MODEL SELECTION IN PARTIALLY NONSTATIONARY VECTOR AUTOREGRESSIVE PROCESSES WITH REDUCED RANK STRUCTURE

BY

JOHN C. CHAO and PETER C. B. PHILLIPS

COWLES FOUNDATION PAPER NO. 992

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2000

http://cowles.econ.yale.edu/

ELSEVIER Journal of Econometrics 91 (1999) 227–271

JOURNAL OF Econometrics

www.elsevier.nl/locate/econbase

Model selection in partially nonstationary vector autoregressive processes with reduced rank structure

John C. Chao[a,*], Peter C.B. Phillips[b,c]

[a] *Department of Economics, University of Maryland, College Park, MD 20742, USA*
[b] *Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520, USA*
[c] *University of Auckland, Private Bag, Auckland, New Zealand*

Received 1 June 1994; received in revised form 1 July 1998

Abstract

The current practice for determining the number of linearly independent cointegrating vectors, or the cointegrating rank, in a vector autoregression (VAR) requires the investigator to perform a sequence of cointegration tests. However, as was shown in Johansen (1992), this type of sequential procedure does not lead to consistent estimation of the cointegrating rank. Moreover, these methods take as given the correct specification of the lag order of the VAR, though in actual applications the true lag length is rarely known. Simulation studies by Toda and Phillips (1994) and Chao (1995), on the other hand, have shown that test performance of these procedures can be adversely affected by lag misspecification.

This paper addresses these issues by extending the analysis of Phillips and Ploberger (1996) on the Posterior Information Criterion (PIC) to a partially nonstationary vector autoregressive process with reduced rank structure. This extension allows lag length and cointegrating rank to be jointly selected by the criterion, and it leads to the consistent estimation of both. In addition, we also evaluate the finite sample performance of PIC relative to existing model selection procedures, BIC and AIC, through a Monte Carlo study. Results here show PIC to perform at least as well and sometimes better than the other two methods in all the cases examined. © 1999 Elsevier Science S.A. All rights reserved.

JEL classifications: C32, C52

Keywords: Cointegrating rank; Information criterion; Order selection; PIC; Reduced rank regression; Vector autoregression

* Corresponding author: Tel: (301) 405-1579; fax: (301) 405-3542; E-mail: chao@econ.umd.edu

0304-4076/99/$ - see front matter
PII: S0304-4076(98)00077-3

1. Introduction

Since the pathbreaking work of Engle and Granger (1987), research in cointegration has become a rapidly expanding industry. Much of the effort has been directed at procedures which will enable the empirical investigator to determine the number of linearly independent cointegrating vectors, or the cointegrating rank, in a general vector autoregressive process. Toward this end, several tests of cointegration have been developed; a nonexhaustive list includes the likelihood ratio tests of Johansen (1988, 1991) and Reinsel and Ahn (1992) and the Stock and Watson (1988) tests for common trends. Most of these procedures are designed to test the null hypothesis that the cointegrating rank is less than or equal to some preassigned value r against the alternative that the cointegrated rank is greater than r. Hence, estimating the number of cointegrating relations requires performing a sequence of such tests for different values of r. One such sequential procedure has recently been proposed by Johansen (1992) who recommends testing from the subhypothesis $r = 0$ onwards.

The sequential procedure, however, does not yield a consistent estimator of the cointegrating rank. As was shown in the Johansen paper (see Theorem 2 of Johansen, 1992), the probability of underestimating the rank under the Johansen procedure goes to zero asymptotically, but the probability of overestimation remains positive in the limit and is constrained by the size of the test. Secondly, this type of procedure assumes that the correct lag length of the vector autoregressive (VAR) process is known. In actual empirical situations, this is almost never the case. Moreover, simulation studies by Toda and Phillips (1994) and Chao (1995) have shown that test performance of these procedures can be adversely affected by lag misspecification.

The present paper offers a fresh perspective on the problem of cointegrating rank determination. We seek to address the issues raised above by reconsidering this problem from the viewpoint of model selection. The Posterior Information Criterion (PIC) put forth recently by Phillips and Ploberger (1994, 1996) is especially useful in this endeavor. Here, we extend the Phillips–Ploberger analysis to a VAR process with reduced rank cointegration structure. This extension enables us to jointly select the lag length and the cointegrating rank in a vector error-correction model. Our criterion has the additional advantage that it carries an implicit penalty function which symmetrizes the costs of under- and over-parameterization. As a result, this approach achieves consistent estimation of both the cointegrating rank and the VAR lag length.

A second objective of this paper is to conduct a Monte Carlo study comparing our criterion to the alternative model selection procedures BIC and AIC. Our results show PIC to perform at least as well and sometimes better than both BIC and AIC in all the cases studied. A likely explanation for the good sampling performance of the PIC procedure is that its penalty function takes into account not only the number of estimated parameters in the model but also the nonstationarity of the regressors associated with some of these parameters.

The paper proceeds as follows. In Section 2, we discuss the model, the data generating process and the associated assumptions. Section 3 is divided into two subsections. In Section 3.1, we describe our model selection procedure PIC and show that estimators of cointegrating rank and lag length which emerge from our procedure are weakly consistent. Section 4 reports a Monte Carlo investigation comparing PIC with alternative model selection procedures. Some concluding thoughts are offered in Section 5, and all proofs and technical material are provided in the appendices.

2. Model and assumptions

Consider the m-dimensional vector autoregressive process of $(p+1)$ order:

$$Y_t = J(L)Y_{t-1} + \varepsilon_t, \tag{1}$$

where $J(L) = \sum_{i=1}^{p+1} J_i L^{i-1}$. We initialize the process denoted by Eq. (1) at $t = -p, \ldots, 0$. Since the values $\{Y_0, Y_{-1}, \ldots, Y_{-p}\}$ do not affect our subsequent asymptotic analysis, we allow them to be any random vector including constants. Alternatively, Eq. (1) can be written in the vector error-correction model (VECM) representation as

$$\Delta Y_t = J^*(L)\Delta Y_{t-1} + J_* Y_{t-1} + \varepsilon_t, \tag{2}$$

where $J_* = J(1) - I_m$ and $J^*(L) = \sum_{i=1}^{p} J_i^* L^{i-1}$ with $J_i^* = -\sum_{\ell=i+1}^{p+1} J_\ell$ with $(i = 1, \ldots, p)$. Moreover, we assume the following conditions:

(i) $\det[I_m - J(L)L] = 0$ implies that either $L = 1$ or $|L| > 1$.
(ii) $J_* = \Gamma_r A_r'$, where Γ_r and A_r are $m \times r$ matrices of full column rank r, $0 \leqslant r \leqslant m$. (If $r = 0$, we take $\Gamma_0 = A_0 = 0$, and if $r = m$, we take $\Gamma_m = J_*$ and $A_m = I_m$.)
(iii) $\Gamma_{\perp,r}'(J^*(1) - I_m)A_{\perp,r}$ is nonsingular for $0 \leqslant r < m$, where $\Gamma_{\perp,r}$ and $A_{\perp,r}$ are $m \times (m-r)$ matrices of full column rank $m-r$ such that $\Gamma_{\perp,r}'\Gamma_r = 0 = A_{\perp,r}'A_r$. (If $r = 0$, we take $\Gamma_{\perp,0} = A_{\perp,0} = I_m$.)
(iv) $\{\varepsilon_t\}_1^T = \text{iid}\,\mathrm{N}(\mathbf{0}, \Omega)$, $\Omega > 0$.

These conditions allow for nonstationarity in the sense that the characteristic polynomial of the VAR model described by Eq. (1) may have roots on the unit circle. Condition (i), however, explicitly excludes explosive processes from our consideration. These conditions also allow for cointegration so that certain linear combinations of Y_t may result in $I(0)$ processes. Condition (ii) specifies the rank of the cointegration space (or the cointegrating rank) to be r. The $m \times r$ matrix A_r in condition (ii) is known as the cointegrating matrix and its columns form a basis for the cointegration space. Note that without further restrictions, Γ_r and A_r in condition (ii) are unidentified. To achieve identification, we follow Ahn and Reinsel (1990) in selecting a normalized parameterization in which $A_r' = [I_r, \bar{A}_r']$. Condition (iii) ensures the application of the Granger representation theorem so that ΔY_t is stationary and has a Wold representation.

Taken together, conditions (i)–(iii) imply that if $r < m$, then $\{Y_t\}$ is an integrated process of order one, or an $I(1)$ process, with $m - r$ common unit root components. Moreover, if $r > 0$, then the number of common unit root components in the multivariate system (1) is less than m, the number of constituent univariate $I(1)$ processes in Y_t, as a result of cointegration. Thus, for $0 < r < m$, we can isolate the $I(0)$ and $I(1)$ components of Y_t by defining the matrix $\underline{A}_r = [A_{\perp,r}, A_r]$ and writing $\underline{A}_r' Y_t = [(A_{\perp,r}'Y_t)', (A_r'Y_t)']'$. Note that here $A_r'Y_t$ is $I(0)$ and has a moving average representation which we shall give in Section A.2 of this paper. $A_{\perp,r}'Y_t$, on the other hand, is $I(1)$ and represents the $m - r$ common unit root components.

Finally, the normality condition (iv) allows us to write down the conditional likelihood function for the model given in Eq. (2) as

$$L_T(\Gamma_r, A_r, J_1^*, \dots, J_p^*, \Omega) = (2\pi)^{-Tm/2} |\Omega|^{-T/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \varepsilon_t' \Omega^{-1} \varepsilon_t \right\}, \tag{3}$$

where $\varepsilon_t = \Delta Y_t - J^*(L)\Delta Y_{t-1} - \Gamma_r A_r' Y_{t-1}$ as can be seen from expression (2). The likelihood function (3) fully specifies, up to the unknown parameters $(\Gamma_r, A_r, J_1^*, \dots, J_p^*, \Omega)$, a VECM with cointegrating rank r and order of lagged differences p, which we shall denote with the symbol $M_{p,r}$. Let $\theta_{p,r} = (\text{vec}(\bar{A}_r)', \text{vec}(\Gamma_r)', \text{vec}(J_1^*)', \dots, \text{vec}(J_p^*)', \omega')'$, where ω is the $m(m+1)/2 \times 1$ vector of nonredundant elements of Ω. We often find it convenient to partition $\theta_{p,r} = (\underline{\theta}_{p,r}', \omega')'$, and we assume that $\theta_{p,r}$ belongs to the parameter space $\Theta_{p,r} = \Theta_{\Gamma_r} \times \Theta_{\bar{A}_r} \times \Theta_{J_1^*} \times \cdots \times \Theta_{J_p^*} \times \Theta_\omega = \underline{\Theta}_{p,r} \times \Theta_\omega$, where $\Theta_{p,r}$ is a subset of $R^{(2mr - r^2 + m^2 p + (1/2)m(m+1))}$ such that Γ_r, A_r, J_1^*, $\dots$, J_p^* satisfy conditions (i)–(iii), and note that the dimension of $\Theta_{p,r}$ depends on the value of p and r.

Our task in this paper is to select a VECM with particular p and r, say $(\hat{p}, \hat{r})$, from amongst a class of these models $(M_{p,r} : r = 0, \dots, m;\ p = 0, \dots, \bar{p})$. For this purpose, we shall assume that there exist $r^0 = (0 \leqslant r^0 \leqslant m)$ and $p^0 = (0 \leqslant p^0 \leqslant \bar{p})$ corresponding to a unique 'true' and 'minimal' model M_{p^0,r^0} with conditional likelihood function $L_T(\theta_{p^0,r^0})$ which depends on the parameter vector $\theta_{p^0,r^0} \in \Theta_{p^0,r^0}$. In addition, as is common in parametric frameworks, we assume that the data generating process is an element of the set of structures defined by the model M_{p^0,r^0}. Thus, let $\theta_{p^0,r^0}^0 = (\text{vec}(\Gamma_{r^0}^0)', \text{vec}(\bar{A}_{r^0}^0)', \text{vec}(J_1^{*0})', \text{vec}(J_p^{*0})', \omega^{0\prime})'$ be the true value of the parameter θ_{p^0,r^0}. Then the data generating process is of the form (2) where the conditions (i)–(iv) are satisfied with $p = p^0$ and $r = r^0$ and where the parameters of the model take on the true value $\theta^0_{p^0,r^0}$.

A few words on notation. In what follows, we let $Y = [Y_1, \dots, Y_T]'$, $Y_{-1} = [Y_0, \dots, Y_{t-1}]'$, $\Delta Y = [\Delta Y_1, \dots, \Delta Y_T]'$ and $W(p) = [W_1(p), \dots, W_T(p)]'$ with $W_t(p) = [\Delta Y'_{t-1}, \dots, \Delta Y'_{t-p}]'$. We shall often wish to partition $W(\bar{p}) = [W(p), W(p^*)]$, where the submatrices $W(p)$ and $W(p^*)$ contain, respectively, the first mp columns and the last $m(\bar{p} - p)$ columns of the $T \times m\bar{p}$ matrix $W(\bar{p})$. $F(r)$ will be used to denote the $m \times (m - r)$ matrix for which $F(r)' = [0, I_{m-r}]$ and $M_X = I_T - X(X'X)^{-1}X'$ is the projection onto the orthogonal complement of the range space of X. In addition, we let $X = [\Delta Y, Y_{-1}, W(p), W(p^*)]$ and $S = X'X$ and write S in partitioned form as:

$$S = \begin{pmatrix} \Delta Y'\Delta Y & \Delta Y'Y_{-1} & \Delta Y'W(p) & \Delta Y'W(p^*) \\ Y'_{-1}\Delta Y & Y'_{-1}Y_{-1} & Y'_{-1}W(p) & Y'_{-1}W(p^*) \\ W(p)'\Delta Y & W(p)'Y_{-1} & W(p)'W(p) & W(p)'W(p^*) \\ W(p^*)'\Delta Y & W(p^*)'Y_{-1} & W(p^*)'W(p) & W(p^*)'W(p^*) \end{pmatrix}$$

$$= \begin{pmatrix} S_{\Delta\Delta} & S_{\Delta y} & S_{\Delta p} & S_{\Delta p^*} \\ S_{y\Delta} & S_{yy} & S_{yp} & S_{yp^*} \\ S_{p\Delta} & S_{py} & S_{pp} & S_{pp^*} \\ S_{p^*\Delta} & S_{p^*y} & S_{p^*p} & S_{p^*p^*} \end{pmatrix}$$

$$= \begin{pmatrix} S_{\Delta\Delta} & S_{\Delta y} & S_{\Delta\bar{p}} \\ S_{y\Delta} & S_{yy} & S_{y\bar{p}} \\ S_{\bar{p}\Delta} & S_{\bar{p}y} & S_{\bar{p}\bar{p}} \end{pmatrix} \quad \text{(say).}$$

Finally, we define[1]

$$S_{ij.k} = S_{ij} - S_{ik}S_{kk}^{-1}S_{kj} \quad \text{for } i, j = \Delta,\ y \text{ and } k = p,\ \bar{p},$$

$$S_{ij.k.l} = S_{ij.k} - S_{il.k}S_{ll.k}^{-1}S_{lj.k} \quad \text{for } i, j = \Delta,\ p^* \text{ and } k, l = y,\ p,\ \bar{p}.$$

[1] Note that for the symbol $S_{ij.k.l}$, the range for its indices is subject to the restriction that $k, l \neq \bar{p}$ unless $i \neq p^*$ and $j \neq p^*$.

