November 27, 1997

Order Selection in Testing for the Cointegrating Rank of a VAR Process*

by

Helmut Lütkepohl

Institut für Statistik und Ökonometrie
Wirtschaftswissenschaftliche Fakultät
Humboldt–University
Spandauer Str. 1
10178 Berlin
GERMANY
Tel.: +49-30-2093-5718
Fax.: +49-30-2093-5712

and

Pentti Saikkonen

Department of Statistics
University of Helsinki
P.O. Box 54
SF-00014 University of Helsinki
FINLAND
+358-9-1918867
+358-9-1918872

Abstract

The impact of the choice of the lag length on tests for the number of cointegration relations in a vector autoregressive (VAR) process is investigated. It is shown that the asymptotic distribution of likelihood ratio (LR) tests for the cointegrating rank remains unchanged if the true data generation process (DGP) is of finite order and a consistent model selection criterion is used for choosing the lag length. A similar result also holds if the true DGP is an infinite order VAR. In a simulation study we find that small sample power and size of LR cointegration tests strongly depend on the choice of the lag order.

*We thank Christian Müller and Kirstin Hubrich for helping with the computations. The Deutsche Forschungsgemeinschaft, SFB 373, provided financial support. Part of this research was done while the second author was visiting the Institute of Statistics and Econometrics at the Humboldt University in Berlin.

1 Introduction

Following the invention of cointegration by Granger (1981, 1986) and Engle & Granger (1987), time series econometrics has changed considerably. In multiple time series analysis investigating the cointegration properties at an early stage of the analysis has become standard practice by now. For this purpose, Johansen's (1988, 1991, 1995) likelihood ratio (LR) tests are used frequently (see also Reinsel & Ahn (1992)). In that approach a vector autoregressive (VAR) process or error correction model (ECM) of some finite order is usually fitted to the data and the tests are then performed conditionally on the order being the true one. In some studies it was found, however, that the choice of the lag order or truncation lag can have an important impact on the outcome of unit root and cointegration tests (see, e.g., Schwert (1989), Ng & Perron (1995), Agiagloglou & Newbold (1996) for unit root tests in univariate time series and Reimers (1992) and Haug (1996) for cointegration tests in vector processes). Therefore it is of interest to investigate the relation between the choice of trancation lag and the properties of the resulting cointegration tests based on a model with a prespecified order.

In practice, the VAR or ECM order is ususally chosen by some criterion based on the available data. In particular, the order is chosen so that the residuals appear to be white noise under some data dependent criterion. The cointegration tests are then performed conditionally on the order obtained in this way. Therefore, a proper overall assessment of the properties of cointegration tests has to take into account the data dependent choice of the order or truncation lag of the underlying model. Also assuming that the true data generation process (DGP) is in fact of finite order may be too limited for capturing all situations of relevance for applied work. This has been acknowledged in the univariate case for unit root tests by Hall (1994), Said & Dickey (1984) and Ng & Perron (1995) who investigate the asymptotic properties of augmented Dickey-Fuller unit root tests for a number of different rules for choosing the truncation lag. They find that the tests maintain their asymptotic properties for quite general DGPs if the truncation lag is chosen by a suitable deterministic rule or by one of the standard lag order selection criteria. For a deterministic rule a similar result was obtained by Saikkonen & Luukkonen (1997) (henceforth SL) for LR cointegration tests.

In this study we will extend these results to data dependent rules for choosing the trun-

cation lag in multivariate models. In particular, it will be shown that if the DGP is in fact a finite order VAR process then any one of the consistent model selection criteria may be used prior to testing for cointegration. The asymptotic properties of the LR tests for the cointegrating rank will in that case be the same as if the true order were known. Furthermore, if the true DGP is an infinite order process similar results are shown to hold. We also report some simulation results to illustrate the small sample problems related to choosing the truncation lag prior to testing for cointegration.

The structure of this study is as follows. In the next section the standard LR approach for testing for the cointegrating rank is presented formally. The model assumptions used for our purposes are presented in Section 3. In Section 4 results for choosing the VAR order in some deterministic fashion are summarized and in Section 5 the consequences of a data dependent order choice are explored. In Sections 2 - 4 we operate under the unrealistic assumption that the DGP has no deterministic terms. This is done for convenience in order to simplify the exposition. In Section 6 the extension to the case where the DGP has a nonzero mean term is discussed. Simulation results are reported in Section 7 and Section 8 concludes. Most proofs of our theoretical results are given in the Appendix.

The following notation is used throughout. The vector $y_t = (y_{1t}, \ldots, y_{nt})'$ denotes an observable n-dimensional set of time series variables. The sample size is signified by T, the symbol K is reserved for the lag order or truncation lag of an ECM and $N = T - K - 1$ is the effective sample size used for estimation and testing. The differencing operator is denoted by Δ, that is, $\Delta y_t = y_t - y_{t-1}$. The symbol $I(d)$ is used to denote a process which is integrated of order d, that is, it is stationary (or asymptotically stationary) after differencing d times while it is still nonstationary after differencing just $d-1$ times. The symbol $\xrightarrow{p}$ signifies convergence in probability and $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for the order of convergence and convergence in probability, respectively, of a sequence. We abbreviate 'independently, identically distributed' in the usual way by i.i.d.. The normal distribution with mean (vector) μ and variance (covariance matrix) Σ is denoted by $N(\mu, \Sigma)$. Moreover, I_n denotes the $(n \times n)$ identity matrix. If A is an $(n \times m)$ matrix we let $A_\perp$ stand for its orthogonal complement. As a general convention, a sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

2 Cointegration Tests

Given a system of n variables $y_t = (y_{1t}, \ldots, y_{nt})'$, the number of linearly independent cointegrating relations among them is usually determined by considering the rank of the matrix Π in the error correction form

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{K} \Gamma_j \Delta y_{t-j} + e_t. \tag{2.1}$$

This is usually done by testing either one of the following two pairs of hypotheses:

$$H_0(r_0) : \mathrm{rk}(\Pi) = r_0 \qquad \text{vs.} \qquad H_1(r_0) : \mathrm{rk}(\Pi) > r_0 \tag{2.2}$$

or

$$H_0(r_0) : \mathrm{rk}(\Pi) = r_0 \qquad \text{vs.} \qquad \bar{H}_1(r_0) : \mathrm{rk}(\Pi) = r_0 + 1. \tag{2.3}$$

Assuming that the error term e_t in (2.1) is Gaussian white noise, the corresponding likelihood ratio statistics as derived by Johansen (1988) may be obtained as follows. For a sample $y_1, \ldots, y_T$, define $z_t' = (\Delta y_{t-1}', \ldots, \Delta y_{t-K}')$ and, using $N = T - K - 1$,

$$M_T = N^{-1} \left[\sum_{t=K+2}^{T} y_{t-1} y_{t-1}' - \sum_{t=K+2}^{T} y_{t-1} z_t' \left(\sum_{t=K+2}^{T} z_t z_t' \right)^{-1} \sum_{t=K+2}^{T} z_t y_{t-1}' \right]. \tag{2.4}$$

Denoting the least squares (LS) residuals from model (2.1) by $\tilde{e}_t$, define

$$\tilde{\Sigma} = N^{-1} \sum_{t=K+2}^{T} \tilde{e}_t \tilde{e}_t'. \tag{2.5}$$

Moreover, let $\tilde{\Pi}$ be the LS estimator of Π from (2.1). Denoting by $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ the ordered generalized eigenvalues obtained as solutions of

$$\det(\tilde{\Pi} M_T \tilde{\Pi}' - \lambda \tilde{\Sigma}) = 0, \tag{2.6}$$

Johansen's trace statistic for testing the pair of hypotheses (2.2) is given by

$$LR_{trace}(r_0) = N \sum_{j=r_0+1}^{n} \log(1 + \tilde{\lambda}_j) \tag{2.7}$$

and the so-called maximum eigenvalue statistic for testing (2.3) is given by

$$LR_{max}(r_0) = N \log(1 + \tilde{\lambda}_{r_0+1}). \tag{2.8}$$

The null distributions of these two test statistics are nonstandard and critical values have been tabulated, e.g., in Johansen (1988, 1995).

In practice it is usually assumed that the maximum lag K in the ECM (2.1) is chosen appropriately which means that it has to be chosen in such a way that the test statistics have the correct asymptotic null distributions. Usually some data-driven procedure is used for choosing K. In the following we will show that under quite general assumptions, the usual model selection criteria may be used for that purpose without affecting the asymptotic properties of the test statistics. In the next section we will spell out the precise assumptions for the data generation process (DGP) which are used in the theoretical analysis. It will be seen that K does not have to be the 'true' lag length or order. In fact, the DGP may have an infinite order VAR representation.

As mentioned in the introduction, in practice there will often be deterministic terms in the ECM (2.1) such as an intercept, seasonal dummies or a linear trend term. To simplify the theoretical analysis we will begin by assuming that no such terms are present. In Section 6 we will comment on the consequences of a nonzero mean term. It will be argued that our results can be extended in a straightforward manner to that case.

3 Model Assumptions

We use the general framework of Saikkonen (1992) and Saikkonen & Lütkepohl (1996) and partition y_t as

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}, \quad t = 1, \ldots, T, \tag{3.1}$$

where y_{it} is $(n_i \times 1)$, $i = 1, 2$, and $n_1 + n_2 = n$. We assume that the DGP is of the following form:

$$y_{1t} = Ay_{2t} + u_{1t}, \tag{3.2a}$$

$$\Delta y_{2t} = u_{2t}. \tag{3.2b}$$

Here $u_t = [u_{1t}', u_{2t}']'$ is a strictly stationary process with $E(u_t) = 0$, positive definite covariance matrix $\Sigma_u = E(u_t u_t')$ and continuous spectral density matrix which is positive definite at zero frequency. These assumptions imply that y_{2t} is $I(1)$ and not cointegrated while y_{1t} and

y_{2t} are cointegrated. Without affecting the subsequent results the initial vector y_0 is assumed to be any random vector with a fixed probability distribution.

It is well-known that the model $(3.2a)/(3.2b)$ may be written in triangular error correction form

$$\Delta y_t = J\Theta' y_{t-1} + v_t, \tag{3.3}$$

where $J' = [-I_{n_1} : 0]$, $\Theta' = [I_{n_1} : -A]$, and $v_t = [v_{1t}', v_{2t}']'$ is a nonsingular linear transformation of u_t given by

$$v_t = \begin{bmatrix} I_{n_1} & A \\ 0 & I_{n_2} \end{bmatrix} u_t$$

(see, e.g., Phillips (1991) and Saikkonen (1992)). The process v_t (and hence u_t) is assumed to have an infinite order VAR representation

$$\sum_{j=0}^{\infty} G_j v_{t-j} = \varepsilon_t, \quad G_0 = I_n, \tag{3.4}$$

where ε_t is a sequence of continuous i.i.d.$(0, \Sigma)$ random vectors with Σ being positive definite. It is also assumed that the ε_t have finite fourth moments and that the $(n \times n)$ coefficient matrices G_j satisfy the summability condition

$$\sum_{j=1}^{\infty} j^a \|G_j\| < \infty \quad \text{for some } a \geq 1. \tag{3.5}$$

This condition restricts the temporal dependence of the process v_t. It is satisfied for all $a \geq 1$ in the important special case where v_t is a vector autoregressive moving average (VARMA) process. Condition (3.5) also implies that the process v_t and, hence, y_t can be approximated by a finite order autoregression. Specifically, using (3.3) and (3.4) it can be shown that

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{K} \Gamma_j \Delta y_{t-j} + e_t, \quad t = K+2, K+3, \ldots \tag{3.6}$$

where

$$e_t = \varepsilon_t - \sum_{j=K+1}^{\infty} G_j v_{t-j}.$$

Thus, the model can be brought in the form of our starting model (2.1). Now the error term is not white noise, though, if some of the G_j are nonzero for $j > K$. Note, however, that our assumptions do not rule out finite order models.

