1999 og fremover er publikasjonene tilgjengelig på www.norges-bank.no

Working papers from Norges Bank, from 1992/1 to 2009/2 can be ordered by e-mail:
servicesenter@norges-bank.no
or from Norges Bank, Subscription service
P.O.Box. 1179 Sentrum
N-0107Oslo, Norway.
Tel. +47 22 31 63 83, Fax. +47 22 41 31 05

Working papers from 1999 onwards are available on www.norges-bank.no

ISSN 1502-8143 (online)
ISBN 978-82-7553-509-0 (online)

Bootstrapping the Likelihood Ratio Cointegration Test in Error Correction Models with Unknown Lag Order *

Christian Kascha[†] Carsten Trenkler[‡]
Norges Bank *University of Mannheim*

July 14, 2009

Abstract

We investigate the small-sample size and power properties of bootstrapped likelihood ratio systems cointegration tests via Monte Carlo simulations when the true lag order of the data generating process is unknown. A recursive bootstrap scheme is employed. We estimate the order by minimizing different information criteria. In comparison to the standard asymptotic likelihood ratio test based on an estimated lag order we found that the recursive bootstrap procedure can lead to improvements in small samples even when the true lag order is unknown while the power loss is moderate.

Keywords: Cointegration Tests, Bootstrapping, Information Criteria
JEL-Codes: C15, C32

1 Introduction

In this note, we compare the performance of likelihood ratio cointegration tests with asymptotical and bootstrap critical values in terms of their size and power in the case in which the true lag order in the vector error correction model (VECM) is not known a priori. To the best of our knowledge, these tests have been compared only in situations in which the true lag order is known as the theory on bootstrapping systems cointegration tests has been developed only recently by Swensen (2006) and extended by Cavaliere, Rahbek & Taylor (2009) and Trenkler (2009). We conduct Monte Carlo experiments using three different data generating processes (DGPs) and a recursive bootstrap procedure. The lag order is estimated by applying different information criteria. In comparison to the asymptotic likelihood ratio test we found that the bootstrap can lead to improvements in small samples even in the unknown lag order case.

*We thank Anders Rygh Swensen and participants of the Econometric Seminar at the University of Kiel for their helpful comments. The views expressed in this paper are our own and do not necessarily reflect the views of Norges Bank.

[†]Norges Bank, Research Department. christian.kascha@norges-bank.no

[‡]Address of corresponding author: University of Mannheim, Department of Economics, Chair of Empirical Economics, 68131 Mannheim, Germany; trenkler@uni-mannheim.de

Most tests for cointegration are formulated in the well-known VECM framework for a n-dimensional time series $y_t = (y_{1,t}, \ldots, y_{n,t})'$ observed for $t = 1, \ldots, T$,

$$\Delta y_t \;=\; \mu_0 + \mu_1 t + \Pi y_{t-1} + \sum_{j=1}^{k} \Gamma_j \Delta y_{t-j} + \epsilon_t, \;\; t = k+2, \ldots, T, \tag{1}$$

where μ_0 and μ_1 are $(n \times 1)$ parameter vectors, Π and $\Gamma_1, \ldots, \Gamma_k$ are $(n \times n)$ parameter matrices. For the error term it is usually assumed that $\epsilon_t \sim$ i.i.d. $(0, \Sigma)$ and $E(\epsilon_t^4) < \infty$. Moreover, k is the lag order of the VECM such that $p = k + 1$ is the order of the underlying vector autoregressive (VAR) process for y_t. We assume that y_t is $I(1)$, i.e. its components are at most integrated of order one. When the matrix Π has rank $r > 0$ the series are said to cointegrate and one can write $\Pi = \alpha\beta'$ with α being a $(n \times r)$ matrix of adjustment coefficients and β being a $(n \times r)$ matrix of cointegrating vectors. The most popular test for cointegration is the likelihood ratio (LR) test for the null hypothesis $\mathcal{H}_0 : r = r_0$ versus $\mathcal{H}_1 : r > r_0$ proposed by Johansen (1988, 1991).

The asymptotic distribution of the LR test was first derived by Johansen (1988) under the assumption that the true lag length is used. Even in this favorable case, many authors such as Toda (1995), Ho & Sorensen (1996) or Gonzalo & Pitarakis (1999) have shown that the size of the LR test in small samples can substantially differ from its nominal value when asymptotic critical values are used. The problem has been addressed in two different ways. The first approach tries to correct or modify the test statistic such that its finite sample distribution is closer to the one obtained from asymptotic theory. Examples are the corrections proposed by Reinsel & Ahn (1992), Reimers (1992) and, in particular, Johansen (2002). The second approach uses bootstrap methods to obtain critical values of the finite sample distribution of the test statistic (see e.g. Swensen 2006).

In practice, however, the lag order k is unknown and has to be estimated prior to testing for cointegration. This is usually done by applying information criteria with respect to unrestricted VAR models fitted to y_t. In addition, it might be that the true lag order in (1) is infinite because the data are generated by a vector autoregressive moving average (VARMA) process. Saikkonen & Luukkonen (1997) showed that the use of Johansen's LR test is justified in the VARMA case in that the test has the same limiting distribution as in the finite order case under some restrictions on the process and the lag order. Lütkepohl & Saikkonen (1999) established the corresponding result in the case where the lag order is estimated by information criteria for finite and infinite-order VECMs.

The estimation of the lag order can have several effects on the performance of the standard LR test. First, if the estimated lag length is shorter than the true lag length, the test is based on an inadequate model. Second, if the used lag length is larger, the test is based on a model that nests the DGP but additional uncertainty is introduced. Thus, it is no surprise if the additional uncertainty stemming from the order estimation aggravates the above mentioned problems. Indeed, there are many simulation studies pointing to intolerable size distortions when the wrong lag order is chosen (see e.g. Boswijk & Franses 1992, Cheung & Lai 1993, Yap & Reinsel 1995). Similar pessimistic results when the lag order is estimated have been obtained by Lütkepohl & Saikkonen (1999).

Swensen (2006) and Trenkler (2009) found encouraging results for some recursive bootstrap systems cointegration tests. In a recursive bootstrap, an i.i.d. sample from the residuals is drawn and the bootstrap data are generated recursively according to the chosen time series model. The situation of an unknown lag order has not been investigated in these papers.[1] The results on the asymptotic LR test suggest, however, that it is crucial to compare bootstrap cointegration tests with the asymptotic LR test in a setup in which the lag order is unknown. In the recursive bootstrap, the lag order estimated from the original data is used in each bootstrap replication when applying the cointegration test to the bootstrapped data. In other words, this procedure treats the estimated lag length as if it was known for the generation of the bootstrap data. This treatment of the lag length might therefore have adverse effects on the performance of the bootstrap. Thus, we analyse via Monte Carlo simulations how the application of lag selection criteria affects the small sample properties of recursive bootstrap cointegration tests. Thereby, we can give some directions for applied researchers on which approach to use.

The rest of the paper is organized as follows. The test procedures are described in Section 2. In Section 3, we explain the design of the Monte Carlo simulations and discuss the results. Section 4 concludes.

