DEPT. OF MATH. UNIVERSITY OF OSLO
STATISTICAL RESEARCH REPORT No. 5
ISSN 0806–3842 APRIL 2004

Bootstrap algorithms for testing and determining the cointegration rank in VAR models.

Anders Rygh Swensen
Department of Mathematics, University of Oslo

[1]

Abstract

In this paper a bootstrap algorithm for a reduced rank VAR-model with a restricted linear trend is analyzed. For testing the cointegration rank the asymptotic distribution under the hypothesis is the same as for the usual likelihood ratio test. It is furthermore shown that a bootstrap procedure for determining the rank is asymptotically consistent in the sense that the probability of choosing the rank too small converges to zero. An empirical illustration is given.

Keywords: VAR models, reduced rank, bootstrap, likelihood ratio test, determination of rank.

JEL Classification: C12, C13, C15, C32.

[1]Address: Department of Mathematics,
University of Oslo, P.O. Box 1053, N-0316 Blindern, Oslo, Norway
Tel: +47 22855871
Tel: +47 22854349
e-mail: swensen@math.uio.no

1 Introduction

There are two aspects of bootstrap methods that are essential for the considerable interest such methods has attracted during the last two and a half decades. The first, which was the original motivation, is that it is possible to find approximate distributions for complicated statistics, where there is no hope of finding exact results and where also asymptotic methods are not feasible. Even if the asymptotic distributions can be found, they often depend on unknown parameters. This means that the statistics are asymptotically non-pivotal, and that asymptotic tables must to be calculated for a large number of values of the parameters. Bootstrap methods can be useful in such situations. An interesting example can be found in the case of fractional cointegration model in Davidson (2002 and 2004).

The second is that the bootstrap approximations can be fairly accurate, even more accurate than the traditional asymptotic approximations of distributions. The possibility of obtaining such improvements is connected to the possibility of basing the bootstrap calculations on functions of the data that are asymptotically pivotal. Another crucial point in the traditional argument for demonstrating the improvements is based on the existence of Edgeworth expansions.

Bootstrap methods for time series containing unit roots have received much attention recently in simulation studies as well as in analytical computations. Typically observations from processes of this type can be written

$$Y_t = \alpha Y_{t-1} + U_t, t = 1, \ldots, T, \tag{1}$$

where $\alpha = 1$, $\{U_t\}$ is a stationary time series with finite second order moments and Y_0 has a specified distribution.

Basawa et al. (1991a) considered the simplest case where the process $\{U_t\}$ consists of independent, identically distributed random variables with finite variance. They showed that the bootstrapped version of the ordinary least squares (OLS) estimator, where the observations are regressed on the lagged values, has an random asymptotic distribution when the OLS estimator is used in the recursion generating the bootstrap sample. In Basawa et al. (1991b) they demonstrated that the asymptotic distribution will be the usual Dickey-Fuller distribution when the bootstrap sample is generated as a random walk using random draws from the residuals from fitting the OLS. Ferretti and Romo (1996) considered a similar setup, but allowed the time series $\{U_t\}$ to be an autoregressive process of order k. In this case the asymptotic distribution of the bootstrapped OLS estimator contains parameters describing the dependence structure in $\{U_t\}$. Psaradakis (2001), Park (2002) and Swensen (2003a) considered the situation where $\{U_t\}$ was approximated by autoregressive processes whose order increases at a suitable rate as the sample size increases. Chang and Park (2002) showed that the bootstrap version of the augmented Dickey-Fuller estimator in such a situation has an asymptotic distribution equal to the usual Dickey-Fuller distribution.

Another way to treat the situation with dependent $\{U_t\}$ is by employing a block bootstrap method instead of approximating by autoregressive processes of increasing length. Swensen (2003a) showed that both procedures result in the same asymptotic distributions for the bootstrapped OLS estimator when the stationary bootstrap is employed. Paparoditis and Politis (2001) used a continuous path block bootstrap. Paparoditis and Politis (2003) studied block bootstrap methods based on residuals from fitting (1).

Some considerations of power have also been carried out. Heimann and Kreiss (1996) demonstrated that when $\alpha = 1$ and the time series $\{U_t\}$ is independent, identically distributed random variables, a test based on bootstrapping the OLS estimator, has a power which tend to one as T increases. Swensen (2003b) considered power in a local to unit root framework, also when the process $\{U_t\}$ consists of independent, identically distributed random variables, and showed that the bootstrap based tests have the same asymptotic power as the ordinary ones. Similar results can be found in Paparoditis and Politis (2003) and Park (2003) for more general processes.

All the references above, except Park (2003), are related to the first aspect of bootstrap methods mentioned earlier. The procedures have the same asymptotic distributions as the original ones. However, due to the versatility of the method, and the intensive use of the data, it can be expected that the approximations are superior in many situations. Park (2003), in fact, derived a stochastic expansion of some bootstrap statistics and explained why an improvement is likely to occur.

In the multivariate situation, allowing for possible cointegrated vectors, there are less results on bootstrapping and most studies are based on simulations. Li and Maddala (1997) considered bootstrapping a cointegrating regression model and Mantalos and Shukur (1998) studied tests for cointegration in a bivariate system based on the coefficient of the error correction term. Gredenhoff and Jacobson (2001) and Fachin (2002) contain material on testing restrictions on the cointegration vectors. More relevant for the present paper are the studies by van Giersbergeen (1996), Harris and Judge (1998) and Mantalos and Shukur (2001) who all studied variants of the recursive bootstrap algorithm for testing the cointegration rank with which we will be concerned. In fact, Harris and Judge (1998) inquired about consistency results similar to those to be presented below.

Consider the error correction form of the p-dimensional VAR model

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{k-1} \Delta X_{t-(k-1)} + \mu_1 t + \mu_0 + \epsilon_t \qquad (2)$$

where the errors ϵ_t are independent, identically distributed random variables with expectation zero and covariance matrix Ω and the initial observations $X_1, \ldots, X_k$ are considered fixed. Furthermore, the fourth moment of each element of ϵ_t is assumed to be finite. The matrix Π will have rank $0 \leq r < p$, and can therefore be written $\Pi = \alpha\beta'$ where α and β are $p \times r$ matrices of full rank r.

It is well known that it is crucial for the analysis of models of type (2) whether μ_1 is contained in the space spanned by the columns of α or not. In the latter case, where there are no restrictions on the linear drift term μ_1, the model (2) is denoted as $H(r)$. The process $\{X_t\}_{t=1}^{\infty}$ will then in general contain a quadratic trend. When the constraint on the linear drift which excludes the quadratic trend, i.e $\mu_1 = \alpha\rho_1$, is satisfied, we denote the model by $H^{\#}(r)$. An appealing feature of this formulation, as stressed by Nielsen and Rahbek (2000) is that the asymptotic distributions of the likelihood ratio tests for cointegration rank is independent of ρ_1. This is also the most general form of the models recommended by Doornik, Hendry and Nielsen (1998) for determining the rank.

We will therefore consider models of this type and suppose that the observations $X_1, \ldots, X_T$ satisfy

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{k-1} \Delta X_{t-(k-1)} + \alpha\rho_1 t + \mu_0 + \epsilon_t. \qquad (3)$$

We will investigate the properties of a bootstrap algorithm for the model (3) when the time series $\{X_t\}$ is integrated of order one, $I(1)$. Here $I(1)$ is used in the sense defined in Johansen (1995), i. e. $\Delta(X_t - E(X_t))$ is a linear process where the sum of the coefficients does not equal the matrix consisting only of zeros. This is the most important situation allowing cointegration, but not the only one. Therefore, before the bootstrap algorithm is employed, the validity of the assumption must be checked to ensure that the appropriate description of the data is chosen.

The bootstrap algorithm we consider is based on the recursive scheme defined (3) using estimated coefficients, errors sampled from the residuals and with the initial observations as starting values. It is thus an example of what Li and Maddala (1997) call a recursive bootstrap and consists of a parametric part reflecting the dependence structure and a nonparametric part due to the sampling of the residuals.

First we will study methods for testing the rank of the matrix Π and show, as one can expect, that the bootstrap based tests have the usual asymptotic distributions. In addition, we will consider the procedure for determining the rank, and show that the bootstrap version is in fact consistent. This problem is more intriguing, and presents some new aspects which are nonstandard in a bootstrap context, since we have to do the resampling for different values of the rank of the estimated reduced rank matrix. The dimension of the cointegration space in the generated observations will therefore not correspond the dimension of the cointegration space in the original data, but the imposed rank in the simulated version of (3). Nevertheless, we shall demonstrate that, as in procedures based directly on the observations, the asymptotic probability of choosing the wrong rank can be bounded.

The plan for the paper is as follows. In the next section we describe the bootstrap algorithm. In section 3 we state the conditions under which the algorithm yields estimators that have the same asymptotic distributions as the estimators constructed directly from the observations. Section 4 treats the procedure to determine the rank and we propose a modification of the algorithm from section 2. In section 5 we consider the Finnish data example from Johansen and Juselius (1990). The proofs can be found in the last section.

If an $m \times n$ matrix a, where $n \leq m$, has full rank, $a_\perp$ denotes an $m \times (m-n)$ matrix of full rank such that $a_\perp' a = 0$. The matrix $a(a'a)^{-1}$ is defined as $\bar{a}$, so that $a'\bar{a} = I_n$ and $\bar{a}a'$ is the projection matrix on the space spanned by the columns of a.

2 Background and bootstrap algorithms

A test for the hypothesis $H^\#(r)$ versus $H^\#(p)$ is to reject for small values of the likelihood ratio $L(H^\#(r))/L(H^\#(p))$. It is well known, see e.g. Johansen (1994 and 1995), that if the errors are Gaussian, the so-called trace statistic, $Q_r{}^\#$, has the form

$$Q_r{}^\# = -2\log(L(H^\#(r)/L(H^\#(p)) = -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i)$$

where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq \hat{\lambda}_{p+1} = 0$ are the ordered eigenvalues of

$$\det[\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}] = 0.$$

The matrices $S_{ij}, i,j = 0,1$ are defined by $S_{ij} = T^{-1}\sum_{t=k+1} R_{it}R_{jt}'$ where R_{0t} and R_{1t} are the residuals of the differences ΔX_t and the variables $X_{t-1}^\# = (X_{t-1}', t)'$ after regressing on $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}$, and 1.

The most common use of the tests based on the trace statistic $Q_r^\#$ is for determining the rank r of the matrix Π, as explained in Johansen (1992) and in Chapter 12 in Johansen (1995). This can be formulated as a multiple decision or estimation problem. If $\hat{R}$ is the estimator for the rank, then the estimator $\hat{R}$ has the consistency property, see Johansen (1992 and 1995),

$$\lim_{T\to\infty} P_{r,T}(\hat{R} = i) = 0, i = 0, 1, \ldots, r-1.$$

The probability measures $P_{r,T}$ denote the distributions in the situations where the matrix Π has rank r and $\alpha_\perp'\mu_1 = 0$. When there is no ambiguity we simply drop the subscripts.

We now consider the following bootstrap algorithm for constructing samples of bootstrapped or pseudo-observations $X_1^*, \ldots, X_T^*$ to test $H^{\#}(r)$ versus $H^{\#}(p)$
.

Algorithm 1

(i) Estimate the model $H^{\#}(r)$ by reduced rank regression obtaining estimates $\hat{\alpha}, \hat{\beta}$ and $\hat{\rho}_1$. For the remaining parameters, use the unrestricted ordinary least squares estimates $\hat{\Gamma}_1, \cdots, \hat{\Gamma}_{k-1}$ and $\hat{\mu}_0$. Compute the unrestricted OLS residuals $\hat{\epsilon}_{k+1}, \ldots, \hat{\epsilon}_T$.