Due to the cointegration assumption, the coefficient matrix Π has reduced rank and, hence, the structure

$$\Pi = \Phi\Theta' = -\sum_{j=0}^{K} G_j J \Theta' \tag{3.7}$$

where the second equality defines the $(n \times n_1)$ matrix Φ which is of full column rank (at least for K large enough). Details of the derivation of (3.6) and (3.7) can be found in Saikkonen (1992) and Saikkonen & Lütkepohl (1996) and are not repeated here. We note, however, that the coefficient matrices Γ_j $(j = 1, \ldots, K)$ are functions of Θ, G_j $(j = 1, 2, \ldots)$ and K and they form an absolutely summable sequence.

In Section 2 we have argued that the (approximate) ECM (3.6) forms a basis for cointegration testing procedures. It may be worth noting that the cointegration tests do not require knowledge on which components of y_t belong to y_{1t} and y_{2t}. In fact, for them it is sufficient to know that the components of y_t can in principle be devided into the two groups, possibly after a suitable linear transformation. The application of the tests does require a suitable choice of the truncation lag or order of truncation, K, however. It is intuitively clear that K should be so large that $G_j \approx 0, j > K$, because then we approximately have $e_t \approx \varepsilon_t$. In particular, to be able to prove useful asymptotic results, one has to assume that the order of truncation increases with the sample size at a suitable rate. Since it is clear that consistent estimators and tests cannot be obtained if the order of truncation increases too fast compared with the sample size the following technical assumption is commonly used in the literature which considers approximating an infinite order model by a finite order VAR or ECM (e.g., LS, Lewis & Reinsel (1985), Lütkepohl & Poskitt (1996), Saikkonen & Lütkepohl (1995, 1996) and Lütkepohl & Saikkonen (1997)).

Assumption 1. K is chosen as a function of T such that $K \to \infty$ and $K^3/T \to 0$ as $T \to \infty$.

Assumption 1 specifies an upper bound for the rate at which the value of K is allowed to tend to infinity with the sample size. In most of the aforementioned related literature a lower bound for the lag order is also imposed. Recently Ng & Perron (1995) showed, however, that the limiting distribution of the univariate unit root tests of Said & Dickey (1984) may be obtained under Assumption 1 without imposing a lower bound condition for the lag length. Ng & Perron (1995) also showed that choosing the order of truncation on the

basis of conventional model selection criteria, like AIC or SC (see Lütkepohl (1991, Chapters 4 and 11)), yields $K = O_p(\log T)$, a choice which is consistent with Assumption 1. In the next sections results similar to those of Ng & Perron (1995) are obtained for the multivariate case.

4 Results with Deterministic Choice of the Truncation Lag

SL show that the LR tests for the cointegrating rank of a system remain valid for processes without deterministic trend if the lag order is chosen according to Assumption 1. We will summarize their result here because it is the basis for studying the consequences of using data dependent rules for the lag order. It is assumed that the tests of the cointegrating rank are based on the estimated version of (3.6),

$$\Delta y_t = \tilde{\Pi} y_{t-1} + \tilde{\Gamma} z_t + \tilde{\varepsilon}_t, \qquad t = K+2, \ldots, T, \tag{4.1}$$

where $z_t = [\Delta y'_{t-1}, \ldots, \Delta y'_{t-K}]'$, as in (2.4), and $\tilde{\Pi}$ and $\tilde{\Gamma} = [\tilde{\Gamma}_1 : \cdots : \tilde{\Gamma}_K]$ are the ordinary least squares (OLS) estimators of the coefficient matrices Π and $\Gamma = [\Gamma_1 : \cdots : \Gamma_K]$, respectively. Moreover, the $\tilde{\varepsilon}_t$ are the OLS residuals. The following theorem states that the LR tests for the cointegrating rank maintain their usual asymptotic properties if the truncation lag is chosen according to some rule which satisfies Assumption 1.

THEOREM 4.1. If the truncation lag is chosen as prescribed in Assumption 1 then LR_{trace} and LR_{max} have the same limiting distribution under the null hypothesis as in the case where the true VAR order is known and finite.

A detailed proof of this result for the test statistic LR_{trace} may be found in SL. The arguments used there may be adapted to prove the theorem also for LR_{max}.

5 Results with Data-Dependent Choice of the VAR Order

In this section we shall study the data-dependent selection of the order of truncation in the unrestricted approximate ECM (3.6). It is assumed that the order of truncation is chosen by minimizing the criterion

$$\log|\tilde{\Sigma}_K| + (K+1)C_T/T, \quad K \leq K_T = o(T^{1/3}), \quad C_T > n^2, \quad C_T/T \to 0, \tag{5.1}$$

where $\tilde{\Sigma}_K$ equals our previous $\tilde{\Sigma}$. The sequences K_T and C_T have to be prescribed. The former provides an upper bound for the considered values of K. Unless otherwise stated it will be assumed that $K_T \to \infty$ so that K_T satisfies the upper bound condition in Assumption 1. This assumption is needed for the results to be proved in the following. The sequence C_T determines the considered criterion. If $C_T = 2n^2$ then (5.1) yields the familiar Akaike information criterion, AIC, choosing $C_T = 2n^2 \log\log T$ gives the Hannan-Quinn criterion, HQ, and if $C_T = n^2 \log T$ then another popular criterion, often referred to as SC is obtained (see, e.g., Lütkepohl (1991, Ch. 4)). We write $\tilde{K}$ for the value that minimizes (5.1).

We now wish to derive the asymptotic properties of the LR tests based on a model with lag order $\tilde{K}$. For this purpose we first consider the infeasible least squares regression

$$\Delta y_t = \hat{\Phi} u_{1,t-1} + \hat{\Gamma} z_t + \hat{\varepsilon}_{Kt}, \quad t = K+2, \cdots, T, \tag{5.2}$$

and define the associated residual covariance matrix by

$$\hat{\Sigma}_K = N^{-1} \sum_{t=K+2}^{T} \hat{\varepsilon}_{Kt} \hat{\varepsilon}'_{Kt}.$$

The following lemma shows that in (5.1) the covariance matrix estimator $\tilde{\Sigma}_K$ can be replaced by $\hat{\Sigma}_K$ without affecting the asymptotic behaviour of the criterion. A proof is given in the appendix.

LEMMA 5.1. Suppose that y_t $(t = 1, \ldots, T)$ is generated by (3.3) and (3.4) and that condition (3.5) holds for some $a > 1$. Suppose further that K_T in (5.1) satisfies Assumption 1. Then, uniformly in $K \leq K_T$, $\tilde{\Sigma}_K = \hat{\Sigma}_K + o_p(K_T/T)$.

Lemma 5.1 is a multivariate extension of Lemma 4.2 of Ng & Perron (1995) where the error term is $o_p(T^{-1/2})$ instead of $o_p(K_T/T)$. This difference of error terms is actually of

importance because the penalty term $(K+1)C_T/T$ in the criterion (5.1) is typically of a lower order of magnitude than $o(T^{-1/2})$. This implies that the error term can be at most of order $o_p(K_T C_T/T)$ if one wishes to conclude that replacing the estimator $\tilde{\Sigma}_K$ by $\hat{\Sigma}_K$ has no effect on the asymptotic behaviour of the criterion.

It is easy to see that equation (5.2) can be reparameterized as a regression of Δy_t on $u_{t-1}, \ldots, u_{t-K}, u_{1,t-K-1}$ (see $(A.1)$ in the Appendix). As far as the minimization of (5.1) is concerned, one can here also replace the regressand by u_t because $\Delta y_{1t} = A u_{2t} + \Delta u_{1t}$ and $\Delta y_{2t} = u_{2t}$. Thus, Lemma 5.1 implies that asymptotic properties of a minimizer of (5.1) can be studied by using the stationary process u_t (or v_t). In the stationary case these properties have been studied extensively, as the monograph of Hannan & Deistler (1988) shows. Although the results in Hannan & Deistler (1988) are formulated without the regressor $u_{1,t-K-1}$ it is clear that this has no effect on the main conclusions. Thus, from Lemma 5.1 and Theorem 7.4.7(b) of Hannan & Deistler (1988) we can, for instance, conclude that, if the order of truncation is chosen by minimizing the AIC or SC criterion we have $\tilde{K} = O_p(\log T)$ if u_t (or v_t) has a finite order VARMA representation which satisfies suitable conditions. For the details of this result see the discussion on p. 334 of Hannan & Deistler (1988).

We shall not provide a detailed discussion of the asymptotic behaviour of a minimizer of (5.1) but only prove the following theorem in the appendix.

THEOREM 5.1. Under the conditions of Lemma 5.1 the following results hold.

(i) If (3.4) is not a finite order autoregression then $\tilde{K} \to \infty$ in probability.

(ii) If (3.4) is an autoregression of a finite order $K_0 \leq K_T$ and $C_T \to \infty$ then $\tilde{K} \xrightarrow{p} K_0$.

The first result of Theorem 5.1 shows that choosing the value of K by conventional model selection criteria is consistent with the upper bound condition in Assumption 1. The following theorem shows that choosing the VAR order by a criterion of the form (5.1) leaves the asymptotic null distributions of the test statistics for the cointegrating rank unchanged. Again, a proof is provided in the appendix.

THEOREM 5.2. If the assumptions of Lemma 5.1 are satisfied, then LR_{trace} and LR_{max}, computed on the basis of a model with lag order $\hat{K}$, have the same limiting distribution under the null hypothesis as in the case where the true VAR order is known and finite.

6 Models with an Intercept

The results of the previous sections can be extended to models with intercept terms in the cointegrating relations. In this case $(3.2a)$ becomes

$$y_{1t} = \mu + Ay_{2t} + u_{1t} \tag{3.2a'}$$

while $(3.2b)$ remains as before. This implies that instead of (3.6) we have

$$\Delta y_t = \nu + \Pi y_{t-1} + \sum_{j=1}^{K} \Gamma_j \Delta y_{t-j} + e_t^*, \qquad t = K+2, K+3, \ldots \tag{3.6'}$$

where $\nu = -\Phi\mu$ and Π has the structure (3.7). Of course, we now have to add intercept terms to the least squares regressions. However, in the same way as in Saikkonen (1992) all the results proved in Sections 4 and 5 still hold provided appropriate modifications are made in their presentation. Details of these modifications are discussed in Saikkonen (1992, Section 5) and are not repeated here. Modifications required in the proofs are briefly discussed in the appendix. In the next section some small sample results are obtained by simulations.

7 Simulation Results

For unit root tests in the context of univariate time series Ng & Perron (1995) and Agiagloglou & Newbold (1996) found that the ADF tests lose power if the lag length is overspecified. Also some size distortion was observed if the lag length is underspecified. Since LR cointegration tests are the corresponding tests in the multivariate case one may expect that they have similar small sample properties. Because in practice the cointegrating rank is usually determined by testing $H(0) : \mathrm{rk}(\Pi) = 0$, $H(1) : \mathrm{rk}(\Pi) = 1$, etc. sequentially until the null hypothesis is rejected for the first time, one would expect that too few cointegration relations are found if a large lag length is chosen. Previous simulation studies which have also considered this aspect of testing for cointegration are Reimers (1992), Cheung & Lai (1993), Yap & Reinsel (1995), Haug (1996) and SL among others. In all these studies it was

confirmed that the lag order choice has a substantial impact on the outcome of cointegration tests in samples of the size commonly used in macroeconometric studies.