2 Test Procedures

In this section, we describe tests for cointegration in the VECM framework given in (1) that are used in the Monte Carlo simulations. We focus here on the VECM with a restricted trend term and r cointegrating relations such that we can write $\mu_1 = \alpha\rho$ with α as above and ρ being a scalar. Model (1) can then be written as

$$\Delta y_t = \mu_0 + \alpha(\beta' y_{t-1} + \rho t) + \sum_{j=1}^{k} \Gamma_j \Delta y_{t-j} + \epsilon_t. \qquad (2)$$

Given the lag order k, denote by R_{0t} and R_{1t} the residuals obtained from regressing Δy_t and $(y_{t-1}, t)'$ on $(1, \Delta y_{t-1}, \dots, \Delta y_{t-k})'$, respectively. Define further $S_{ij} = T^{-1} \sum_{t=k+2}^{T} R_{it} R'_{jt}$ and denote by $(\hat{\lambda}_{r_0+1}, \dots, \hat{\lambda}_n)$ the $n - r_0$ smallest eigenvalues in $|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$. For simplicity, the dependence on the lag order is omitted for all these quantities but it is retained

[1] Only van Giersbergen (1996) has examined whether a stationary bootstrap can help to reduce distortions due to lag order misspecification. However, the stationary bootstrap is of limited applicability here because it relies on critical auxiliary parameters which are difficult to estimate in practice.

for the test statistics. Johansen's LR trace test statistic for $\mathcal{H}_0 : r = r_0$ versus $\mathcal{H}_1 : r > r_0$ is

$$LR_{r_0}(k) \quad = \quad -(T-k-1) \sum_{i=r_0+1}^{n} \log(1-\hat{\lambda}_i). \tag{3}$$

This is the most commonly used test statistic and its (non-standard) asymptotic distribution can be found in Johansen (1988, 1996). Here we use asymptotical critical values computed by Doornik (1998). As pointed out in the introduction, Reinsel & Ahn (1992) and Reimers (1992), among others, have suggested modified LR statistics to improve the size properties of the standard LR test with mixed success (Cheung & Lai 1993, Hubrich, Lütkepohl & Saikkonen 2001). Furthermore, Johansen (2002) proposed a small sample correction following Bartlett (1937). Swensen (2006), however, showed that bootstrap approximations work better than the correction suggested by Johansen (2002) in a number of situations. Our own simulations have confirmed this. Moreover, we have found that the LR test outperforms the Bartlett-corrected test version in most of our simulation setups. Therefore, we focus on the former one in the following and do not report results on the Bartlett-corrected LR test. The use of the true lag order will be explicitly indicated by the notation $LR_{r_0}(k_{true})$.

In practice, the lag order in (1) is of course unknown. A researcher might therefore use some information criterion, $IC(k)$, and choose $\hat{k}_{IC} = \text{argmin}_k\, IC(k)$, where the minimization is over $k = 0, \ldots, k_T$ and k_T is a given upper bound on the possible lag orders. Paulsen (1984) shows that the standard order selection criteria are consistent for multivariate autoregressive processes with unit roots. Lütkepohl & Saikkonen (1999) show specifically for cointegration tests that $k_T \to \infty$ and $k_T^3/T \to 0$ as $T \to \infty$ suffice to show that the LR test with a lag order selected by an information criterion has asymptotically the same distribution as $LR_{r_0}(k_{true})$. Therefore, we use $k_{50} = 3$, $k_{100} = 4$ and $k_{200} = 5$, which corresponds to taking the largest integers such that $k_T \leq T^{1/3}$. Using larger values for k_T can lead to excessive size distortions when the sample size is small (e.g. $T = 50$). The information criteria take the general form

$$IC(k) = \ln|\hat{\Sigma}(k)| + C_T \frac{kn^2}{N}, \tag{4}$$

where $N = T - k_T - 1$, $\hat{\Sigma}(k)$ is an estimate of the covariance matrix $\hat{\Sigma}(k) = \sum_{t=k_T+2}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$ and the $\hat{\epsilon}_t$ are obtained by estimating an unrestricted VECM, i.e. a VAR model of order $k+1$, on $y_{k_T+2}, \ldots, y_T$. We consider Akaike's (1973) information criterion (AIC) with $C_T = 2$, the Hannan-Quinn (HQ) criterion (Hannan & Quinn 1979) with $C_T = \ln \ln N$ and Schwarz's (1978) Bayesian information criterion (SC), $C_T = \ln N$. In addition, we employ the modified Akaike information criterion (MAIC) of Qu & Perron (2007) that imposes r_0 at the lag specification stage in order to obtain a better estimate of the Kullback-Leibler divergence in small samples. It is given by

$$MAIC(k, r_0) = \ln|\hat{\Sigma}(k, r_0)| + 2\frac{\widetilde{LR}_{r_0}(k) + kn^2}{N}, \tag{5}$$

where $\widetilde{LR}_{r_0}(k) = -N\sum_{i=r_0+1}^{n} \ln(1 - \hat{\lambda}_i)$ is just the usual LR test statistic but obtained by maximizing the likelihood of $y_{k_T+2}, \ldots, y_T$ in the VECM. Thus, the MAIC is almost identical to the AIC but includes the extra term $\widetilde{LR}_{r_0}(k)$. Also, the notation $\hat{\Sigma}(k, r_0)$ emphasizes that the estimate of Σ is based on the VECM form (2) with imposed cointegrating rank r_0.

When the LR test in (3) is used with a lag order selected by one of the information criteria we denote it by $LR_{r_0}(\hat{k}_{AIC})$, $LR_{r_0}(\hat{k}_{HQ})$ and so forth. When we refer to these tests in general we use the notation $LR_{r_0}(\hat{k}_{IC})$.

The bootstrap procedure calculates p-values as follows.

1. Compute $\hat{k}_{IC}$ and estimate model (2) under $\mathcal{H}_0 : r = r_0$ to obtain estimates $\hat{\mu}_0, \hat{\rho}, \hat{\alpha}, \hat{\beta}$ and $\hat{\Gamma}_i$, $i = 1, \ldots, \hat{k}_{IC}$, and the residuals $\hat{\epsilon}_{\hat{k}_{IC}+2}, \ldots, \hat{\epsilon}_T$.

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

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

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

3. Compute B bootstrap replications of y_t^*, $t = \hat{k}_{IC} + 2, \ldots, T$, recursively by

$$\Delta y_t^* = \hat{\mu}_0 + \hat{\alpha}(\hat{\beta}' y_{t-1}^* + \hat{\rho}t) + \sum_{j=1}^{\hat{k}_{IC}} \hat{\Gamma}_j \Delta y_{t-j} + \epsilon_t^*,$$

   where the ϵ_t^* are drawn with replacement from the residuals $\hat{\epsilon}_{\hat{k}_{IC}+2}, \ldots, \hat{\epsilon}_T$. The starting values of the recursion, $y_1^*, \ldots, y_{\hat{k}_{IC}+1}^*$ are set equal to $y_1, \ldots, y_{\hat{k}_{IC}+1}$.

4. For each replication $b = 1, \ldots, B$, given $\hat{k}_{IC}$, estimate model (2) under r_0 and compute the LR test as in (3). Denote the bootstrap statistics by $LR_{r_0}^*(\hat{k}_{IC})_b$.

5. Estimate the p-value of the test statistic as

$$p^*(LR_{r_0}(\hat{k}_{IC})) = \frac{1}{B}\sum_{b=1}^{B} I(LR_{r_0}^*(\hat{k}_{IC})_b > LR_{r_0}(\hat{k}_{IC})), \tag{6}$$

   where I denotes the indicator function.