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

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

are equal to 1 or outside the unit circle and whether $\hat{\alpha}'_\perp \hat{\Gamma} \hat{\beta}_\perp$ is non- singular, where $\hat{\Gamma} = I - \hat{\Gamma}_1 - \cdots - \hat{\Gamma}_{k-1}$.

(iii) If so, compute $X_t^*, t = k+1, \ldots, T$ recursively from

$$\Delta X_t^* = \hat{\alpha}\hat{\beta}' X_{t-1}^* + \hat{\Gamma}_1 \Delta X_{t-1}^* + \cdots + \hat{\Gamma}_{k-1} \Delta X_{t-(k-1)}^* + \hat{\alpha}\hat{\rho}_1 t + \hat{\mu}_0 + \epsilon_t^*. \quad (4)$$

with sampled residuals ϵ_t^* drawn with replacement from the estimated residuals $\hat{\epsilon}_{k+1}, \ldots, \hat{\epsilon}_T$ and with $X_1, \ldots, X_k$ as starting values. Set $X_1^*, \ldots, X_k^*$ equal to $X_1, \ldots, X_k$.

The residuals and pseudo-observations depend on T, but we do not indicate this dependency when no confusion can arise. As usual in the literature, we use an $*$ to denote the bootstrap distributions, so P_r^* is the conditional distribution of $X_1^*, \ldots, X_T^*$ given the observations $X_1, \ldots, X_T$. The subscript indicates that the pseudo-observations $X_1^*, \ldots, X_T^*$ are generated imposing the rank equal to r. In some cases, where there are no possibility for confusion, we drop the subscript.

Remark 1. The requirement (ii) is due to the fact that the estimates are not necessarily corresponding to characteristic polynomials having zeros outside the unit circle. This can happen although not very often, see Johansen (1995, p. 71), and corresponds to the processes being explosive.

The purpose of the checking the conditions in (ii) is to ensure that the pseudo-observations from the recursive scheme are in fact $I(1)$ variables. In case the conditions are not satisfied it is an indication that another more appropriate recursive scheme reflecting the properties of the observed data should be used. From a numerical point of view the requirement that $\hat{\alpha}'_\perp \hat{\Gamma} \hat{\beta}_\perp$ is non-singular is innocuous. Since $\hat{\Gamma}$ is the unrestricted OLS estimate, it will be true with probability 1. However, to stress the point that we want the bootstrap samples to be $I(1)$ we keep the requirement.

For scalar time series it is possible to dispense with (ii) using the Yule-Walker estimators which are known to imply that the estimated roots are outside the unit circle. However, that is not necessarily an advantage. If the estimated roots correspond to explosive roots, we would like to discover it.

Remark 2. In addition to a constant term it is in many situations natural to include centered seasonal dummies in the model. If so, they should be treated in the bootstrap algorithm in the same way as the constant function.

Remark 3. Remark that we do the resampling from the residuals from fitting an unrestricted VAR model of order k using OLS. Also the unrestricted OLS estimators $\hat{\Gamma}_1, \cdots, \hat{\Gamma}_{k-1}$ and $\hat{\mu}_0$ are used in the recursion (4). An alternative is to use the residuals and estimators after having imposed the reduced rank r.

The contrast between using the unrestricted estimators and imposing all restrictions corresponds in the univariate case to base the bootstrap replications either on the OLS residuals or on the differences of the observations, i.e. the restricted or unrestricted residuals in the terminology of Li and Maddala (1997). The recent paper by Paparoditis and Politis (2003) indicates that there may be certain advantages using the unrestricted ones.

Remark 4. As pointed out in the introduction variants of Algorithm 1 have been studied earlier. van Giersbergeen (1996) sampled the residuals using a stationary bootstrap to allow for possible misspecification. Harris and Judge (1998) and Mantalos and Shukur (2001) sampled the residuals independently, but used what we called restricted residuals in Remark 3.

The bootstrap version of the statistic $Q_r^{\#}$, which we denote by $Q_r^{\#*}$, is now computed by regressing ΔX_t^* and $X_{t-1}^{\#*} = (X_{t-1}^{*\prime}, t)'$ on $\Delta X_{t-1}^*, \ldots, \Delta X_{t-k+1}^*$ and 1. Denote the moment matrices computed from the residuals $S_{ij}^*, i, j = 0, 1$ where the value 0 as before refers to differences and value 1 to levels. Then $\hat{\lambda}_1^* \geq \cdots \geq \hat{\lambda}_p^* \geq \hat{\lambda}_{p+1}^* = 0$ are the ordered eigenvalues of

$$\det[\lambda S_{11}^* - S_{10}^* S_{00}^{*-1} S_{01}^*] = 0$$

and

$$Q_r^{\#*} = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i^*).$$

The bootstrap distribution of $Q_r^{\#*}$ is the conditional distribution of $Q_r^{\#*}$ given the observations $X_1, \ldots, X_T$. Denote the cumulative distribution function by $F_{r,T}^*$. The bootstrap test of $H^{\#}(r)$ versus $H^{\#}(p)$ at level γ consists of rejecting $H^{\#}(r)$ if $Q_r^{\#}$ is larger than the $(1-\gamma)$- quantile of the bootstrap distribution, that is if

$$1 - F_{r,T}^*(Q_r^{\#}) \leq \gamma. \tag{5}$$

Since the bootstrap distribution is usually a very complicated function of the observations, it must be approximated. A feasible bootstrap test consists of rejecting if

$$\frac{1}{B}\#\{Q_r^{\#*} > Q_r^{\#}\} \leq \gamma, \tag{6}$$

where $\#\{\cdot\}$ denotes the number of elements in the set $\{\cdot\}$ and B denotes the number of pseudo-samples that are generated. As B increases, the expression (6) will converge toward (5). It is important to choose B large enough, so that a good approximation is achieved. For a recent treatment of this issue, see Davidson and MacKinnon (2000).

3 Asymptotic distributions of the bootstrap likelihood ratio tests

We shall in this section consider the asymptotic distribution of the bootstrap distribution introduced in the previous section for the case that $\{X_t\}$ is an $I(1)$ process defined by (3). More precisely, we shall assume

Assumption 1

i) Assume that the roots of the characteristic polynomial

$$A(z) = (1-z)I_p - \Pi z - \Gamma_1(1-z)z - \cdots - \Gamma_{k-1}(1-z)z^{k-1}$$

are located outside the complex unit circle or at 1. Also assume that the matrices α and β have full rank r and that

$$\alpha_{\perp}'\Gamma\beta_{\perp}$$

has full rank $p-r$, where $\Gamma = I_p - \Gamma_1 - \cdots \Gamma_{k-1}$.

ii) The random variables $\epsilon_i, i = 0, \pm 1, \ldots$ are independent, identically distributed with expectation 0 and covariance matrix Ω.

From Johansen (1994 and 1995) it follows that under Assumption 1, X_t has the representation

$$X_t = C\sum_{i=k+1}^{t}\epsilon_i + \tau(t-k) + Y_t + A \tag{7}$$

where $C = \beta_{\perp}(\alpha_{\perp}'\Gamma\beta_{\perp})^{-1}\alpha_{\perp}'$, $\tau = \bar{\beta}_{\perp}(\bar{\alpha}_{\perp}'\Gamma\bar{\beta}_{\perp})^{-1}\bar{\alpha}_{\perp}'\mu_0 + (\bar{\beta}_{\perp}(\bar{\alpha}_{\perp}'\Gamma\bar{\beta}_{\perp})^{-1}\bar{\alpha}_{\perp}'\Gamma - I)\bar{\beta}\rho_1 = C\mu_0 + (C\Gamma - I)\bar{\beta}\rho_1$, Y_t is a stationary process and A is a term depending only on the initial values so that $\beta' A = 0$. This implies, in particular, that if $\beta^{\#} = (\beta', \rho_1)'$, the process $\beta^{\#\prime}X_t^{\#}$ will be stationary for a suitable initial distribution.

The random process $\{\frac{1}{\sqrt{T}}\sum_{i=1}^{[uT]}\epsilon_i, 0 \leq u \leq 1\}$, where $[uT]$ is the integer value of uT, converges weakly toward the Brownian motion $\{W(u), 0 \leq u \leq 1\}$ with covariance matrix Ω. From this result the asymptotic behavior of X_t can be deduced.

The asymptotic distribution of the likelihood ratio test is derived in Johansen (1994 and 1995). Since we do not assume Gaussian errors, $\Sigma_{00}, \Sigma_{0\beta}$ and $\Sigma_{\beta\beta}$ have to be defined as the probability limits of $S_{00}, S_{01}\beta^{\#}$ and $\beta^{\#\prime}S_{11}\beta^{\#}$ respectively.

Theorem 1 *Assume that Assumption 1 is satisfied. As $T \to \infty$ for all x,*

$$P_{r,T}(Q_r^{\#} \leq x) \to P(tr(\int_0^1 (dB)F'[\int_0^1 FF']^{-1}\int_0^1 F(dB)') \leq x), \qquad (8)$$

where B is a $p-r$ dimensional standard Brownian motion on the unit interval and $F(u) = ((B(u) - \int_0^1 B(u)du)', u - 1/2)'$.

Now, let us consider the likelihood ratio test for $H^{\#}(r)$ vs. $H^{\#}(p)$ constructed on the pseudo-data generated by Algorithm 1. By "weak convergence in P_T probability" it is meant that the conditional distribution, given the observations, converges in P_T probability.

It is not surprising that a representation of the form (7) is valid for each bootstrap replication. The remainder terms will depend on the realization, however, so in a bootstrap context it is necessary to pay closer attention to such terms. The following representation can be established by an elaboration of the proof of (7).

Lemma 1 *Under Assumption 1 and if in addition the fourth moment of the errors $\epsilon_1, \ldots$ exists, the generated pseudo-observations have the following representation*

$$X_t^* = \hat{C}\sum_{i=k+1}^{t} \epsilon_i^* + \hat{\tau}(t-k) + \sqrt{T}R_{t,T}^*, t = k+1, \ldots, T,$$

where for all $\eta > 0$, $P^*(max_{k+1\leq t\leq T}|R_{t,T}^*| > \eta) \to 0$ in $P_{r,T}$ probability as $T \to \infty$, $|\cdot|$ is the usual Euclidean distance, $\hat{C} = \hat{\beta}_\perp(\hat{\alpha}_\perp'\hat{\Gamma}\hat{\beta}_\perp)^{-1}\hat{\alpha}_\perp'$ and $\hat{\tau} = \hat{C}\hat{\mu}_0 + (\hat{C}\hat{\Gamma} - I)\bar{\hat{\beta}}\hat{\rho}_1$. Furthermore, $E^*[\epsilon_i^*\epsilon_i^{*\prime}] = \Omega_T \to \Omega$ in $P_{r,T}$ probability as $T \to \infty$.

Therefore, X_t^* consists of a stochastic trend and a deterministic trend in addition to the stationary part. The random process $\{\frac{1}{\sqrt{T}}\sum_{i=k+1}^{[uT]} \epsilon_i^*, 0 \leq u \leq 1\}$ converges in P^* distributions as functions in $D[0,1]^p$ toward a p-dimensional Brownian motion, W, with covariance matrix Ω. By following the arguments in Johansen (1994 and 1995), one has

Proposition 1 *Let the bootstrap samples be generated by Algorithm 1. Then, under Assumption 1 and if in addition the fourth moment of the errors* $\epsilon_1, \ldots$ *exists,*

$$P^*(Q_r^{*\#} \leq x) \to L(x)$$

in $P_{r,T}$ *probability for all* x *as* $T \to \infty$, *where* L *is the distribution defined in (8) in Theorem 1.*

The essential point in proving Proposition 1 is to show that

Lemma 2 *Under the same assumptions as in Proposition 1 and if* $\eta > 0$,

$$P^*(\| S^*_{00} - \Sigma_{00} \| > \eta) \to 0,$$
$$P^*(\| S^*_{01}\hat{\beta}^{*\#} - \Sigma_{0\beta} \| > \eta) \to 0,$$

and

$$P^*(\| \hat{\beta}^{*\#\prime} S^*_{11}\hat{\beta}^{*\#} - \Sigma_{\beta\beta} \| > \eta) \to 0$$

in P_T *probability where* $\| \, . \, \|$ *denote the matrix norm* $\| C \| = [tr(C'C)]^{1/2}$.