In particular, Reimers found that choosing an unnecessarily large lag order results in size distortions and power reductions. In his simulation study the empirical size does not necessarily decline with increasing lag length. In contrast, the rejection rate in some cases exceeded the nominal one substantially when the lag order was overspecified. In his experiment the SC criterion worked best for choosing the VAR order prior to using LR tests for the cointegrating rank. This may be a consequence of the specific processes used. In particular, he considered only VAR processes of orders 1 and 2 (ECMs of order $K = 0$ and 1). Clearly, for low order processes parsimonious criteria such as SC may have an advantage over more lavish criteria such as AIC.

Cheung & Lai (1993) investigated the impact of the lag length on the size of LR cointegration tests in a bivariate setting and found that underspecifying the true lag length can lead to massive size distortions while overspecifying the lag order may be less problematic. Severe size problems were also found for infinite order processes. Similar conclusions were also reached by Haug (1996) for bivariate processes and Yap & Reinsel (1995) and SL for three-dimensional processes. In fact, Haug (1996, p 113) concludes that "the study of Ng and Perron should be extended to cointegration tests because the experiments with various lag lengths ... indicate that additional lags decrease size distortions dramatically, however, the loss in power may also be large." Following this proposal we have performed a Monte Carlo experiment which focusses on the specific impact of the lag length on size and power of LR tests for the cointegrating rank.

In most of the aforementioned studies the properties of the tests for specific null hypotheses are investigated whereas in practice the aforementioned sequential procedure is commonly used. We will therefore investigate the properties of the sequential procedure which tests $H(0) : \mathrm{rk}(\Pi) = 0$, $H(1) : \mathrm{rk}(\Pi) = 1$, etc. and terminates when the null hypothesis is rejected for the first time. We will use the DGPs from SL and focus our study on the following questions: What is the impact of the lag order on the distribution of the cointegrating ranks determined by the LR tests? If the lag length is chosen by some model selection criterion, what is the impact of the model selection criterion on the properties of the LR tests? In the previous sections we have seen that asymptotically the choice of the

lag length has no impact on the LR tests for the cointegrating rank provided the simple condition for the upper bound of the lag order in Assumption 1 is observed. Of course, the situation may be quite different in small samples.

Although our DGPs have zero mean, we will only present results for tests which allow for a nonzero mean term. The reason is that assuming a zero mean is rather unusual in practice. For simplicity we focus on LR_{trace}. Critical values are taken from Johansen & Juselius (1990, Table A.2).

As mentioned earlier, our simulations are based on the DGPs used in SL. The first one is a VAR(2) process which has an EC representation of order $K_0 = 1$,

$$\Delta y_t = P^{-1} \left(\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} - I_3 \right) P y_{t-1} + \begin{bmatrix} -0.080 & 0.224 & -0.152 \\ 0.177 & 0.046 & -0.254 \\ 0.000 & -0.102 & 0.129 \end{bmatrix} \Delta y_{t-1} + \varepsilon_t \quad (7.1)$$

with $\varepsilon_t \sim$ i.i.d. $N(0, \Sigma)$. Here

$$P = \begin{bmatrix} -0.29 & -0.47 & -0.57 \\ -0.01 & -0.85 & 1.00 \\ -0.75 & 1.39 & -0.55 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 0.47 & 0.20 & 0.18 \\ 0.20 & 0.32 & 0.27 \\ 0.18 & 0.27 & 0.30 \end{bmatrix}.$$

The second DGP is a mixed VARMA process which was also used by Yap & Reinsel (1995),

$$\Delta y_t = P^{-1} \left(\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} - I_3 \right) P y_{t-1} + \varepsilon_t - P_\theta \begin{bmatrix} 0.297 & 0 & 0 \\ 0 & -0.202 & 0 \\ 0 & 0 & \lambda_\theta \end{bmatrix} P_\theta^{-1} \varepsilon_{t-1}, \quad (7.2)$$

where P and ε_t are as in (7.1) and

$$P_\theta = \begin{bmatrix} -0.816 & -0.657 & -0.822 \\ -0.624 & -0.785 & 0.566 \\ -0.488 & 0.475 & 0.174 \end{bmatrix}.$$

Yap & Reinsel (1995) also considered a process very similar to (7.1) in their Monte Carlo experiment. Using the two processes (7.1) and (7.2) allows us to obtain results for finite order as well as infinite order VAR processes. We have chosen three-dimensional processes because, given the empirical studies reported in the literature, this dimension may be regarded as moderate. In any case, it turned out to be large enough to study important features related

to variations in the cointegrating rank. The values of the λ_i, $i = 1, 2, 3$, determine the cointegration properties of the processes. More precisely, the number of λ_i with absolute value less than one is just the cointegrating rank of the system. The precise values used in the simulations will be given later when we discuss the results. The size of λ_θ determines to some extent how well the mixed VARMA process can be approximated by a low order pure VAR model. A λ_θ close to zero ensures that a low order VAR provides a good approximation because the other eigenvalues of the MA coefficient matrix are also small (0.297 and -0.202). A large λ_θ value, on the other hand, requires a larger VAR order for a good approximation. To ensure invertibility of the MA part, $|\lambda_\theta|$ has to be less than one. We have chosen $\lambda_\theta = 0$ and $\lambda_\theta = \pm 0.5$ in the simulations and report some of the results in the following.

The number of replications is 1000 and we have used sample sizes of $T = 100$ and $T = 200$. The effective sample size used in a specific situation depends on the VAR order, of course, as in the theoretical derivations of the previous sections. The maximal orders K_T are chosen to be $K_{100} = 4$ and $K_{200} = 5$, that is, K_T is the largest integer which is smaller than $T^{1/3}$. In (5.1) K_T is required to be of smaller order than $T^{1/3}$. This, however, is only an asymptotic condition. In principle this does not mean that the order cannot be greater than $T^{1/3}$ for any given finite T. The maximal lag lengths used here turn out to be sufficient to study the implications of a relatively large and potentially overspecified lag order.

In Table 1 some of the results for samples of size $T = 100$ obtained for the pure VAR process (7.1) are given. It is obvious from the table that the choice of the lag order has an impact on both the size and the power of the cointegration tests. For the DGP with cointegrating rank $r = 0$, the impact of the lag order on the size of the test is seen most easily. In this case the empirical size is grossly distorted if the test is based on a zero order ECM and hence, the order is underspecified. Instead of the nominal 5% the actual rejection rate of $H(0) : \mathrm{rk}(\Pi) = 0$ is almost five times as large. On the other hand, the rejection rate also tends to increase with the lag order if the latter is overspecified. For an increasing lag order the sampling uncertainty increases which affects the performance of the LR test for the cointegrating rank. Note, however, that even for the true lag order $K = 1$ the actual rejection rate of the true null hypothesis $\mathrm{rk}(\Pi) = 0$ exceeds the nominal 5% considerably in this case. When $r > 0$, the power of the LR test is much better for an underspecified lag length than for an overspecified order. Of course, the former result is a reflection of the massive size

distortion for an underspecified lag order. The power deteriorates with increasing lag order K. In other words, too small a rank is chosen with an increasing probability if K increases. Generally the choice of the cointegrating rank is more diverse if the lag order is increased and the true rank is greater than one.

For the present process, using order selection criteria such as AIC, HQ or SC amounts to choosing an order around the true lag order of $K_0 = 1$ with a high probability and consequently, selecting the lag length with any one of these criteria overall results in a better performance of the LR tests for the cointegrating rank than for a deterministically chosen order of about $T^{1/3}$. Although the choice of the selection criterion has some impact on the outcome of the cointegration tests, none of the criteria is generally superior to its competitors. In particular, none of the model selection criteria leads to generally superior performance of the cointegration tests. Note, however, that the tests appear to have very little power and do not find the correct rank with much certainty if the rank is greater than one, say. With a sample size of $T = 100$, which is the order of magnitude often encountered in macroeconometric studies with quarterly data, it is obviously difficult for the LR test to find the correct rank if that is greater than one, even under the present artificial conditions. Still it may be worth investing some effort in choosing a reasonable lag order.

We have repeated this experiment with samples of size $T = 200$ and also for other λ_i values. The results are not shown because they are qualitatively similar to those for $T = 100$ although the reliability of the tests in finding the true cointegrating rank improves if $T = 200$ and a small VAR order is used. For instance, if model selection criteria are employed, a true rank of one is found in more than 90% of the replications. However, even with 200 observations a true rank of 2 or 3 is not found with a satisfactory frequency. Generally the performance of the LR tests in terms of power and size deteriorates for increasing VAR order.

In Table 2 some results for VARMA processes are given. The sample size underlying the table is $T = 100$. Because the characteristics of the MA part now determine which lag order is necessary for a good approximation of the DGP it is not surprising that the frequency distributions of the cointegrating ranks selected now also depends on the MA characteristics. Using a very small lag order may result in a quite poor performance of the LR tests both in terms of size and power. In addition, the performance of the LR tests again

deteriorates eventually for increasing lag length. Using the proportion of correct choices of the cointegrating rank as a criterion, it is clearly helpful to apply order selection criteria. However, even with this device a choice of a correct rank $r > 1$ is not very likely.

As can be seen in Table 3, the situation improves slightly for samples of size $T = 200$. Even then the success rate is not impressive, though. Note also that the model selection criteria do not necessarily find the optimal lag order for the purposes of testing for the cointegrating rank. For instance, in Table 3 for a DGP with $r = 1$ and $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = 0.9$, $\lambda_\theta = -0.5$, the correct rank is found by the testing procedure in 92.4% of the replications if $K = 2$, whereas less than 90% correct decisions on the cointegrating rank are made if any of the order selection criteria is used. In most cases, however, using model selection criteria results in correct decisions with a probability close to the best one obtained with any one fixed order. Hence, on the basis of this limited evidence using model selection criteria seems to be a good idea. In most cases AIC and HQ have a slight advantage over the very parsimonious SC criterion. This is in line with simulation results by Agiagloglou & Newbold (1996) for univariate unit root tests but contrasts with findings by Reimers (1992). Of course, it is not clear that the proportion of correct choices of the cointegrating rank is necessarily the best performance criterion here. Therefore the full frequency distributions of the selected ranks are given in the tables. Obviously the frequency distributions tend to be more concentrated on small ranks if the lag order increases.

As mentioned earlier, we have also used other Monte Carlo designs. They led to qualitatively similar results and are therefore not shown in order to save space. The general conclusion from the simulations is that the choice of the lag order has a massive impact on the cointegrating rank determined in the usual squential manner on the basis of LR tests. Choosing too small an order as well as overspecifying the order both lead to size distortions and loss in power. Unfortunately, it turns out that even for the simple processes considered in our simulation experiment a correct cointegrating rank greater than one is not found very often for samples of the size typically available in macroeconometric studies.

8 Conclusions

In this study we have investigated the impact of the choice of the lag order on tests for the number of cointegrating relations in a VAR or ECM framework. It is found that the

asymptotic distribution of LR tests for the cointegrating rank remains unchanged if the true DGP is of finite order and a consistent model selection criterion is used for choosing the lag order. In fact, the asymptotic distribution of the LR tests remains even valid if the true VAR order is infinite as, for instance, in VARMA processes. In other words, from an asymptotic point of view, the common practice of choosing the lag order with one of the model selection criteria is justified.