The bootstrap test versions with the corresponding criterion are denoted by $BOOT_{r_0}(\hat{k}_{AIC})$,

$BOOT_{r0}(\hat{k}_{HQ})$ et cetera.[2]

Four remarks on the bootstrap procedure are in order. First, step (2) assures that the generated pseudo observations are indeed $I(1)$. The crucial requirement is that the characteristic equation has no explosive roots. If this condition is not satisfied, one may refer to a more appropriate resampling scheme as pointed out by Swensen (2006, Remark 1).[3]

Second, for the case of a known lag order, Swensen (2006) suggested estimating a VAR model without imposing the cointegrating rank null hypothesis to obtain residuals and estimates of μ_0 and Γ_i, $i = 1, \ldots, k$. Hence, a combination of estimates from two different models would be applied. However, this combination can cause inferior small-sample properties in the case of nonzero deterministic terms as pointed out by Trenkler (2009). Therefore, we do not consider this bootstrap version in our context.

Third, Swensen (2006) proved the asymptotic validity of the bootstrap in the case of a known lag order. Our simulation results do not give rise to concerns when pre-estimating the lag length. However, it is not trivial to rigorously prove the asymptotic validity of the bootstrap test in this case. A proof is beyond the scope of the paper and is left for future research. Some comments on the asymptotic problems follow.

The VECM parameter estimators on which the bootstrap data generation is based are consistent when applying SC, HQ or AIC in conjunction with a fixed lag order upper bound. This follows from Pötscher (1991, Lemmata 1 and 2). Hence, the corresponding bootstrap data would asymptotically satisfy the $I(1)$ condition in step 2 of the bootstrap algorithm, which is crucial for the validity of the bootstrap, compare Swensen (2006). However, one should not conclude from the consistency of the estimators that an invariance principle automatically holds for the bootstrap error term vector ϵ_t^*. As pointed out by Leeb & Pötscher (2005), the unconditional distributions arising in the post-model-selection stage can be quite different from the distributions obtained for a fixed model setup. However, an invariance principle has been proven for a sieve bootstrap framework by Chang, Park & Song (2006) and Palm, Smeekes & Urbain (2009) in related systems cointegration setups. The asymptotic results of these papers could be a promising starting point for designing a sieve bootstrap cointegration rank test for which the asymptotic distribution can be derived. Note that by letting the maximum order increase at a suitable rate, we not only assure asymptotic validity of the standard LR rank test but our implementation of the bootstrap corresponds to the idea of the sieve bootstrap.

Fourth, although the bootstrap procedure does not address the uncertainty stemming from

[2]We also considered the fast double bootstrap (FDB) of Davidson & MacKinnon (2007) as a potential refinement of the bootstrap procedure. The results for the FDB were, however, very similar to the results of the standard bootstrap and are therefore not discussed. Note that Ahlgren & Antell (2008) reported slight improvements when applying the FDB in case of a known lag order.

[3]The condition was violated in a very small number of cases, typically less than 10 out of 5000 simulations. In these cases, the null hypothesis was most often not rejected and we set rejections to non-rejections in the remaining cases.

lag length estimation, it might still yield improvements in finite samples over the standard LR test, which uses the estimated lag length together with the asymptotical critical values. In a different context, Kilian (1998) suggested a so-called endogenous lag order bootstrap which re-estimates the lag order at each bootstrap replication. He does so in order to take account of the additional uncertainty due to the lag order estimation when computing confidence intervals. However, it is not clear whether the endogenization of the lag order choice is beneficial for our test setup in which only the right-hand-side of a distribution is considered. While some improvements could be obtained with the $MAIC$, the endogenous lag order bootstrap is generally inferior when applied with the AIC, HQ, and SC. Therefore, we do not discuss the endogenous bootstrap further.

3 Monte Carlo Simulations

We simulate three different DGPs for sample sizes $T = 50,\ 100$, and $T = 200$. These sample sizes are typical for macroeconomic applications.[4] The number of replications is $R = 5000$. For determining the quantiles of the empirical bootstrap distributions, we use $B = 1000$ bootstrap replications. We believe that these numbers of replications are large enough to obtain sufficiently precise estimates of the tests' true rejection frequencies. Since the overall replications are Bernoulli trials, the standard deviation of a rejection frequency is limited by $\sqrt{p(1-p)/R} \leq \sqrt{1/4R} = 0.007$. Note, however, that this limit ignores the uncertainty involved in the bootstrap simulations and in the simulation of the critical values.

The DGPs have mainly been chosen because they were used in the literature to exemplify the size distortions of the LR test. The first DGP was suggested by Toda (1994, 1995)

$$x_t = \begin{bmatrix} a_1 & 0 \\ 0 & 1 \end{bmatrix} x_{t-1} + \varepsilon_t, \ \ \varepsilon_t \sim \text{iid } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix} \right). \tag{7}$$

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 the simulations, we use $\theta = 0.8$. 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. Other bivariate VAR(1) processes of interest can be obtained from (7) by linear transformations which leave the LR tests invariant, compare (Toda 1994, 1995).

Since the process in (7) is a rather simple one, we also use a more complex, data-based DGP by referring to an empirical study of King, Plosser, Stock & Watson (1991) (KPSW). King et al. (1991) analyse a small macroeconomic model for the U.S. which consists of the

[4]The computations are performed using programs written in GAUSS V8 for Windows. The RNDNS function with a fixed seed has been used to generate standard normally distributed random numbers.

logarithms of per-capita private real GNP, per-capita real consumption, and per-capita gross private domestic fixed investment. We estimate a subset-VECM with one lag and two restricted cointegrating relationships using quarterly data in logarithms for the period 1949:1-1988:4. Subset restrictions have been imposed by using a *Top-Down* strategy employing the AIC.[5] We obtain the following process

$$\Delta y_t = \begin{bmatrix} -0.038 \\ -0.186 \\ 0.032 \end{bmatrix} + \begin{bmatrix} 0 & -0.026 \\ 0.217 & -0.150 \\ 0.126 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 & 0.154 \\ 0 & 0.282 & 0.660 \\ 0.272 & 0.162 & 0 \end{bmatrix} \Delta y_{t-1} + \epsilon_t, \quad (8)$$

where $\epsilon_t \sim$ i.i.d. $N(0, \Sigma)$ with

$$\Sigma = 10^{-4} \begin{bmatrix} 0.588 & 0.821 & 0.465 \\ & 4.870 & 1.688 \\ & & 1.376 \end{bmatrix}.$$

As starting values, we chose the corresponding empirical data. The same process was used by Trenkler (2009) in a related study. It turned out that both asymptotic and bootstrap cointegration tests displayed inferior small-sample properties for this process. Hence, we may regard the DGP (8) as a demanding reference for the test procedures.