Furthermore, from the representation in Lemma 1 it follows that

Lemma 3 *Under the same assumptions as in Proposition 1 with norming matrix* $C_T = \begin{pmatrix} \bar{\beta}_\perp & 0 \\ -\tau'\bar{\beta}_\perp & T^{-1/2} \end{pmatrix}$,

$$\frac{1}{T} C_T' S^*_{11} C_T \to \int_0^1 GG'$$

and

$$\frac{1}{T} C_T' (S^*_{10} - S^*_{11}\hat{\beta}^{\#}\hat{\alpha}') \to \int_0^1 G(dW')$$

where $\bar{W} = \int_0^1 W(u)du$, $G(u) = ((W(u) - \bar{W})'\hat{C}'\bar{\beta}_\perp, u - 1/2)'$ *and the convergence is weakly in* P_T *probability.*

Proof of Proposition 1. The Proposition can be proved arguing in the same way as in the proof of Theorem 11.1 in Johansen (1995) so we only sketch the modifications that are necessary. The proof in Johansen (1995) is based on a model with unrestricted constant and no linear drift, i.e. assuming $\rho_1 = 0$ in model (3) . From Lemma 2 it follows that the $p - r + 1$ smallest solutions of

$$\det[\lambda S^*_{11} - S^*_{10} S^{*-1}_{00} S^*_{01}] = 0$$

converge to zero.

To find the asymptotic distribution we use Lemma 3 which implies that the $p-r+1$ smallest solutions multiplied by T converge weakly in P_T probability toward the distribution of the solutions of

$$|\rho \int_0^1 GG' - \int_0^1 G(dW)'\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' \int_0^1 (dW)G'| = 0. \tag{9}$$

Proposition 1 will now follow by defining $B = (\alpha_\perp'\Omega\alpha_\perp)^{-1/2}\alpha_\perp' W$, which is a standard Brownian motion.

Remark 5. Combining Theorem 1 and Proposition 1 it follows that for all x when $\alpha_\perp'\mu_1 = 0$,

$$P^*(Q_r^\# \leq x) - P_{r,T}(Q_r^\# \leq x) \to 0 \tag{10}$$

in $P_{r,T}$ probability as $T \to \infty$.

The limit (10) is the bootstrap justification for the approximation to the probabilities $P_{r,T}(Q_r^\# \leq x)$ which are the probabilities we are interested in and try to approximate.

Remark 6. To establish a relation like (10) it is of course sufficient to prove that the terms have the same asymptotic distribution. The statistics we have considered so far have asymptotic distributions which do not depend on any unknown parameters. However, once this is relaxed a lot of other test statistics are possible to use. For example, a test can be based on the moment matrices $M_{i,j}, i,j = 0,1$ of the levels and differences in the same way as on the corresponding matrices $S_{i,j}, i,j = 0,1$ formed by the residuals after regressing on $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}$. In view of Theorem 4.2 in Johansen (1995) one can expect that the asymptotic distributions are equal for the observed data and the generated pseudo-data. The asymptotic distributions are much more complicated however, and we see no point in demonstrating this in detail, since the important property in a bootstrap context is (10).

Remark 7. There also exist other statistics for testing the rank order of VAR models. In particular we would like to mention the so-called λ_{max} statistic which is defined as $\lambda_{max,r} = -T\log(1-\hat{\lambda}_{r+1})$ and is the log likelihood ratio for testing $H^\#(r)$ versus $H^\#(r+1)$. By the arguments used to prove Proposition 1 it follows that the asymptotic distribution of the bootstrap version of this test also remains the same as for the corresponding test based on the original observations.

Remark 8. By inspecting the proof of Proposition 1 one can see that the result is still valid when the pseudo-observations are generated by using the restricted residuals and estimators described in Remark 3. What is needed is that the mean of the squared residuals converges in probability toward Ω and that the estimators are consistent. These properties are true for the restricted residuals and estimators.

Remark 9. In bootstrap testing it is important to know how the generated pseudo-series behaves when the cointegration rank in the observations is different from the rank imposed under the null hypothesis. This question has not been addressed in the present section. However, it will follow from Lemmas 5 and 6 in the next section, that when the true rank is larger than the imposed one, the cointegration rank of the generated pseudo-observation will in fact be rank postulated under the null hypothesis.

4 Asymptotic consistency of the procedure to determine the rank

As explained in section 2 the most important use of the likelihood ratio test for the cointegration rank is in the procedure to determine the rank. To consider this problem the results from the previous section need some modification because the pseudo-samples have to be constructed also for the ranks $0, 1, \ldots, p-1$, not only for rank equal to r, which is the rank for the original observations. Therefore, this is not the standard situation for bootstrapping time series, since the pseudo-observations and the actual observations are generated by different schemes.

Since pseudo-samples are produced for each rank, Algorithm 1 must be modified. We propose to start out with the unrestricted OLS estimators.

Algorithm 2

(i) Let $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}, \hat{\mu}_0$ *be the unrestricted OLS estimators and* $\hat{\epsilon}_{k+1}, \ldots, \hat{\epsilon}_T$ *the OLS residuals, i.e estimate (3) with* $r = p$.

(ii) Let $\hat{\beta}_j^{\#} = (\hat{\beta}_j', \hat{\rho}_{1j})', j = 0, \ldots, p-1$ *be the estimators of the cointegration vectors assuming the reduced rank is* j *and imposing the condition* $\alpha_\perp \mu_1' = 0$ *as defined in Johansen (1995, p. 97). Define* $\hat{\alpha}_j = S_{01}\hat{\beta}_j^{\#}(\hat{\beta}_j^{\#\prime} S_{11}\hat{\beta}_j^{\#})^{-1}$ *and* $\hat{\Pi}_j^{\#} = \hat{\alpha}_j \hat{\beta}_j^{\#\prime}, \hat{\Pi}_j = \hat{\alpha}_j \hat{\beta}_j', j = 0, \ldots, p-1$.

(iii) Starting with $j = 0$*, check whether the roots of the characteristic equation* $\det[\hat{A}_j(z)] = 0$ *where*

$$\hat{A}_j(z) = (1-z)I_p - \hat{\Pi}_j z - \hat{\Gamma}_1(1-z)z - \cdots - \hat{\Gamma}_{k-1}(1-z)z^{k-1}$$

are equal to 1 or located outside the unit circle and whether $\hat{\alpha}_{j\perp}' \hat{\Gamma} \hat{\beta}_{j\perp}$ *is non-singular with* $\hat{\Gamma} = I - \hat{\Gamma}_1 - \cdots - \hat{\Gamma}_{k-1}$.

(iv) If so, compute $X_{j,t}^*, t = k+1, \ldots, T$ *recursively by the recursion*

$$\Delta X_{j,t}^* = \hat{\Pi}_j X_{j,t-1}^* + \hat{\Gamma}_1 \Delta X_{j,t-1}^* + \cdots + \hat{\Gamma}_{k-1} \Delta X_{j,t-(k-1)}^* + \hat{\mu}_0 + \hat{\alpha}_j \hat{\rho}_{1j} t + \epsilon_t^*, t = k+1, \ldots, T \tag{11}$$

with the sampled residuals ϵ_t^* drawn with replacement from the estimated residuals $\hat{\epsilon}_{k+1}, \ldots, \hat{\epsilon}_T$. Use $X_1, \ldots, X_k$ as starting values. Set $X_{j,1}^*, \ldots, X_{j,k}^*$ equal to $X_1, \ldots, X_k$.

(v) *Check if the fraction $\frac{1}{B}\#\{Q_j^{\#*} > Q_j^{\#}\}$ exceeds a fixed level γ. If yes, estimate the rank as j and stop, else perform step (iii)-(v) with rank equal to $j+1$.*

(vi) *If none of the fractions $\frac{1}{B}\#\{Q_0^{\#*} > Q_0^{\#}\}, \ldots, \frac{1}{B}\#\{Q_{p-1}^{\#*} > Q_{p-1}^{\#}\}$ exceed γ, let the estimated rank be p.*

Remark 10. As in Algorithm 1, the purpose of (iii) is to check that the process from which the pseudo-observations will be generated, is in fact $I(1)$. If the answer is negative another model might be more appropriate, or that data generating process is in fact $I(1)$, but the rank of the model that we are about to generate pseudo-observations from, is of higher rank than the actual. As we show below, the asymptotic probability of the latter event is bounded by γ.

This bootstrap procedure has the following consistency property in common with the usual procedure to determine the rank.

Proposition 2 *Let the bootstrap samples or pseudo-observations be generated by Algorithm 2. Then, under Assumption 1 and if in addition the fourth moment of the errors $\epsilon_1, \ldots$ exists,*

$$F_{j,T}^*(Q_j^{\#}) \to 1, j = 0, \ldots, r-1$$

in $P_{r,T}$ probability as $T \to \infty$.

In other words, the probability of estimating the rank as $0, 1, \cdots, r-1$ when it actually is r, tends to zero as T increases.

Remark 11. If R^* is the estimator of the rank described in Algorithm 2, we have the following under Assumption 1

$$\begin{aligned}
\lim_{T\to\infty} P_{r,T}(R^* = j) &= 0 && \textit{for } j = 0, \ldots, r-1 \\
\lim_{T\to\infty} P_{r,T}(R^* = r) &= 1-\gamma && \textit{for } j = r \\
\limsup_{T\to\infty} P_{r,T}(R^* = j) &\leq \gamma && \textit{for } j = r+1, \ldots, p.
\end{aligned}$$

The statement in the first line is a reformulation of Proposition 2, and the statement in the second line is the probability of not rejecting $H^{\#}(r)$ against $H^{\#}(p)$, and follows from Proposition 1. Since $\sum_{j=0}^{p} P_{r,T}(R^* = j) = 1$ the statement in the third line must also hold.

Proposition 2 follows from the fact that $Q_0^{\#}, \ldots, Q_{r-1}^{\#} \to \infty$ in $P_{r,T}$ probability, since all these statistics contain a contribution from an estimated eigenvalue which is positive in the limit. This is proved in Lemma 12.4 in Johansen (1995). Hence the proposition will follow if we show the random variables $Q_0^{\#*}, \ldots, Q_{r-1}^{\#*}$ converge weakly in probability.

To do that we proceed as in section 3 using the following generalization of Lemma 1

Lemma 4 *Under the same assumptions as in Proposition 2 the generated pseudo-observations from Algorithm 2 have the following representations for each $j = 0, \ldots, r-1$*

$$X_{j,t}^* = \hat{C}_j \sum_{i=k+1}^{t} \epsilon_i^* + \hat{\tau}_j(t-k) + \sqrt{T} R_{j,t,T}^*, t = k+1, \ldots, T,$$

where for all $\eta > 0$, $P^*(max_{k+1 \leq t \leq T} |R_{j,t,T}^*| > \eta) \to 0$ in $P_{r,T}$ probability as $T \to \infty$, $|\cdot|$ is the usual Euclidean distance, $\hat{C}_j = \hat{\beta}_{j\perp}(\hat{\alpha}_{j\perp}' \hat{\Gamma} \hat{\beta}_{j\perp})^{-1} \hat{\alpha}_{j\perp}'$ and $\hat{\tau}_j = \hat{C}_j \hat{\mu}_0 + (\hat{C}_j \hat{\Gamma} - I) \hat{\bar{\beta}}_j \hat{\rho}_{1j}$.