Using simulations we found, however, that the small sample properties of the cointegration tests are strongly dependent on the choice of the lag length. Choosing a very small lag length which results in a poor approximation of the true DGP may equally well result in major size distortions and reduced power of the tests as a large lag length which introduces substantial sampling uncertainty into the estimated model. Generally, increasing the lag length eventually results in size and power erosions. Therefore, choosing the lag length with order selection criteria which tend to find a balance between a good approximation of the DGP and an efficient use of the sample information seems to be a good strategy for applied work.

Our simulation results are exclusively based on three-dimensional processes which may be viewed as restrictive. However, from other studies with a different focus it appears that small variations in the dimension are likely to result in qualitatively similar findings. That is, for processes with dimension two, for example, we expect to also find size and power distortions for increasing lag length. Of course, in such processes there are fewer possibilities to underestimate the cointegrating rank if the process is stationary, say.

In this study we have exclusively focussed on processes without deterministic trend terms. Given the importance of processes with deterministic linear trends in applied work, an extension of the present results to this case is desirable. We have not considered it here because it appears to be nontrivial at least as far as the asymptotic theory is concerned. Similar remarks are true for other deviations from the simple standard case considered here. For instance, investigating processes with structural shifts or heavy tailed, ARCH type residuals may be of interest from a practical point of view.

Unfortunately, even without such complications the performance of the LR tests is overall not satisfactory if the true cointegrating rank is greater than one. This result suggests that it may be worthwhile to consider alternative tests such as those proposed by SL.

Appendix. Proofs

A.1 Preliminaries and Intermediate Results

Following Saikkonen (1992) and Saikkonen & Lütkepohl (1996) we shall first reparameterize equation (3.6) as

$$\Delta y_t = \sum_{j=1}^{K} \underline{\Gamma}_j u_{t-j} + \underline{\Gamma}_{K+1,1} u_{1,t-K-1} + \underline{\Pi}_2 y_{2,t-1} + e_t, \qquad t = K+2, \ldots, T, \tag{A.1}$$

where $\underline{\Pi}_2 = 0$ and $\underline{\Gamma}_j = [\underline{\Gamma}_{j1} : \underline{\Gamma}_{j2}]$ with $\underline{\Gamma}_{11} = \Phi + \Gamma_{11}$, $\underline{\Gamma}_{j1} = \Gamma_{j1} - \Gamma_{j-1,1}$, $j = 2, \ldots, K$, $\underline{\Gamma}_{K+1,1} = -\Gamma_{K1} = 0$ and $\underline{\Gamma}_{j2} = \Gamma_{j1} A + \Gamma_{j2}$, $j = 1, \ldots, K$. Define

$$q_t' = \left[u_{t-1}', \ldots, u_{t-K}', u_{1,t-K-1}'\right], \quad p_t' = \left[q_t' : y_{2,t-1}'\right]$$

and

$$\Xi = \left[\underline{\Gamma}_1 : \cdots : \underline{\Gamma}_K : \underline{\Gamma}_{K+1,1}\right], \quad \Lambda = \left[\Xi : \underline{\Pi}_2\right].$$

With these definitions we can write $(A.1)$ as

$$\Delta y_t = \Lambda p_t + e_t, \qquad t = K+2, \ldots, T. \tag{A.2}$$

Let $\tilde{\Lambda} = [\tilde{\Xi} : \underline{\tilde{\Pi}}_2]$ be the least squares estimator of Λ obtained from $(A.2)$. Then it follows from the definitions that

$$(\tilde{\Lambda} - \Lambda) D_T^{-1} = \sum_{t=K+2}^{T} e_t p_t' D_T \left[D_T \sum_{t=K+2}^{T} p_t p_t' D_T \right]^{-1} \tag{A.3}$$

where $D_T = \mathrm{diag}[N^{-1/2} I_{nK+n_1} : N^{-1} I_{n_2}]$. As in Saikkonen (1992) we have to study the asymptotic properties of the right hand side of $(A.3)$. For this purpose it is convenient to introduce the matrix norm $\|C\|_1 = \sup\{\|Cx\| : \|x\| \leq 1\}$ where the symbol $\|\cdot\|$ signifies the Euclidean norm. The useful inequality

$$\|C_1 C_2\| \leq \|C_1\| \|C_2\|_1 \tag{A.4}$$

is known to hold for any conformable matrices (see, e.g., Lütkepohl (1996, Chapter 8)) and will be frequently used without explicit reference. Next define

$$\hat{R} = D_T \sum_{t=K+2}^{T} p_t p_t' D_T \quad \text{and} \quad R = \mathrm{diag}\left[\Gamma_{qq} : N^{-1} \sum_{t=K+2}^{T} y_{2,t-1} y_{2,t-1}'\right],$$

where $\Gamma_{qq} = E(q_t q_t')$. For the inverses of these matrices we have

$$\|\hat{R}^{-1} - R^{-1}\|_1 = O_p(K/N^{1/2}). \tag{A.5}$$

This result follows directly from Lemmas A2 - A4 of Saikkonen (1991) by observing that the proofs of these lemmas are based on moment calculations which require only Assumption 1. This fact will also be used in subsequent derivations without mentioning the difference in assumptions. We partition

$$\hat{R}^{-1} = [\hat{R}^{ij}]_{i,j=1,2}$$

conformably with the partition of R and prove the following auxiliary result.

LEMMA A.1. Suppose that y_t is generated by (3.3) and (3.4) and that condition (3.5) holds. Then, as $T \to \infty$, uniformly in $K \leq K_T = o(T^{1/3})$,

(i) $\|\hat{R}^{11}\|_1 = O_p(1)$

(ii) $\|\hat{R}^{11} - \Gamma_q^{-1}\|_1 = O_p(K_T/N^{1/2})$

(iii) $\|\hat{R}^{22} - (N^{-2}\sum_{t=K+2}^{T} y_{2,t-1}y_{2,t-1}')^{-1}\|_1 = O_p(K_T/N)$

(iv) $\|\hat{R}^{12}\|_1 = O_p((K_T/N)^{1/2})$

Proof. We shall first give a proof for any chosen $K \leq K_T$. From the inversion formula of a partitioned matrix one obtains

$$\begin{aligned}(\hat{R}^{11})^{-1} - & \quad N^{-1}\sum_{t=K+2}^{T} q_t q_t' \\ & = N^{-3/2}\sum_{t=K+2}^{T} q_t y_{2,t-1}'(N^{-2}\sum_{t=K+2}^{T} y_{2,t-1}y_{2,t-1}')^{-1} N^{-3/2}\sum_{t=K+2}^{T} y_{2,t-1}q_t'.\end{aligned}$$

The inverse on the r.h.s. is of order $O_p(1)$ by well-known properties of integrated processes whereas

$$\|N^{-3/2}\sum_{t=K+2}^{T} y_{2,t-1}q_t'\| = O_p((K_T/N)^{1/2}) \tag{A.6}$$

by arguments used to prove Lemma A2 of Saikkonen (1991) and by the assumption $K \leq K_T$. Hence, since $\|\cdot\|_1 \leq \|\cdot\|$, we have

$$\|(\hat{R}^{11})^{-1} - N^{-1}\sum_{t=K+2}^{T} q_t q_t'\|_1 = O_p(K_T/N). \tag{A.7}$$

From Lemmas A3 and A4 of Saikkonen (1991) we find that the $\|\cdot\|_1$ norm of the inverse of the latter matrix on the l.h.s. is of order $O_p(1)$ and this holds even uniformly in $K \leq K_T$. Thus, from Lemma A2 of Saikkonen & Lütkepohl (1996) it follows that $(A.7)$ also holds for the corresponding inverses and, furthermore, that $\|\hat{R}^{11}\|_1 = O_p(1)$ for any $K \leq K_T$. The same arguments and Lemma A2 of Saikkonen (1991) yield the second assertion and, after changing the roles of q_t and $y_{2,t-1}$, also the third one, for any $K \leq K_T$. Finally, since

$$\hat{R}^{12} = -\left(N^{-1}\sum_{t=K+2}^{T} q_t q_t'\right)^{-1}\left(N^{-3/2}\sum_{t=K+2}^{T} q_t y_{2,t-1}'\right)\hat{R}^{22}$$

one can similarly show that the fourth result of the lemma holds for any $K \leq K_T$.

To complete the proof, we have to establish uniformity in $K \leq K_T$. This, however, only requires straightforward modifications to the above arguments. First note that, since $K_T = o(T^{1/3})$, it is easy to show that changing the range of summation from $t = K+2, \ldots, T$ to $t = K_T + 2, \ldots, T$ does not change the above conclusions and this holds uniformly in $K \leq K_T$. This means that we have to establish the desired uniformity with respect to the dimension of q_t. This, however, follows because the above proof applies with $K = K_T$ and because the norm of a matrix does not decrease when its dimension is increased. Thus, the l.h.s. of $(A.6)$ for example, is dominated by the corresponding quantity with $K = K_T$ plus a term which is of order $O_p((K/N)^{1/2})$ uniformly in $K \leq K_T$. Hence, Lemma A.1 is established. □

In the following it is convenient to define

$$e_{1t} = -\sum_{j=K+1}^{\infty} G_j v_{t-j}$$

so that $e_t = \varepsilon_t + e_{1t}$. For e_{1t} we have, uniformly in t,

$$E\|e_{1t}\|^2 \leq c\left(\sum_{j=K+1}^{\infty}\|G_j\|\right)^2 = o(K^{-2}), \tag{A.8}$$

where the first relation is given in (A12) of Saikkonen (1992) and the second one is an immediate consequence of condition (3.5). Here as well as below the symbol c signifies a finite positive constant (not necessarily the same throughout). We shall next prove the following lemma.