The third DGP is a mixed VARMA process which was also used by Yap & Reinsel (1995) and Lütkepohl & Saikkonen (1999). This process allows us to obtain results for infinite order VAR processes. It is given by

$$\Delta y_t = P^{-1} \left(\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} - I_3 \right) P y_{t-1} + \epsilon_t - P_\theta \begin{bmatrix} \lambda_\theta & 0 & 0 \\ 0 & 0.297 & 0 \\ 0 & 0 & -0.202 \end{bmatrix} P_\theta^{-1} \epsilon_{t-1}, \quad (9)$$

where $\epsilon_t \sim$ i.i.d. $N(0, \Sigma)$,

$$P = \begin{bmatrix} -0.29 & -0.47 & -0.57 \\ -0.01 & -0.85 & 1.00 \\ -0.75 & 1.39 & -0.55 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.47 & 0.20 & 0.18 \\ & 0.32 & 0.27 \\ & & 0.30 \end{bmatrix}, P_\theta = \begin{bmatrix} -0.816 & -0.657 & -0.822 \\ -0.624 & -0.785 & 0.566 \\ -0.488 & 0.475 & 0.174 \end{bmatrix}.$$

The values of the λ_i, $i = 1, 2, 3$, determine the cointegration properties of the series. That is, the number of λ_i with $|\lambda_i| < 1$ is the cointegrating rank of the system. The precise values are given in the tables later on. Note that the size of λ_θ determines how well the VARMA can be approximated by a VAR. A low value for λ_θ in modulus implies that all eigenvalues of the moving-average matrix are small and a finite-order VAR should be able to capture the true

[5] Computations have been done using JMulTi (Lütkepohl & Krätzig 2004, Chapter 3).

dynamics well since the other two eigenvalues are small as well. If λ_θ is large in modulus the moving-average part has one large eigenvalue and a VAR with a larger lag order is needed to approximate the DGP. In the simulations we use $\lambda_\theta = -0.5, \lambda_\theta = 0$, and $\lambda_\theta = 0.5$.

The results on the size and power of the different test procedures are given in Tables 1 - 5 and Figures 1 - 3. We only report here results for a nominal size of 0.05. Figures 1 and 2 show for the Toda DGP (7) how frequent a certain order was chosen by the information criteria. Note again that $MAIC$ can suggest different lag orders depending on the rank tested under the null hypothesis (Figure 2). Figure 3 comprises graphs that highlight different properties of testing for cointegration using either the AIC or the MAIC for prior model selection.

For the simple bivariate Toda-DGP, Table 1 gives a comparison of the empirical size of the different test procedures. The cointegrating rank is $r = 0$ in Panel A and $r = 1$ in Panels B and C. We can see that using AIC, HQ, or SC to determine the lag choice prior to employing either asymptotic tests or the bootstrap leads to higher empirical size values compared to applying the corresponding test with the true lag order if $T = 50$. Regarding the larger sample sizes we only observe an upward size effect for AIC, while for HQ and SC the empirical sizes are rather similar to the ones obtained when $k_{true} = 0$ is used. This results from the fact that the correct lag order is suggested in about 98% and more of the replications by HQ and SC if $T = 100$ and $T = 200$. By contrast, the fraction of correct suggestions is only between 0.80 and 0.85 for AIC, compare also Figure 1. Since $k_{true} = 0$, a too high lag order is chosen in the remaining replications. Thus, an overestimation of the lag length leads to larger size values and not to smaller ones for the standard criteria. Interestingly, the size-increasing effect of estimating the lag order is much stronger for the asymptotic tests. In comparison to these tests, application of the bootstrap reduces the size but much less so when the corresponding asymptotic test with a particular lag selection criterion is only slightly oversized. Thus, the bootstrap correction is not mechanically reducing sizes but is sensitive to the size distortion of the corresponding asymptotic tests. Accordingly, choosing the bootstrap can be very useful to avoid or reduce excessive size distortions, which we observe for the standard criteria in a number of cases, compare Panel A and C of Table 1.

$MAIC$ behaves very differently from the other information criteria in that the empirical size values clearly fall for the asymptotic and bootstrap tests compared to using the true lag order. As a result, the test procedures are very conservative. From Figure 1 we see that $MAIC$ does not overestimate k excessively more often than AIC. Yet, the correlations of the AIC's and $MAIC$'s estimates of k are relatively weak. To be precise, the correlations are between 0.4 and 0.5 for the processes with $r = 1$ and between 0 and 0.2 for the process with $r = 0$. Obviously, this low correlation must explain the different rejection frequencies between $MAIC$ and the other criteria. Furthermore, note that overestimation of k leads to a size drop when using $MAIC$. Hence, no general conclusion can be drawn on whether overfitting results in size increases or decreases.

Table 2 provides some results on the power of the tests. The bootstrap tests using AIC, HQ, or SC have a rather similar power as the bootstrap test with true order. More importantly, the power is only slightly smaller than those of the asymptotic tests with AIC, HQ, or SC. Hence, there is no relevant price to pay for bootstrapping in the case of an unknown lag order for the the Toda-DGP (7).

Given the results for the empirical size, it may not be surprising that $BOOT_{r_0}(\hat{k}_{MAIC})$ has a very low power. One should remember again that $MAIC$ can choose different lag orders for $\mathcal{H}_0 : r = 0$ than for $\mathcal{H}_0 : r = 1$, which is the case considered in Panels B and C of Table 1. The lag order choice of $MAIC$ under $\mathcal{H}_0 : r = 0$ fortifies the relative differences between $BOOT_{r_0}(\hat{k}_{MAIC})$ and $LR_{r_0}(\hat{k}_{MAIC})$. For $\mathcal{H}_0 : r = 0$, $MAIC$ overestimates the lag order clearly more strongly and more often than for $\mathcal{H}_0 : r = 1$, as illustrated in Figure 2. Thus, we should have an additional downward pressure on the power since model overfitting usually leads to power losses.

Table 3 provides results for the more complex KPSW-DGP (8) with a true cointegrating rank of two. Thus, Panel A ($\mathcal{H}_0 : r_0 = 2$) gives results on the tests' size while Panels B and C ($\mathcal{H}_0 : r_0 = 1$ and $\mathcal{H}_0 : r_0 = 0$) give results on the power of the tests. Since we now have one lag in the VECM, underestimation of k can occur. In fact, the information criteria, in particular HQ and SC, underestimate k quite often for $T = 50$. This may explain why we observe higher rejection frequencies for the bootstrap tests with estimated lag order compared to $BOOT_{r_0}(k_{true})$ if $T = 50$ or $T = 100$. If $T = 200$, slightly lower rejection frequencies are obtained, in particular for HQ and SC, compare Panels A and B of Table 3. The low empirical size values seen in Panel A seem to be the result of distortions due to parameter estimation as even $LR_{r_0}(k_{true})$ is undersized and SC and HQ estimate the true lag length at $T = 200$. This may also have an negative effect on the tests' power. In sum, in the case of the KPSW-DGP the overall effect of the bootstrap on the size of the tests is most often advantageous for $T = 100$ and $T = 200$ but small in any case.

Comparing the power of corresponding bootstrap and asymptotic tests, we see that a price in terms of power loss has to be paid when bootstrapping in the case of $\mathcal{H}_0 : r = 0$ if the sample size is small. However, the relative power loss is somewhat lower than in the case of using a true lag order, in particular for SC with $T = 50$. As a result, the introduction of lag order uncertainty tends to favor the bootstrap in relative terms. Finally, we note that all procedures perform rather poorly in the current setup unless a large sample size is available.