By inspecting the proof of Lemma 1 it can be seen that it is essential to verify that the equation $\det[\hat{A}_j(z)] = 0$ corresponding to the recursion (11), which is used to generate the pseudo-data, has roots in 1 or located outside the unit circle, that $\hat{\Pi}_j$ has full rank j and that $\hat{\alpha}_{j\perp}' \hat{\Gamma} \hat{\beta}_{j\perp}$ has full rank $p - j$, when T is large enough. These properties will also guarantee that the bootstrapped data is $I(1)$ and has cointeration rank $j, j = 0, \ldots, r-1$. For a proof, use the consistency of the least squares estimators $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}, \hat{\mu}_0$, see e.g. Sims, Stock and Watson (1990), and the following lemma

Lemma 5 *Let Assumption 1 be satisfied, let $\rho_1 > \ldots > \rho_r > 0$ be the nonzero eigenvalues of*

$$\det[\rho \Sigma_{00} - (\Sigma_{00} - \Omega)] = 0,$$

let $w_1, \ldots, w_r$ be the corresponding eigenvectors satisfying $W' \Sigma_{00} W = I$, where W is the matrix with columns $w_1, \ldots, w_r$, let W_j be the $p \times j$ matrix having columns $w_1, \ldots, w_j$ and let $\Pi^{\#} = (\Pi, \alpha \rho_1)$.

Then, as $T \to \infty$,

$$\hat{\Pi}_j^{\#} \to (\Sigma_{00} - \Omega) W_j [W_j' (\Sigma_{00} - \Omega) W_j]^{-1} W_j' \Pi^{\#} = \Pi_j^{\#}$$

in $P_{r,T}$ probability, $j = 0, 1, \ldots, r-1$.

Then it suffices to show that the equation $\det[A_j(z)] = 0$, where

$$A_j(z) = (1-z) I_p - \Pi_j z - \Gamma_1 (1-z) z - \cdots - \Gamma_{k-1} (1-z) z^{k-1}$$

has roots which are either equal to 1 or located outside the unit circle for $j = 0, \ldots, r-1$. $\Pi_j^{\#}$ is defined in Lemma 5 and can be expressed as $\Pi_j^{\#} = \alpha_j \beta_j^{\#\prime}$, where $\beta_j^{\#} = (\beta_j', \rho_{1j})'$. Then $\Pi_j = \alpha_j \beta_j'$. Furthermore, $\alpha_{j\perp}' \Gamma \beta_{j\perp}$ must have full rank. These facts follows from

Lemma 6 *Let Assumption 1 be satisfied. Then the roots of the equations* $\det[A_j(z)] = 0$ *are either equal to* 1 *or have modules larger than* 1 *and* $\alpha_{j\perp}' \Gamma \beta_{j\perp}$ *has full rank* $p-j$, $j = 0, 1, \ldots, r-1$.

Proposition 2 will then follow by using the representation in Lemma 4 and an argument similar to the one used in Lemma 2 and 3 to prove that

Lemma 7 *Under the same assumptions as in Proposition 2*

$$P^*(Q_j^{*\#} \leq x) \rightarrow L_j(x)$$

in $P_{r,T}$ *probability for all* x *and all* $j = 0, \ldots, r-1$ *as* $T \rightarrow \infty$, *where* L_j *is the distribution defined in (8) in Theorem 1 when* B *is a* $p-j$ *dimensional standard Brownian motion.*

Remark 12. In Remark 8 we pointed out that the result in Proposition 1 is still true if we use the estimators and residuals from the reduced rank model. However, the corresponding result about the validity of Proposition 2 in such a case, seems more difficult to establish by the machinery used here. What must be shown is that the recursions using the probability limits of the estimators computed for rank j, when the actual rank is r, satisfy the property described in Lemma 6.

5 An example

To gain some insight in the performance of the bootstrap methods we apply the procedures to an example with data on demand for money in Finland, which has been treated previously by Johansen and Juselius (1990). The data consists of quarterly observations of the variables $(m1, y, \Delta p, i^m)$ where $m1$ is log money stock, y is log real income, Δp is log inflation rate and i^m is the marginal rate of interest of the Bank of Finland. The observation period is 1958:1 to 1984:3. Johansen and Juselius (1990) estimated a VAR-model of order two containing an unrestricted constant term and centered seasonal dummies. They used rank order three, opening for the possibility that the appropriate rank is two. Reanalyzing the data with the procedure for determining the rank Johansen (1992) found the a reduced rank equal to two.

We first report the result from estimating using a restricted drift term in addition, i.e. the model given by (3) plus centered seasonal dummies. The non-zero eigenvalues are 0.3425, 0.2541, 0.0921 and 0.0444 corresponding to $Q_0^{\#} = 88.85, Q_1^{\#} = 45.25, Q_2^{\#} = 14.77, Q_3^{\#} = 4.72$ and $\lambda_{max,0} = 43.60, \lambda_{max,1} = 30.48, \lambda_{max,2} = 10.05, \lambda_{max,3} = 4.72$. Determining the rank using the critical values from Table 2* in Osterwald-Lenum (1992) and a significance level equal to 5% still gives a rank equal to two. But it is worth pointing out that the 95% quantile for the trace statistic, $Q_1^{\#}$, is 42.44 and the 97.5% quantile is 45.42, so the alternatives $r = 1$ and $r = 2$ are fairly close.

We will look at some bootstrap distributions for both rank equal to one and rank equal to two. The bootstrap samples were first generated according to Algorithm 1, where we used random numbers generated by the procedure ran2 in Press at. al (1992). Figure 1 shows histograms based on $B = 1000$ replications. The figures to the left and right are based on using the $\lambda^*_{max,1}$ and $\lambda^*_{max,2}$ statistics for rank=1 and rank=2 respectively, while the figure in the middle shows the bootstrap distribution for the trace statistic, $Q_1^{\#*}$, when the rank=1. The fractions for which the statistics yield bootstrapped values larger than the observed value are 0.030, 0.164 and 0.776.

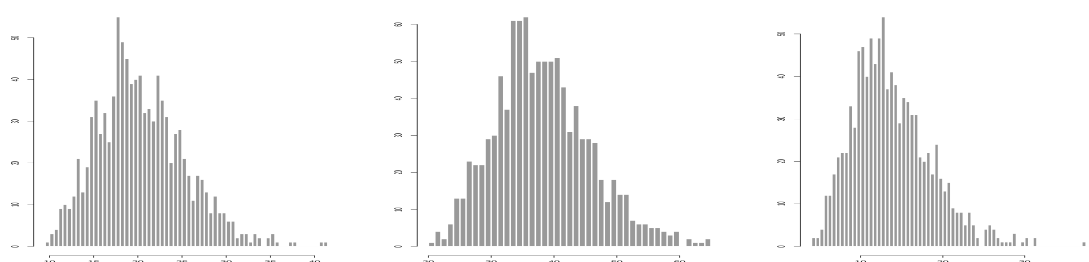

Figure 1: Histogram for bootstrap distributions of $\lambda^*_{max,1}, Q_1^{*\#}$ and $\lambda^*_{max,2}$ based on 1000 replications using Algorithm 1. The two leftmost figures are bootstrapped using rank=1, the figure to the right using rank=2.

We then use Algorithm 2 to determine the rank. If the procedure is based on the trace statistics the sequence of fractions is 0.000, 0.164 and 0.788 so that we estimated the rank as one with levels 0.05 or 0.10.

We now comment upon the fact mentioned in Remark 8, that we may resample using the restricted residuals and estimators for approximating the distributions of the trace and λ_{max} statistics. The figures corresponding to those reported using Algorithm 1, which is based on the unrestricted OLS residuals and estimators, are 0.040, 0.122, and 0.790. Figure 2 shows a more detailed picture. Here the quantiles of $\lambda^*_{max,1}, Q_1^{\#*}$ and $\lambda^*_{max,2}$ from using restricted residuals/estimators and unrestricted residuals/estimators are plotted against each other with the quantiles from the restricted versions as the abscissa. Different seeds for initiating the random number generator were used in the two cases.

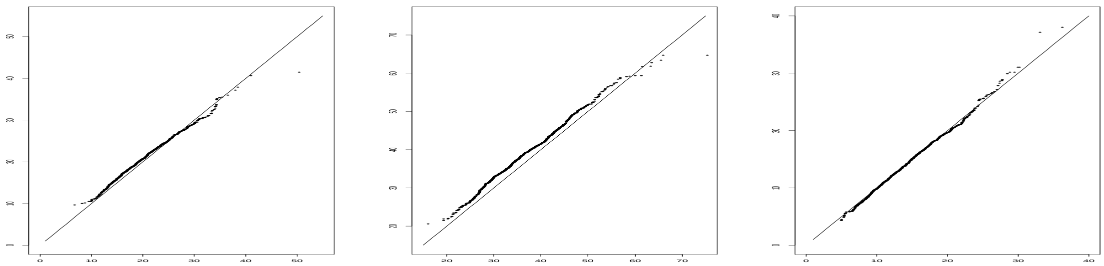

Figure 2: Q-Q plot of quantiles of $\lambda^*_{max,1}$, $Q_1^{\#*}$ and $\lambda^*_{max,2}$ using unrestricted residuals/estimators and restricted residuals/estimators. The two leftmost figures are bootstrapped using rank=1, the figure to the right using rank=2. The abscissa shows the quantiles from the restricted versions. The figures are based on 1000 replications.

As one can see the distributions are fairly similar. The most noticeable difference is that the quantiles of the $Q_1^{\#*}$ statistic based on the unrestricted residuals are a bit larger than those of the statistic based on the restricted ones.

6 Proofs

Proof of Lemma 1. The first part of the proof follows the proof of Theorem 4.2 in Johansen (1995). We shall provide some details, but for a full account we refer to Johansen (1995).

We use the result from Lemma 4.1 in Johansen (1995): if $\Psi(z) = \sum_{i=0}^{\infty} \Psi_i z^i$ is convergent for $|z| < 1 + \delta$ for some $\delta > 0$, and $\Psi_i^{\#}$ and $\Psi^{\#}(z)$ are defined by $\Psi_i^{\#} = \sum_{j=i+1}^{\infty} \Psi_j$ and $\Psi^{\#}(z) = \sum_{i=0}^{\infty} \Psi_i^{\#} z^i$ respectively, then $\Psi^{\#}(z)$ is also convergent for $|z| < 1+\delta$ and $\Psi(z) = \Psi(1) + (1-z)\Psi^{\#}(z)$. If $\Psi(z)$ is a polynomial, so is $\Psi^{\#}(z)$.

From the proof of Theorem 4.2 in Johansen (1995) it also follows that if $\tilde{X}_t^* = (X_t^{*\prime}\hat{\beta}, \Delta X_t^{*\prime}\hat{\beta}_{\perp})'$, the recursion defined in Algorithm 1 may be expressed as

$$\tilde{\hat{A}}(L)\tilde{X}_t^* = (\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_{\perp})'(\epsilon_t^* + \hat{\mu}_0 + \hat{\alpha}\hat{\rho}_1 t)$$

for a suitable polynomial.