LEMMA A.2. Under the conditions of Lemma A.1,

(i) $N^{-1}\sum_{t=K+2}^{T} e_{1t}y'_{2,t-1} = o_p(K^{-1})$

(ii) $\|N^{-1}\sum_{t=K+2}^{T} e_{1t}q'_t\| = o_p(K^{-\frac{1}{2}})$

Proof: Denote the typical components of the vectors y_{2t} and e_{1t} by y_{2it} and e_{1jt}, respectively, and notice that

$$\begin{aligned} E\|N^{-1}\sum_{t=K+2}^{T} e_{1t}y'_{2,t-1}\|^2 &= N^{-2}\sum_{t=K+2}^{T}\sum_{s=K+2}^{T} E(y'_{2,t-1}y_{2,s-1}e'_{1s}e_{1t}) \\ &= N^{-2}\sum_{i=1}^{n_2}\sum_{j=1}^{n}\sum_{t=K+2}^{T}\sum_{s=K+2}^{T} E(y_{2i,t-1}y_{2i,s-1}e_{1js}e_{1jt}). \end{aligned} \tag{A.9}$$

In what follows we shall make the initial value assumption $y_{20} = 0$ which is easily seen to have no effect on asymptotic results. With this assumption we have the well-known identity

$$\begin{aligned} E(y_{2i,t-1}y_{2i,s-1}e_{1js}e_{1jt}) &= E(y_{2i,t-1}y_{2i,s-1})E(e_{1js}e_{1jt}) + E(y_{2i,t-1}e_{1jt})E(y_{2i,s-1}e_{1js}) \\ &\quad + E(y_{2i,t-1}e_{1js})E(y_{2i,s-1}e_{1jt}) + \mathrm{cum}(y_{2i,t-1}, y_{2i,s-1}, e_{1jt}, e_{1js}) \end{aligned} \tag{A.10}$$

where $\mathrm{cum}(\cdot,\cdot,\cdot,\cdot)$ denotes the fourth order cumulant of the indicated random variables [see Stuart & Ord (1987, p. 439)]. Well-known properties of integrated processes imply that $|E(y_{2i,t-1}y_{2i,s-1})| \leq c\min\{(t-1, s-1)\}$ which in conjunction with $(A.8)$ shows that the contribution of the first term on the r.h.s. of $(A.10)$ to $(A.9)$ is of order $o(K^{-2})$. Next note that, since the covariance function of the process v_t is absolutely summable, we have

$$\begin{aligned} \|E(e_{1s}y'_{2,t-1})\| &= \|E\left(\sum_{i=K+1}^{\infty} G_i v_{s-i}\sum_{j=1}^{t-1} v'_{2j}\right)\| \\ &\leq \sum_{i=K+1}^{\infty}\|G_i\|\sum_{j=1}^{t-1}\|E(v_{s-i}v'_{2j})\| \\ &\leq c\sum_{i=K+1}^{\infty}\|G_i\|. \end{aligned}$$

By $(A.8)$ the last quantity is of order $o(K^{-1})$ and it follows that the contribution of the second and third terms on the r.h.s. of $(A.10)$ to $(A.9)$ is of order $o(K^{-2})$. Finally, since cumulants are linear in each of their arguments and the fourth order cumulant function of the process v_t is absolutely summable, one readily finds that, uniformly in t and s,

$$|\mathrm{cum}(y_{2i,t-1}, y_{2i,s-1}, e_{1jt}, e_{1js})| = o(K^{-2}).$$

Hence, the contribution of the fourth term on the r.h.s. of $(A.10)$ to $(A.9)$ is also of order $o(K^{-2})$. Altogether we have thus shown that $(A.9)$ is of order $o(K^{-2})$ so that the first assertion of the lemma follows from Markov's inequality.

As for the second assertion, we can use the argument in (2.9) of Lewis & Reinsel (1985, p. 397) and conclude that

$$E\left\|N^{-1}\sum_{t=K+2}^{T} e_{1t}q_t'\right\| \leq cK^{\frac{1}{2}}\sum_{j=K+1}^{\infty}\|G_j\|.$$

The r.h.s. is of order $o(K^{-\frac{1}{2}})$ which yields the desired result. □

We also need some intermediate results for estimators using the restriction $\underline{\Pi}_2 = 0$ in $(A.1)$. Therefore we write $(A.2)$ as

$$\Delta y_t = \Xi(K)q_t + \varepsilon_{Kt}, \qquad t = K+2,\ldots,T, \tag{A.11}$$

where

$$\Xi(K) = E(\Delta y_t q_t')E(q_t q_t')^{-1} \stackrel{def}{=} \Gamma_{\Delta yq}\Gamma_{qq}^{-1}$$

and ε_{Kt} is defined to make the identity hold. In other words, ε_{Kt} is the one-step ahead prediction error of the best linear predictor of Δy_t based on $u_{t-1},\ldots,u_{t-K},u_{1,t-K-1}$. By stationarity, $\Xi(K)$ is independent of t and, by the definitions, $E(q_t\varepsilon_{Kt}') = 0$. Let $\hat{\Xi}(K)$ be the least squares estimator of $\Xi(K)$ obtained from $(A.11)$. Since this least squares regression is a reparameterized form of (5.2) we have $\hat{\varepsilon}_{Kt} = \Delta y_t - \hat{\Xi}(K)q_t$. Furthermore,

$$\hat{\Xi}(K) - \Xi(K) = \left(N^{-1}\sum_{t=K+2}^{T}\varepsilon_{Kt}q_t'\right)\left(N^{-1}\sum_{t=K+2}^{T}q_t q_t'\right)^{-1}. \tag{A.12}$$

Since $\underline{\Pi}_2 = 0$ we can augment $(A.11)$ to

$$p_t = \Lambda(K)q_t + \varepsilon_{Kt}, \qquad t = K+2,\ldots,T, \tag{A.13}$$

where $\Lambda(K) = [\Xi(K) : \underline{\Pi}_2]$. Viewing $\tilde{\Xi}$ as an estimator of $\Xi(K)$ we thus have

$$\tilde{\Xi} - \Xi(K) = \left(N^{-1}\sum_{t=K+2}^{T}\varepsilon_{Kt}q_t'\right)\hat{R}^{11} + \left(N^{-3/2}\sum_{t=K+2}^{T}\varepsilon_{Kt}y_{2,t-1}'\right)\hat{R}^{21}, \tag{A.14}$$

where $\hat{R}^{11}$ and $\hat{R}^{21}$ are as in Lemma A.1. In order to study the difference between $\tilde{\Xi}$ and $\hat{\Xi}(K)$ and further to prove Lemma 5.1 we need the following results.

LEMMA A.3. Suppose that the assumptions of Lemma 5.1 hold. Then,

(i) $\|N^{-1}\sum_{t=K+2}^{T}\varepsilon_{Kt}q_t'\| = O_p(K_T/N^{1/2})$ uniformly in $K \leq K_T = o_p(T^{1/3})$ and

(ii) $N^{-1}\sum_{t=K+2}^{T}\varepsilon_{Kt}y'_{2,t-1} - N^{-1}\sum_{t=K+2}^{T}\varepsilon_t y'_{2,t-1} = O_p(1)\sum_{j=K+1}^{\infty} j^a\|G_j\| + o_p(K_T^{-1})$

where the terms $O_p(1)$ and $o_p(K_T^{-1})$ are uniform in $K \leq K_T$.

Proof: By the definitions, $\varepsilon_{Kt} = \Delta y_t - \Xi(K)q_t$ and hence

$$\begin{aligned}\|N^{-1}\textstyle\sum_{t=K+2}^{T}\varepsilon_{Kt}q'_t\| &= \|N^{-1}\textstyle\sum_{t=K+2}^{T}\Delta y_t q'_t - \Xi(K)N^{-1}\sum_{t=K+2}^{T}q_t q'_t\| \\ &\leq \|N^{-1}\textstyle\sum_{t=K+2}^{T}\Delta y_t q'_t - \Xi(K)\Gamma_{qq}\| + \|\Xi(K)(N^{-1}\sum_{t=K+2}^{T}q_t q'_t - \Gamma_{qq})\| \\ &\leq \|N^{-1}\textstyle\sum_{t=K+2}^{T}\Delta y_t q'_t - \Gamma_{\Delta yq}\| + \|\Xi(K)\|\|N^{-1}\sum_{t=K+2}^{T}q_t q'_t - \Gamma_{qq}\|_1.\end{aligned}$$

The first norm in the last expression is of order $O_p((K_T/T)^{1/2})$ for all $K \leq K_T$, because each element of the involved matrix has mean square of order $O(N^{-1})$ uniformly in the row and column index (see the proof of Lemma A2 of Saikkonen (1991)). Further, arguments used in the proof of Lemma A.1(ii) show that that the $\|\cdot\|_1$ norm in the last expression is of order $O_p(K_T/N^{1/2})$ for all $K \leq K_T$. Thus to prove the first result, it suffices to show that $\|\Xi(K)\| = O(1)$ uniformly in $K \leq K_T$. To see this, notice that $\|\Xi(K)\| \leq \|\Gamma_{\Delta yq}\|\|\Gamma_{qq}^{-1}\|_1$ where the latter norm on the r.h.s. is of order $O(1)$ uniformly in $K \leq K_T$ (see, e.g., the proof of Lemma A3 of Saikkonen (1991)). The former norm also has the same property because $\Delta y_{1t} = Au_{2t} + \Delta u_{1t}$ and $\Delta y_{2t} = u_{2t}$ and the covariance function of u_t is absolutely summable.

To prove the second assertion, observe that

$$\varepsilon_{Kt} - \varepsilon_t = e_{1t} - (\Xi(K) - \Xi)q_t. \tag{A.15}$$

Thus, we can prove the result in two parts and first consider

$$-N^{-1}\sum_{t=K+2}^{T} e_{1t}y'_{2,t-1} = N^{-1}\sum_{t=K+2}^{T}\sum_{j=K+1}^{K_T+1} G_j v_{t-j}y'_{2,t-1} + N^{-1}\sum_{t=K+2}^{T}\sum_{j=K_T+2}^{\infty} G_j v_{t-j}y'_{2,t-1}.$$

Using $(A.8)$ it is straightforward to check that changing the range of summation from $t = K+2,\ldots,T$ to $t = K_T,\ldots,T$ has an effect which is at most of order $o_p(K_T^{-1})$. Thus, from Lemma A.2(i) it follows that the latter quantity on the r.h.s. is of order $o_p(K_T^{-1})$ uniformly in $K \leq K_T$ and for the first one we can consider

$$\begin{aligned}&\|N^{-1}\textstyle\sum_{t=K_T+2}^{T}\sum_{j=K+1}^{K_T+1} G_j v_{t-j}y'_{2,t-1}\| \\ &\quad\leq \textstyle\sum_{j=K+1}^{K_T+1} j^a\|G_j\|\|j^{-a}N^{-1}\sum_{t=K_T+2}^{T} v_{t-j}y'_{2,t-1}\| \\ &\quad\leq \max_{1\leq j\leq K_T+1}\|j^{-a}N^{-1}\textstyle\sum_{t=K_T+2}^{T} v_{t-j}y'_{2,t-1}\|\sum_{i=K+2}^{\infty} i^a\|G_i\|.\end{aligned}$$

We have to show that the first factor in the last expression is of order $O_p(1)$. For $M > 0$ consider

$$\begin{aligned} & P\left\{\max_{1\le j\le K_T+1} \|j^{-a}N^{-1}\textstyle\sum_{t=K_T+2}^{T} v_{t-j}y'_{2,t-1}\| > M\right\} \\ & \le \textstyle\sum_{j=1}^{K_T+1} P\left\{\|N^{-1}\sum_{t=K_T+2}^{T} v_{t-j}y'_{2,t-1}\| > j^a M\right\} \\ & \le M^{-2}\textstyle\sum_{j=1}^{K_T+1} j^{-2a} E\|N^{-1}\sum_{t=K_T+2}^{T} v_{t-j}y'_{2,t-1}\|^2, \end{aligned}$$

where the last relation follows from Markov's inequality and the expectation therein is of order $O(1)$ uniformly in j (see the proof of Lemma A2 of Saikkonen (1991)). Thus, the last expression above can be bounded by cM^{-2} and, since this holds for any $0 < M < \infty$, the desired result follows.