3. Order selection in a partially nonstationary VAR

3.1. Posterior information criterion and consistent order estimation

Our object is to jointly estimate the cointegrating rank r and the order of lagged differences of the VECM Eq. (2) using the Posterior Information Criterion (PIC) developed in Phillips and Ploberger (1994, 1996). More specifically, we propose to select $(\hat{p},\hat{r})$ as follows:

$$(\hat{p},\hat{r}) = \arg\min \mathrm{PIC}(p,r), \tag{4}$$

where

$$\begin{aligned}\mathrm{PIC}(p,r) = {} & \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p,r) - \hat{J}_*(p))S_{yy.p}(\tilde{J}_*(p,r) - \hat{J}_*(p))']\} \\ & \times \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}\hat{J}^*(p^*)S_{p^*p^*.y.p}\hat{J}^*(p^*)']\} \\ & \times [|\hat{\Omega}^{-1}\otimes S_{pp}|^{1/2}/|\hat{\Omega}^{-1}\otimes S_{\bar{p}\bar{p}}|^{1/2}] \\ & \times [|\tilde{H}(p,r)(\hat{\Omega}^{-1}\otimes S_{yy.p})\tilde{H}(p,r)'|^{1/2}/|\hat{\Omega}^{-1}\otimes S_{yy.\bar{p}}|^{1/2}].\end{aligned} \tag{5}$$

Here $\tilde{J}_*(p,r) = (\hat{\Gamma}(p,r), \hat{\Gamma}(p,r)\hat{\bar{A}}(p,r)')$ where $\hat{\Gamma}(p,r)$ and $\hat{\bar{A}}(p,r)$ are the Gaussian maximum likelihood estimators of the reduced rank parameters Γ and $\bar{A}$ when the cointegrating rank is assumed to be r and the order of lagged differences is assumed to be p. These estimators are obtained from a Newton–Raphson procedure which we describe in Section A.1 of Appendix A. We let $\hat{J}_*(p) = S_{\Delta y.p}(S_{yy.p})^{-1}$ and $\hat{J}^*(p^*) = S_{\Delta p^*.y.p}(S_{p^*p^*.y.p})^{-1}$ denote, respectively, the least square estimator of J_* in a VECM of lag order p and the least square estimator of the last $m(\bar{p}-p)$ columns of J^* (or the coefficients of the last $\bar{p}-p$ lagged differences) in a VECM of lag order $\bar{p}$. In addition, $\hat{\Omega} = S_{\Delta\Delta.y.\bar{p}}/T$ is the maximum likelihood estimator of Ω in the case where the model given by Eq. (2) has the highest possible order in our setup, i.e., $r = m$ and $p = \bar{p}$; and the $(2mr - r^2)\times m^2$ matrix $\tilde{H}(p,r)$ is defined as

$$\tilde{H}(p,r) = \begin{bmatrix} (\hat{\Gamma}(p,r)'\otimes F(r)') \\ (I_m\otimes(I_r,\hat{\bar{A}}(p,r)')) \end{bmatrix} \tag{6}$$

While criterion (5) appears complicated, it has a simple intuitive interpretation as a combination of likelihood ratio statistics, which test the fit of the reduced rank model given by Eq. (2), and penalty terms, which reflect the complexity of the model. To see this, note first that we have written Eq. (5) as the product of four terms. The trace expression in the exponent of the first term, i.e.,

$$\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p,r) - \hat{J}_*(p))S_{yy.p}(\tilde{J}_*(p,r) - \hat{J}_*(p))'] \tag{7}$$

is, in fact, the likelihood ratio statistics for testing the null hypothesis that the cointegrating rank equals r against the general alternative that the rank is m (cf. Reinsel and Ahn, 1992). Likewise, the trace expression in the exponent of the second term, i.e.,

$$\mathrm{tr}[\hat{\Omega}^{-1}\hat{J}^*(p^*)S_{p^*p^*.y.p}\hat{J}^*(p^*)'] \tag{8}$$

can be easily seen to be the likelihood ratio statistic for testing the null hypothesis that the VECM given by Eq. (2) has p lags against the alternative that it has $\bar{p} > p$ lags. Moreover, the third and the fourth terms are terms which, ceteris paribus, penalize models for having higher lag order and/or greater cointegrating rank. We shall discuss these penalty terms in more detail in Remarks 3.2(i)–(ii) below but note for the time being that, unlike AIC, BIC and other information criteria, whose penalty function depends on a simple parameter count, the penalty terms of PIC compare the determinant of the Fisher information matrix of the larger model with that of the smaller model. In this sense, it is closely related to the Fisher Information Criterion (FIC), which was independently developed and analyzed for the univariate case by Wei (1992). Both Wei (1992) and Phillips and Ploberger (1994) have argued that this penalty function, which uses the redundant information introduced by a spurious regressor to penalize excess parameterization, has the particular desirable feature that, in making model comparisons, it takes into consideration not only the number of regressors included in the alternative models but also the magnitude of the regressors and the sample information accumulated in the data about the models' parameters. Hence, one would expect the criterion given by Eq. (5) to perform well when applied to partially nonstationary VARs, as such models involve $I(1)$ and $I(0)$ components of vastly different magnitudes.

In the next subsection, we give results showing that the PIC criterion given by Eq. (5) can be derived using a combination of Bayesian and classical ideas. Our main justification for proposing PIC is based not on Bayesian foundational arguments but on the criterion's good sampling properties, both in small and large samples, and the fact that it delivers jointly consistent estimates of cointegrating rank and VAR lag order. The Monte Carlo simulation results are presented in Section 4. Below is a formal statement of the weak consistency property of PIC in joint order selection of p and r.

Theorem 3.1. Suppose the true data generating process belongs to the set of structures defined by the model M_{p^o,r^o} of the form given by Eq. (2) and satisfies *assumptions (i)–(iv), with lag order $0 \leqslant p^0 \leqslant \bar{p}$ and cointegrating rank $0 \leqslant r^0 \leqslant m$. Suppose $(\hat{p}, \hat{r})$ is selected in accordance with the criterion given by Eq. (4). Then,*

$$\begin{pmatrix} \hat{p} \\ \hat{r} \end{pmatrix} \to \begin{pmatrix} p^0 \\ r^0 \end{pmatrix} \quad \textit{in probability as } T \to \infty .$$

Remark 3.2.

(i) To provide some intuition on the weak consistency of PIC, take the special case where, under the null hypothesis, $p = \bar{p}$ and $r < \bar{r} = m$. First, notice that in this case expression (5) reduces to

$$\begin{aligned} &\mathrm{PIC}(\bar{p}, r) \\ &\quad = |\tilde{H}(\bar{p}, r)(\hat{\Omega}^{-1} \otimes S_{yy.\bar{p}})\tilde{H}(\bar{p}, r)'|^{1/2} |\hat{\Omega}^{-1} \otimes S_{yy.\bar{p}}|^{-1/2} \\ &\quad\quad \times \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(\bar{p}, r) - \hat{J}_*(\bar{p}))S_{yy.\bar{p}}(\tilde{J}_*(\bar{p}, r) - \hat{J}_*(\bar{p}))']\}. \end{aligned} \tag{9}$$

Taking the logarithm of Eq. (9) and multiplying by 2, we have

$$\begin{aligned} &2 \ln \mathrm{PIC}(\bar{p}, r) \\ &\quad = \mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(\bar{p}, r) - \hat{J}_*(\bar{p}))S_{yy.\bar{p}}(\tilde{J}_*(\bar{p}, r) - \hat{J}_*(\bar{p}))'] \\ &\quad\quad + \ln[|\tilde{H}(\bar{p}, r)(\hat{\Omega}^{-1} \otimes S_{yy.\bar{p}})\tilde{H}(\bar{p}, r)'| / |\hat{\Omega}^{-1} \otimes S_{yy.\bar{p}}|]. \end{aligned} \tag{10}$$

Now, observe that the first term is simply the likelihood ratio statistic for testing the null hypothesis that the cointegrating rank equals r against the general alternative that the rank is m (cf., Reinsel and Ahn, 1992). To analyze the second term, we need to determine the orders of magnitude of the elements of the matrices that appear in the determinants in this term. Rotating the regressor space to isolate components of $S_{yy.\bar{p}}$ of different orders of magnitude (see Phillips (1988) for details of how to do this), we find that under the null hypothesis,

$$|\hat{\Omega}^{-1} \otimes S_{yy.\bar{p}}| \equiv O_p(T^{2m^2 - rm}), \tag{11}$$

$$|\tilde{H}(\bar{p}, r)(\hat{\Omega}^{-1} \otimes S_{yy.\bar{p}})\tilde{H}(\bar{p}, r)'| \equiv O_p(T^{(3mr - 2r^2)}). \tag{12}$$

Since $2m^2 - rm - (3mr - 2r^2) = 2m^2 - 4mr + 2r^2 = 2(m - r)^2 > 0$ for all $r < m$, the second term of Eq. (10) will be negative for large T whenever $r < m$. Hence, the criterion penalizes the alternative when the null hypothesis is correct. Recall that for Johansen type sequential procedures, as mentioned in the Introduction, the probability of overestimation never vanishes, not even in infinite samples. The PIC procedure corrects for upward bias by imposing a penalty on overparameterization. Moreover, the penalty does not contribute to a Type II error in the limit because, being a logarithmic function, it changes more slowly than the likelihood ratio statistic with an increase in the sample size.

A similar analysis can be carried out for the case where under the null hypothesis $r = \bar{r}$ and $p < \bar{p}$. Here, two times the logarithmic transformation of the criterion given by Eq. (5) reduces to

$$2\ln \mathrm{PIC}(p,\bar{r}) = \mathrm{tr}[\hat{\Omega}^{-1}\hat{J}^{*}(p^{*})S_{p^{*}p^{*}.y.p}\hat{J}^{*}(p^{*})'] + \ln[|\hat{\Omega}^{-1}\otimes S_{pp}|/|\hat{\Omega}^{-1}\otimes S_{\bar{p}\bar{p}}|] + \ln[|\hat{\Omega}^{-1}\otimes S_{yy.p}|/|\hat{\Omega}^{-1}\otimes S_{yy.\bar{p}}|]. \quad (13)$$

Note that Eq. (13) is expressed as the sum of a likelihood ratio statistic, a penalty function, and a third term which we will show to be insignificant asymptotically. The LR statistic tests the null hypothesis that the true lag length is p against the alternative that the lag order is greater than p. The remaining terms can be analyzed by noting that under the null hypothesis,

$$|\hat{\Omega}^{-1}\otimes S_{pp}| \equiv \mathrm{O_p}(T^{m^2p}), \quad (14)$$

$$|\hat{\Omega}^{-1}\otimes S_{\bar{p}\bar{p}}| \equiv \mathrm{O_p}(T^{m^2\bar{p}}), \quad (15)$$

and

$$|\hat{\Omega}^{-1}\otimes S_{yy.p}|/|\hat{\Omega}^{-1}\otimes S_{yy.\bar{p}}| = |\hat{\Omega}^{-1}\otimes(T^{-2})S_{yy.p}|/|\hat{\Omega}^{-1}\otimes(T^{-2})S_{yy.\bar{p}}| \xrightarrow{\mathrm{p}} 1. \quad (16)$$

Hence, the last term converges in probability to zero. The second term, on the other hand, converges in probability to $-\infty$, thus, eliminating the possibility of committing a Type I error in the limit.

In the general case where $r \neq \bar{r}$ and $p \neq \bar{p}$, both lag and rank overspecification will be penalized. This modification of the traditional likelihood ratio test is what drives the consistency result in Theorem 3.1.

(ii) We can also find an approximation for our criterion which facilitates a direct comparison of its penalty function with that of BIC. First rewrite criterion (5) in the equivalent form

$$\mathrm{PIC}(p,r) = \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')' M_{W(p)}(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')]\}$$

$$\times \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\Delta Y - Y_{-1}\hat{J}_*(\bar{p})')' M_{W(\bar{p})}(\Delta Y - Y_{-1}\hat{J}_*(\bar{p})')]\}$$

$$\times [|\hat{\Omega}^{-1} \otimes S_{pp}|^{1/2} / |\hat{\Omega}^{-1} \otimes S_{\bar{p}\bar{p}}|^{1/2}]$$

$$\times [|\tilde{H}(p,r)(\hat{\Omega}^{-1} \otimes S_{yy.p})\tilde{H}(p,r)'|^{1/2} / |\hat{\Omega}^{-1} \otimes S_{yy.\bar{p}}|^{1/2}]. \quad (17)$$

Now, multiply the logarithmic transformation of Eq. (17) by $2/T$ and ignoring those terms that do not involve p and r, we see that minimizing Eq. (17) with respect to p and r is identical to minimizing

$$\tfrac{1}{T}\ln|\hat{\Omega}^{-1} \otimes S_{pp}| + \tfrac{1}{T}\ln|\tilde{H}(p,r)(\hat{\Omega}^{-1} \otimes S_{yy.p})\tilde{H}(p,r)'| + \mathrm{tr}[\hat{\Omega}^{-1}\tilde{\Omega}_{p,r}], \quad (18)$$

where $\tilde{\Omega}_{p,r} = [(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')' M_{W(p)}(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')]/T$. To rewrite $\mathrm{tr}[\hat{\Omega}^{-1}\tilde{\Omega}_{p,r}]$ in a form closer to BIC, we make use of the first-order Taylor expansion:

$$\ln|\tilde{\Omega}_{p,r}| \sim \ln|\hat{\Omega}| + \mathrm{tr}[\hat{\Omega}^{-1}(\tilde{\Omega}_{p,r} - \hat{\Omega})]$$

so that minimizing Eq. (18) is seen to be asymptotically equivalent to minimizing

$$\ln|\tilde{\Omega}_{p,r}| + \tfrac{1}{T}\ln|\hat{\Omega}^{-1} \otimes S_{pp}| + \tfrac{1}{T}\ln|\tilde{H}(p,r)(\hat{\Omega}^{-1} \otimes S_{yy.p})\tilde{H}(p,r)'|. \quad (19)$$

Finally, in light of Eqs. (12) and (14) of the last Remark, an approximation to Eq. (19) that takes into account the orders of magnitude of the data matrices as $T \to \infty$ (without characterizing their partially random limit) is

$$\ln|\tilde{\Omega}_{p,r}| + \tfrac{1}{T}\ln T^{m^2 p} + \tfrac{1}{T}\ln T^{(2r(m-r)+mr)}$$

$$= \ln|\tilde{\Omega}_{p,r}| + (m^2 p + 2r(m-r) + mr)\,\tfrac{1}{T}\ln T. \quad (20)$$

We can compare Eq. (20) to the BIC criterion

$$\mathrm{BIC}(p,r) = \ln|\tilde{\Omega}_{p,r}| + (m^2 p + r(m-r) + mr)\,\tfrac{1}{T}\ln T \quad (21)$$

given in Phillips (1993) for VECMs. We see that BIC penalizes all parameters in the same way, while PIC attaches twice as great a penalty to the $r(m-r)$

parameters of the cointegrating matrix than it does the parameters associated with stationary regressors. Hence, the PIC criterion takes into account not only the number of parameters but also the potential rates of convergence of the estimators of these parameters. As Wei (1992) pointed out, the inclusion of excess nonstationary regressors should be more heavily penalized as it leads to a greater increase in prediction error than over-parameterization with respect to stationary regressors when the inclusion of these regressors is incorrect.