We then use the fact that the zeros of $\det[\hat{A}(z)] = 0$ are equal to 1 or are outside the unit circle, when T is large enough. This follows from the fact that

Assumption 1 implies that $\det[A(z)] = 0$ has exactly $p - r$ solutions at 1 and the rest of the solutions are outside the unit circle, see Corollary 4.3 in Johansen (1995). The definition of $\hat{A}(z)$ implies that $\det[\hat{A}(z)] = 0$ has at least $p - r$ solutions at 1, see Johansen (1995, p. 16). Since the estimators of the coefficients are consistent, the solutions of $\det[\hat{A}(z)] = 0$ must converge to those of $\det[A(z)] = 0$. Thus when T is large enough, $p - r$ of the solutions $\det[\hat{A}(z)] = 0$ are equal to 1 and the rest are outside the unit circle.

Lemma 4.1 of Johansen (1995) can now be applied to $\tilde{\hat{C}}(z) = \tilde{\hat{A}}(z)^{-1}$. The reason is that $\det[\hat{A}(z)] = 0$ and $\det[\tilde{\hat{A}}(z)] = 0$ have the same roots except for $z = 1$ and that $\det[\tilde{\hat{A}}(1)] \neq 0$. An argument is in Johansen (1995). Since, as noticed above, $\det[\hat{A}(z)] = 0$ has no root inside the unit circle, all the roots of $\det[\tilde{\hat{A}}(z)] = 0$ must be outside the unit circle when T is large enough. Now, since $\det[\tilde{\hat{A}}(z)]$ is a polynomial, $\det[\tilde{\hat{A}}(z)] = 0$ has a finite number of solutions. Since all are outside the unit circle, there must exist a $\delta > 0$ such that $\det[\tilde{\hat{A}}(z)] \neq 0$ when $|z| < 1 + \delta$. Hence, we can argue as in Theorem 11.3.1 in Brockwell and Davis (1991) to verify that $\tilde{\hat{C}}(z) = \tilde{\hat{A}}(z)^{-1}$ satisfy the condition of Lemma 4.1 of Johansen (1995).

Then,

$$\begin{aligned}
\tilde{X}_t^* &= \begin{pmatrix} \hat{\beta}' X_t^* \\ \hat{\beta}_\perp' \Delta X_t^* \end{pmatrix} = \tilde{\hat{A}}(L)^{-1}(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\epsilon_t^* + \hat{\mu}_0 + \hat{\alpha}\hat{\rho}_1 t) \\
&= \tilde{\hat{C}}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\epsilon_t^* + \hat{\mu}_0 + \hat{\alpha}\hat{\rho}_1 t) \qquad (12) \\
&= \tilde{\hat{C}}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\epsilon_t^* + \hat{\mu}_0 + \hat{\alpha}\hat{\rho}_1 t) + \tilde{\hat{C}}^{\#}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\Delta(\epsilon_t^* + \hat{\mu}_0 + \hat{\alpha}\hat{\rho}_1 t).
\end{aligned}$$

We first consider the linear part and start with the decomposition

$$X_t^* = \bar{\hat{\beta}}_\perp \hat{\beta}_\perp' \sum_{i=k+1}^{t} \Delta X_i^* + \bar{\hat{\beta}}\hat{\beta}' X_t^* + \bar{\hat{\beta}}_\perp \hat{\beta}_\perp' X_k^*. \qquad (13)$$

Since $(0, \bar{\hat{\beta}}_\perp)'\tilde{\hat{C}}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)' = \hat{C} = \hat{\beta}_\perp(\hat{\alpha}_\perp' \hat{\Gamma} \hat{\beta}_\perp)^{-1}\hat{\alpha}_\perp'$, the first term in (13) is equal to

$$\begin{aligned}
\bar{\hat{\beta}}_\perp \hat{\beta}_\perp' \sum_{i=k+1}^{t} \Delta X_i^* &= \hat{C} \sum_{i=k+1}^{t} (\epsilon_i^* + \hat{\mu}_0 + \hat{\alpha}\hat{\rho}_1 t) + (t-k)(0, \bar{\hat{\beta}}_\perp)\tilde{\hat{C}}^{\#}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\hat{\alpha}\hat{\rho}_1 \\
&+ (0, \bar{\hat{\beta}}_\perp)\tilde{\hat{C}}^{\#}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\epsilon_t^* - \epsilon_k^*). \qquad (14)
\end{aligned}$$

For $z \neq 1$ and $\tilde{\hat{A}}(z)$ non-singular,

$$\tilde{\hat{C}}^{\#}(z)\tilde{\hat{A}}(z) = \frac{\tilde{\hat{A}}(z)^{-1} - \tilde{\hat{A}}(1)^{-1}}{(1-z)}\tilde{\hat{A}}(z) = \qquad (15)$$

$$\frac{I - \tilde{\hat{A}}(1)^{-1}\tilde{\hat{A}}(z)}{(1-z)} \quad = \quad \frac{-\tilde{\hat{A}}(1)^{-1}[\tilde{\hat{A}}(z) - \tilde{\hat{A}}(1)]}{(1-z)} = \tilde{\hat{A}}(1)^{-1}\tilde{\hat{A}}^{\#}(z).$$

Therefore $\tilde{\hat{C}}^{\#}(1) = \tilde{\hat{A}}(1)^{-1}\tilde{\hat{A}}^{\#}(1)\tilde{\hat{A}}(1)^{-1}$. But $\tilde{\hat{A}}^{\#}(1) = -\frac{d}{dz}\tilde{\hat{A}}(z)|_{z=1} = -\dot{\tilde{\hat{A}}}(1)$. The expression $(0, \bar{\hat{\beta}}_\perp)\tilde{\hat{C}}^{\#}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\hat{\alpha}\hat{\rho}_1$ equals $\hat{C}\hat{\Gamma}\hat{\bar{\beta}}\hat{\rho}_1$, which follows from equation 37 in Rahbek and Mosconi (1999). The contribution to the linear term from (14) is thus $\hat{C}\hat{\mu}_0 + \hat{C}\hat{\Gamma}\hat{\bar{\beta}}\hat{\rho}_1$, and the contribution from $\bar{\hat{\beta}}\hat{\beta}'X_t^*$ is $(\bar{\hat{\beta}}, 0)\tilde{\hat{C}}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\hat{\alpha}\hat{\rho}_1 = \bar{\hat{\beta}}\bar{\hat{\alpha}}'(\hat{\Gamma}\hat{C} - I)\hat{\alpha}\hat{\rho}_1 = -\bar{\hat{\beta}}\hat{\rho}_1$. Summing, this is seen to equal $\hat{\tau}$.

We now have considered the linear part of the lemma and have shown that

$$X_t^* = \hat{C}\sum_{i=k+1}^{t}\epsilon_i^* + \hat{\tau}(t-k) + \sqrt{T}R_{t,T}^*$$

where

$$\begin{aligned}
\sqrt{T}R_{t,T}^* &= (0, \bar{\hat{\beta}}_\perp)\tilde{\hat{C}}^{\#}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\epsilon_t^* - \epsilon_k^*) + (\bar{\hat{\beta}}, 0)\tilde{\hat{C}}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\epsilon_t^* + \hat{\mu}_0) \\
&+ (\bar{\hat{\beta}}, 0)\tilde{\hat{C}}^{\#}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\Delta(\epsilon_t^* + \hat{\mu}_0 + \hat{\alpha}\hat{\rho}_1 t) + \bar{\hat{\beta}}_\perp\hat{\beta}_\perp' X_k^* \\
&= (0, \bar{\hat{\beta}}_\perp)\tilde{\hat{C}}^{\#}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\epsilon_t^* - \epsilon_k^*) + (\bar{\hat{\beta}}, 0)\tilde{\hat{C}}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\epsilon_t^* \\
&+ (\bar{\hat{\beta}}, 0)\tilde{\hat{C}}^{\#}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\Delta\epsilon_t^* \\
&+ (\bar{\hat{\beta}}, 0)\tilde{\hat{C}}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\hat{\mu}_0 + (\bar{\hat{\beta}}, 0)\tilde{\hat{C}}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)\hat{\alpha}\hat{\rho}_1 + \bar{\hat{\beta}}_\perp\hat{\beta}_\perp' X_k^*.
\end{aligned}$$

Next consider the remainder term $R_{t,T}^*$. The last two terms involving $\hat{\mu}_0$ and $\hat{\rho}_1$ in the previous expression are bounded in probability since the estimators are consistent. Since the terms involving ϵ_t^* have the same form as if the recursion in Algorithm 1 contains no deterministic terms, we may therefore in the rest of the proof assume that $\mu_0 = \mu_1 = 0$. Thus

$$\begin{aligned}
\sqrt{T}R_{t,T}^* &= (0, \bar{\hat{\beta}}_\perp)\tilde{\hat{C}}^{\#}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\epsilon_t^* - \epsilon_k^*) + (\bar{\hat{\beta}}, 0)\tilde{\hat{C}}(1)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\epsilon_t^* \\
&+ (\bar{\hat{\beta}}, 0)\tilde{\hat{C}}^{\#}(L)(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\Delta\epsilon_t^* + \bar{\hat{\beta}}_\perp\hat{\beta}_\perp' X_k^*.
\end{aligned}$$

Now,

$$\begin{aligned}
\sqrt{T}R_{t,T}^* &= (0, \bar{\hat{\beta}}_\perp)\tilde{\hat{C}}^{\#}(L)\tilde{\hat{A}}(L)(\tilde{X}_t^* - \tilde{X}_k^*) + \bar{\hat{\beta}}\hat{\beta}'X_t^* + \bar{\hat{\beta}}_\perp\hat{\beta}_\perp' X_k^* \\
&= \bar{\hat{\beta}}\hat{\beta}'X_t^* + (0, \bar{\hat{\beta}}_\perp)\tilde{\hat{A}}(1)^{-1}\tilde{\hat{A}}^{\#}(L)(\tilde{X}_t^* - \tilde{X}_k^*) + \bar{\hat{\beta}}_\perp\hat{\beta}_\perp' X_k^*,
\end{aligned}$$

where the first equality follows from (12) and the second from (15).

Note that $\tilde{\hat{A}}^{\#}(z)$ is a polynomial, where the coefficients are functions of the parameters so that $R_{t,T}$ does only involve a finite number of the generated X_t^* as $T \to \infty$.

Hence, the first part of the lemma follows if, for all $\eta > 0$,

$$P^*(\max_{k+1 \le t \le T} |\tilde{X}_t^*|/\sqrt{T} > \eta) \tag{16}$$

tends to zero in probability. By a well known argument, see e.g. Hall and Heyde (1980, p. 53),

$$\begin{aligned} & P^*(\max_{k+1 \le t \le (T-k)} |\tilde{X}_t^*|/\sqrt{(T-k)} > \eta) \\ = \ & P^*(\frac{1}{T-k} \sum_{t=k+1}^{T} \tilde{X}_t^{*\prime} \tilde{X}_t^* I[|\tilde{X}_t^*| > \eta\sqrt{(T-k)}] > \eta^2) \\ \le \ & \frac{1}{\eta^2} E^*[\frac{1}{T-k} \sum_{t=k+1}^{T} \tilde{X}_t^{*\prime} \tilde{X}_t^* I(|\tilde{X}_t^*| > \eta\sqrt{(T-k)})] \le \frac{1}{\eta^4 (T-k)^2} \sum_{t=k+1}^{T} E^*(\tilde{X}_t^{*\prime} \tilde{X}_t^*)^2, \end{aligned}$$

so it suffices to show that that

$$\frac{1}{(T-k)^2} \sum_{t=k+1}^{T} E^*(\tilde{X}_t^{*\prime} \tilde{X}_t^*)^2 \to 0 \tag{17}$$

in probability.