To complete the proof of (ii) in Lemma A.4 we still have to consider the second part related to the latter quantity on the r.h.s. of $(A.15)$. This means that we have to find an appropriate bound for

$$\left\|(\Xi(K)-\Xi)N^{-1}\sum_{t=K+2}^{T} q_t y'_{2,t-1}\right\| \le \|(\Xi(K)-\Xi)\| \left\|N^{-1}\sum_{t=K+2}^{T} q_t y'_{2,t-1}\right\|. \qquad (A.16)$$

Consider the first norm on the r.h.s. and, analogously to the definition of Ξ, write $\Xi(K) = [\underline{\Gamma}_1(K) : \ldots : \underline{\Gamma}_K(K) : \underline{\Gamma}_{K+1,1}(K)]$. Then,

$$\begin{aligned} \|\Xi(K)-\Xi\| &\le \textstyle\sum_{j=1}^{K} \|\underline{\Gamma}_j(K)-\underline{\Gamma}_j\| + \|\underline{\Gamma}_{K+1,1}(K)-\underline{\Gamma}_{K+1,1}\| \\ &\le c\textstyle\sum_{j=K+1}^{\infty}\|G_j\|. \end{aligned} \qquad (A.17)$$

Here the first inequality follows because the Euclidean norm is dominated by the L_1-norm. To justify the second inequality, notice that, since $\Delta y_{1t} = Au_{2t} + \Delta u_{1t}$ and $\Delta y_{2t} = u_{2t}$, the coefficient matrices involved in $\Xi(K)$ and Ξ are simple tranformations of analogous coefficient matrices obtained from the infinite order autoregressive representation of u_t. Thus, since u_t is a linear transformation of v_t, it follows that we need to justify the last inequality for corresponding coefficient matrices obtained from G_j, the coefficient matrices of the infinite order autoregressive representation of v_t. After noticing this, the required result follows from Theorem 6.6.12 of Hannan & Deistler (1988, see also p. 271 after the theorem).

Thus, $(A.17)$ and condition (3.5) imply that an appropriate upper bound for the r.h.s. of $(A.16)$ is obtained by showing that

$$K^{-a}\|N^{-1}\sum_{t=K+2}^{T} q_t y'_{2,t-1}\| = O_p(1) \qquad (A.18)$$

uniformly in $K \leq K_T$. In the same way as in the proof of the first part of the lemma it is again straightforward to check that K in the summation can be replaced by K_T and that, for any $M > 0$,

$$P\left\{\max_{1\leq K\leq K_T} K^{-a}\|N^{-1}\sum_{t=K_T+2}^{T} q_t y'_{2,t-1}\| > M\right\} \leq M^{-2}\sum_{K=1}^{K_T} K^{-2a}E\|N^{-1}\sum_{t=K_T+2}^{T} q_t y'_{2,t-1}\|^2$$

where the expectation is of order $O(K)$ (see the proof of Lemma A2 of Saikkonen (1991)). Thus, we can conclude that the last expression is bounded by cM^{-2}. This implies $(A.16)$ and completes the proof of Lemma A.3. □

Lemma A.3 is used to prove the following result.

LEMMA A.4. Suppose the assumptions of Lemma 5.1 hold. Then, uniformly in $K \leq K_T$,

(i) $\tilde{\underline{\Pi}}_2 = O_p(N^{-1})$

(ii) $\|\tilde{\Xi} - \hat{\Xi}(K)\| = o_p(K_T/N)$

(iii) $\|\hat{\Xi}(K) - \Xi(K)\| = O_p(K_T/N^{1/2})$.

Proof: Viewing $\tilde{\underline{\Pi}}_2$ as an estimator of $\underline{\Pi}_2$ in $(A.12)$ yields

$$N\tilde{\underline{\Pi}}_2 = N^{-\frac{1}{2}}\sum_{t=K+2}^{T} \varepsilon_{Kt}q'_t\hat{R}^{12} + N^{-1}\sum_{t=K+2}^{T} \varepsilon_{Kt}y'_{2,t-1}\hat{R}^{22}. \qquad (A.19)$$

Here the first term on the r.h.s. is of order $O_p(K_T)O_p((K_T/N)^{1/2}) = o_p(1)$ uniformly in $K \leq K_T$ by Lemmas A.1(iv) and A.3(i). That the second one is of order $O_p(1)$ uniformly in $K \leq K_T$ can be seen from Lemmas A.1(iii) and A.3(ii) and the fact that changing the range of summation from $t = K+2, \ldots, T$ to $t = K_T+2, \ldots, T$ does not change the conclusion.

To prove (ii), notice that from $(A.12)$ and $(A.14)$ it follows that

$$\begin{aligned}
\|\tilde{\Xi} - \hat{\Xi}(K)\| &\leq \|N^{-1}\textstyle\sum_{t=K+2}^{T} \varepsilon_{Kt}q'_t\|\|\left(N^{-1}\sum_{t=K+2}^{T} q_t q'_t\right)^{-1} - \hat{R}^{11}\|_1 \\
&\quad + \|N^{-3/2}\textstyle\sum_{t=K+2}^{T} \varepsilon_{Kt}y'_{2,t-1}\|\|\hat{R}^{21}\|_1 \\
&= O_p(K_T/N^{1/2})O_p(K_T/N) + O_p(N^{-1/2})O_p((K_T/N)^{1/2}) \\
&= o_p(K_T/N)
\end{aligned}$$

uniformly in $K \leq K_T$. Here the first equality can be justified by using $(A.7)$ and Lemmas A.1(iv) and A.3. Finally, (iii) is an immediate consequence of $(A.12)$, Lemma A.4(i) and the

fact mentioned after $(A.7)$. □

Now we are able to prove Lemma 5.1.

A.2 Proof of Lemma 5.1

First note that

$$\tilde{\varepsilon}_t = \hat{\varepsilon}_{Kt} + (\hat{\Xi}(K) - \tilde{\Xi})q_t - \underline{\tilde{\Pi}}_2 y_{2,t-1}.$$

Thus, since $\hat{\varepsilon}_{Kt}$ and q_t are orthogonal,

$$\begin{aligned}\tilde{\Sigma}_K - \hat{\Sigma}_K = \quad & \left(\hat{\Xi}(K) - \tilde{\Xi}\right) N^{-1} \textstyle\sum_{t=K+2}^{T} q_t q_t' \left(\hat{\Xi}(K) - \tilde{\Xi}\right)' + \underline{\tilde{\Pi}}_2 N^{-1} \sum_{t=K+2}^{T} y_{2,t-1} y_{2,t-1}' \underline{\tilde{\Pi}}_2' \\ & - \left(\hat{\Xi}(K) - \tilde{\Xi}\right) N^{-1} \textstyle\sum_{t=K+2}^{T} q_t y_{2,t-1}' \underline{\tilde{\Pi}}_2' + \underline{\tilde{\Pi}}_2 N^{-1} \sum_{t=K+2}^{T} y_{2,t-1} q_t' \left(\hat{\Xi}(K) - \tilde{\Xi}\right)' \\ & - N^{-1} \textstyle\sum_{t=K+2}^{T} \hat{\varepsilon}_{Kt} y_{2,t-1}' \underline{\tilde{\Pi}}_2' - \underline{\tilde{\Pi}}_2 N^{-1} \sum_{t=K+2}^{T} y_{2,t-1} \hat{\varepsilon}_{Kt}'.\end{aligned}$$

It is not difficult to see that $\|\Gamma_{qq}\|_1 = O(1)$ uniformly in $K \leq K_T$, so that, using arguments similar to those in the proof of Lemma A.3(i), it can be shown that the $\|\cdot\|_1$-norm of the matrix in the middle of the first term on the r.h.s. is of order $O_p(1)$ uniformly in $K \leq K_T$. After dividing by N, a similar result clearly holds for the matrix in the middle of the second term on the r.h.s.. Using these facts, $(A.6)$ and Lemma A.4 it can be seen that the first four terms on the r.h.s. are at most of order $o_p(K_T/N)$ uniformly in $K \leq K_T$. To show that this is also the case for the last two terms, notice that

$$\begin{aligned}N^{-1} \textstyle\sum_{t=K+2}^{T} \hat{\varepsilon}_{Kt} y_{2,t-1}' \underline{\tilde{\Pi}}_2' = \quad & N^{-1} \textstyle\sum_{t=K+2}^{T} \varepsilon_{Kt} y_{2,t-1}' \underline{\tilde{\Pi}}_2' \\ & - \left(\hat{\Xi}(K) - \Xi(K)\right) N^{-1} \textstyle\sum_{t=K+2}^{T} q_t y_{2,t-1}' \underline{\tilde{\Pi}}_2'.\end{aligned}$$

The desired result readily follows from this, Lemma A.3(ii), Lemma A.4 and $(A.6)$. □

A.3 Proof of Theorem 5.1

For the first assertion we have to show that $P\{\tilde{K} > K\}$ tends to unity for every fixed K. Observe that $\tilde{K} > K$ is implied by

$$\log|\tilde{\Sigma}_K| - \log|\tilde{\Sigma}_{K+k}| - kC_T/T > 0$$

for some positive integer k. Denote $E(\varepsilon_{Kt}\varepsilon_{Kt}')$ by Σ_K. Using the result of Lemma 5.1 and arguments similar to those in its proof one can readily check that the l.h.s. of the above

inequality converges in probability to $\log|\Sigma_K| - \log|\Sigma_{K+k}|$. When (3.4) is not a finite order autoregression this difference is strictly positive for some $k > 0$ and the required result follows.

To prove the second assertion, we first note that the above proof implies that we must have $\tilde{K} \geq K_0$ in probability. Thus, it suffices to show that here strict inequality is not possible. First observe that now $T^{1/2}\sum_{j=K+1}^{\infty}\|G_j\| \to 0$ as $K \to \infty$ and $T \to \infty$. Thus, we have $(\tilde{\Lambda} - \Lambda)D_T^{-1} = O_p(K^{1/2})$ for any $K_0 \leq K \leq K_T$ (see Saikkonen (1992, p. 21)). Standard arguments used to derive the limiting distribution of LR tests in VAR models readily show that $T(\log|\tilde{\Sigma}_K| - \log|\tilde{\Sigma}_{K_0}|) = O_p(K)$. The desired result follows from this fact and the assumption $C_T \to \infty$.

A.4 Proof of Theorem 5.2

We shall first prove an auxiliary result about the estimators $\underline{\tilde{\Pi}}_2$, $\tilde{\Phi}$ and $\tilde{\Sigma}$ based on the estimated VAR order $\tilde{K}$. Notice that here $\tilde{\Phi} = \tilde{\Pi}_1$ is obtained from $\tilde{\Pi} = [\tilde{\Pi}_1 : \tilde{\Pi}_2]$. Since the theoretical counterpart of $\tilde{\Phi}$ depends on the VAR order (see (3.7)) it is convenient to consider the limiting version

$$\Phi_\infty = -\sum_{j=0}^{\infty} G_j J.$$

Now we can prove the following result.

LEMMA A.5. Suppose the assumptions of Theorem 5.2 hold and that the estimators $\underline{\tilde{\Pi}}_2$, $\tilde{\Phi}$ and $\tilde{\Sigma}$ are based on the lag order $\tilde{K}$ selected by minimizing (5.1). Then,

(i) $N\underline{\tilde{\Pi}}_2 = N^{-1}\sum_{t=K_T+2}^{T}\varepsilon_t y_{2,t-1}'(N^{-2}\sum_{t=K_T+2}^{T} y_{2,t-1}y_{2,t-1}')^{-1} + o_p(1)$

(ii) $\tilde{\Phi} = \Phi_\infty + o_p(1)$

(iii) $\tilde{\Sigma} = \Sigma + o_p(1)$

Proof: To prove (i), note first that we again have $(A.19)$ but with K on the r.h.s. replaced by $\tilde{K}$. Since we noticed below $(A.19)$ that the first term on the r.h.s. is of order $o_p(1)$ uniformly in $K \leq K_T$ the same order result is also obtained when K is replaced by $\tilde{K}$. From

this fact and Lemma A.1(iii) we find that

$$N\underline{\tilde{\Pi}}_2 = N^{-1}\sum_{t=\tilde{K}+2}^{T}\varepsilon_{\tilde{K}t}y'_{2,t-1}\left(N^{-2}\sum_{t=\tilde{K}+2}^{T}y_{2,t-1}y'_{2,t-1}\right)^{-1} + o_p(1).$$

By Lemma A.3(ii) and Theorem 5.1 we here have

$$N^{-1}\sum_{t=\tilde{K}+2}^{T}\varepsilon_{\tilde{K}t}y'_{2,t-1} = N^{-1}\sum_{t=\tilde{K}+2}^{T}\varepsilon_t y'_{2,t-1}+O_p(1)\sum_{j=\tilde{K}+1}^{\infty}j^a\|G_j\|+o_p(K_T^{-1}) = N^{-1}\sum_{t=\tilde{K}+2}^{T}\varepsilon_t y'_{2,t-1}+o_p(1).$$

Thus, we have the equation stated in (i) except that the summation is started at $\tilde{K}+2$ instead of K_T+2. However, the stated result follows because, in the same way as in previous similar cases, it can be shown that changing the range of summation from $K+2,\ldots,T$ to $K_T+2,\ldots,T$ with a nonstochastic choice of K has an asymptotically negligible effect uniformly in $K \le K_T$.