3.2. *A partially Bayesian interpretation of PIC*

The formula for criterion (5) can be derived using a combination of Bayesian and frequentist ideas. Let $L_T(\theta_{p,r})$ be the likelihood function of the model $M_{p,r}$ as described in Section 2, and note that $L_T(\theta_{p,r})$ has the form given by Eq. (3). Let $\underline{\theta}_{p,r}$ be as defined in Section 2 and let $\pi_{p,r}(\underline{\theta}_{p,r})$ be a (possibly improper) prior density on $\underline{\theta}_{p,r}$. Then, PIC is based on the mixture density

$$\Pi_T(M_{p,r}|\Omega,Y) = \int_{\underline{\Theta}_{p,r}} \pi(\underline{\theta}_{p,r}) L_T(\underline{\theta}_{p,r},\Omega)\,\mathrm{d}\underline{\theta}_{p,r}. \tag{22}$$

In the special case where Ω is known and the prior density $\pi(\underline{\theta}_{p,r})$ is proper, expression (22) is, in fact, proportional to the posterior probability of $M_{p,r}$ and ratios of this integral can be used to test hypotheses within the traditional Bayesian framework of posterior odds. In practice, of course, Ω is never known and the conventional Bayesian approach is to define a joint prior over $\underline{\theta}_{p,r}$ and Ω and to integrate with respect to both. Thus, expression (22) highlights two ways in which our approach departs from Bayesian inference based on posterior odds. First, our treatment of the nuisance parameter Ω is classical in the sense that we estimate it using a consistent estimator (to be discussed more fully below) and conduct inference conditioned on this estimate. Second, in the actual derivation of our criterion, we adopt an improper uniform prior for $\underline{\theta}_{p,r}$, and, in consequence, mixture (22) defines a σ-finite measure rather than a proper probability measure, as was discussed in Phillips and Ploberger (1996). We do not see these deviations from the Bayesian posterior odds paradigm as invalidating our approach, which has its own asymptotic justification. Moreover, we have found that the sampling performance of our criterion is better when we condition on a consistent estimate of Ω (cf. the results and discussion in Phillips (1995a) regarding this treatment of the scale parameter in the univariate case). Further, in many practical applications, it is difficult to justify the imposition of any particular proper prior density on $\underline{\theta}_{p,r}$, especially in situations where prior knowledge of cointergrating rank and lag length is very limited.

To find an explicit form for Eq. (22), note that the nonlinear reduced rank restriction $J_* = \Gamma_r A'_r$ resulting from cointegration precludes exact computation of the integral (22) in the cases where $0 < r < m$. We therefore develop an asymptotic approximation for the integral (22) using the Laplace's method.

Theorem 3.3. Let $\pi(\underline{\theta}_{p,r})$ be a diffuse prior density such that $\pi(\underline{\theta}_{p,r}) - (2\pi)^{-(1/2)(m^2p+2mr-r^2)}$ for $\underline{\theta}_{p,r} \in \underline{\Theta}_{p,r}$ and suppose that the covariance matrix Ω is known. Then,

$$\frac{\Pi_T(M_{p,r}|\Omega,Y)}{\hat{\Pi}_T(M_{p,r}|\Omega,Y)} \to 1 \text{ in probability as } T \to \infty,$$

where $\Pi_T(M_{p,r}|\Omega,Y)$ is as given in expression (22) and where

$$\begin{aligned}\hat{\Pi}_T(M_{p,r}|\Omega,Y) &= (2\pi)^{-Tm/2}|\Omega|^{-T/2}|\Omega \otimes S_{pp}|^{-1/2} \\ &\quad \times |\tilde{H}(p,r)(\Omega^{-1} \otimes S_{yy.p})\tilde{H}(p,r)'|^{-1/2} \\ &\quad \times \exp\{-\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')' M_{W(p)} \\ &\quad \times (\Delta Y - Y_{-1}\tilde{J}_*(p,r)')]\} \end{aligned} \tag{23}$$

with $\tilde{H}(p,r)$ and $\tilde{J}_*(p,r) = (\hat{\Gamma}(p,r), \hat{\Gamma}(p,r)\hat{\tilde{A}}(p,r)')$ as defined in Section 3.1 above.

Since Ω is usually unknown, we advocate plugging the consistent estimator $\hat{\Omega} = S_{\Delta\Delta.y.\hat{p}}/T$ into expression (23) and selecting the order of lagged differences p and the cointegrating rank r by minimizing the ratio

$$\frac{\hat{\Pi}_T(M_{\hat{p},\hat{r}}|\hat{\Omega},Y)}{\hat{\Pi}_T(M_{p,r}|\hat{\Omega},Y)}, \tag{24}$$

This, of course, simply results in the procedure as described by expressions (4) and (5) in Section 3.1 earlier. The result below shows that our plug-in procedure is asymptotically equivalent to conditioning on a known Ω.

Corollary 3.4. Suppose that the conditions of Theorem 3.3 hold except that Ω is now unknown and let $\hat{\Omega} = S_{\Delta\Delta.y.\hat{p}}/T$. Then,

$$\frac{\Pi_T(M_{\hat{p},\hat{r}}|\Omega,Y)}{\Pi_T(M_{p,r}|\Omega,Y)} \Bigg/ \frac{\hat{\Pi}_T(M_{\hat{p},\hat{r}}|\hat{\Omega},Y)}{\hat{\Pi}_T(M_{p,r}|\hat{\Omega},Y)} \to 1 \text{ in probability as } T \to \infty,$$

where $\Pi_T(M_{p,r}|\Omega, Y)$ *is as defined in expression* (22) *and where*

$$\hat{\Pi}_T(M_{p,r}|\hat{\Omega},Y) = (2\pi)^{-Tm/2}|\hat{\Omega}|^{-T/2}|\hat{\Omega}\otimes S_{pp}|^{-1/2}|\tilde{H}(p,r)(\hat{\Omega}^{-1}\otimes S_{yy.p})\tilde{H}(p,r)'|^{-1/2}$$

$$\times \exp\{-\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')'$$

$$\times M_{W(p)}(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')]\} \qquad (25)$$

again with $\tilde{H}(p,r)$ *and* $\tilde{J}_*(p,r)$ *as defined in Section* 3.1.

Remark 3.5. (i) In the special case where there is sharp prior information about the nuisance parameter Ω (i.e., Ω is known a priori), the procedure is similar to a posterior odds comparison of a family of models indexed by p and r, with the important qualification that it uses an improper prior on $\underline{\theta}_{p,r}$.

(ii) It has been known since the discussion in Bartlett (1957) that the use of an improper diffuse prior in Bayesian tests of models of different dimensions leads to an arbitrary scale effect in that the height of an improper prior density can be made to be as large or small as one desires. If we follow this interpretation of our criterion, the implied diffuse prior has height $\pi(\underline{\theta}_{p,r}) = (2\pi)^{-(1/2)(m^2p+2mr-r^2)}$, which corresponds to the normalization constant in a multivariate normal distribution of dimension $m^2p + 2mr - r^2$. This height was chosen primarily out of convenience so no rescaling was needed during the course of the Laplace approximation. It is therefore indeed subject to the criticism of arbitrariness if one follows a Bayesian interpretation of the criterion. However, other interpretations, such as prequential odds are possible and these are discussed at length in Phillips (1996), so that it is not necessary to rely on the Bayesian approach in justifying a criterion like Eq. (24), especially when we condition on an initial set of observed data. Further, the choice of constant prior here is, in our view, no more arbitrary than many proper prior densities used in Bayesian empirical work applying the posterior odds ratio since those priors are also frequently chosen out of computational convenience and not because they properly model subjective prior information.

(iii) In addition, we emphasize that however arbitrary the scale effect of an improper diffuse prior may be, its effect is asymptotically of a lower stochastic order than both the 'likelihood ratio' and the penalty function components of the criterion. To see this, suppose we set $\pi(\underline{\theta}_{p,r}) = c_{p,r}$ (for $0 < p \leqslant \bar{p}$ and $0 < r \leqslant \bar{r}$), where the $c_{p,r}$'s are positive real constants, then following the same arguments as that employed in the proofs of Theorem 3.3 and Corollary 3.4, we can obtain the alternative criterion

$$\mathrm{PIC}^*(p,r) = K(p,r)$$

$$\times \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')' M_{W(p)}(\Delta Y - Y_{-1}\tilde{J}_*(p,r)')]\}$$

$$\times \exp\{-\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\Delta Y - Y_{-1}\hat{J}_*(\bar{p})')' M_{W(\bar{p})}(\Delta Y - Y_{-1}\hat{J}_*(\bar{p})')]\}$$

$$\times [|\hat{\Omega}^{-1} \otimes S_{pp}|^{1/2} / |\hat{\Omega}^{-1} \otimes S_{\bar{p}\bar{p}}|^{1/2}]$$

$$\times [|\tilde{H}(p,r)(\hat{\Omega}^{-1} \otimes S_{yy.p})\tilde{H}(p,r)'|^{1/2} / |\hat{\Omega}^{-1} \otimes S_{yy.\bar{p}}|^{1/2}], \quad (26)$$

where $K(p,r) = (c_{\bar{p},\bar{r}}/c_{p,r})(2\pi)^{(1/2)[m^2(\bar{p}-p)+(m-r)^2]}$. Note that PIC*$(p,r)$ differs from PIC(p,r) as given by expression (17), by only the factor $K(p,r)$. Now, arguing as in Remark 3.2(ii), we see that minimizing Eq. (26) with respect to p and r is (asymptotically) the same as minimizing

$$\ln|\tilde{\Omega}_{p,r}| + \frac{1}{T}\ln|\hat{\Omega}^{-1} \otimes S_{pp}| + \frac{1}{T}\ln|\tilde{H}(p,r)(\hat{\Omega}^{-1} \otimes S_{yy.p})\tilde{H}(p,r)'|$$

$$+ \frac{2}{T}\ln K(p,r). \quad (27)$$

Note that the first term of Eq. (27) is $O_p(1)$. The second and third terms, which are the primary penalty terms of this criterion, are each $O_p(\ln T/T)$ while the term involving the factor $K(p,r)$ is only $O_p(1/T)$. Hence, while the height of the prior density will certainly have an impact in small samples, as the sample size becomes large its effect will diminish relative to that of the first three terms of our criterion. Note further that our choice of prior density height, i.e., $\pi(\underline{\theta}_{p,r}) = c_{p,r} = (2\pi)^{-(1/2)(m^2p+2mr-r^2)}$, is tantamount to setting $K(p,r) = 1$ in expression (27) and, thus, effectively ignoring the last term.

(iv) In a stimulating recent paper, Kleibergen and van Dijk (1994) present a Bayesian study of a possibly cointegrated VAR system. With an emphasis that differs from that of the present work, Kleibergen and van Dijk (1994) focus on posterior distributions derived under the diffuse and Jeffreys priors and point out that a diffuse prior specification may lead to nonintegrable marginal posterior distributions for some parameters under their specification of the cointegrated system. As discussed in Remark 3.5(i) above, our own approach is one of model selection and bears a closer resemblance to posterior odds analysis. Also, Kleibergen and van Dijk (1994) take the lag order of their VAR as given, while this paper explicitly considers the problem of jointly estimating the cointegrating rank and the VAR lag order. Finally, whereas Kleibergen and van Dijk (1994) adopt a Bayesian perspective, our paper is also concerned with the sampling properties of our procedure, and shows that the use of PIC leads to consistent cointegrating rank and lag order estimation.

(v) Note also that the normality condition (iv) given in Section 2 is not needed for either Theorem 3.3 or the consistency result, Theorem 3.1, of the last subsection. Instead, we can obtain these results under the weaker assumptions that $\{\varepsilon_t\}_1^T \equiv \text{iid}\,(\mathbf{0}, \Omega)$, $\Omega > 0$, and $\mathrm{E}|\varepsilon_{it}|^{2+\delta} < \infty$ for some $\delta > 0$ $(i = 1, \ldots, m)$, where ε_{it} denotes the ith component of the disturbance vector ε_t. Such error assumptions are common in the study of VARs and RRRs.

(vi) It is possible to extend our methods of consistent cointegrating rank estimation to time series models of reduced rank that allow for general weakly dependent error processes. One way of pursuing such an extension is to follow Phillips (1991a,b) in setting up a frequency domain version of the likelihood and basing the order selection criterion for cointegrating rank on a penalized version of this likelihood. This approach was adopted in a research note by Phillips (1991c) and a formal extension of the methods and limit theory of the present paper to this general environment is now being conducted by the authors.