In equation 4.13 in Johansen (1995) it is shown that the relation between $\hat{A}(z)$ and $\tilde{\hat{A}}(z)$ is given by $(\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)' \hat{A}(z) = \tilde{\hat{A}}(z)(\hat{\beta}, \hat{\beta}_\perp(1-z))'$ when $z \neq 1$. Therefore, $\tilde{\hat{A}}(0) = (\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\hat{\beta}, \hat{\beta}_\perp)^{-1} = (\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'(\bar{\hat{\beta}}, \bar{\hat{\beta}}_\perp)$, so that $\tilde{\hat{A}}(L)\tilde{X}_t^* = (\bar{\hat{\alpha}}, \bar{\hat{\alpha}}_\perp)'\epsilon_t^*$ may alternatively be represented as $B(L)\tilde{X}_t^* = (\hat{\beta}, \hat{\beta}_\perp)'\epsilon_t^*$ where $\hat{B}(z)$ is a polynomial of order k with $\hat{B}(0) = I$.

Introduce $\underline{\tilde{X}}_t^* = (\tilde{X}_t^{*\prime}, \ldots, \tilde{X}_{t-k+1}^{*\prime})'$, $\kappa_t^* = (\hat{\beta}, \hat{\beta}_\perp)'\epsilon_t^*$, $\underline{\kappa}_t^* = (\kappa_t^{*\prime}, 0, \ldots, 0)'$, and the matrix

$$\hat{B} = \begin{pmatrix} \hat{B}_1 & & \cdots & & \hat{B}_k \\ I & 0 & \cdots & & 0 \\ \vdots & & & & \vdots \\ 0 & & \cdots & I & 0 \end{pmatrix}.$$

Then

$$\underline{\tilde{X}}_t^* = \hat{B}\underline{\tilde{X}}_{t-1}^* + \underline{\kappa}_t^*, t = k+1, \ldots, T.$$

Solving the equation backward, we therefore have

$$\underline{\tilde{X}}_t^* = \sum_{j=0}^{t-(k+1)} \hat{B}^j \underline{\kappa}_{t-j}^* + \hat{B}^{(t-k)} \underline{\tilde{X}}_k^* = \underline{V}_{1,t}^* + \underline{V}_{2,t}.$$

Since $E^*(\tilde{X}_t^{*\prime}\tilde{X}_t^*)^2 \leq E^*(\underline{\tilde{X}}_t^{*\prime}\underline{\tilde{X}}_t^*)^2 \leq 4(E^*(\underline{V}_{1,t}^{*\prime}\underline{V}_{1,t}^*)^2 + (\underline{V}_{2,t}^{\prime}\underline{V}_{2,t})^2)$, (17) and hence (16) will follow from

$$\frac{1}{(T-k)^2}\sum_{t=k+1}^{T} E^*(\underline{V}_{l,t}^{*\prime}\underline{V}_{l,t}^*)^2 \to 0, l = 1, 2 \tag{18}$$

in probability. Consider first the case $l = 1$.

The equations $\det[\hat{B}(z)] = 0$ and $\det[\hat{\tilde{A}}(z)] = 0$ have the same solutions, and the eigenvalues of $\hat{B}$ equal the inverse of these solutions. Therefore all the eigenvalues of $\hat{B}$ must be smaller in modulus than 1 when T is large enough, since the coefficients of the polynomial $\hat{\tilde{A}}(z)$ tend to those of $\tilde{A}(z)$, and the solutions of $\det[\tilde{A}(z)] = 0$ have moduli larger than 1 by assumption, see e.g. Johansen (1995, p. 51).

Then

$$E^*(\underline{V}_{1,t}^{*\prime}\underline{V}_{1,t}^*)^2 \leq E^*(\sum_{i,j=0}^{t-(k+1)} \underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^i\underline{\kappa}_{t-i}^*)^2 \tag{19}$$

$$\leq E^*(\sum_{j=0}^{t-(k+1)} \underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^j\underline{\kappa}_{t-j}^*\underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^j\underline{\kappa}_{t-j}^* + 2\sum_{i<j}^{t-(k+1)} \underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^i\underline{\kappa}_{t-i}^*\underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^i\underline{\kappa}_{t-i}^*$$

$$+2\sum_{i<j}^{t-(k+1)} \underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^i\underline{\kappa}_{t-i}^*\underline{\kappa}_{t-i}^{*\prime}\hat{B}^{\prime i}\hat{B}^j\underline{\kappa}_{t-j}^* + 2\sum_{i<j}^{t-(k+1)} \underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^j\underline{\kappa}_{t-j}^*\underline{\kappa}_{t-i}^{*\prime}\hat{B}^{\prime i}\hat{B}^i\underline{\kappa}_{t-i}^*).$$

For a vector $x = (x_1, \ldots, x_n)'$, consider the norm $\|x\|_\infty = \max_{1\leq i\leq n}|x_i|$ and for a $n \times n$ matrix $C = \{c_{ij}\}$ let $\|C\|_\infty$ be the induced norm, which equals $\max_{1\leq i\leq n}\sum_{j=1}^n |c_{ij}|$. Then $\|x\|_\infty^2 \leq x'x \leq n\|x\|_\infty^2$ and if all the eigenvalues of C have modulus less than one, $\|C^m\|_\infty \leq const.|\lambda|^{m/2}$ where λ is an eigenvalue having maximal modulus, see Corollary A.2 in Johansen (1995).

Since in our case $n = pk$,

$$E^*(\underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^j\underline{\kappa}_{t-j}^*\underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^j\underline{\kappa}_{t-j}^*) = E^*[(\underline{\kappa}_{t-j}^{*\prime}\hat{B}^{\prime j}\hat{B}^j\underline{\kappa}_{t-j}^*)^2] \leq (pk)^2E^*(\|\hat{B}^j\underline{\kappa}_{t-j}^*\|_\infty^4) \leq$$

$$(pk)^2\|\hat{B}\|_\infty^{4j}E^*(\|\underline{\kappa}_{t-j}^*\|_\infty^4) \leq (pk)^2\|\hat{B}\|_\infty^{4j}E^*[(\underline{\kappa}_{t-j}^{*\prime}\underline{\kappa}_{t-j}^*)^2] = (pk)^2\|\hat{B}\|_\infty^{4j}E^*(|\underline{\kappa}_{t-j}^*|^4).$$

But, $\|\hat{B}\|_\infty^{4j}$ is less than the product of the eigenvalue of $\hat{B}$ having the largest modulus to the power $2j$ and a constant, and this eigenvalue value converges in probability to a number with modulus less than 1. The existence of the fourth moments of the elements of ϵ_t will assure that $E^*(|\underline{\kappa}_t^*|^4)$ converges and therefore that (18) holds. A similar argument works for the case $l = 2$, since the eigenvalues of $\hat{B}$ have moduli that are less than 1.

The last statement in the Lemma follows by the consistency of the estimators since $E^*[\epsilon_t^*\epsilon_t^{*\prime}] = \frac{1}{T-k}\sum_{i=T-k+1}^{T}\hat{\epsilon}_i\hat{\epsilon}_i^{\prime}$.

Proof of Lemma 2. We only consider the first part. The others are proved in a similar manner. Note that $S_{00}^* = M_{00}^* - M_{02}^* M_{22}^{*-1} M_{20}^*$, where

$$M^* = \begin{pmatrix} M_{00}^* & M_{02}^* \\ M_{20}^* & M_{22}^* \end{pmatrix}.$$

is the $(pk+1)\times(pk+1)$ matrix $\sum_{t=k+1}^{T} Z_t^* Z_t^{*\prime}/(T-k)$ and $Z_t^* = (\Delta X_t^{*\prime}, \ldots, \Delta X_{t-k+1}^{*\prime}, 1)'$, $t = k+1, \ldots, T$. Let M be defined similarly in terms of the original observations, where $Z_t = (\Delta X_t', \ldots, \Delta X_{t-k+1}', 1)', t = k+1, \ldots, T$. By the ergodic theorem $M \to \Sigma_M$, say, in probability, as $T \to \infty$. The lemma will follow if, for all $\eta > 0$,

$$P^*(\|M^* - \Sigma_M\| > \eta) \to 0 \tag{20}$$

in probability.

We first note that it is sufficient to consider a situation without a constant. Also, since $\|M^* - \Sigma_M\|^2$ is equal to the sum of squares of the elements of $M^* - \Sigma_M$, it is sufficient to prove that (20) is valid for each element.

We now verify that the variables Z_t and Z_t^* have a moving average representation.

As pointed out by Hansen and Johansen (1999, p. 311), the stochastic vector $(X_t'\beta, \Delta X_{t-1}', \cdots, \Delta X_{t-k+1}')'$ can be represented as an $AR(1)$ process using the matrix

$$\begin{pmatrix} \beta'\alpha + I & \beta'\Gamma_1 & \cdots & & \beta'\Gamma_{k-1} \\ \alpha & \Gamma_1 & \cdots & & \Gamma_{k-1} \\ 0 & I & & & \\ \vdots & & & & \vdots \\ 0 & & \cdots & I & 0 \end{pmatrix}. \tag{21}$$

By expanding the state space also $(X_t'\beta, \Delta X_{t-1}', \cdots, \Delta X_{t-k}')'$ has such a representation. The process is stationary. Here we need the assumption that $\alpha_\perp'\Gamma\beta_\perp$ has full rank $p - r$. Furthermore, $(X_t'\beta, \Delta X_{t-1}', \cdots, \Delta X_{t-k}')'$ can be represented in terms of the errors $\epsilon_t, \epsilon_{t-1}, \ldots$. To see that we note that the term Y_t in (7) corresponds to $\bar{\beta}\Delta Z_t$ in equation (4.7) in the proof of Theorem 4.2 in Johansen (1995). Furthermore, from equation (4.16) in the same proof Z_t, and therefore Y_t, is part of an autoregressive process which can be expressed by the errors $\epsilon_t, \epsilon_{t-1}, \ldots$. This implies that all terms, apart from A, in (7) can be represented in this way. But A cancels in $\Delta X_{t-j}, j = 1, \ldots, k$, and since $\beta' A = 0$, all terms in $(X_t'\beta, \Delta X_{t-1}', \cdots, \Delta X_{t-k}')'$ can be represented in terms of the errors $\epsilon_t, \epsilon_{t-1}, \ldots$. The process is therefore causal, and has a characteristic polynomial with determinant having zeros outside the unit circle see e.g. Theorem 3.1.1 and 11.3.1 in Brockwell and Davis (1991). Since the eigenvalues of the matrix in the $AR(1)$ representation are the inverses of these roots, $(X_t'\beta, \Delta X_{t-1}', \cdots, \Delta X_{t-k}')'$ has a moving average representation of the form $(X_t'\beta, \Delta X_{t-1}', \cdots, \Delta X_{t-k}')' = \sum_{i=0}^{\infty} \zeta_i \eta_{t-i}$

where $\eta_t = (\epsilon_t'\beta, \epsilon_t', 0, \cdots, 0)'$, $\zeta_0 = I$ and $\zeta_1, \cdots$ are matrices such that the maximal eigenvalue of ζ_i is bounded by λ^i for some $0 < \lambda < 1$, uniformly in $i = 1, \ldots$. Thus, $Z_t = \sum_{i=0}^{\infty} G\zeta_i\eta_{t-i}$ for a suitable matrix G and each element of Z_t, which we also denote by Z_t, dropping the index, has a moving average representation

$$Z_t = \sum_{i=0}^{\infty} \xi_i'\epsilon_{t-i} \tag{22}$$

where $|\xi_i| < const.\lambda^i$ uniformly in $i = 1, \ldots$. Similarly,

$$Z_t^* = \sum_{i=0}^{\infty} \hat{\xi}_i{}' \epsilon_{t-i}^* \tag{23}$$

where $|\hat{\xi}_i| < const.\lambda^i$ uniformly in $i = 1, \ldots$ when T is large enough.