Lastly, we consider two versions of the VARMA-DGP (9) for which Tables 4 and 5 present the outcomes. First, Table 4 shows results for the size of the tests when $\lambda_1 = \lambda_2 = \lambda_3 = 1$ such that the true cointegrating rank is $r = 0$ and $\mathcal{H}_0 : r_0 = 0$ is tested. Since there is no true finite lag order, only results for tests that estimate the order are presented. The tests using asymptotic critical values conditional on an estimated lag length are quite oversized for different sample sizes and different values of λ_θ. It appears that the introduction of a moving-average component with small eigenvalues in modulus can already lead to severe size distortions. An exception is $LR_{r_0}(\hat{k}_{MAIC})$ which is much less oversized and almost correctly sized for $T = 200$, see Panels A and B. This might be because the $MAIC$ is designed to take the possibility of an underlying VARMA-DGP explicitly into account when testing for cointegration, see also Qu & Perron (2007). The tests are less size distorted for $\lambda_\theta = -0.5$ and $\lambda_\theta = 0$ which imply mostly non-negative eigenvalues of the moving-average part. We do not think that the differences in the lag order estimates with respect to λ_θ play a crucial role in terms of the size distortion since no clear pattern could be detected.

The application of the bootstrap reduces the empirical sizes clearly, in particular if AIC is used. Note also that $BOOT_{r_0}(\hat{k}_{MAIC})$ can be even a bit undersized. As before, we also note that the bootstrap correction is sensitive to the degree of size distortion of the corresponding asymptotic tests. However, even though the bootstrap reduces excessive size distortions, it is no panacea in this case as even the bootstrap tests are still quite oversized.

Table 5 shows results regarding the size of the tests for $\lambda_1 = 1$, $\lambda_2 = 0.8$ and $\lambda_3 = 0.7$. Hence the true cointegrating rank is two. All tests are very conservative for $T = 50$ and their empirical size increases with the sample size such that they are usually oversized for $T = 200$, especially in the case of $\lambda_\theta = 0.5$. There is a clear tendency for all information criteria to suggest larger models, the larger the sample size is. This may have caused the increasing size values.

The effect of applying the bootstrap procedure on the size depends on the value of λ_θ and the sample size. However, the bootstrap generally corrects the size of the asymptotic tests towards the nominal size depending on whether the $LR_{r_0}(\hat{k}_{IC})$ tests are under- or oversized, respectively. Hence, for the current DGP, the size of the bootstrap tests may also systematically increase compared to the corresponding asymptotic tests. This is in contrast to the Toda- and KPSW-DGPs, where the empirical sizes of the bootstrap tests fall in general.[6]

The results of the simulations can be summarized as follow

- When the lag order is not known a priori, the recursive bootstrap remains advantageous in that it can bring empirical sizes closer to the nominal ones both when the asymptotic tests are over- or undersized.
- Hence, the introduction of lag order uncertainty does not impair the relative performance of the bootstrap. In fact, it even tends to favor the bootstrap in some cases.
- The bootstrap is, however, no panacea in the case of very large size distortions such as for the VARMA-DGPs.
- In the particular case of the VARMA models, the MAIC is a good choice for large sample sizes. However, the use of the MAIC leads to far too conservative tests and severe power loss for other sample sizes and DGPs.

[6]Due to the observed strong size distortions for the VARMA processes, we do not evaluate the tests' small sample power.

A final remark is in order here on the different behavior of the LR test when applied with the standard information criteria and the $MAIC$. Figure 3 exemplifies for the Toda DGP (7) that the $MAIC$ induces an inward shift of the statistic's distribution compared to the one obtained by applying the AIC, as expected from the foregoing results.

A likely reason is that the $MAIC$ introduces a different correlation pattern between the estimated lag order and test statistic which leads to different test decisions. Consider again the formula for the $MAIC$

$$\begin{aligned} MAIC(k, r_0) &= \ln|\hat{\Sigma}(k, r_0)| + 2\frac{\widetilde{LR}_{r_0}(k) + kn^2}{N} \cong \ln|\hat{\Sigma}(k)| + 2\frac{kn^2}{N} + 2\frac{\widetilde{LR}_{r_0}(k)}{N} \\ &= AIC(k) + 2\frac{\widetilde{LR}_{r_0}(k)}{N}. \end{aligned}$$

Since $\widetilde{LR}_{r_0}(k)$ is included in the minimization, $\hat{k}_{MAIC}$ should be associated with a smaller test statistic than $\hat{k}_{AIC}$ on average. To be precise, we should have $E[LR_{r_0}(\hat{k}_{MAIC})] \leq E[LR_{r_0}(\hat{k}_{AIC})]$ for a given DGP. In Figure 3, Panels (a) and (b) confirm that claim and also explain that even though AIC and $MAIC$ choose the same lengths with nearly the same overall frequency, the resulting rejection frequencies differ significantly.

4 Conclusion

In this paper, we investigated the properties of bootstrap LR cointegration tests in terms of size and power in the vector error correction model when the lag order is not known a priori and has to be estimated by some information criterion. We performed Monte Carlo experiments using three different data generating processes and compared the effects of lag order estimation on a recursive bootstrap procedure that uses the estimated lag order for all bootstrap replications and on the corresponding asymptotic test procedures.

We found that pre-estimating the lag order by some information criterion has qualitatively similar effects on the size of the bootstrap tests as it has on the size of the asymptotic tests. The lag order estimation does not induce power losses which may be a result specific to the simulated DGPs. We find that the recursive bootstrap remains advantageous in that it can bring empirical sizes closer to the nominal ones both when the asymptotic tests are over- or undersized, even when the lag order is not known a priori. Hence, the introduction of lag order uncertainty does not impair the relative performance of the bootstrap. In fact, it even tends to favor the bootstrap in some cases. The bootstrap is, however, no panacea in the case of very large size distortions which we observed in particular for our VARMA-DGPs. Therefore, an interesting topic for future research is to test for cointegration in a VARMA framework that allows for lag order uncertainty such that a finite-order VAR model is contained as a special case, for example, along the lines of Bauer & Wagner (2009). Such a procedure would ideally perform well both in the VAR as well as in the more general VARMA case.

References

Ahlgren, N. & Antell, J. (2008), 'Bootstrap and fast double bootstrap tests of cointegration rank with financial time series', *Computational Statistics & Data Analysis* **52**(10), 4754–4767.

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrov & F. Csaki, eds, '2nd International Symposium on Information Theory', Akademia Kiado, Budapest.

Bartlett, M. S. (1937), 'Properties of sufficiency and statistical tests', *Proceedings of the Royal Society of London, Series A* **160**(901), 268–282.

Bauer, D. & Wagner, M. (2009), 'Using subspace algorithm cointegration analysis: Simulation performance and application to the term structure', *Computational Statistics & Data Analysis* **53**(6), 1954–1973.

Boswijk, P. & Franses, P. H. (1992), 'Dynamic specification and cointegration', *Oxford Bulletin of Economics and Statistics* **54**(3), 369–81.

Cavaliere, G., Rahbek, A. & Taylor, A. (2009), 'Testing for co-integration in vector autoregressions with non-stationary volatility', *Journal of Econometrics,* **forthcoming**.

Chang, Y., Park, J. Y. & Song, K. (2006), 'Bootstrapping cointegrating regressions', *Journal of Econometrics* **133**(2), 703–739.

Cheung, Y.-W. & Lai, K. S. (1993), 'Finite-sample sizes of Johansen's likelihood ration tests for conintegration', *Oxford Bulletin of Economics and Statistics* **55**(3), 313–28.

Davidson, R. & MacKinnon, J. G. (2007), 'Improving the reliability of bootstrap tests with the fast double bootstrap', *Computational Statistics and Data Analysis* **51**(7), 3259–3281.

Doornik, J. A. (1998), 'Approximations to the asymptotic distributions of cointegration tests', *Journal of Economic Surveys* **12**(5), 573–93.