4. Monte Carlo results

This section reports the results of a simulation study comparing the finite sample performance of PIC with the alternative model selection procedures BIC and AIC in VAR models with some unit roots. Eight experiments were conducted; in each case the data generating process is assumed to be a trivariate VAR with Gaussian disturbances, and the sample size is $T = 150$. The precise descriptions of these experiments are as follows:

Experiment 1

Cointegrating rank and lag order: $p = 0$, $r = 1$,
Roots of $\det[I_m - J(L)L] = 0$: $L = 1.00, 1.00, 1.01$,
Error-correction form: $\Delta Y_t = \Gamma_1 A_1' Y_{t-1} + \varepsilon_t$,

$$\Gamma_1 A_1' = \begin{bmatrix} -0.01 \\ 0 \\ 0.23 \end{bmatrix} \begin{bmatrix} 1 & -1.5 & 0 \end{bmatrix}, \qquad \Omega = \begin{bmatrix} 0.64 & 1.68 & 1.36 \\ 1.68 & 4.66 & 5.17 \\ 1.36 & 5.17 & 14.34 \end{bmatrix}.$$

Experiment 2

Cointegrating rank and lag order: $p = 0$, $r = 1$,
Roots of $\det[I_m - J(L)L] = 0$: $L = 1.00, 1.00, 2.00$,
Error-correction form: $\Delta Y_t = \Gamma_1 A_1' Y_{t-1} + \varepsilon_t$,

$$\Gamma_1 A_1' = \begin{bmatrix} -0.5 \\ 0 \\ -0.04 \end{bmatrix} \begin{bmatrix} 1 & 0.8 & 0 \end{bmatrix},$$

$$\Omega = \begin{bmatrix} 3.25 & -0.24 & -0.074 \\ -0.24 & 5.76 & 0.048 \\ -0.074 & 0.048 & 4.842 \end{bmatrix}$$

Experiment 3

Cointegrating rank and lag order: $p = 0$, $r = 2$,
Roots of $\det[I_m - J(L)L] = 0$: $L = 1.00, 1.01, 1.05$,
Error-correction form: $\Delta Y_t = \Gamma_2 A_2' Y_{t-1} + \varepsilon_t$,

$$\Gamma_2 A'_2 = \begin{bmatrix} 0 & 0.1 \\ 0 & -0.05 \\ -0.2 & 0.3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0.05 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\Omega = \begin{bmatrix} 1.30 & 0.99 & 0.641 \\ 0.99 & 0.81 & 0.009 \\ 0.641 & 0.009 & 5.85 \end{bmatrix}.$$

Experiment 4

Cointegrating rank and lag order: $p = 0$, $r = 2$,
Roots of $\det[I_m - J(L)L] = 0$: $L = 1.00, 2.00, 2.00$,
Error-correction form: $\Delta Y_t = \Gamma_2 A_2' Y_{t-1} + \varepsilon_t$,

$$\Gamma_2 A_2{}' = \begin{bmatrix} 0 & 0.1 \\ 0 & -0.5 \\ 0.1 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 0 \end{bmatrix}.$$

$$\Omega = \begin{bmatrix} 9.61 & -0.62 & 0.155 \\ -0.62 & 2.00 & 0.018 \\ 0.155 & 0.018 & 2.563 \end{bmatrix}.$$

Experiment 5

Cointegrating rank and lag order: $p = 1$, $r = 1$,
Roots of $\det[I_m - J(L)L] = 0$: $L = 1.00, 1.00, 1.01, 1.01, 1.05, 1.05$,
Error-correction form: $\Delta Y_t = J_1^* \Delta Y_{t-1} + \Gamma_1 A_1' Y_{t-1} + \varepsilon_t$,

$$\Gamma_1 A_1' = \begin{bmatrix} 0 \\ -0.01 \\ 0 \end{bmatrix} [1 \quad 0.25 \quad 0.8], \qquad J_1^* = \begin{bmatrix} 0.99 & 0 & 0 \\ 0 & 0.9025 & 0 \\ 0 & 0 & 0.99 \end{bmatrix},$$

$$\Omega = \begin{bmatrix} 2.25 & 2.55 & 1.95 \\ 2.55 & 3.25 & 2.81 \\ 1.95 & 2.81 & 2.78 \end{bmatrix}.$$

Experiment 6

Cointegrating rank and lag order: $p = 1$, $r = 1$,
Roots of $\det[I_m - J(L)L] = 0$: $L = 1.00, 1.00, 2.00, 2.00, 33.33, 50.00$,
Error-correction form: $\Delta Y_t = J_1^* \Delta Y_{t-1} + \Gamma_1 A_1' Y_{t-1} + \varepsilon_t$,

$$\Gamma_1 A_1' = \begin{bmatrix} 0 \\ 0.5 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -0.5 & 0.4 \end{bmatrix}, \qquad J_1^* = \begin{bmatrix} 0.02 & 0 & 0 \\ -0.5 & 0.25 & -0.2 \\ 0 & 0 & 0.03 \end{bmatrix},$$

$$\Omega = \begin{bmatrix} 4.00 & 3.60 & 4.40 \\ 3.60 & 3.40 & 4.20 \\ 4.40 & 4.20 & 5.24 \end{bmatrix}.$$

Experiment 7

Cointegrating rank and lag order: $p = 1$, $r = 2$,
Roots of $\det[I_m - J(L)L] = 0$: $L = 1.00, 1.01, 1.05, 1.05, 1.11, 1.11$,
Error-correction form: $\Delta Y_t = J_1^* \Delta Y_{t-1} + \Gamma_2 A_2' Y_{t-1} + \varepsilon_t$,

$$\Gamma_2 A_2' = \begin{bmatrix} -0.05 & 0 \\ 0 & 0 \\ 0 & -0.1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 6 \\ 0 & 1 & 0.05 \end{bmatrix}.$$

$$J_1^* = \begin{bmatrix} 0.855 & 0 & 0 \\ 0 & 0.99 & 0 \\ 0 & 0 & 0.855 \end{bmatrix}, \qquad \Omega = \begin{bmatrix} 17.64 & 10.08 & 10.92 \\ 10.08 & 6.40 & 7.20 \\ 10.92 & 7.20 & 8.24 \end{bmatrix}.$$

Experiment 8

Cointegrating rank and lag order: $p = 1$, $r = 2$,
Roots of $\det[I_m - J(L)L] = 0$: $L = 1.00, 2.00, 2.00, 2.00, 2.00, 10.00$,
Error-correction form: $\Delta Y_t = J_1^* \Delta Y_{t-1} + \Gamma_2 A_2' Y_{t-1} + \varepsilon_t$,

$$\Gamma_2 A_2{}' = \begin{bmatrix} -0.25 & 0 \\ 1.2 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.5 \end{bmatrix}.$$

$$J_1^* = \begin{bmatrix} 0.25 & 0 & 0 \\ -1.2 & 0.1 & 0 \\ 0 & -0.5 & 0.25 \end{bmatrix}, \quad \Omega = \begin{bmatrix} 5.76 & 10.08 & -8.16 \\ 10.08 & 18.00 & -15.18 \\ -8.16 & -15.18 & 13.90 \end{bmatrix}.$$

Table 1
Results of Experiment 1 ($r = 1$, $p = 0$)

PIC

r \ p	0	1	2	3	4	5	6
0	31	0	0	0	0	0	0
1	9811	1	0	0	0	0	0
2	157	0	0	0	0	0	0
3	0	0	0	0	0	0	0

BIC

r \ p	0	1	2	3	4	5	6
0	7	0	0	0	0	0	0
1	9663	0	0	0	0	0	0
2	312	0	0	0	0	0	0
3	18	0	0	0	0	0	0

AIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	4179	71	4	1	0	0	0
2	4562	91	5	0	0	0	0
3	1068	18	1	0	0	0	0

Note: Number of replications = 10,000; sample size $T = 150$.

The experiments are chosen to allow for data generating processes with different lag and rank order, ranging from VECMs with $p = 0$ and $r = 1$ as represented by Experiments 1 and 2 to VECMs with $p = 1$ and $r = 2$ as represented by the Experiments 7 and 8. Moreover, Experiments 1, 3, 5, and 7 were designed so that the 'stationary' roots of the characteristic polynomial of the VAR model lie in the range 0.90–0.99 (corresponding to the roots of $\det[I_m - J(L)L] = 0$ being in the range 1.01–1.11), which are considerably closer to the unit circle than the 'stationary' roots of the characteristic polynomial of the model represented by Experiments 2, 4, 6, and 8; which, in turn, are in the 0.02–0.5 range (which corresponds to the roots of $\det[I_m - J(L)L] = 0$ being in the range 2–50). Note that the maximum lag and rank order considered in these

Table 2
Results of Experiment 2 ($r = 1$, $p = 0$)

PIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	9966	0	0	0	0	0	0
2	34	0	0	0	0	0	0
3	0	0	0	0	0	0	0

BIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	9862	0	0	0	0	0	0
2	130	0	0	0	0	0	0
3	8	0	0	0	0	0	0

AIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	6166	90	9	1	0	0	0
2	3162	61	1	1	0	0	0
3	499	10	0	0	0	0	0

Note: Number of replications = 10,000; sample size $T = 150$.

experiments are $\bar{p} = 6$ and $\bar{r} = m = 3$. In addition, choices of p and r from BIC and AIC were obtained by minimizing the criteria:

$$\mathrm{BIC}(p,r) = \ln|\hat{\Omega}(p,r)| + \{m^2 p + mr + r(m-r)\}\ln(T)/T,$$

$$\mathrm{AIC}(p,r) = \ln|\hat{\Omega}(p,r)| + \{m^2 p + mr + r(m-r)\}2/T,$$

where $\hat{\Omega}(p,r)$ is the residual covariance matrix from a fitted reduced rank regression.

Results from the eight experiments based on 10,000 replications are presented in Tables 1–8. Table 9 reports the average bias and standard deviations of rank

Table 3
Results of Experiment 3 ($r = 2$, $p = 0$)
PIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	588	0	0	0	0	0	0
2	9400	0	0	0	0	0	0
3	12	0	0	0	0	0	0

BIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	250	0	0	0	0	0	0
2	9129	0	0	0	0	0	0
3	621	0	0	0	0	0	0

AIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	8	0	0	0	0	0	0
2	7082	139	12	5	0	1	0
3	2684	64	5	0	0	0	0

Note: Number of replications = 10,000; sample size $T = 150$.

and lag estimation (where averages are taken over the 8 experiments) as computed from the empirical distributions generated by our experiments. In all eight experiments, PIC outperforms both BIC and AIC in cointegrating rank selection although BIC produces a correct lag choice with slightly greater frequency than PIC. Overall, the probability of a correct model choice (i.e., correct choice of both the lag length and the cointegrating rank) by PIC exceeds that of BIC by about 0.04 on average and that of AIC by about 0.39 on average. Moreover, PIC also exhibits the least variation in rank selection with an average standard deviation of 0.144 over the eight experiments as opposed to 0.238 for BIC and 0.514 for AIC. Clearly, AIC is the worst performer in terms of both

Table 4
Results of Experiment 4 ($r = 2$, $p = 0$)

PIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	9900	0	0	0	0	0	0
3	100	0	0	0	0	0	0

BIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	9677	0	0	0	0	0	0
3	323	0	0	0	0	0	0

AIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	7939	136	8	1	1	0	0
3	1889	26	0	0	0	0	0

Note: Number of replications = 10,000; sample size $T = 150$.

Table 5
Results of Experiment 5 ($r = 1$, $p = 1$)
PIC

r \ p	0	1	2	3	4	5	6
0	0	742	2	0	0	0	0
1	0	9248	0	0	0	0	0
2	0	8	0	0	0	0	0
3	0	0	0	0	0	0	0

BIC

r \ p	0	1	2	3	4	5	6
0	0	43	0	0	0	0	0
1	0	7825	0	0	0	0	0
2	0	1917	0	0	0	0	0
3	0	215	0	0	0	0	0

AIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	834	18	1	0	0	0
2	0	6539	176	15	1	0	0
3	0	2326	74	13	3	0	0

Note: Number of replications = 10,000; sample size $T = 150$.

rank and lag selection. Note that relative to PIC and BIC, AIC shows a much greater tendency to overestimate both the cointegrating rank and the order of lagged differences. With respect to lag order estimation, our Monte Carlo evidence is entirely in accord with the asymptotic analyses of Shibata (1976) and Tsay (1984), which show AIC to be inconsistent in the sense that the probability of overestimation under this criterion does not approach zero as sample size approaches infinity. That our experiments also find AIC to overestimate the cointegrating rank with great regularity leads us to conjecture that it is similarly inconsistent for cointegrating rank estimation.

A surprising result from these experiments is that while a priori we would expect BIC to perform well relative to PIC in cases where the ‘stationary’ roots

Table 6
Results of Experiment 6 ($r = 1$, $p = 1$)

PIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	9963	0	0	0	0	0
2	0	37	0	0	0	0	0
3	0	0	0	0	0	0	0

BIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	9858	0	0	0	0	0
2	0	138	0	0	0	0	0
3	0	4	0	0	0	0	0

AIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	5921	116	10	0	0	0
2	0	3311	61	5	2	0	0
3	0	565	8	1	0	0	0

Note: Number of replications = 10,000; sample size $T = 150$.

are closer to the unit circle; given that, ceteris paribus, the latter tends to favor specifications with fewer cointegration relationships; the opposite seems to hold true in our experiments. The two experiments where PIC has most dramatically outperformed BIC are experiments 5 and 7, where the ‘stationary’ roots are in the 0.90–0.99 range and where BIC, counter-intuitively, has shown a heightened tendency to overselect the cointegrating rank. It turns out that in these cases reduced rank regression given the correct cointegrating rank often does not result in a good fit; in fact, overparameterizing the number of cointegrating relationships often results in a better fit. Moreover, while the penalty function of PIC is strong enough to overcome the inclination to overfit in these cases, that

Table 7
Results of Experiment 7 ($r = 2$, $p = 1$)
PIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	724	2	0	0	0	0
2	0	9273	1	0	0	0	0
3	0	0	0	0	0	0	0

BIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	3	0	0	0	0	0
2	0	8764	0	0	0	0	0
3	0	1233	0	0	0	0	0

AIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	0	6177	162	9	0	2	0
3	0	3559	84	5	1	0	1

Note: Number of replications = 10,000; sample size $T = 150$.

of BIC is not, thus resulting in more incorrect choices by BIC in the direction of overselection.

These results speak favorably of our criterion. We attribute the good performance of PIC to a penalty function that takes into account not only the number of parameters but also the nonstationarity of the regressors associated with some of the parameters.