Now, $E(M) = E(Z_t^2) = \sum_{i=0}^{\infty} \xi_i'\Omega\xi_i$ and $E^*(M^*) = E^*(Z_t^{*2}) = \sum_{i=0}^{\infty} \hat{\xi}_i'\Omega_T\hat{\xi}_i$ so that $E^*(M^*) \to E(M)$ in probability. Since $\|M^* - \Sigma_M\| \leq \|M^* - E^*(M^*)\| + \|E^*(M^*) - \Sigma_M\|$, (20) will follow from

$$P^*(\|M^* - E^*(M^*)\| > \eta) \to 0$$

in probability. Hence, by Chebychev's inequality it suffices to show that

$$\frac{\mathrm{Var}^*(\sum_{t=k+1}^{T} Z_i^{*2})}{(T-k)^2} = \frac{\sum_{t=-T+k-1}^{T-k+1}(1-\frac{j}{T})\mathrm{Cov}^*(Z_0^{*2}, Z_t^{*2})}{(T-k)} \to 0 \tag{24}$$

in probability as $T \to \infty$.

But,

$$\begin{aligned} \mathrm{Cov}^*(Z_0^{*2}, Z_t^{*2}) &\leq Var^*(Z_0^{*2}) \leq E^*(Z_0^{*4}) \\ &= E^*[(\sum_{i=0}^{\infty}(\hat{\xi}_i'\epsilon_{-i}^*)^4 + 2\sum_{0\leq i<j}(\hat{\xi}_i'\epsilon_{-i}^*)^2(\hat{\xi}_j'\epsilon_{-j}^*)^2] \end{aligned}$$

Now, we use that $(\hat{\xi}_i'\epsilon_{-i}^*)^2 \leq |\hat{\xi}_i|^2|\epsilon_{-i}^*|^2$, that $|\hat{\xi}_i| < const.\lambda^i, i = 1, \ldots$ when T is large enough, that $E^*(\|\epsilon_t^*\|)^4 \to E^*(\|\epsilon_t\|)^4 < \infty$ in probability by the weak law of large numbers and the existence of the fourth moments of the elements of ϵ_t. Then (24) will follow.

Proof of Lemma 3. The proof relies on the continuous mapping theorem in the same way as the proof in Lemma 2 in Johansen (1994). The same functionals are involved so it suffice to prove that the process $\{X_{[uT]}^*/\sqrt{T} : 0 \leq u \leq 1\}$ converges weakly in $P_{r,T}$ probability as element in $D[0,1]^p$.

By the the results from Lemma 1 it follows that the remainder term R_T^* vanishes, and it is sufficient to consider the linear part and $S_T^*(u) = \sum_{i=k+1}^{[uT]} \epsilon_i^*/\sqrt{T}$. The linear term is treated as in Lemma 2 in Johansen (1994).

To prove that $S_T^* \to W$ weakly in probability, it is convenient to follow the approach of Pollard (1984), and exploit that the limit is continuous. Hence, one can work with the uniform norm in $D[0,1]^p$ and show that

$$E^*[f(S_T^*)] \to E[f(W)]$$

in $P_{r,T}$ probability for all bounded continuous functions f. This is explained in more detail for the one dimensional case in Swensen (2003a).

Proof of Lemma 4. This is proved in the same way as Lemma 1 using the result stated in Lemma 6.

Proof of Lemma 5. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq \hat{\lambda}_{p+1} = 0$ be the ordered eigenvalues of

$$\det[\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}] = 0,$$

and $\hat{v}_1, \ldots, \hat{v}_{p+1}$ the corresponding eigenvectors such that $\hat{V}' S_{11} \hat{V} = I$ where $\hat{V}$ is the matrix with columns $\hat{v}_1, \ldots, \hat{v}_{p+1}$. Then

$$S_{11} \hat{V} \hat{\Lambda}_{p+1} = S_{10} S_{00}^{-1} S_{01} \hat{V} \tag{25}$$

where $\hat{\Lambda}_{p+1}$ is a diagonal matrix with diagonal elements $\hat{\lambda}_1, \ldots, \hat{\lambda}_{p+1}$. Define the $(p+1) \times p$ matrices $\hat{\Lambda}^- = (\hat{\Lambda}_p^{-1}, 0)'$ and $\hat{\Lambda}^{--1/2} = (\hat{\Lambda}_p^{-1/2}, 0)'$. Using (25) the p first columns of V may be written

$$S_{11}^{-1} S_{10} S_{00}^{-1} S_{01} \hat{V} \hat{\Lambda}^- = S_{11}^{-1} S_{10} S_{00}^{-1} S_{01} \hat{V} \hat{\Lambda}^{--1/2} \hat{\Lambda}_p^{-1/2} = S_{11}^{-1} S_{10} \hat{W} \hat{\Lambda}_p^{-1/2} \tag{26}$$

where $\hat{W} = S_{00}^{-1} S_{01} \hat{V} \hat{\Lambda}^{--1/2}$. But from

$$S_{00} \hat{W} \hat{\Lambda}_p = S_{01} \hat{V} \hat{\Lambda}^{--1/2} \hat{\Lambda}_p$$

and

$$\begin{aligned} S_{01} S_{11}^{-1} S_{10} \hat{W} &= S_{01} S_{11}^{-1} S_{10} S_{00}^{-1} S_{01} \hat{V} \hat{\Lambda}^{--1/2} \\ &= S_{01} S_{11}^{-1} S_{11} \hat{V} \hat{\Lambda}_{p+1} \hat{\Lambda}^{--1/2} \\ &= S_{01} \hat{V} \hat{\Lambda}_{p+1} \hat{\Lambda}^{--1/2}, \end{aligned}$$

it follows that

$$S_{00} \hat{W} \hat{\Lambda}_p = S_{01} S_{11}^{-1} S_{10} \hat{W}. \tag{27}$$

Also using the definition of $\hat{W}$ and (25), $\hat{W}' S_{00} \hat{W} = \hat{\Lambda}^{--1/2\prime} \hat{V}' S_{10} S_{00}^{-1} S_{00} S_{00}^{-1} S_{01} \hat{V} \hat{\Lambda}^{--1/2} = \hat{\Lambda}^{--1/2\prime} \hat{V}' S_{11} \hat{V} \hat{\Lambda}_{p+1} \hat{\Lambda}^{--1/2} = I$.

Since $\hat{\beta}_j^{\#}$ is defined as the first j columns of $\hat{V}$, it follows from (26) that it may alternatively be expressed as $\hat{\beta}_j^{\#} = S_{11}^{-1} S_{10} \hat{W}_j \hat{\Lambda}_j^{-1/2}$, where $\hat{W}_j$ is the $p \times j$

matrix consisting of the j first columns of $\hat{W}$, and $\hat{\Lambda}_j$ is the $j \times j$ diagonal matrix with diagonal elements $\hat{\lambda}_1, \cdots, \hat{\lambda}_j$. Then $\hat{\alpha}_j = S_{01}\hat{\beta}_j^{\#}(\hat{\beta}_j^{\#\prime} S_{11}\hat{\beta}_j^{\#})^{-1}$ and

$$\hat{\Pi}_j^{\#} = \hat{\alpha}_j \hat{\beta}_j^{\#\prime} = S_{01}S_{11}^{-1}S_{10}\hat{W}_j(\hat{W}_j' S_{01}S_{11}^{-1}S_{10}\hat{W}_j)^{-1}\hat{W}_j' S_{01}S_{11}^{-1}. \tag{28}$$

Equation (27) shows that $\hat{\lambda}_1, \cdots, \hat{\lambda}_p$ and $\hat{W}$ are defined from the equation

$$\det[\lambda S_{00} - S_{01}S_{11}^{-1}S_{10}] = 0$$

and the normalization $\hat{W}' S_{00}\hat{W} = I$. Note that when $T \to \infty$, $S_{00} \to \Sigma_{00}$ in probability and that the OLS estimator $S_{01}S_{11}^{-1}$ of $\Pi^{\#}$ is consistent. The conclusion of the Lemma will follow from (28) if

$$S_{01}S_{11}^{-1}S_{10} \to \Sigma_{00} - \Omega \text{ in } P_{r,T} \text{ probability as } T \to \infty. \tag{29}$$

But $S_{00} - S_{01}S_{11}^{-1}S_{10} = \frac{1}{T-k}\sum_{t=k+1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$ where $\hat{\epsilon}_t, t = k+1, \ldots, T$ are the OLS residuals and by the consistency of the OLS estimators $\frac{1}{T-k}\sum_{t=k+1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t' \to \Omega$ in probability as $T \to \infty$.

In fact, a bit more can be shown. Since $\hat{\beta}_j^{\#} = S_{11}^{-1}S_{10}\hat{W}_j\hat{\Lambda}_j^{-1/2}$ and $\hat{\alpha}_j = S_{01}\hat{\beta}_j^{\#}(\hat{\beta}_j^{\#\prime} S_{11}\hat{\beta}_j^{\#})^{-1}$, it follows from the previous results that $\hat{\beta}_j^{\#} \to \Pi^{\#\prime} W_j \Lambda_j^{-1/2}$ and $\hat{\alpha}_j \to (\Sigma_{00} - \Omega)W_j[W_j'(\Sigma_{00} - \Omega)W_j']^{-1}\Lambda_j^{1/2}$ in $P_{r,T}$ probability.

Proof of Lemma 6. Since we consider the zeroes of the determinants of characteristic polynomials, we may assume that $\mu_1 = \mu_0 = 0$.

Write $A_j(z) = (1-z)I_p - \Pi_j z - \Gamma_1(1-z)z - \cdots - \Gamma_{k-1}(1-z)z^{k-1}$ as $A_j(z) = I - A_{1,j}z - A_2 z^2 - \cdots - A_k z^k$ where $A_{1,j} = I + \Pi_j + \Gamma_1$, $A_i = \Gamma_i - \Gamma_{i-1}, i = 2, \cdots, k-1$ and $A_k = -\Gamma_{k-1}$.

For each j the solutions to the equation $\det[A_j(z)] = 0$ are equal to the inverses of the eigenvalues of the matrix

$$\mathcal{A}_j = \begin{pmatrix} A_{1,j} & A_2 & \cdots & & A_k \\ I & 0 & \cdots & & 0 \\ \vdots & & & & \vdots \\ 0 & & \cdots & I & 0 \end{pmatrix}.$$

Now, suppose that $a = (a_1', \ldots, a_k')'$ is an eigenvector corresponding to an eigenvalue equal to 1. Then

$$\begin{aligned} A_{1,j}a_1 + A_2 a_2 \cdots + A_k a_k &= a_1, \\ a_i &= a_{i-1}, i = 2, \ldots, k \end{aligned}$$

so that $A_j(1)a_1 = a_1 - A_{1,j}a_1 - A_2 a_1 \cdots - A_k a_1 = 0$. But $A_j(1) = -\Pi_j$, which has rank j. The space of solutions of $A_j(1)a_1 = 0$ has therefore dimension $p - j$

and the equation $\det[A_j(z)] = 0$ has $pk - (p-j) = p(k-1) + j$ solutions which are different from 1. We will show that they all have modulus larger than 1.

Let $\Pi_j = \alpha_j \beta_j'$ for some $p \times j$ matrices α_j and β_j of rank j. As in the proof of Lemma 2, consider the stochastic process $(X_t'\beta_j, \Delta X_{t-1}', \cdots, \Delta X_{t-k+1}')'$, which can be represented as an $AR(1)$ process using the matrix

$$\begin{pmatrix} \beta_j'\alpha_j + I & \beta_j'\Gamma_1 & \cdots & & \beta_j'\Gamma_{k-1} \\ \alpha & \Gamma_1 & \cdots & & \Gamma_{k-1} \\ 0 & I & & & \\ \vdots & & & & \vdots \\ 0 & & \cdots & I & 0 \end{pmatrix}. \tag{30}$$

The $AR(1)$ processes in question are stationary for all j since columns of β_j belong to the cointegration space $sp(\beta)$. By the same argument as in the proof of Lemma 2 it follows that the processes are causal, and that the eigenvalues of the matrix in (30) must be nonzero and have moduli less than 1. Let λ be such an eigenvalue and $b = (b_1', \ldots, b_k')'$ the corresponding eigenvector.