Gonzalo, J. & Pitarakis, J. Y. (1999), Dimensionality effect in cointegration analysis, *in* R. F. Engle & H. White, eds, 'Cointegration, Causality and Forecasting: Festschrift in Honour of Clive Granger', Oxford University Press, Oxford, pp. 212–229.

Hannan, E. J. & Quinn, B. G. (1979), 'The determination of the order of an autoregression', *Journal of the Royal Statistical Society. Series B (Methodological)* **41**(2), 190–195.

Ho, M. S. & Sorensen, B. E. (1996), 'Finding cointegration rank in high dimensional systems using the Johansen test: An illustration using data based monte carlo simulations', *The Review of Economics and Statistics* **78**(4), 726–32.

Hubrich, K., Lütkepohl, H. & Saikkonen, P. (2001), 'A review of systems cointegration tests', *Econometrics Reviews* **20**(3), 247–318.

Johansen, S. (1988), 'Statistical analysis of cointegration vectors', *Journal of Economic Dynamics and Control* **12**(2-3), 231–254.

Johansen, S. (1991), 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica* **59**(6), 1551–1580.

Johansen, S. (1996), *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**(5), 1929–1961.

Kilian, L. (1998), 'Accounting for lag order uncertainty in autoregressions: the endogenous lag order bootstrap algorithm', *Journal of Time Series Analysis* **19**(5), 531–548.

King, R. G., Plosser, C. I., Stock, J. H. & Watson, M. W. (1991), 'Stochastic trends and economic fluctuations', *American Economic Review* **81**(4), 819–840.

Leeb, H. & Pötscher, B. M. (2005), 'Model selection and inference: Facts and fictions', *Econometric Theory* **21**(1), 21–59.

Lütkepohl, H. & Krätzig, M. (2004), *Applied Time Series Econometrics*, Cambridge University Press.

Lütkepohl, H. & Saikkonen, P. (1999), Order selection in testing for the cointegrating rank of a var process, *in* R. F. Engle & H. White, eds, 'Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger', Oxford University Press, Oxford.

Palm, F. C., Smeekes, S. & Urbain, J.-P. (2009), 'A sieve bootstrap test for cointegration in a conditional error correction model', *Econometric Theory,* **forthcoming**.

Paulsen, J. (1984), 'Order determination of multivariate autoregressive time series with unit roots', *Journal of Time Series Analysis* **5**(2), 115–127.

Pötscher, B. M. (1991), 'Effects of model selection on inference', *Econometric Theory* **7**(2), 163–185.

Qu, Z. & Perron, P. (2007), 'A modified information criterion for cointegration tests based on a var approximation', *Econometric Theory* **23**(4), 638–685.

Reimers, H. (1992), 'Comparisons of tests for multivariate cointegration', *Statistical Papers* **33**(1), 335–359.

Reinsel, G. C. & Ahn, S. K. (1992), 'Vector AR models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting', *Journal of Time Series Analysis* **13**(4), 353–375.

Saikkonen, P. & Luukkonen, R. (1997), 'Testing cointegration in infinite order vector autoregressive processes', *Journal of Econometrics* **81**(1), 93–126.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**(2), 461–464.

Swensen, A. R. (2006), 'Bootstrap algorithms for testing and determining the cointegration rank in VAR models', *Econometrica* **74**(6), 1699–1714.

Toda, H. Y. (1994), 'Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present', *The Review of Economics and Statistics* **76**(1), 66–79.

Toda, H. Y. (1995), 'Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions', *Econometric Theory* **11**(5), 1015–1032.

Trenkler, C. (2009), 'Bootstrapping systems cointegration tests with a prior adjustment for deterministic terms', *Econometric Theory* **25**(1), 243–269.

van Giersbergen, N. P. A. (1996), 'Bootstrapping the trace statistic in var models: Monte carlo results and applications', *Oxford Bulletin of Economics and Statistics* **58**(2), 391–408.

Yap, S. F. & Reinsel, G. C. (1995), 'Estimation and testing for unit roots in a partially non-stationary vector autoregressive moving average model', *Journal of the American Statistical Association* **90**(429), 253–267.

Table 1: Rejection Frequencies of Tests for Bivariate Toda DGP (7).

	Panel A: $a_1 = 1$ $(r = 0)$, $\mathcal{H}_0 : r = 0$			Panel B: $a_1 = 0.9$ $(r = 1)$, $\mathcal{H}_0 : r = 1$			Panel C: $a_1 = 0.7$ $(r = 1)$, $\mathcal{H}_0 : r = 1$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
$BOOT_{r_0}(k_{true})$	0.0478	0.0502	0.0504	0.0148	0.0336	0.0618	0.0474	0.0530	0.0602
$BOOT_{r_0}(\hat{k}_{AIC})$	0.0834	0.0668	0.0566	0.0198	0.0370	0.0618	0.0542	0.0572	0.0616
$BOOT_{r_0}(\hat{k}_{HQ})$	0.0656	0.0496	0.0492	0.0168	0.0322	0.0602	0.0508	0.0564	0.0574
$BOOT_{r_0}(\hat{k}_{SC})$	0.0534	0.0486	0.0502	0.0154	0.0334	0.0624	0.0478	0.0536	0.0584
$BOOT_{r_0}(\hat{k}_{MAIC})$	0.0184	0.0200	0.0294	0.0052	0.0184	0.0444	0.0184	0.0338	0.0488
$LR_{r_0}(k_{true})$	0.0584	0.0528	0.0532	0.0152	0.0344	0.0682	0.0562	0.0600	0.0646
$LR_{r_0}(\hat{k}_{AIC})$	0.1206	0.0800	0.0622	0.0222	0.0408	0.0698	0.0668	0.0658	0.0668
$LR_{r_0}(\hat{k}_{HQ})$	0.0848	0.0564	0.0546	0.0182	0.0352	0.0690	0.0610	0.0618	0.0644
$LR_{r_0}(\hat{k}_{SC})$	0.0658	0.0532	0.0534	0.0162	0.0344	0.0684	0.0574	0.0602	0.0646
$LR_{r_0}(\hat{k}_{MAIC})$	0.0284	0.0232	0.0338	0.0070	0.0202	0.0512	0.0230	0.0392	0.0530

Note: The table shows rejection frequencies for replications of the Monte Carlo simulation. The number of simulations is 5000. The true cointegrating rank is r. The nominal significance level is 0.05. In the table, LR is Johansen's likelihood ratio test and $BOOT$ denotes the bootstrap version of the LR test. See section 2 for explanation.

Table 2: Rejection Frequencies of Tests for Bivariate Toda DGP (7) with $r = 1$.