5. Conclusion

This paper takes a model selection approach to the problem of determining the cointegrating rank. More specifically, we extend the analysis of Phillips and

Table 8
Results of Experiment 8 ($r = 2$, $p = 1$)

PIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	0	9987	1	0	0	0	0
3	0	12	0	0	0	0	0

BIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	0	9664	0	0	0	0	0
3	0	336	0	0	0	0	0

AIC

r \ p	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	0	7935	145	14	2	1	0
3	0	1855	45	3	0	0	0

Note: Number of replications = 10,000; sample size $T = 150$.

Table 9
Avg. Bias and Std. Deviation of Rank and Lag Selection

Method	Avg. Bias of $\hat{r}$[a]	Avg. Std. Dev. of $\hat{r}$[a]	Avg. Bias of $\hat{p}$[a]	Avg. Std. Dev. of $\hat{p}$[a]
PIC	0.030	0.144	0.0001	0.006
BIC	0.065	0.238	0	0
AIC	0.342	0.514	0.024	0.169

[a]Average is taken over the eight experiments.

Ploberger (1996) to a vector autoregressive process with reduced rank structure. There are four principal advantages to this approach. First, it provides a coherent framework under which the VECM lag order p and the cointegrating rank r can be jointed selected. Secondly, it leads to consistent estimation of both p and r. Third, the method is extremely easy to use in practice,[2] involves minimal computation and does not require the use of complicated statistical tables. Finally, the penalty function implicit in our criterion takes into account not only the number of parameters but also the nonstationarity of the regressors associated with some of the parameters. This latter attribute, we believe, explains why our criterion performed well relative to BIC and AIC in the simulation experiments presented in Section 4.

The methods given here can be generalized in several ways. The time series model investigated in this paper has neither a deterministic nor a moving-average component. However, these components are important in some econometric models. A natural extension of the methods allows decisions to be made with respect to the trend degree and the order of the moving average component. Models for scalar time series with these features were studied using similar model selection methods in Phillips and Ploberger (1994). As indicated in Remark 3.5(vi), extensions of the approach given here to semiparametric models of reduced rank with general time series errors are also of interest. We hope to report at a later time some progress on extensions along these lines.

Acknowledgements

The authors would like to thank Don Andrews, Oliver Linton, Chris Sims, and two anonymous referees for helpful comments and suggestions. This paper is a revised version of Chapter 1 of Chao's Yale University Ph.D. Dissertation and was first drafted in January 1994. Chao thanks the Cowles Foundation for financial support during his time as a graduate student, and thanks the second author for his continuing guidance during the course of this project. The paper was revised while Phillips visited the University of Auckland in the (southern) summer and fall of 1997 and Phillips thanks the University for its hospitality during this visit. Chao's research was supported by the Office of Graduate Studies and Research of the University of Maryland. Phillips's research was supported by the NSF under grant no.'s SES 91-22142 and SES 94-22922.

---

[2] Some empirical applications in ex ante macroeconomic forecasting are reported in Phillips (1995b).

Appendix A

For notational simplicity, we shall, throughout this and the subsequent appendices, suppress the indices p and r and write $\Gamma, \bar{A}, F$, and W instead of Γ_r, $\bar{A}_r$, $F(r)$, and $W(p)$, whenever we are not using the same symbols to denote parameter or data matrices of different dimensions.

A.1. Maximum likelihood estimation

In this section we will briefly describe our procedure for obtaining maximum likelihood estimators for the model described by Eq. (2), assuming that Ω is known. The maximization is carried out in stages. First, note that we can maximize the likelihood (as given by expression (3)) with respect to $J^* = (J_1^*, \dots, J_p^*)$ and obtained the concentrated log-likelihood:

$$\ell(\Gamma, \bar{A}|\Omega, \text{data}) = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T} [u_t'\Omega^{-1}u_t - u_t'(\Omega^{-1}\otimes W_t') \times \sum_{t=1}^{T} (\Omega^{-1}\otimes W_t W_t'))^{-1} \sum_{t=1}^{T} (\Omega^{-1}\otimes W_t)u_t], \tag{A.1}$$

where

$$W_t = (\Delta Y_{t-1}', \dots, \Delta Y_{t-p}')',$$

$$u_t = \Delta Y_t - (I_m \otimes Y_{1t-1}')\text{vec}\,\Gamma - (I_m \otimes Y_{2t-1}')(I_m \otimes \bar{A})\text{vec}\,\Gamma$$

$$= \Delta Y_t - (I_m \otimes Y_{1t-1}')\text{vec}\,\Gamma - (I_m \otimes Y_{2t-1}')(\Gamma \otimes I_{m-r})\text{vec}\,\bar{A}'.$$

Next, following Ahn and Reinsel (1990) we note that the parameter vector $\beta = [(\text{vec}\,\bar{A}')', (\text{vec}\,\Gamma)']'$ can be estimated using the approximate Newton–Raphson relations:[3]

$$\hat{\beta}_{(i+1)} = \hat{\beta}_{(i)} + (H(\Omega^{-1}\otimes Y_{-1}'M_W Y_{-1})H')_{\hat{\beta}_{(i)}}^{-1}(H(\Omega^{-1}\otimes Y_{-1}'M_W)\text{vec}\,U')_{\hat{\beta}_{(i)}}, \tag{A.2}$$

[3] As our estimation procedure parallel that of Ahn and Reinsel (1990), only a very abbreviated account is given here. Interested readers are referred to Ahn and Reinsel (1990) and an earlier version of this paper, Chao and Phillips (1997), for more details.

where

$$U' = [u_1, \dots, u_T],$$

$$H = \begin{pmatrix} (\Gamma' \otimes F') \\ (I_m \otimes (I_r, \overline{A}')) \end{pmatrix}.$$

A.2. Review of the relevant background asymptotics

Here, we review some properties of the VECM given by Eq. (2) in Section 2 as well as give some asymptotic results derived under the assumption that the model given by Eq. (2) is correctly specified with respect to p and r. The discussion here is useful in the development of our own asymptotic analysis. The treatment here follows that of Toda and Phillips (1993) and Ahn and Reinsel (1990). To begin, we define the $m \times 1$ vector $Z_t = (Z'_{1t}, Z'_{2t})' = (Y'_t A_\perp, Y'_t A)'$. Write $v_t = (\varepsilon'_t, \Delta Z'_{1t}, Z'_{2t}, W'_t)'$ and define the long-run covariance matrix Σ such that

$$\Sigma = \Sigma^* + \Lambda + \Lambda', \tag{A.3}$$

where

$$\Sigma^* = \mathrm{E}(v_t v'_t),$$

$$\Lambda = \sum_{j=1}^{\infty} \mathrm{E}(v_t v'_{t+j}).$$

We often find it convenient to partition Σ, Σ^*, and Λ conformably with v_t so, for example, we can write

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{pmatrix}, \tag{A.4}$$

where the indices '1', '2', '3', and '4' correspond to $\varepsilon_t, \Delta Z_{1t}, Z_{2t}$, and W_t, respectively. Note in particular that $\Sigma_{11} = \Omega$ since $\mathrm{E}(\varepsilon_t \varepsilon_{t+j}) = 0$ for all $j \geqslant 1$. Note further that making use of Eq. (2), we can write $(Z'_{2t+1}, W'_{t+1})'$ as the first order system

$$\begin{pmatrix} Z_{2t+1} \\ W_{t+1} \end{pmatrix} = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} = \begin{pmatrix} Z_{2t} \\ W_{2t} \end{pmatrix} + \begin{pmatrix} A' \\ e_{p-1} \otimes I_m \end{pmatrix} \varepsilon_t, \tag{A.5}$$

where e_{p-1} is a $(p-1)\times 1$ vector such that $e_{p-1} = (1, 0, \dots, 0)'$ and where

$$\Phi_{11} = A'\Gamma + I_r,$$

$$\Phi_{12} = (A'J_1^*, \dots, A'J_p^*)$$

$$\Phi_{21} = \begin{pmatrix} \Gamma \\ 0 \end{pmatrix},$$

$$\Phi_{22} = \begin{pmatrix} J_1^*, \dots, J_{p-1}^* & J_p^* \\ I_{m(p-1)} & 0 \end{pmatrix}.$$

Since Z_{2t} and W_t are $I(0)$ processes, the eigenvalues of the matrix

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}$$

must be outside the unit circle, and we can write Eq. (A.5) in the moving average representation

$$\begin{aligned} \begin{pmatrix} Z_{2t} \\ W_t \end{pmatrix} &= \begin{pmatrix} \Theta_{11}(L) & \Theta_{12}(L) \\ \Theta_{21}(L) & \Theta_{22}(L) \end{pmatrix} \begin{pmatrix} A' \\ e_{p-1} \otimes I_m \end{pmatrix} \varepsilon_{t-1} \\ &= \Theta(L) \Xi \varepsilon_{t-1} \\ &= \sum_{i=0}^{\infty} \Phi^j \Xi \varepsilon_{t-1-j} \end{aligned} \tag{A.6}$$

where

$$\Xi = \begin{pmatrix} A' \\ e_{p-1} \otimes I_m \end{pmatrix}.$$

We shall next discuss a few lemmas which are used in the proofs of Theorems 3.1 and 3.3 and Corollary 3.4. Before proceeding, however, let us first introduce some more notations. First, define $\underline{A}' = [A_\perp, A]$ and partition $\underline{A}'$ further as

$$\underline{A}' = \begin{bmatrix} A_{\perp_1} & A_1 \\ A_{\perp_2} & A_2 \end{bmatrix} = \begin{bmatrix} A_{\perp_1} & I_r \\ A_{\perp_2} & A_2 \end{bmatrix},$$

where $A_{\perp_1}$, $A_{\perp_2}$, and A_2 are, respectively, $r\times(m-r)$, $(m-r)\times(m-r)$, and $(m-r)\times r$. Also define P to be the inverse of $\underline{A}$ so that $P\underline{A}=\underline{A}P=I_m$ and partition P conformably to $\underline{A}'$ as

$$P=\begin{bmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{bmatrix}.$$

In addition, let $Z_1=Y_{-1}A_\perp=[Z_{10},\dots,Z_{1T-1}]'$ and $Z_2=Y_{-1}A=[Z_{20},\dots,Z_{2T-1}]'$ and let $W_d(s)$ $(s\in[0,1])$ be a d-dimensional standard Brownian motion.

Lemma A.1. Let data be generated by a process of the form (2) *under assumptions* (*i*)–(*iv*) *given in Section* 2, *then the following convergence results hold as* $T\to\infty$.

$$\text{(a)}\ \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\varepsilon_t\Rightarrow B_0(s)\equiv\Omega^{1/2}W(s),$$

$$\text{(b)}\ \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\Delta Z_{1t}\Rightarrow A'_\perp[I_m+(\Gamma,J_1^*,\dots,J_{p-1}^*)\Theta(1)\psi]B_0(s)\equiv B_1(s),$$

$$\text{(c)}\ \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\begin{bmatrix}Z_{2t}\\ W_t\end{bmatrix}\Rightarrow\begin{bmatrix}(\Theta_{11}(1)A'+\Theta_{12}(1)(e_{p-1}\otimes I_m))B_0(s)\\ (\Theta_{21}(1)A'+\Theta_{22}(1)(e_{p-1}\otimes I_m))B_0(s)\end{bmatrix}\equiv\begin{bmatrix}B_2(s)\\ B_3(s)\end{bmatrix}.$$

Proof. See Lemma 1 of Toda and Phillips (1993). □

Lemma A.2. Under the same assumptions as Lemma 1, the following convergence results hold as $T\to\infty$:

$$\text{(a)}\ T^{-2}Z_1'M_WZ_1\Rightarrow\int_0^1 B_1(s)B_1(s)'\,\mathrm{d}s,$$

$$\text{(b)}\ T^{-2}F'Y'_{-1}M_WY_{-1}F\Rightarrow P_{21}\int_0^1 B_1(s)B_1(s)'\,\mathrm{d}s\,P'_{21},$$

$$\text{(c)}\ T^{-1}U'M_WZ_1\Rightarrow\left\{\int_0^1 B_1(s)\,\mathrm{d}B_0(s)'\right\}',$$

(d) $T^{-1}U'M_W Y_{-1}F \Rightarrow \left\{P_{21}\int_0^1 B_1(s)\,\mathrm{d}B_0(s)'\right\}'$,

(e) $Z_2'Z_2/T \xrightarrow{p} \Sigma_{33}^*$,

(f) $Z_2'W/T \xrightarrow{p} \Sigma_{34}^*$,

(g) $W'W/T \xrightarrow{p} \Sigma_{44}^*$,

(h) $\mathrm{vec}(U'M_W Z_2/\sqrt{T}) \Rightarrow \underset{\sim}{\mathrm{N}}(0\ ,(\Omega \otimes \Sigma_{33.4}^*))$,

where $\Sigma_{33.4}^* = \Sigma_{33}^* - \Sigma_{34}^*\Sigma_{44}^{*-1}\Sigma_{43}^*$,

(i) $F'Y_{-1}'M_W Y_{-1}\hat{A}/T^{3/2} \xrightarrow{p} 0$.