To prove (ii) first observe that

$$\tilde{\Phi}-\Phi_\infty = \tilde{\Phi}-\Phi+\sum_{j=\tilde{K}+1}^{\infty}G_jJ = \tilde{\Phi}-\Phi+o_p(1),$$

where the latter equality follows from Theorem 5.1. Here Φ depends on $\tilde{K}$ and, by the definitions,

$$\tilde{\Phi}-\Phi = \sum_{j=1}^{\tilde{K}+1}(\tilde{\underline{\Gamma}}_{j1}-\underline{\Gamma}_{j1}) = (\tilde{\Xi}-\Xi)J_{\tilde{K}},$$

where $J_K = [-J':\ldots:-J':I_{n_1}]'$ is a $((Kn+n_1)\times n_1)$ matrix and $\|J_K\|_1 = (K+1)^{\frac{1}{2}}$. To prove the assertion, write

$$\begin{aligned}
\|\tilde{\Phi}-\Phi\| &= \|(\tilde{\Xi}-\Xi)J_{\tilde{K}}\| \\
&\le \|(\tilde{\Xi}-\hat{\Xi}(\tilde{K}))J_{\tilde{K}}\|+\|(\hat{\Xi}(\tilde{K})-\Xi(\tilde{K}))J_{\tilde{K}}\|+\|(\Xi(\tilde{K})-\Xi)J_{\tilde{K}}\| \\
&\le o_p(K_T/N)O_p(K_T^{1/2})+O_p(K_T/N^{1/2})O_p(K_T^{1/2})+c\tilde{K}^{1/2}\textstyle\sum_{j=\tilde{K}+1}^{\infty}\|G_j\| \\
&= o_p(1).
\end{aligned}$$

Here the third relation is based on Lemma A.4 and $(A.17)$ and the fourth one on Theorem 5.1 and the assumption $K_T = o(T^{1/3})$.

As for (iii), conclude first from Lemma 5.1 that it suffices to show that $\hat{\Sigma}_{\tilde{K}} = \Sigma + o_p(1)$, where $\hat{\Sigma}_{\tilde{K}}$ is obtained from the residuals $\hat{\varepsilon}_{\tilde{K}t} = \varepsilon_{\tilde{K}t}-(\hat{\Xi}(\tilde{K})-\Xi(\tilde{K}))q_t$. Using Lemmas A.3(i) and A.4(ii) in conjunction with $(A.4)$ and the fact that the $\|\cdot\|_1$-norm of the matrix of second

sample moments of q_t is of order $O_p(1)$ (cf. the proof of Lemma 5.1) it is straightforward to conclude from this fact that

$$\hat{\Sigma}_{\tilde{K}} = N^{-1} \sum_{t=\tilde{K}+1}^{T} \varepsilon_{\tilde{K}t}\varepsilon'_{\tilde{K}t} + o_p(1).$$

From this and $(A.15)$ it follows that

$$\begin{aligned}
&\hat{\Sigma}_{\tilde{K}} - N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} \varepsilon_t\varepsilon'_t \\
&\quad = (\Xi(\tilde{K}) - \Xi)N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} q_t q'_t(\Xi(\tilde{K}) - \Xi)' \\
&\qquad +(\Xi(\tilde{K}) - \Xi)N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} q_t\varepsilon'_t - N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} \varepsilon_t q'_t(\Xi(\tilde{K}) - \Xi)' \\
&\qquad -N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} e_{1t}q'_t(\Xi(\tilde{K}) - \Xi)' - (\Xi(\tilde{K}) - \Xi)N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} q_t e'_{1t} \\
&\qquad +N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} \varepsilon_t e'_{1t} + N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} e_{1t}\varepsilon'_t + N^{-1}\textstyle\sum_{t=\tilde{K}+1}^{T} e_{1t}e'_{1t} + o_p(1).
\end{aligned} \tag{A.20}$$

From Theorem 5.1, $(A.4)$, $(A.17)$ and the above mentioned fact about the matrix of second sample moments of q_t it follows that the first term on the r.h.s. is of order $o_p(1)$. Since the second sample moments between the components of ε_t and q_t are readily seen to be of order $O_p(N^{-1})$ uniformly in the dimension of q_t (when the dimension is supposed to be nonstochastic) it similarly follows that the second and third terms on the r.h.s. of $(A.20)$ are of order $o_p(1)$.

The remaining terms, which involve e_{1t}, are somewhat more complicated to deal with. Consider the last one of these and, for simplicity, denote its Euclidean norm by $Z_T(K)$ when $\tilde{K}$ is replaced by K. Let ϵ be arbitrary and K_0 an integer to be determined below. Then note that

$$\begin{aligned}
P\left\{\sup_{K\geq K_0} Z_T(K) > \epsilon\right\} &\leq \textstyle\sum_{K=K_0}^{\infty} P\left\{Z_T(K) > \epsilon\right\} \\
&\leq \textstyle\sum_{K=K_0}^{\infty} P\left\{N^{-1}\sum_{t=K+1}^{T} \|e_{1t}\|^2 > \epsilon\right\} \\
&\leq \epsilon^{-1}c\textstyle\sum_{K=K_0}^{\infty} K^{-2}.
\end{aligned} \tag{A.21}$$

Here the second inequality follows from the definition of $Z_T(K)$ and the third one from Markov's inequality and $(A.8)$. For any $\epsilon > 0$ the last quantity can be made arbitrarily small by taking K_0 large enough. Next note that

$$\begin{aligned}
P\left\{Z_T(\tilde{K}) > \epsilon\right\} &= P\left\{Z_T(\tilde{K}) > \epsilon; \tilde{K} \geq K_0\right\} + P\left\{Z_T(\tilde{K}) > \epsilon; \tilde{K} < K_0\right\} \\
&\leq P\left\{\sup_{K\geq K_0} Z_T(K) > \epsilon\right\} + P\left\{\tilde{K} < K_0\right\}.
\end{aligned}$$

From $(A.21)$ and Theorem 5.1 it can be seen that the last expression can be made arbitrarily small by taking K_0 large enough. Hence, $Z_T(\tilde{K}) = o_p(1)$ or, in other words, the term preceding $o_p(1)$ on the r.h.s. of $(A.20)$ is of order $o_p(1)$. Combining this result with those obtained

earlier in the proof and using $(A.4)$ and the Cauchy-Schwarz inequality it is straightforward to show that the remaining terms on the r.h.s. of $(A.20)$ are of order $o_p(1)$. Thus the last assertion follows because $\varepsilon_t\varepsilon_t'$ obeys a (weak) law of large numbers. □

Now we can prove Theorem 5.2. No detailed proof will be given because a proof can be obtained by following the proof of Theorem 3.1 of SL with appropriate modifications. Since SL assumed a deterministic order selection rule they were able to employ improved versions of Lemma A.5(ii) and (iii), where orders of consistency were also given. However, since these orders of consistency were actually not needed in the proof of SL, the results of Lemma A.5(ii) and (iii) are sufficient in this respect.

As in SL we first note that $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ are identical to the eigenvalues of the matrix $\underline{\mathrm{M}}_T\tilde{\underline{\Pi}}'\tilde{\Sigma}^{-1}\tilde{\underline{\Pi}}$, where $\underline{\mathrm{M}}_T = \underline{\mathrm{A}}M_T\underline{\mathrm{A}}'$ and $\tilde{\underline{\Pi}} = \tilde{\Pi}\underline{\mathrm{A}}^{-1}$ with $\underline{\mathrm{A}} = [\Theta : J_\perp]'$. Notice that $\tilde{\underline{\Pi}} = [\tilde{\Phi} : \tilde{\underline{\Pi}}_2]$ and let $\underline{\mathrm{M}}_T = [\underline{\mathrm{M}}_{ij,T}]_{i,j=1,2}$ be a conformable partition of $\underline{\mathrm{M}}_T$. The next step in the proof of SL was to obtain some results, (A.13) - (A.15) in their paper, about the asymptotic behaviour of the matrices $\underline{\mathrm{M}}_{ij,T}$. The proofs of these results were based on asymptotic properties of second sample moments of $y_{2,t-1}$ and q_t with the dimension of q_t depending on a deterministically chosen order $K = o(T^{1/3})$. Since the results of Lemma A.1 and arguments used in their proofs hold uniformly in $K \leq K_T$ it is straightforward to check that (A.13) - (A.15) of SL also hold in the present context, provided the order K is replaced by the maximum order K_T. After this the previous proof can be repeated in an obvious way to show that the $n - r_0$ smallest eigenvalues $\tilde{\lambda}_{r_0+1} \geq \cdots \geq \tilde{\lambda}_n$ are asymptotically equivalent to the solutions of the generalized eigenvalue problem

$$N^2\left(\tilde{\underline{\Pi}}_2'\tilde{\Sigma}^{-1}\tilde{\underline{\Pi}}_2 - \tilde{\underline{\Pi}}_2'\tilde{\Sigma}^{-1}\tilde{\Phi}(\tilde{\Phi}'\tilde{\Sigma}^{-1}\tilde{\Phi})^{-1}\tilde{\Phi}'\tilde{\Sigma}^{-1}\tilde{\underline{\Pi}}_2 - \lambda\sum_{t=\tilde{K}+2}^{T} y_{2,t-1}y_{2,t-1}'\right) = 0.$$

From Lemma A.5 and the arguments used in its proof it can be seen that in this eigenvalue problem the data dependent order $\tilde{K}$ can be replaced by the maximum order without changing asymptotic results. This means that the eigenvalues $\tilde{\lambda}_{r_0+1} \geq \cdots \geq \tilde{\lambda}_n$ are asymptotically equivalent to those in SL so that the desired result follows.

A.5 Models with an Intercept

When an intercept is included in the least squares regression (4.1) the estimator $\check{\Lambda}$ obtained from the least squares regression of $(A.2)$ is defined by using mean corrected observations. This means that we still have $(A.3)$ but with p_t measured as a deviation from its sample mean $\bar{p} = [\bar{q}' : \bar{y}_2']'$ with obvious notation. As pointed out in Saikkonen (1992, p. 22) we have $\|\bar{q}\| = O_p((K/N)^{1/2})$ and $\bar{y}_2 = O_p(N^{1/2})$ implying $\|D_T\bar{p}\| = O_p(N^{-1/2})$. (Note that in Saikkonen (1992, p. 22) this order is erroneously $O_p(K^{1/2}/N)$ but fortunately this has no effect on the subsequent conclusions of that paper.) Using these properties of sample means it is not difficult to check that the results of Lemmas A.1 - A.5 hold even when q_t and $y_{2,t-1}$ are replaced by their mean corrected versions. In the same way one can also readily show that the mean correction has no effect on the other conclusions made above. Details are straightforward but somewhat tedious and will be omitted.