Then

$$\begin{aligned} (\lambda - 1)b_1 - \beta_j'\alpha_j b_1 - \beta_j'\Gamma_1 b_2 - \cdots - \beta_j'\Gamma_{k-1}b_k &= 0, \\ -\alpha b_1 + (\lambda I - \Gamma_1)b_2 - \Gamma_2 b_3 - \cdots - \Gamma_{k-1}b_k &= 0, \\ \lambda b_i - b_{i-1} &= 0, i = 3, \ldots, k. \end{aligned}$$

This means that $b_i = \lambda^{-i+2}b_2, i = 3, \ldots, k$ and the two first equations above may be written

$$(\lambda - 1)b_1 - \beta_j'\alpha_j b_1 - \beta_j'\Gamma_1 b_2 - \cdots - \beta_j'\Gamma_{k-1}b_2\lambda^{-k+2} = 0 \tag{31}$$

and

$$-\alpha_j b_1 + (\lambda I - \Gamma_1)b_2 - \Gamma_2 b_2 \lambda^{-1} - \cdots - \Gamma_{k-1}b_2\lambda^{-k+2} = 0. \tag{32}$$

Rearranging (32), pre-multiplying by β_j' and inserting (31), yield

$$\lambda\beta_j' b_2 = (\lambda - 1)b_1.$$

Therefore, if $\lambda \neq 1$, (32) may be written

$$-(\lambda/(\lambda-1))\alpha_j\beta_j' b_2 + (\lambda I - \Gamma_1)b_2 - \Gamma_2 b_2\lambda^{-1} - \cdots - \Gamma_{k-1}b_2\lambda^{-k+2} = 0,$$

which after multiplying with $\lambda - 1$ and rearranging becomes

$$-\lambda\alpha_j\beta_j' b_2 + (\lambda - 1)\lambda b_2 - (\lambda - 1)[\Gamma_1 + \Gamma_2\lambda^{-1} + \cdots + \Gamma_{k-1}\lambda^{-k+2}]b_2 = 0.$$

This equals

$$\lambda^2 b_2 - \lambda(\alpha_j\beta_j' + I + \Gamma_1)b_2 + (\Gamma_1 - \Gamma_2)b_2 + \cdots + \lambda^{-k+3}(\Gamma_{k-2} - \Gamma_{k-1})b_2 + \lambda^{-k+2}\Gamma_{k-1}b_2 = 0. \tag{33}$$

Remark, however that $A_j(z)$ can also be expressed as

$$\begin{aligned} A_j(z) &= (1-z)I - \Pi_j z - \sum_{i=1}^{k-1}\Gamma_i(1-z)z^i \\ &= I - (I + \Pi_j + \Gamma_1)z - (\Gamma_2 - \Gamma_1)z^2 - \cdots - (\Gamma_{k-1} - \Gamma_{k-2})z^{k-1} + \Gamma_{k-1}z^k \end{aligned}$$

which has the same form as (33). Hence, for eigenvalues of the matrix (30) not equal to 0 or 1, $\lambda^2 A_j(1/\lambda)b_2 = 0$, so that $1/\lambda$ is a solution of $\det[A_j(z)] = 0$. But as already pointed out these eigenvalues all have modulus less than 1. Also there are $p \times (k-1) + j$ such eigenvalues. Hence, we have demonstrated that all solutions of $\det[A_j(z)] = 0$ which are not equal to 1, have modulus larger than 1.

That $\alpha_{j\perp}\Gamma\beta_{j\perp}$ is nonsingular for each $j = 0, \ldots, r-1$ follows from the necessary condition for stationarity in Theorem 4.2 in Johansen (1995). The argument is as follows. We start out with an $I(1)$ process which can be represented as described in Theorem 4.2 in Johansen (1995) using the reduced rank matrix Π. Then we consider new processes $X_{j,t}$ with reduced rank matrices $\Pi_j = \alpha_j\beta_j'$. They are not necessarily $I(1)$. But, due to the way there are generated, $(\beta_j' X_{j,t}, \Delta X_{j,t-1}, \ldots, \Delta X_{j,t-k+1})$ may be represented by an $AR(1)$ process arguing as for (21). To conclude that this representation corresponds to a stationary process we note that the matrix (21) in this case is exactly the matrix (30), which, as we have already argued, corresponds to a stationary process. Thus, $\beta_j' X_{j,t}$, $j = 0, \ldots, r-1$ are stationary processes and hence $\alpha_{j\perp}\Gamma\beta_{j\perp}$ is nonsingular for each $j = 0, \ldots, r-1$.

Proof of Lemma 7. This follows by results analogous to those of Lemma 2 and 3, using the norming matrix $\hat{C}_{j,T} = \begin{pmatrix} \bar{\hat{\beta}}_{j\perp} & 0 \\ -\hat{\tau}_j'\bar{\hat{\beta}}_{j\perp} & T^{-1/2} \end{pmatrix}$, and arguing as in the proof of Proposition 1.

7 Conclusion

In this paper we have considered a recursive bootstrap algorithm for testing and determining the rank in an reduced rank VAR-model. Assuming that the data generating process is I(1) and that the cointegration rank is r, we have demonstrated that the bootstrap version of the likelihood ratio test for the hypothesis that the rank is r versus the alternative that there is no reduced rank, has the usual asymptotic distribution. Furthermore, the bootstrapped version of the procedure to determine or estimate the rank has beeen proved to be asymptotically consistent in the sense that the probability of selecting a rank less than the true rank tends to zero, and the asymptotically probability of selecting a rank larger than the true rank can be bounded. Applying the bootstrap methods to the data on demand for money in Finland yields sensible results.

Acknowledgements

A very preliminary version of the paper was presented at the first conference of the European Science Foundation network: Econometric Methods for the Modelling of Nonstationary Data, Policy Analysis and Forecasting (EMM) in Arona, Italy, 19-21 September 2002. Helpful comments from the participants and partial financial support from the EMM network are gratefully acknowledged.

References

[1] Basawa, I.V., A.K. Mallik, W.P. McCormic, J.H. Reeves, and R.L. Taylor, 1991a, Bootstrapping unstable first-order autoregressive processes, Annals of Statistics 19, 1098-1101.

[2] Basawa, I.V., A.K. Mallik, W.P. McCormic, J.H. Reeves, and R.L. Taylor, 1991b, Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes, Communications in Statistics A 20, 1015-1026.

[3] Brockwell, P.J. and R. A. Davis, 1991, Time series: theory and methods (Springer, New York).

[4] Chang, Y. and J.Y. Park, 2003, A sieve bootstrap for the test of a unit root, Journal of Time Series Analysis, 24, 379-400.

[5] Davidson, J., 2002, A model of fractional cointegration , and tests for cointegration using the bootstrap, Journal of Econometrics 110, 187-212.

[6] Davidson, J., 2003, Alternative bootstrap procedures for testing cointegration in fractionally integrated processes, Journal of Econometrics forthcoming.

[7] Davidson, R. and J. MacKinnon, 2000, Bootstrap tests: How many bootstraps? Econometric Reviews 19, 55-69. 110, 187-212.

[8] Doornik, J.A., Hendry D.F. and B. Nielsen, 1998, Inference in cointegrated models: UK M1 revisited, Journal of Economic Surveys 12, 533-572.

[9] Fachin, S., 2000, Bootstrap and asymptotic tests of long-run relationships in cointegrated systems, Oxford Bulletin of Economics and Statistics 62, 543-551.

[10] Ferretti, N. and J. Romo, 1996, Unit root bootstrap tests for AR(1) models, Biometrika 83, 849-860.

[11] Gredenhoff, M. and T. Jacobson, 2001, Bootstrap testing linear restrictions on cointegrating vectors, Journal of Business & Economic Statistics 19, 63-72.

[12] Hall, P. and C.C. Heyde, 1980, Martingale limit theory and its applications (Academic Press, New York).

[13] Hansen, H. and S. Johansen, 1999, Some tests for parameter constancy in cointegrated VAR-models, Econometrics Journal 2, 306-333.

[14] Harris, R.I.D. and G. Judge, 1998, Small sample testing for cointegration using the bootstrap approach, Economics Letters 58, 31-37.

[15] Heimann, G. and J-P. Kreiss, 1996, Bootstrapping general first order autoregressions, Statistics and Probability Letters 30, 87-98.

[16] Johansen, S., 1992, Determination of cointegration rank in the presence of a linear trend, Oxford Bulletin of Economics and Statistics 54, 383-397.

[17] Johansen, S., 1994, The role of the constant and linear terms in cointegration analysis of nonstationary variables, Econometric Reviews 13, 205-229.

[18] Johansen, S., 1995, Likelihood-based inference in cointegrated vector autoregressive models (Oxford University Press, Oxford).

[19] Johansen, S. and K.Juselius, 1990, Maximum likelihood estimation and inference on cointegration - with application to the demand for money, Oxford Bulletin of Economics and Statistics 52, 169-210.

[20] Li, H. and G.S. Maddala, 1997, Bootstrapping cointegrating regressions, Journal of Econometrics 80, 297-318.

[21] Mantalos, P. and G. Shukur, 1998, Size and power of the error correction model cointegration test. A bootstrap approach, Oxford Bulletin of Economics and Statistics 60, 249-255.

[22] Mantalos, P. and G. Shukur, 2001, Bootstrapped Johansen tests for cointegartion relationships: a graphical analysis. Journal of Statistical Computation and Simulation 68, 351-371.

[23] Nielsen, B. and A. Rahbek, 2000, Similarity issues in cointegration analysis, Oxford Bulletin of Economics and Statistics 62, 5-22.

[24] Osterwald-Lenum, M., 1992, A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics, Oxford Bulletin of Economics and Statistics 54, 461-471.

[25] Paparoditis, E and D.N. Politis, 2001, The continuous-path block bootstrap, in: Madan Puri, ed., Asymptotics in statistics and probability, (VSP publications, Zeist (The Netherlands)).

[26] Paparoditis, E and D.N. Politis, 2003, Residual-based block bootstrap for unit root testing, Econometrica 71, 813-855. .

[27] Park, J.Y., 2002, An invariance principle for sieve bootstrap in time series, Econometric Theory 18, 469-490.

[28] Park, J.Y., 2003, Bootstrap unit root tests, Forthcoming Econometrica.

[29] Pollard, D. 1984, Convergence of stochastic processes (Springer, New York).

[30] Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, 1992, Numerical recipes in Fortran (2nd edn.) (Cambridge University Press,Cambridge).

[31] Psaradakis, Z., 2001, Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors, Journal of Time Series Analysis 22, 577-594.

[32] Rahbek, A. and R. Mosconi, 1999, Cointegration rank inference with stationary regressors in VAR models, Econometrics Journal 2, 76-91.

[33] Sims, C.A., J.H. Stock and M.W. Watson, 1990, Inference in linear time series models with some unit roots, Econometrica 58, 113-144.

[34] Swensen, A.R., 2003a, Bootstrapping unit root tests for integrated processes, Journal of Time Series Analysis 24, 99-126.

[35] Swensen, A. R., 2003b, A note on the power of bootstrap unit root tests, Econometric Theory 19, 32-48.

[36] van Giersbergen, N.P.A., 1996, Bootstrapping the trace statistics in VAR models: Monte Carlo results and applications, Oxford Bulletin of Economics and Statistics 58, 391-408.