Proof. All results follow directly from Lemma A.1, the continuous mapping theorem, and arguments analogous to those used in Lemma 2.1 of Phillips and Park (1989). □

Lemma A.3. Let $\hat{\beta} = [(\mathrm{vec}\,\hat{\bar{A}}')', (\mathrm{vec}\,\hat{\Gamma})']'$ *be the Gaussian maximum likelihood estimator generated by the iterative relation* (*A.2*), *then*

$$\sqrt{T}(\mathrm{vec}\,\hat{\Gamma} - \mathrm{vec}\,\Gamma^0) \Rightarrow \underset{\sim}{\mathrm{N}}(0\ ,(\Omega \otimes \Sigma_{33.4}^{*-1})),$$

$$T(\hat{\bar{A}} - \bar{A}^0)' \Rightarrow (\Gamma^{0\prime}\Omega^{0^{-1}}\Gamma^0)^{-1}\Gamma^{0\prime}\Omega^{0^{-1}}\left(\int_0^1 B_1(s)\,\mathrm{d}B_0(s)'\right)'$$

$$\left(\int_0^1 B_1(s)B_1(s)'\,\mathrm{d}s\right)^{-1} P_{21}^{-1}$$

Proof. See Theorem 2 of Ahn and Reinsel (1990). □

Lemma A.4. *Let $\hat{J}_* = \Delta Y' M_W Y_{-1}(Y'_{-1} M_W Y_{-1})^{-1}$ be the least squares estimator for the model described by Eq. (2), then*

$$(\hat{J}_* - J_*^0)PD \Rightarrow [R, S],$$

where

$$D = \mathrm{diag}(TI_{m-r}, \sqrt{T}I_r),$$

$$R = \left\{\int_0^1 B_1(s)\,\mathrm{d}B_0(s)'\right\}'\left(\int_0^1 B_1(s)B_1(s)'\,\mathrm{d}s\right)^{-1},$$

$$S = \underset{\sim}{\mathrm{N}}(0, (\Omega \otimes \Sigma_{33.4}^{*-1})).$$

Proof. See Theorem 1 of Ahn and Reinsel (1990). □

Lemma A.5. *Consider the model given by Eq. (2) under Assumptions (i)–(iv); the likelihood ratio statistic for testing the null hypothesis that the cointegrating rank = r has the asymptotic distribution given by*

$$\mathrm{tr}\{\hat{\Omega}^{-1}(\tilde{J}_*(p,r) - \hat{J}_*(p))Y'_{-1}M_{W(p)}Y_{-1}(\tilde{J}_*(p,r) - \hat{J}_*(p))'\}$$

$$\Rightarrow \mathrm{tr}\left\{\left(\int_0^1 W_{m-r}(s)\,\mathrm{d}W_{m-r}(s)'\right)'\left(\int_0^1 W_{m-r}(s)\,\mathrm{d}W_{m-r}(s)'\,\mathrm{d}s\right)^{-1}\right.$$

$$\left.\times\left(\int_0^1 W_{m-r}(s)\,\mathrm{d}W_{m-r}(s)'\right)\right\},$$

where $\tilde{J}_*(p,r) = (\hat{\Gamma}(p,r), \hat{\Gamma}(p,r)\hat{\hat{A}}(p,r)')$ with $\hat{\Gamma}(p,r)$ and *$\hat{\hat{A}}(p,r)$ being given by the iterative relations in Eq. (A.2) and $\hat{J}_*(p) = \Delta Y' M_{W(p)} Y_{-1}(Y'_{-1}M_{W(p)}Y_{-1})^{-1}$.*

Proof. See Theorem 1 of Reinsel and Ahn (1992). □

Appendix B

Proof of Theorem 3.1. To show that $(\hat{p}, \hat{r}) \xrightarrow{p} (p^0, r^0)$, we need to show that for all $p \neq p^0$ and/or $r \neq r^0$

$$\mathbb{P}(\hat{\Pi}_T(M_{p^0,r^0}|\hat{\Omega}, Y)/\hat{\Pi}_T(M_{p,r}|\hat{\Omega}, Y) > 1) \rightarrow 1 \quad \text{as } T \rightarrow \infty.$$

This will certainly be true if for all $p \neq p^0$ and $r \neq r^0$,

$$\frac{\hat{\Pi}_T(M_{p^0,r^0}\,|\,\hat{\Omega},Y)}{\hat{\Pi}_T(M_{p,r}\,|\,\hat{\Omega},Y)} \to \infty \quad \text{in probability as } T \to \infty. \tag{B.1}$$

We shall check this divergence only for cases where either $p \neq p^0$ or $r \neq r^0$, as the analysis for cases where $p \neq p^0$ and $r \neq r^0$ follow analogously.

Now consider the case where $r > r^0$ and $p = p^0$. From expression (5), we have

$$\begin{aligned}\frac{\hat{\Pi}_T(M_{p^0,r^0}\,|\,\hat{\Omega},Y)}{\hat{\Pi}_T(M_{p^0,r}\,|\,\hat{\Omega},Y)} &= |\tilde{H}(p^0,r^0)(\hat{\Omega}^{-1}\otimes Y'_{-1}M_{W(p^0)}Y_{-1})\tilde{H}(p^0,r^0)|^{-1/2}\\ &\times|\tilde{H}(p^0,r)(\hat{\Omega}^{-1}\otimes Y'_{-1}M_{W(p^0)}Y_{-1})\tilde{H}(p^0,r)|^{1/2}\\ &\times\exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\Delta Y - Y_{-1}\tilde{J}_*(p^0,r)')'\\ &\times M_{W(p^0)}(\Delta Y - Y_{-1}\tilde{J}_*(p^0,r)')]\}\\ &\times\exp\{-\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\Delta Y - Y_{-1}\tilde{J}_*(p^0,r^0)')'\\ &\times M_{W(p^0)}(\Delta Y - Y_{-1}\tilde{J}_*(p^0,r^0)')]\},\end{aligned}$$

where $\tilde{H}(.,.)$ is as defined earlier in Section 3. It follows that we can rewrite the expression above as:

$$\begin{aligned}\frac{\hat{\Pi}_T(M_{p^0,r^0}\,|\,\hat{\Omega},Y)}{\hat{\Pi}_T(M_{p^0,r}\,|\,\hat{\Omega},Y)} &= (|\hat{\Gamma}(p^0,r^0)'\hat{\Omega}^{-1}\hat{\Gamma}(p^0,r^0)\otimes F(r^0)'Y'_{-1}M_{W(p^0)}Y_{-1}F(r^0)|\\ &\times|(\hat{\Omega}^{-1}\otimes[I_{r^0},\hat{\bar{A}}(p^0,r^0)']Y'_{-1}M_{W(p^0)}Y_{-1}[I_{r^0},\hat{\bar{A}}(p^0,r^0)']')\\ &-([\hat{\Omega}^{-1}\hat{\Gamma}(p^0,r^0)(\hat{\Gamma}(p^0,r^0)'\hat{\Omega}^{-1}\hat{\Gamma}(p^0,r^0))^{-1}\hat{\Gamma}(p^0,r^0)'\hat{\Omega}^{-1}]\\ &\otimes[I_{r^0},\hat{\bar{A}}(p^0,r^0)']Y'_{-1}M_{W(p^0)}Y_{-1}F(r^0)\\ &\times(F(r^0)'Y'_{-1}M_{W(p^0)}Y_{-1}F(r^0))^{-1}F'(r^0)\\ &\times Y'_{-1}M_{W(p^0)}Y_{-1}[I_{r^0},\hat{\bar{A}}(p^0,r^0)']')|)_1^{-1/2}\\ &\times(|\hat{\Gamma}(p^0,r)'\hat{\Omega}^{-1}\hat{\Gamma}(p^0,r)\otimes F(r)'Y'_{-1}M_{W(p^0)}Y_{-1}F(r)|\\ &\times|(\hat{\Omega}^{-1}\otimes[I_r,\hat{\bar{A}}(p^0,r)']\end{aligned}$$

$$\times Y'_{-1}M_{W(p^0)}Y_{-1}[I_r, \hat{\bar{A}}(p^0,r)']') - ([\hat{\Omega}^{-1}\hat{\Gamma}(p^0,r)(\hat{\Gamma}(p^0,r)'$$

$$\times \hat{\Omega}^{-1}\hat{\Gamma}(p^0,r))^{-1}\hat{\Gamma}(p^0,r)'\hat{\Omega}^{-1}]\otimes[I_r, \hat{\bar{A}}(p^0,r)']$$

$$\times Y'_{-1}M_{W(p^0)}Y_{-1}F(r)(F(r)'Y'_{-1}M_{W(p^0)}Y_{-1}F(r))^{-1}$$

$$\times F(r)'Y'_{-1}M_{W(p^0)}Y_{-1}[I_r, \hat{\bar{A}}(p^0,r)']')|)_2^{1/2}$$

$$\times \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0,r) - \hat{J}_*(p^0))Y'_{-1}M_{W(p^0)}Y_{-1}$$

$$\times (\tilde{J}_*(p^0,r) - \hat{J}_*(p^0))']\}$$

$$\times \exp\{-\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0,r^0) - \hat{J}_*(p^0))Y'_{-1}M_{W(p^0)}Y_{-1}$$

$$\times (\tilde{J}_*(p^0,r^0) - \hat{J}_*(p^0))']\}, \tag{B.2}$$

where $\tilde{J}_*(.,.)$ and $\hat{J}_*(\cdot)$ are as defined earlier in Section 3. By Lemma A.2

$$T^{-2}(F(r^0)'Y'_{-1}M_{W(p^0)}Y_{-1}F(r^0)) \equiv O_p(1),$$

$$[I_{r^0}, \hat{\bar{A}}(p^0,r^0)']T^{-1}(Y'_{-1}M_{W(p^0)}Y_{-1})[I_{r^0}, \hat{\bar{A}}(p^0,r^0)']' \equiv O_p(1),$$

$$[I_{r^0}, \hat{\bar{A}}(p^0,r^0)']T^{-3/2}(Y'_{-1}M_{W(p^0)}Y_{-1})F(r^0))(T^{-2}(F(r^0)'Y'_{-1}M_{W(p^0)}Y_{-1}F(r^0)))$$

$$\times T^{-3/2}(F(r^0)'Y'_{-1}M_{W(p^0)}Y_{-1})[I_{r^0}, \hat{\bar{A}}(p^0,r^0)']' \equiv o_p(1).$$

We now write the complex expression (B.2) in the symbolic form

$$(\cdot)_1^{-1/2}(\cdot)_2^{1/2}\exp\{(1/2(\cdot)_a\}\exp\{-(1/2)(\cdot)_b\}, \tag{B.3}$$

where $(\cdot)_1$ and $(\cdot)_2$ represent the numbered bracketed factors that appear in Eq. (B.2) and $(\cdot)_a$ and $(\cdot)_b$ denote the tr$[\cdot]$ expressions that appear in the final two exponential terms, respectively, of Eq. (B.2).

Some simple scaling manipulations confirm that

$$(\cdot)_1 = O_p(T^{2(m-r^0)r^0+mr^0}) = O_p(T^{3mr^0-2r^{02}}). \tag{B.4}$$

To evaluate the order of $(\cdot)_2$, we need to transform the regressor space to isolate components of different orders of magnitude. For instance, since $r > r^0$ we know that the term

$$[I_r, \hat{\bar{A}}(p^0,r)']Y'_{-1}M_{W(p^0)}Y_{-1}[I_r, \hat{\bar{A}}(p^0,r)']'$$

has a first diagonal $r^0 \times r^0$ sub-block of $O_p(T)$ and a second diagonal sub-block of $O_p(T^2)$, corresponding to the limiting cointegrating submatrix of $[I_r, \hat{\underline{A}}(p^0, r)']$ of order $r^0 \times m$ and its complement, respectively. (Any rotation of the coordinate system that is used to achieve this will not affect the orders of magnitude of the final determinantal form). Proceeding in this way with each element of $(\cdot)_2$ and using the methods outlined in Phillips (1988) we obtain

$$(\cdot)_2 = O_p(T^{\{2(m-r)r+[r^0-2(r-r^0)]m\}}) = O_p(T^{4mr-2r^2-r^0m}). \tag{B.5}$$

Combining Eqs. (B.4) and (B.5) we have

$$(\cdot)_1^{-1}(\cdot)_2 = O_p(T^{4mr-4mr^0-2r^2+2r^{02}}).$$

We observe that the exponent in this order of magnitude is

$$4m(r-r^0) - 2(r-r^0)(r+r^0) = 2(r-r^0)\{2m-(r+r^0)\} > 0$$

for all $r > r^0$. Thus the 'penalty' term in Eq. (B.2) is

$$\{(\cdot)_1^{-1}(\cdot)_2\}^{1/2} = O_p(T^{(r-r^0)\{2m-(r+r^0)\}}) \tag{B.6}$$

which diverges to ∞ in probability for all $r > r^0$.

Finally, we consider the expressions in the exponents of the exponential factors of Eqs. (B.2) and (B.3). We start with $(\cdot)_b$. Note that by Lemma A.5,

$$(\cdot)_b = \mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0, r^0) - \hat{J}_*(p^0))Y'_{-1}M_{W(p^0)}Y_{-1}(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))']$$
$$\equiv O_p(1).$$

Next consider $(\cdot)_a$. We have

$$(\cdot)_a = \mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))Y'_{-1}M_{W(p^0)}Y_{-1}(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))']$$
$$= \mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))PD(D^{-1}\underline{A}Y'_{-1}M_{W(p^0)}Y_{-1}\underline{A}'D^{-1})$$
$$\times DP'(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))']$$

where $D = \mathrm{diag}(TI_{m-r^0}, \sqrt{T}I_{r^0})$.

Note that by the arguments of Theorem 1 of Ahn and Reinsel (1990)

$$D^{-1}\underline{A}Y'_{-1}M_{W(p^0)}Y_{-1}\underline{A}'D^{-1} \equiv O_p(1). \tag{B.7}$$

Add and subtract J_*^0 from $(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))PD$ and get

$$(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))PD = (\tilde{J}_*(p^0, r) - J_*^0)PD + (J_*^0 - \hat{J}_*(p^0))PD.$$

Again, by the arguments of Theorem 1 of Ahn and Reinsel (1990), we see that

$$(J_*^0 - \hat{J}_*(p^0))PD \equiv O_p(1). \tag{B.8}$$

Next, partition

$$\hat{\Gamma}(p^0, r) = \left[\underset{m\times r^0}{\hat{\Gamma}_*}, \underset{m\times(r-r^0)}{\hat{\Gamma}_{**}}\right] \quad \text{and} \quad \hat{A}(p^0, r) = \left[\underset{m\times r^0}{\hat{A}_*}, \underset{m\times(r-r^0)}{\hat{A}_{**}}\right]$$

in a conformable way corresponding to the true number of columns (r^0) and supplementary columns $r - r^0$. The columns in these partitions are ordered according to the size of the corresponding eigenvalues in the associated reduced rank regression in the usual way. We decompose $(\tilde{J}_*(p^0, r) - J_*^0)PD$ as follows:

$$\begin{aligned}(\tilde{J}_*(p^0, r) - J_*^0)PD &= (\hat{\Gamma}_*\hat{A}_*' + \hat{\Gamma}_{**}\hat{A}_{**}' - \Gamma^0 A^{0\prime})PD \\ &= [(\hat{\Gamma}_* - \Gamma^0)A^{0\prime} + \Gamma^0(\hat{A}_* - A^0)' + \hat{\Gamma}_{**}\hat{A}_{**}' \\ &\quad + \mathrm{op}(T^{-1})]PD.\end{aligned}$$

Now, $A^{0\prime}PD = [0, T^{1/2}I_{r^0}]$, so that $(\hat{\Gamma}_* - \Gamma^0)A^{0\prime}PD = O_p(1)$. Also, since $T(\hat{A}_* - A^0) = O_p(1)$ we have $\Gamma^0(\hat{A}_* - A^0)'PD = O_p(1)$. Finally, $\hat{A}_{**} = O_p(1)$, just as in the spurious regression analysis of Phillips (1986), and $\hat{\Gamma}_{**} = O_p(T^{-1})$ (being the coefficient of $\hat{A}_{**}'Y_{t-1}$, which is an $I(1)$ regressor with random coefficients in the limit). Thus $\hat{\Gamma}_{**}\hat{A}_{**}'PD = O_p(1)$, and we have

$$(\tilde{J}_*(p^0, r) - J_*^0)PD = O_p(1). \tag{B.9}$$

Combining Eqs. (B.7), (B.8) and (B.9) we find that $(\cdot)_a = O_p(1)$. Thus both $(\cdot)_a$ and $(\cdot)_b$ are $O_p(1)$. The penalty term (B.6) therefore dominates when $r > r^0$ and we deduce that

$$\frac{\hat{\Pi}_T(M_{p^0,r^0} \mid \hat{\Omega}, Y)}{\hat{\Pi}_T(M_{p^0,r} \mid \hat{\Omega}, Y)} \to \infty \quad \text{in probability as } T \to \infty,$$

as required.