References

Agiagloglou, C. & P. Newbold (1996). The balance between size and power in Dickey-Fuller tests with data-dependent rules for the choice of truncation lag. *Economics Letters*, 52, 229 - 234.

Cheung, Y.-W. & K.S. Lai (1993). Finite-sample sizes of Johansen's likelihood ratio tests for cointegration, *Oxford Bulletin of Economics and Statistics*, 55, 313-328.

Engle, R.F. & C.W.J. Granger (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica*, 55, 251-276.

Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, 16, 121 - 130.

Granger, C.W.J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48, 213 - 228.

Hall, A. (1994). Testing for a unit root in time series with pretest data-based model selection. *Journal of Business & Economic Statistics*, 12, 461 - 470.

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

Haug, A.A. (1996). Tests for cointegration: A Monte Carlo comparison, *Journal of Econometrics*, 71, 89-115.

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

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-1581.

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

Johansen, S. & K. Juselius (1990). Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52, 169-210.

Lewis, R. & G.C. Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* 16, 393-411.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis.* Berlin: Springer-Verlag.

Lütkepohl, H. (1996). *Handbook of Matrices.* Chichester: John Wiley & Sons.

Lütkepohl, H. & D.S. Poskitt (1996). Testing for causation using infinite order vector autoregressive processes. *Econometric Theory*, 12, 61-87.

Lütkepohl, H. & P. Saikkonen (1997). Impulse response analysis in infinite order cointegrated vector autoregressive processes. *Journal of Econometrics*, 81, 127-157.

Ng, S. & P. Perron (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268-281.

Phillips, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283-306.

Reimers, H.-E. (1992). Comparisons of tests for multivariate cointegration. *Statistical Papers*, 33, 335-359.

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

Said, S.E. & D.A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599-607.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration vectors. *Econometric Theory* 7, 1-21.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8, 1-27.

Saikkonen, P. & H. Lütkepohl (1995). Asymptotic inference on nonlinear functions of the coefficients of infinite order cointegrated VAR processes. *Discussion Paper*, Humboldt-Universität, Berlin, Institut für Statistik und Ökonometrie.

Saikkonen, P. & H. Lütkepohl (1996). Infinite order cointegrated vector autoregressive processes: Estimation and inference. *Econometric Theory*, 12, 814-844.

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

Stuart, A. & J.K. Ord (1987). *Kendall's Advanced Theory of Statistics*, Vol. 1. (5th ed.), Charles Griffin, London.

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

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

Table 1. Frequency Distributions of Cointegrating Ranks Selected for VAR DGP with Order $K_0 = 1$ Based on Sample Size 100

	lag order K							
rank	0	1	2	3	4	AIC	HQ	SC
	characteristics of DGP: $r = 0, \lambda_1 = \lambda_2 = \lambda_3 = 1$							
0	0.768	0.909	0.887	0.861	0.844	0.869	0.839	0.779
1	0.215	0.080	0.103	0.131	0.141	0.118	0.146	0.204
2	0.013	0.010	0.009	0.007	0.013	0.011	0.011	0.013
3	0.004	0.001	0.001	0.001	0.002	0.002	0.004	0.004
	characteristics of DGP: $r = 1, \lambda_1 = \lambda_2 = 1, \lambda_3 = 0.9$							
0	0.005	0.368	0.446	0.499	0.520	0.305	0.177	0.028
1	0.921	0.574	0.489	0.434	0.407	0.625	0.748	0.897
2	0.070	0.052	0.058	0.061	0.067	0.063	0.069	0.070
3	0.004	0.006	0.007	0.006	0.006	0.007	0.006	0.005
	characteristics of DGP: $r = 2, \lambda_1 = 1, \lambda_2 = \lambda_3 = 0.9$							
0	0.000	0.024	0.106	0.202	0.293	0.027	0.020	0.010
1	0.650	0.837	0.759	0.681	0.592	0.817	0.778	0.684
2	0.324	0.127	0.121	0.105	0.096	0.146	0.189	0.282
3	0.026	0.012	0.014	0.012	0.019	0.010	0.013	0.024
	characteristics of DGP: $r = 2, \lambda_1 = 1, \lambda_2 = \lambda_3 = 0.8$							
0	0.000	0.000	0.001	0.052	0.145	0.002	0.000	0.000
1	0.264	0.693	0.738	0.731	0.678	0.678	0.645	0.478
2	0.676	0.280	0.237	0.193	0.155	0.292	0.324	0.479
3	0.060	0.027	0.024	0.024	0.022	0.028	0.031	0.043
	characteristics of DGP: $r = 3, \lambda_1 = \lambda_2 = \lambda_3 = 0.9$							
0	0.472	0.509	0.545	0.564	0.590	0.486	0.460	0.443
1	0.393	0.336	0.317	0.312	0.306	0.357	0.381	0.407
2	0.099	0.120	0.112	0.097	0.084	0.121	0.122	0.111
3	0.036	0.035	0.026	0.027	0.020	0.036	0.037	0.039

Table 2. Frequency Distributions of Cointegrating Ranks Selected for VARMA DGP Based on Sample Size 100

	lag order K							
rank	0	1	2	3	4	AIC	HQ	SC
	characteristics of DGP: $r = 0, \lambda_1 = \lambda_2 = \lambda_3 = 1, \lambda_\theta = 0$							
0	0.627	0.864	0.875	0.871	0.845	0.792	0.687	0.633
1	0.349	0.124	0.112	0.117	0.137	0.190	0.289	0.343
2	0.020	0.011	0.013	0.010	0.016	0.017	0.021	0.020
3	0.004	0.001	0.000	0.002	0.002	0.001	0.003	0.004
	characteristics of DGP: $r = 0, \lambda_1 = \lambda_2 = \lambda_3 = 1, \lambda_\theta = -0.5$							
0	0.614	0.853	0.869	0.868	0.849	0.826	0.762	0.664
1	0.353	0.133	0.119	0.118	0.129	0.158	0.218	0.306
2	0.025	0.013	0.012	0.012	0.020	0.015	0.016	0.023
3	0.008	0.001	0.000	0.002	0.002	0.001	0.004	0.007
	characteristics of DGP: $r = 1, \lambda_1 = \lambda_2 = 1, \lambda_3 = 0.9, \lambda_\theta = 0$							
0	0.043	0.311	0.328	0.411	0.462	0.250	0.173	0.070
1	0.606	0.586	0.590	0.499	0.458	0.573	0.562	0.595
2	0.323	0.087	0.076	0.083	0.072	0.156	0.243	0.308
3	0.028	0.016	0.006	0.007	0.008	0.021	0.022	0.027
	characteristics of DGP: $r = 1, \lambda_1 = \lambda_2 = 1, \lambda_3 = 0.9, \lambda_\theta = -0.5$							
0	0.038	0.101	0.201	0.284	0.367	0.107	0.083	0.060
1	0.547	0.776	0.689	0.606	0.537	0.739	0.696	0.597
2	0.366	0.107	0.099	0.101	0.085	0.135	0.197	0.303
3	0.049	0.016	0.011	0.009	0.011	0.019	0.024	0.040
	characteristics of DGP: $r = 2, \lambda_1 = 1, \lambda_2 = \lambda_3 = 0.8, \lambda_\theta = 0$							
0	0.000	0.002	0.007	0.053	0.126	0.001	0.000	0.000
1	0.567	0.595	0.660	0.663	0.633	0.577	0.560	0.562
2	0.418	0.378	0.308	0.251	0.218	0.398	0.420	0.420
3	0.015	0.025	0.025	0.033	0.023	0.024	0.020	0.018
	characteristics of DGP: $r = 2, \lambda_1 = 1, \lambda_2 = \lambda_3 = 0.8, \lambda_\theta = -0.5$							
0	0.000	0.001	0.010	0.050	0.137	0.004	0.000	0.000
1	0.875	0.511	0.702	0.663	0.630	0.564	0.628	0.811
2	0.117	0.462	0.261	0.259	0.212	0.409	0.347	0.173
3	0.008	0.026	0.027	0.028	0.021	0.023	0.025	0.016

Table 3. Frequency Distributions of Cointegrating Ranks Selected for VARMA DGP Based on Sample Size 200

	lag order K								
rank	0	1	2	3	4	5	AIC	HQ	SC
	characteristics of DGP: $r = 0, \lambda_1 = \lambda_2 = \lambda_3 = 1, \lambda_\theta = 0$								
0	0.588	0.884	0.915	0.910	0.900	0.893	0.886	0.852	0.705
1	0.386	0.107	0.077	0.084	0.088	0.098	0.104	0.137	0.273
2	0.023	0.008	0.007	0.005	0.011	0.007	0.009	0.009	0.018
3	0.003	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.004
	characteristics of DGP: $r = 0, \lambda_1 = \lambda_2 = \lambda_3 = 1, \lambda_\theta = -0.5$								
0	0.560	0.876	0.914	0.910	0.894	0.888	0.886	0.874	0.833
1	0.421	0.116	0.078	0.084	0.095	0.104	0.104	0.118	0.159
2	0.016	0.008	0.007	0.006	0.010	0.006	0.010	0.008	0.008
3	0.003	0.000	0.001	0.000	0.001	0.002	0.000	0.000	0.000
	characteristics of DGP: $r = 1, \lambda_1 = \lambda_2 = 1, \lambda_3 = 0.9, \lambda_\theta = 0$								
0	0.000	0.010	0.013	0.029	0.060	0.132	0.012	0.010	0.005
1	0.637	0.876	0.914	0.894	0.863	0.788	0.884	0.857	0.752
2	0.338	0.099	0.064	0.067	0.069	0.070	0.094	0.119	0.227
3	0.025	0.015	0.009	0.010	0.008	0.010	0.010	0.014	0.016
	characteristics of DGP: $r = 1, \lambda_1 = \lambda_2 = 1, \lambda_3 = 0.9, \lambda_\theta = -0.5$								
0	0.000	0.000	0.003	0.011	0.022	0.054	0.002	0.000	0.000
1	0.571	0.890	0.924	0.911	0.895	0.862	0.897	0.889	0.853
2	0.379	0.095	0.064	0.070	0.074	0.076	0.089	0.096	0.127
3	0.050	0.015	0.009	0.008	0.009	0.008	0.012	0.015	0.020
	characteristics of DGP: $r = 2, \lambda_1 = 1, \lambda_2 = \lambda_3 = 0.8, \lambda_\theta = 0$								
0	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
1	0.033	0.062	0.110	0.180	0.256	0.300	0.073	0.048	0.036
2	0.937	0.903	0.855	0.777	0.697	0.651	0.890	0.917	0.931
3	0.030	0.035	0.035	0.043	0.047	0.048	0.037	0.035	0.033
	characteristics of DGP: $r = 2, \lambda_1 = 1, \lambda_2 = \lambda_3 = 0.8, \lambda_\theta = -0.5$								
0	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
1	0.408	0.030	0.169	0.191	0.281	0.337	0.075	0.041	0.194
2	0.564	0.936	0.796	0.767	0.674	0.617	0.889	0.925	0.773
3	0.028	0.034	0.035	0.042	0.045	0.045	0.036	0.034	0.033