	Panel A: $a_1 = 0.9, \mathcal{H}_0 : r = 0$			Panel B: $a_1 = 0.7, \mathcal{H}_0 : r = 0$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
$BOOT_{r0}(k_{true})$	0.1192	0.3888	0.9248	0.7016	0.9986	1.0000
$BOOT_{r0}(\hat{k}_{AIC})$	0.1666	0.4048	0.9172	0.6878	0.9868	0.9996
$BOOT_{r0}(\hat{k}_{HQ})$	0.1404	0.3876	0.9222	0.7040	0.9984	1.0000
$BOOT_{r0}(\hat{k}_{SC})$	0.1250	0.3876	0.9254	0.7028	0.9986	1.0000
$BOOT_{r0}(\hat{k}_{MAIC})$	0.0326	0.1390	0.5480	0.1168	0.4384	0.9420
$LR_{r0}(k_{true})$	0.1342	0.4024	0.9294	0.7312	0.9988	1.0000
$LR_{r0}(\hat{k}_{AIC})$	0.2094	0.4304	0.9248	0.7408	0.9902	0.9998
$LR_{r0}(\hat{k}_{HQ})$	0.1634	0.4086	0.9288	0.7410	0.9984	1.0000
$LR_{r0}(\hat{k}_{SC})$	0.1426	0.4034	0.9294	0.7326	0.9988	1.0000
$LR_{r0}(\hat{k}_{MAIC})$	0.0486	0.1726	0.6070	0.1902	0.5592	0.9624

Note: The table shows rejection frequencies for replications of the Monte Carlo simulation. The number of simulations is 5000. The true cointegrating rank is $r = 1$. The nominal significance level is 0.05. In the table, LR is Johansen's likelihood ratio test and $BOOT$ denotes the bootstrap version of the LR test. See section 2 for explanation.

Table 3: Rejection Frequencies of Tests for KPSW DGP (8) with $r = 2$.

	Panel A: $\mathcal{H}_0 : r = 2$			Panel B: $\mathcal{H}_0 : r = 1$			Panel C: $\mathcal{H}_0 : r = 0$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
$BOOT_{r_0}(k_{true})$	0.0036	0.0114	0.0298	0.0166	0.0756	0.3264	0.1192	0.4686	0.9914
$BOOT_{r_0}(\hat{k}_{AIC})$	0.0042	0.0122	0.0316	0.0230	0.0798	0.3286	0.1778	0.4788	0.9896
$BOOT_{r_0}(\hat{k}_{HQ})$	0.0046	0.0122	0.0286	0.0250	0.0756	0.3268	0.1890	0.4708	0.9914
$BOOT_{r_0}(\hat{k}_{SC})$	0.0040	0.0104	0.0302	0.0232	0.0728	0.3278	0.2250	0.4964	0.9918
$BOOT_{r_0}(\hat{k}_{MAIC})$	0.0008	0.0088	0.0264	0.0022	0.0288	0.2296	0.0366	0.1678	0.7440
$LR_{r_0}(k_{true})$	0.0038	0.0104	0.0300	0.0382	0.0976	0.3598	0.2916	0.5990	0.9956
$LR_{r_0}(\hat{k}_{AIC})$	0.0084	0.0116	0.0300	0.0692	0.1040	0.3598	0.4102	0.6120	0.9946
$LR_{r_0}(\hat{k}_{HQ})$	0.0070	0.0104	0.0300	0.0542	0.0982	0.3598	0.3552	0.6038	0.9956
$LR_{r_0}(\hat{k}_{SC})$	0.0040	0.0110	0.0300	0.0394	0.0930	0.3596	0.3366	0.6146	0.9958
$LR_{r_0}(\hat{k}_{MAIC})$	0.0012	0.0076	0.0252	0.0088	0.0428	0.2616	0.1160	0.2964	0.8382

Note: The table shows rejection frequencies for replications of the Monte Carlo simulation. The number of simulations is 5000. The true cointegrating rank is $r = 2$. The nominal significance level is 0.05. In the table, *LR* is Johansen's likelihood ratio test and *BOOT* denotes the bootstrap version of the LR test. See section 2 for explanation.

Table 4: Rejection Frequencies for VARMA DGP (9) with $\lambda_1 = \lambda_2 = \lambda_3 = 1$ $(r = 0)$.

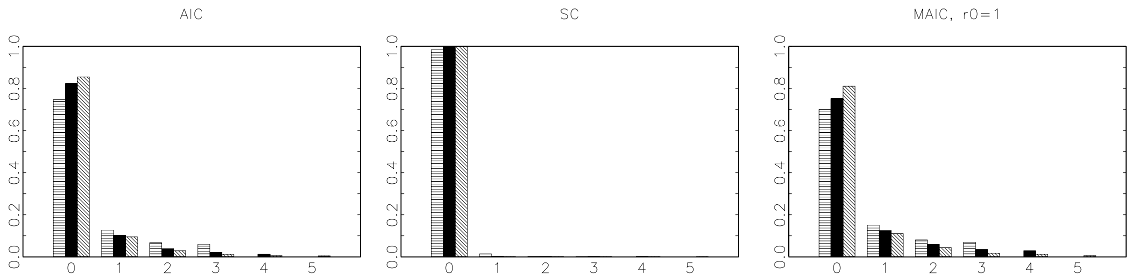
	Panel A: $\lambda_\theta = -0.5,\ \mathcal{H}_0 : r = 0$			Panel B: $\lambda_\theta = 0,\ \mathcal{H}_0 : r = 0$			Panel C: $\lambda_\theta = 0.5,\ \mathcal{H}_0 : r = 0$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
$BOOT_{r_0}(\hat{k}_{AIC})$	0.1654	0.0930	0.0862	0.2052	0.1608	0.0786	0.4126	0.2470	0.1924
$BOOT_{r_0}(\hat{k}_{HQ})$	0.2404	0.1088	0.0908	0.2178	0.2474	0.1258	0.5238	0.4652	0.2272
$BOOT_{r_0}(\hat{k}_{SC})$	0.3118	0.2228	0.0886	0.2158	0.2920	0.2752	0.5710	0.7080	0.3302
$BOOT_{r_0}(\hat{k}_{MAIC})$	0.0144	0.0276	0.0356	0.0218	0.0306	0.0414	0.0238	0.0436	0.0490
$LR_{r_0}(\hat{k}_{AIC})$	0.3614	0.1592	0.1184	0.3318	0.2012	0.0990	0.5476	0.3066	0.2242
$LR_{r_0}(\hat{k}_{HQ})$	0.3634	0.1636	0.1168	0.2816	0.2814	0.1402	0.6084	0.5004	0.2616
$LR_{r_0}(\hat{k}_{SC})$	0.3982	0.2706	0.1160	0.2642	0.3176	0.2882	0.6344	0.7306	0.3564
$LR_{r_0}(\hat{k}_{MAIC})$	0.0750	0.0574	0.0510	0.0618	0.0550	0.0548	0.0834	0.0776	0.0714

Note: The table shows rejection frequencies for replications of the Monte Carlo simulation. The number of simulations is 5000. The true cointegrating rank is $r = 0$. The nominal significance level is 0.05. In the table, LR is Johansen's likelihood ratio test and $BOOT$ denotes the bootstrap version of the LR test. See section 2 for explanation.

Table 5: Rejection Frequencies of Tests for VARMA DGP (9) with $\lambda_1 = 1, \lambda_2 = 0.8, \lambda_3 = 0.7$ $(r = 2)$.