Now, if $r < r^0$ we again write

$$(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))PD = (\tilde{J}_*(p^0, r) - J^0_*)PD + (J^0_* - \hat{J}_*(p^0))PD.$$

We have $(J^0_* - \hat{J}_*(p^0))PD = O_p(1)$ just as in Eq. (B.8). We partition

$$\Gamma^0 = \left[\underset{m\times r}{\Gamma^0_*}, \underset{m\times(r^0-r)}{\Gamma^0_{**}}\right] \quad \text{and} \quad A^0 = \left[\underset{m\times r}{A^0_*}, \underset{m\times(r^0-r)}{A^0_{**}}\right]$$

conformably and then

$$\begin{aligned}(\tilde{J}_*(p, r) - J^*_0)PD &= [\hat{\Gamma}\hat{A}' - \Gamma^0_* A^{0\prime}_* - \Gamma^0_{**}A^{0\prime}_{**}]PD \\ &= [(\hat{\Gamma} - \Gamma^0_*)A^{0\prime}_* + \Gamma^0_*(\hat{A} - A^0_*)' + \mathrm{op}(T^{-1})] \\ &\quad \times PD - \Gamma^0_{**}A^{0\prime}_{**}PD = O_p(1) + O_p(T).\end{aligned}$$

It follows that

$$\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0, r) - J_*(p^0))Y'_{-1}M_{W(p^0)}Y_{-1}(\tilde{J}_*(p^0, r) - \hat{J}_*(p^0))] \equiv O_p(T^2). \tag{B.10}$$

Thus, once again

$$\frac{\hat{\Pi}_T(M_{p^0,r^0} \mid \hat{\Omega}, Y)}{\hat{\Pi}_T(M_{p^0,r} \mid \hat{\Omega}, Y)} \to \infty \quad \text{in probability as } T \to \infty.$$

This time (when $r < r_0$) the exponential term dominates the asymptotic behavior of our criterion.

Now, if instead we have the case where $r = r^0$ but $p > p^0$, then partition

$$W(p) = \left[\underset{T\times mp^0}{W(p^0)}, \underset{T\times m(p-p^0)}{W(*)}\right]$$

and we can write the PIC as

$$\begin{aligned}\frac{\hat{\Pi}_T(M_{p^0,r^0} \mid \hat{\Omega}, Y)}{\hat{\Pi}_T(M_{p,r^0} \mid \hat{\Omega}, Y)} &= \Psi_1(p, p^0, r^0)\Psi_2(p, p^0) \\ &\quad \times \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p, r^0) - \hat{J}_*(p)) \\ &\quad Y'_{-1}M_{W(p)}Y_{-1}(\tilde{J}_*(p, r^0) - \hat{J}_*(p))']\} \\ &\quad \times \exp\{-\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0, r^0) - \hat{J}_*(p^0))\end{aligned}$$

$$\times Y'_{-1}M_{W(p^0)}Y_{-1}(\tilde{J}_*(p^0,r^0)-\hat{J}_*(p^0))']\}$$

$$\times\exp\{-\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\hat{J}^*(*)'W(*)'M_{(W(p^0),Y_{-1})}W(*)\hat{J}^*(*)]\} \tag{B.11}$$

where

$$\hat{J}^*(*)=(W(*)'M_{(W(p^0),Y_{-1})}W(*))^{-1}W(*)'M_{(W(p^0),Y_{-1})}\Delta Y,$$

$$\Psi_1(p,p^0,r^0)=|\hat{\Gamma}(p^0,r^0)'\hat{\Omega}^{-1}\hat{\Gamma}(p^0,r^0)$$

$$\otimes(F(r^0)'Y'_{-1}M_{W(p^0)}Y_{-1}F(r^0))|^{-1/2}$$

$$\times|(\hat{\Omega}^{-1}\otimes[I_{r^0},\hat{\hat{A}}(p^0,r^0)'](Y'_{-1}M_{W(p^0)}Y_{-1})[I_{r^0},\hat{\hat{A}}(p^0,r^0)']')$$

$$-(\hat{\Omega}^{-1}\hat{\Gamma}(p^0,r^0)((\hat{\Gamma}(p^0,r^0)'\hat{\Omega}^{-1}\hat{\Gamma}(p^0,r^0))^{-1}\hat{\Gamma}(p^0,r^0)'\hat{\Omega}^{-1}$$

$$\otimes[I_{r^0},\hat{\hat{A}}(p^0,r^0)'](Y'_{-1}M_{W(p^0)}Y_{-1}F(r^0))(F(r^0)'Y'_{-1}M_{W(p^0)}Y_{-1}F(r^0))^{-1}$$

$$\times(F(r^0)'Y'_{-1}M_{W(p^0)}Y_{-1})[I_{r^0},\hat{\hat{A}}(p^0,r^0)']')|^{-1/2}|\hat{\Gamma}(p,r^0)'\hat{\Omega}^{-1}\hat{\Gamma}(p,r^0)$$

$$\otimes(F(r^0)'Y'_{-1}M_{W(p)}Y_{-1}F(r^0))|^{1/2}|\hat{\Omega}^{-1}\otimes[I_{r^0},\hat{\hat{A}}(p,r^0)']$$

$$\times(Y'_{-1}M_{W(p)}Y_{-1})[I_{r^0},\hat{\hat{A}}(p,r^0)']')$$

$$-(\hat{\Omega}^{-1}\hat{\Gamma}(p,r^0)(\hat{\Gamma}(p,r^0)'\hat{\Omega}^{-1}\hat{\Gamma}(p,r^0))^{-1}\hat{\Gamma}(p,r^0)'\hat{\Omega}^{-1}$$

$$\otimes[I_{r^0},\hat{\hat{A}}(p,r^0)'](Y'_{-1}M_{W(p)}Y_{-1}F(r^0))((F(r^0)'Y'_{-1}M_{W(p)}Y_{-1}F(r^0)))^{-1}$$

$$\times(F(r^0)'Y'_{-1}M_{W(p)}Y_{-1})[I_{r^0},\hat{\hat{A}}(p,r^0)']')|^{1/2},$$

$$\Psi_2(p,p^0)=|\hat{\Omega}^{-1}\otimes(W(p^0)'W(p^0))|^{-1/2}|\hat{\Omega}^{-1}\otimes(W(p)'W(p))|^{1/2}$$

Applying Lemma A.2, we see that for $p>p^0$

$$\Psi_1(p,p^0,r^0)\equiv O_p(1).$$

and

$$\Psi_2(p,p^0)=T^{(1/2)m^2(p-p^0)}|\hat{\Omega}^{-1}\otimes T^{-1}(W(p^0)'W(p^0))|^{-1/2}|\hat{\Omega}^{-1}$$

$$\otimes T^{-1}(W(p)'W(p))|^{1/2}\xrightarrow{p}\infty \tag{B.12}$$

as $T \to \infty$. Moreover, note that

$$\mathrm{tr}[\hat{\Omega}^{-1}\hat{J}^*(*)'W(*)'M_{(W(p^0),Y_{-1})}W(*)\hat{J}^*(*)]$$

$$= \mathrm{tr}[\hat{\Omega}^{-1}T^{1/2}\hat{J}^*(*)'T^{-1}(W(*)M_{(W(p^0),Y_{-1})}W(*))T^{1/2}\hat{J}^*(*)]. \tag{B.13}$$

By Lemma A.2

$$T^{-1}(W(*)'M_{(W(p^0),Y_{-1})}W(*)) \equiv \mathrm{O_p}(1). \tag{B.14}$$

Also,

$$\hat{J}^*(*) = (W(*)'M_{(W(p^0),Y_{-1})}W(*))^{-1}W(*)'M_{(W(p^0),Y_{-1})}E,$$

where $E = [\varepsilon_1, \dots, \varepsilon_T]'$. Hence,

$$T^{1/2}\hat{J}^*(*) = (T^{-1}(W(*)'M_{W(p^0)}W(*)))^{-1}T^{-1/2}(W(*)'M_{W(p^0)}E) + \mathrm{o_p}(1)$$

$$\equiv \mathrm{O_p}(1) \tag{B.15}$$

by standard regression theory for stationary processes. From Eqs. (B.14) and (B.15), and the continuous mapping theorem we conclude that

$$\mathrm{tr}[\hat{\Omega}^{-1}T^{1/2}\hat{J}^*(*)'T^{-1}(W(*)'M_{(W(p^0),Y_{-1})}W(*))T^{1/2}\hat{J}^*(*)] \equiv \mathrm{O_p}(1). \tag{B.16}$$

Furthermore, we note that

$$\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))Y'_{-1}M_{W(p)}Y_{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))'] \equiv \mathrm{O_p}(1). \tag{B.17}$$

Putting together Eqs. (B.12), (B.16) and (B.17) and noting that all the other terms in Eq. (B.11) are $\mathrm{O_p}(1)$ as argued earlier, we have the required result that (for $p > p^0$)

$$\frac{\hat{\Pi}_T(M_{p^0,r^0} \mid \hat{\Omega}, Y)}{\hat{\Pi}_T(M_{p,r^0} \mid \hat{\Omega}, Y)} \to \infty \quad \text{in probability as } T \to \infty$$

by application of the continuous mapping theorem.

Similarly, for the case $p < p^0$, we can partition

$$W(p^0) = \left[\underset{T \times mp}{W(p)}, \underset{T \times m(p^0 - p)}{W_{(**)}}\right]$$

and write PIC as

$$\frac{\hat{\Pi}_T(M_{p^0,r^0}\,|\,\hat{\Omega}, Y)}{\hat{\Pi}_T(M_{p,r^0}\,|\,\hat{\Omega}, Y)} = \Psi_1(p,p^0,r^0)\Psi_2(p,p^0)\times \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))$$

$$\times Y'_{-1}M_{W(p)}Y_{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))']\}$$

$$\times \exp\{-\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0,r^0) - \hat{J}_*(p^0))$$

$$\times Y'_{-1}M_{W(p^0)}Y_{-1}(\tilde{J}_*(p^0,r^0) - \hat{J}_*(p^0))']\}$$

$$\times \exp\{\tfrac{1}{2}\mathrm{tr}[\hat{\Omega}^{-1}\hat{J}^*(**)'W(**)'M_{(W(p),Y_{-1})}W(**)\hat{J}^*(**)]\}, \tag{B.18}$$

where

$$\hat{J}^*(**) = (W(**)'M_{(W(p),Y_{-1})}W(**))^{-1}W(**)'M_{(W(p),Y_{-1})}\Delta Y.$$

From standard regression theory with stationary regressors we know that

$$\hat{J}^*(**)'W(**)'M_{(W(p),Y_{-1})}W(**)\hat{J}^*(**) \equiv O_p(T). \tag{B.19}$$

Moreover, write

$$\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))Y'_{-1}M_{W(p)}Y_{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))']$$

$$= \mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))PD(D^{-1}\underline{A}Y'_{-1}M_{W(p)}Y_{-1}\underline{A}'D^{-1})$$

$$\times DP(\tilde{J}_*(p,r^0) - \hat{J}_*(p))'].$$

Add and substract J_*^0 from $(\tilde{J}_*(p,r^0) - \hat{J}_*(p))PD$ and we get

$$(\tilde{J}_*(p,r^0) - \hat{J}_*(p))PD = (\tilde{J}_*(p,r^0) - J_*^0)PD + (J_*^0 - \hat{J}_*(p))PD.$$

Note that following arguments similar to that given in the proof of Theorem 1 of Ahn and Reinsel (1990), we have

$$(J_*^0 - \hat{J}_*(p))PD \quad = \; -\,[J^*(**)'T^{-1}(W(**)'M_{W(p)}Z_1)(T^{-2}(Z_1'Z_1))^{-1}$$

$$+ (T^{-1}U'Z_1)(T^{-2}(Z_1'Z_1))^{-1},\; J^*(**)'T^{-1/2}(W(**)'$$

$$\times M_{W(p)}Z_2)(T^{-1}(Z_2'M_{W(p)}Z_2))^{-1} + T^{-1/2}$$

$$\times (U'M_{W(p)}Z_2)(T^{-1}(Z_2'M_{W(p)}Z_2))^{-1}] + o_p(1)$$

$$\equiv -[O_p(1), O_p(T^{1/2})], \tag{B.20}$$

where $Z_1 = Y_{-1}A_{\perp,r^0}$ and $Z_2 = Y_{-1}A_{r^0}$. In addition, since $\hat{\Gamma}(p,r^0) - \Gamma^0 \equiv O_p(1)$

$$(\tilde{J}_*(p,r^0) - J_*^0(p))PD = \{T[\hat{\Gamma}(p,r^0) - \Gamma^0, \hat{\Gamma}(p,r^0)\hat{\bar{A}}(p,r^0)' - \Gamma^0\bar{A}^{0\prime}]P_1,$$

$$\times T^{1/2}[\hat{\Gamma}(p,r^0) - \Gamma^0, \hat{\Gamma}(p,r^0)\hat{\bar{A}}(p,r^0)' - \Gamma^0\bar{A}^{0\prime}]P_2\}$$

$$\equiv \{O_p(T), O_p(T^{1/2})\}. \tag{B.21}$$

Furthermore, Lemma A.2 gives

$$D^{-1}\underline{A}Y_{-1}'M_{W(p)}Y_{-1}\underline{A}'D^{-1} \equiv O_p(1). \tag{B.22}$$

Combining Eqs. (B.20), (B.21) and (B.22) we conclude that

$$\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))Y_{-1}'M_{W(p)}Y_{-1}(\tilde{J}_*(p,r^0) - \hat{J}_*(p))'] \equiv O_p(T^2). \tag{B.23}$$

Since

$$\mathrm{tr}[\hat{\Omega}^{-1}(\tilde{J}_*(p^0,r^0) - \hat{J}_*(p^0))Y_{-1}'M_{W(p^0)}Y_{-1}(\tilde{J}_*(p^0,r^0) - \hat{J}_*(p^0))'] \equiv O_p(1)$$

By Lemma A.5 and since the asymptotic behavior of Eq. (B.18) is dominated by the exponential terms, we deduce on the basis of Eqs. (B.19) and (B.23) that for $p < p^0$

$$\frac{\hat{\Pi}_T(M_{p^0,r^0} \mid \hat{\Omega}, Y)}{\hat{\Pi}_T(M_{p,r^0} \mid \hat{\Omega}, Y)} \to \infty \quad \text{in probability as } T \to \infty. \quad \square$$

Proof of Theorem 3.3. The proof of this theorem follows from the general idea of that of Theorem 2.1 of Phillips and Ploberger (1996) and Theorem 3.1 of Kim (1998) and is, thus, omitted. The proof is available in an earlier version of our paper, Chao and Phillips (1997), a copy of which can be obtained from us upon request.

Proof of Corollary 3.4. Since by Theorem 3.3

$$\frac{\Pi_T(M_{\bar{p},\bar{r}}\,|\,\Omega, Y)}{\Pi_T(M_{p,r}\,|\,\Omega, Y)}\Bigg/\frac{\hat{\Pi}_T(M_{\bar{p},\bar{r}}|\Omega, Y)}{\hat{\Pi}_T(M_{p,r}|\Omega, Y)} \xrightarrow{p} 1 \quad \text{as } T\to\infty,$$

it is sufficient that we show here that

$$\frac{\hat{\Pi}_T(M_{\bar{p},\bar{r}}\,|\,\Omega, Y)}{\hat{\Pi}_T(M_{p,r}\,|\,\Omega, Y)}\Bigg/\frac{\hat{\Pi}_T(M_{\bar{p},\bar{r}}\,|\,\hat{\Omega}, Y)}{\hat{\Pi}_T(M_{p,r}\,|\,\hat{\Omega}, Y)} \xrightarrow{p} 1 \quad \text{as } T\to\infty.$$

To proceed, write

$$\frac{\hat{\Pi}_T(M_{\bar{p},\bar{r}}\,|\,\Omega, Y)}{\hat{\Pi}_T(M_{p,r}\,|\,\Omega, Y)}\Bigg/\frac{\hat{\Pi}_T(M_{\bar{p},\bar{r}}\,|\,\hat{\Omega}, Y)}{\hat{\Pi}_T(M_{p,r}\,|\,\hat{\Omega}, Y)}$$

$$= |\Omega^{-1}\otimes W(p)'W(p)|^{1/2}/|\hat{\Omega}^{-1}\otimes W(p)'W(p)|^{1/2}$$

$$\times |\Omega^{-1}\otimes W(\bar{p})'W(\bar{p})|^{-1/2}/|\hat{\Omega}^{-1}\otimes W(\bar{p})'W(\bar{p})|^{-1/2}$$

$$\times [|\tilde{H}(p,r)(\Omega^{-1}\otimes Y'_{-1}M_{w(p)}Y_{-1})\tilde{H}(p,r)'|^{1/2}/$$

$$\times |\tilde{H}(p,r)(\hat{\Omega}^{-1}\otimes Y'_{-1}M_{w(p)}Y_{-1})\tilde{H}(p,r)'|^{1/2}]$$

$$\times [|\Omega^{-1}\otimes Y'_{-1}M_{w(\bar{p})}Y_{-1}|^{-1/2}/|\hat{\Omega}^{-1}\otimes Y'_{-1}M_{w(\bar{p})}Y_{-1}|^{-1/2}]$$

$$\times \exp\left\{\frac{1}{2}\mathrm{tr}[(\Omega^{-1}-\hat{\Omega}^{-1})(\tilde{J}_*(p,r)\right.$$

$$\left. -\hat{J}_*(p))Y'_{-1}M_{w(p)}Y_{-1}(\tilde{J}_*(p,r)-\hat{J}_*(p))']\right\}$$

$$\times \exp\left\{\frac{1}{2}\mathrm{tr}[(\Omega^{-1}-\hat{\Omega}^{-1})(\hat{J}^*(p^*)W(p^*)'M_{(Y_{-1},w(p))}W(p^*)\hat{J}^*(p^*)']\right\}.$$

Note first

$$|\Omega^{-1}\otimes W(p)'W(p)|^{(1/2)}/|\hat{\Omega}^{-1}\otimes W(p)'W(p)|^{1/2}$$

$$= |\Omega|^{-(1/2)mp}/|\hat{\Omega}|^{-(1/2)mp} \xrightarrow{p} 1 \quad \text{as } T\to\infty \tag{B.24}$$

by the consistency of $\hat{\Omega}$ and the Slutsky Theorem. Similarly, we deduce that as $T \to \infty$

$$|\Omega^{-1} \otimes W(\bar{p})' W(\bar{p})|^{-1/2} / |\hat{\Omega}^{-1} \otimes W(\bar{p})' W(\bar{p})|^{-1/2} \xrightarrow{p} 1 \tag{B.25}$$

and

$$|\Omega^{-1} \otimes Y'_{-1} M_{w(\bar{p})} Y_{-1}|^{-1/2} / |\hat{\Omega}^{-1} \otimes Y'_{-1} M_{w(\bar{p})} Y_{-1}|^{-1/2} \xrightarrow{p} 1. \tag{B.26}$$

Next, we write

$$\begin{aligned}
&|\tilde{H}(p,r)(\Omega^{-1} \otimes Y'_{-1} M_{w(p)} Y_{-1}) \tilde{H}(p,r)'| \\
&\quad = |\hat{\Gamma}(p,r)' \Omega^{-1} \hat{\Gamma}(p,r) \otimes F(r)' Y'_{-1} M_{w(p)} Y_{-1} F(r)| \\
&\qquad \times |\Omega^{-1} \otimes (I_r, \hat{\bar{A}}(p,r)') Y'_{-1} M_{w(p)} Y_{-1} (I_r, \hat{\bar{A}}(p,r)')'|(1 + \mathrm{op}(1)).
\end{aligned}$$

Since $|\tilde{H}(p,r)(\hat{\Omega}^{-1} \otimes Y'_{-1} M_{w(p)} Y_{-1}) \tilde{H}(p,r)|$ can be written similarly, we see that

$$\begin{aligned}
&|\tilde{H}(p,r)(\Omega^{-1} \otimes Y'_{-1} M_{w(p)} Y_{-1}) \tilde{H}(p,r)'|^{1/2} / \\
&|\tilde{H}(p,r)(\hat{\Omega}^{-1} \otimes Y'_{-1} M_{w(p)} Y_{-1}) \tilde{H}(p,r)'|^{1/2} \\
&\quad = |\hat{\Gamma}(p,r)' \Omega^{-1} \hat{\Gamma}(p,r)|^{(1/2)(m-r)} / |\hat{\Gamma}(p,r)' \hat{\Omega}^{-1} \hat{\Gamma}(p,r)|^{(1/2)(m-r)} \\
&\qquad \times |\Omega|^{-(1/2)r} / |\hat{\Omega}|^{-(1/2)r} (1 + \mathrm{op}(1)).
\end{aligned}$$

It follows from the consistency of $\hat{\Omega}$ and the Slutsky Theorem that as $T \to \infty$

$$\begin{aligned}
&|\tilde{H}(p,r)(\Omega^{-1} \otimes Y'_{-1} M_{w(p)} Y_{-1}) \tilde{H}(p,r)'|^{1/2} / |\tilde{H}(p,r)(\hat{\Omega}^{-1} \otimes Y'_{-1} M_{w(p)} Y_{-1}) \\
&\quad \tilde{H}(p,r)'|^{1/2} \xrightarrow{p} 1.
\end{aligned} \tag{B.27}$$

Finally, note that under the null hypothesis that the cointegrating rank and the ECM lag order equal r and p respectively, we have by Lemma A.5 that

$$(\tilde{J}_*(p,r) - \hat{J}_*(p)) Y'_{-1} M_{w(p)} Y_{-1} (\tilde{J}_*(p,r) - \hat{J}_*(p))' \equiv O_p(1)$$

and

$$\hat{J}^*(p^*) W(p^*)' M_{(Y_{-1}, w(p))} W(p^*) \hat{J}^*(p^*)' \equiv O_p(1)$$

It then follows from the consistency of $\hat{\Omega}$ and the continuous mapping theorem that as $T \to \infty$

$$\exp\left\{\frac{1}{2}\mathrm{tr}[(\Omega^{-1} - \hat{\Omega}^{-1})(\tilde{J}_*(p,r) - \hat{J}_*(p))Y'_{-1}M_{w(p)}Y_{-1}(\tilde{J}_*(p,r) - \hat{J}_*(p))']\right\}$$

$$\xrightarrow{p} 1 \tag{B.28}$$

and

$$\exp\left\{\frac{1}{2}\mathrm{tr}[(\Omega^{-1} - \hat{\Omega}^{-1})\hat{J}^*(p^*)W(p^*)'M_{(Y_{-1},w(p))}W(p^*)\hat{J}^*(p^*)']\right\} \xrightarrow{p} 1. \tag{B.29}$$

Putting Eqs. (B.24), (B.25), (B.26), (B.27), (B.28) and (B.29) together, we deduce the result

$$\frac{\hat{\Pi}_T(M_{\bar{p},\bar{r}}\,|\,\Omega,Y)}{\hat{\Pi}_T(M_{p,r}\,|\,\Omega,Y)}\Big/\frac{\hat{\Pi}_T(M_{\bar{p},\bar{r}}\,|\,\hat{\Omega},Y)}{\hat{\Pi}_T(M_{p,r}\,|\,\hat{\Omega},Y)} \xrightarrow{p} 1 \quad \text{as } T \to \infty$$

via the continuous mapping theorem.

References

Ahn, S.K., Reinsel, G.C., 1990. Estimation of partially nonstationary multivariate autoregressive models. Journal of the American Statistical Association 85, 813–823.

Bartlett, M.S., 1957. A comment on D.V. Lindley's statistical paradox. Biometrika 44, 533–534.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Chao J., 1995. Some simulation results of likelihood ratio tests for cointegrating ranks in misspecified vector autoregressions. Unpublished manuscript, Pennsylvania State University.

Chao, J., Phillips, P.C.B., 1997. Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. Cowles Foundation Discussion Paper No. 1155.

Engle, R.F., Granger, C.W.J., 1987. Cointegration and error-correction: representation, estimation, and testing. Econometrica 55, 251–276.

Geweke, J., 1994. Bayesian comparison of econometric models. Working Paper No. 532, Federal Reserve Bank of Minneapolis.

Geweke, J., 1996. Bayesian reduced rank regression in econometrics. Journal of Econometrics 75, 121–146.

Heyde, C., Johnstone, I., 1979. On asymptotic posterior normality for stochastic processes. Journal of Royal Statistical Society, Series B 41, 184–189.

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 cointegrating vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.

Johansen, S., 1992. Determination of cointegrating rank in the presence of a linear trend. Oxford Bulletin of Economics and Statistics 54, 383–397.

Kim, J.Y., 1994. Bayesian asymptotic theory in a time series model with a possible nonstationary process. Econometric Theory 10, 764–773.

Kim, J.Y., 1998. Large sample properties of posterior densities, Bayesian Information Criterion and the Likelihood Principle in nonstationary time series models. Econometrica 66, 359–380.

Kleibergen, F., van Dijk, H., 1994. On the shape of the likelihood/posterior in cointegrated models. Econometric Theory 10, 514–551.

Phillips, P.C.B., 1986. Understanding spurious regression in econometrics. Journal of Econometrics 33, 311–340.

Phillips, P.C.B., 1988. Multiple regression with integrated processes. In: Prabhu, N.U. (Ed.), Statistical Inference from Stochastic Processes. Contemporary Mathematics, vol. 80, pp. 79–106.

Phillips, P.C.B., 1991a. Optimal inference in cointegrated systems. Econometrica 59, 283–306.

Phillips, P.C.B., 1991b. Spectral regression for cointegrated time series. In: Barnett, W., Powell, J., Tauchen, G. (Eds.), Nonparametric and Semiparametric Methods in Economics and Statistics. Cambridge University Press, New York, pp. 413–435.

Phillips, P.C.B., 1991c. The posterior odds specification test for cointegration: modification, correction and extension. Yale University, Mimeographed.

Phillips, P.C.B., 1992. Bayes methods for trending multiple time series with an empirical application to the U.S. economy. Cowles Foundation Discussion Paper No. 1025.

Phillips, P.C.B., 1993. Forward exchange market unbiasedness: The case of the Australian dollar since 1984. Cowles Foundation Discussion Paper No. 1055.

Phillips, P.C.B., 1995a. Bayesian predictions: a response. Journal of Econometrics 69, 351–365.

Phillips, P.C.B., 1995b. Automated forecasts of Asia Pacific economic activity. Asia Pacific Economic Review 1, 92–102.

Phillips, P.C.B., 1996. Econometric model determination. Econometrica 64, 763–812.

Phillips, P.C.B., 1998. Impulse response and forecast error variance asymptotics in nonstationary VAR's. Journal of Econometrics 83, 21–56.

Phillips, P.C.B., Park, J.Y., 1989. Statistical inference in regressions with integrated processes: Part 2. Econometric Theory 5, 95–131.

Phillips, P.C.B., Ploberger, W., 1994. Posterior odds testing for a unit root with data-based model selection. Econometric Theory 10, 774–808.

Phillips, P.C.B., Ploberger, W., 1996. An asymptotic theory of Bayesian inference for time series. Econometrica 64, 381–412.

Reinsel, G.C., Ahn, S.K., 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.

Schotman, P., Van Dijk, H, 1991. A Bayesian analysis of the unit root in real exchange rates. Journal of Econometrics 49, 195–238.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's Information Criterion. Biometrika 63, 117–126.

Stock, J., Watson, M., 1988. Testing for common trends. Journal of American Statistical Association 83, 1097–1107.

Toda, H.Y., Phillips, P.C.B., 1993. Vector autoregressions and causality. Econometrica 61, 1367–1393.

Toda, H.Y., Phillips, P.C.B., 1994. Vector autoregression and causality: a theoretical overview and simulation study. Econometric Reviews 13, 259–285.

Tsay, R.S., 1984. Order selection in nonstationary autoregressive models. Annals of Statistics 12, 1425–1433.

Wei, C.Z., 1992. On predictive least squares principles. Annals of Statistics 20, 1–42.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.