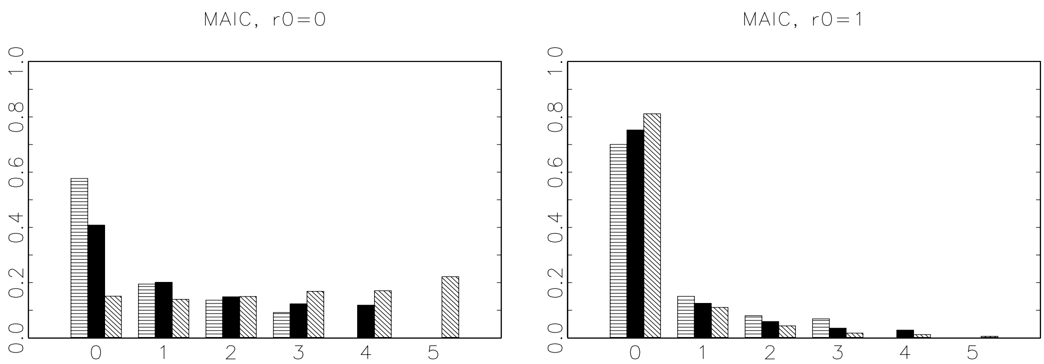
	Panel A: $\lambda_\theta = -0.5,\ \mathcal{H}_0 : r = 2$			Panel B: $\lambda_\theta = 0,\ \mathcal{H}_0 : r = 2$			Panel C: $\lambda_\theta = 0.5,\ \mathcal{H}_0 : r = 2$		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
$BOOT_{r_0}(\hat{k}_{AIC})$	0.0062	0.0364	0.0684	0.0082	0.0328	0.0490	0.0410	0.1602	0.1998
$BOOT_{r_0}(\hat{k}_{HQ})$	0.0056	0.0408	0.0872	0.0084	0.0356	0.0498	0.0552	0.3076	0.2616
$BOOT_{r_0}(\hat{k}_{SC})$	0.0052	0.0326	0.0868	0.0086	0.0354	0.0504	0.0574	0.4240	0.4226
$BOOT_{r_0}(\hat{k}_{MAIC})$	0.0024	0.0164	0.0368	0.0022	0.0160	0.0400	0.0072	0.0412	0.0850
$LR_{r_0}(\hat{k}_{AIC})$	0.0080	0.0308	0.0700	0.0080	0.0308	0.0508	0.0364	0.1518	0.2022
$LR_{r_0}(\hat{k}_{HQ})$	0.0060	0.0344	0.0876	0.0070	0.0330	0.0522	0.0456	0.2992	0.2634
$LR_{r_0}(\hat{k}_{SC})$	0.0050	0.0294	0.0882	0.0068	0.0338	0.0524	0.0486	0.4132	0.4242
$LR_{r_0}(\hat{k}_{MAIC})$	0.0022	0.0122	0.0376	0.0024	0.0128	0.0412	0.0062	0.0386	0.0836

Note: The table shows rejection frequencies for replications of the Monte Carlo simulation. The number of simulations is 5000. The true cointegrating rank is $r = 2$. The nominal significance level is 0.05. In the table, LR is Johansen's likelihood ratio test and $BOOT$ denotes the bootstrap version of the LR test. See section 2 for explanation.

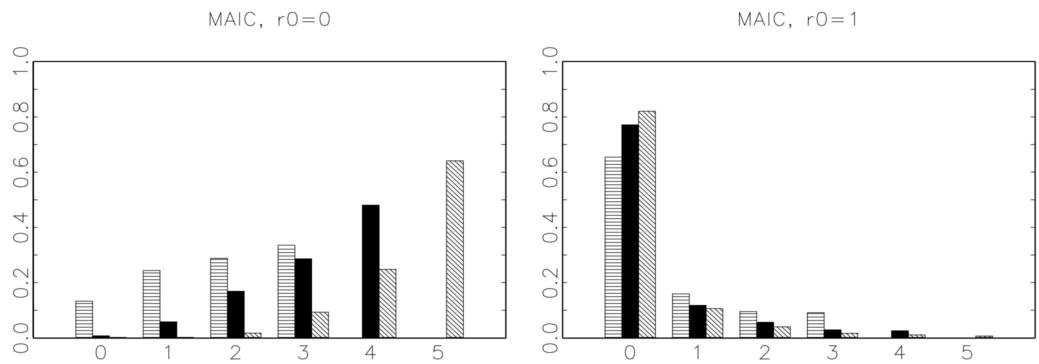


(a) Toda-DGP (7), $a_1 = 0.9$ ($r = 1$)

Figure 1: The figure shows frequencies of lag choices for different criteria and DGPs for $T = 50$ (hatched bar), $T = 100$ (solid bar) and $T = 200$ (diagonally hatched bar).

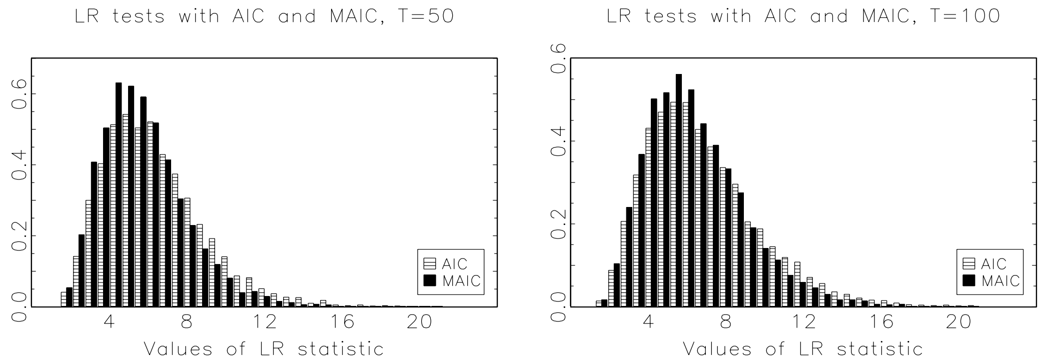


(a) Toda-DGP (7), $a_1 = 0.9$ ($r = 1$)

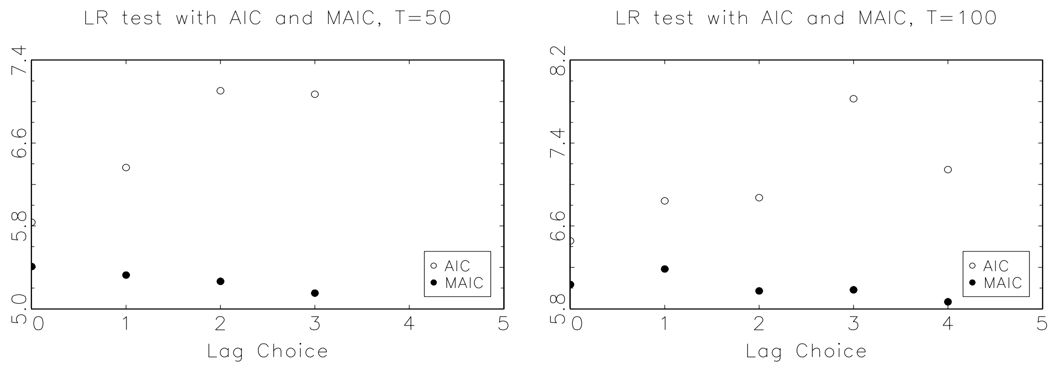


(b) Toda-DGP (7), $a_1 = 0.7$ ($r = 1$)

Figure 2: The figure shows frequencies of lag choices for different criteria and DGPs for $T = 50$ (hatched bar), $T = 100$ (solid bar) ant $T = 200$ (diagonally hatched bar).



(a) Distribution of Test Statistic of the Asymptotic Tests.



(b) Average Test Statistic for Different Lag Lengths

Figure 3: Comparison of the LR test with AIC and $MAIC$. All graphs refer to the Toda-DGP (7), $a_1 = 0.9$ ($r = 1$).