Econometric Theory, **20**, 2004, 735–742. Printed in the United States of America.
DOI: 10.1017/S0266466604204066

THE ASYMPTOTIC DISTRIBUTION OF THE COINTEGRATION RANK ESTIMATOR UNDER THE AKAIKE INFORMATION CRITERION

George Kapetanios
Queen Mary, University of London

We derive the asymptotic distribution of the estimate of the cointegration rank of a multivariate model when Akaike's information criterion is used. It is shown that the use of this criterion is ill-advised given that the estimate is severely upward biased even asymptotically.

1. INTRODUCTION

The determination of the cointegration rank of a multivariate cointegrated system has attracted considerable attention in the econometric literature for the past 15 years. The most widely used procedures for determining cointegration rank are those proposed by Johansen (1988). Alternative testing procedures have been suggested by, among others, Phillips and Ouliaris (1988), Stock and Watson (1988), Snell (1999), and Bierens (1997).

In this paper we start by reviewing the formal justification for the application of model selection criteria in selecting the cointegration rank. We note that the standard necessary and sufficient conditions for a criterion to be weakly consistent in lag order selection extend to the determination of the cointegration rank. The main result of the paper involves the derivation of the asymptotic distribution of the cointegrating rank estimate when the inconsistent Akaike information criterion (AIC) is used. Unlike with stationary models, where AIC approximates the Kullback–Leibler distance between the estimated model and the data generation process, there is no compelling theoretical reason for its use in rank selection in nonstationary cointegration models. It is shown that the use of this criterion is ill-advised given that the estimate is severely upward biased even asymptotically. These results point toward the use of other criteria such as the Bayesian information criterion (BIC) and the posterior information criterion (PIC).

I thank the editor, Professor Peter Phillips, and three anonymous referees for comments and suggestions that improved this paper significantly. All remaining errors are my own. Address correspondence to: George Kapetanios, Department of Economics, Queen Mary, University of London, Mile End Rd., London E1 4NS, UK; e-mail: G.Kapetanios@qmul.ac.uk.

 0266-4666/04 $12.00

2. WEAK CONSISTENCY OF INFORMATION CRITERIA

We assume that the multivariate system may be written as an m-dimensional VAR(k) process given by

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_k \boldsymbol{y}_{t-k} + \boldsymbol{\epsilon}_t, \qquad t = 1,\ldots,T, \tag{1}$$

where the error term $\boldsymbol{\epsilon}_t$ is a zero mean independent and identically distributed (i.i.d.) vector with finite positive definite covariance matrix. This VAR(k) process will be referred to as cointegrated of rank r if $\boldsymbol{\Pi} = \boldsymbol{I} - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_k$ has rank r. In this case the matrix $\boldsymbol{\Pi}$ may be decomposed as $\boldsymbol{\Pi} = \boldsymbol{\alpha\beta}'$ where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are matrices of dimension $m \times r$. The error correction representation[1] of the system is given by

$$\Delta \boldsymbol{y}_t = \boldsymbol{\mu} - \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{\Psi}_1 \Delta \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Psi}_{k-1} \Delta \boldsymbol{y}_{t-k+1} + \boldsymbol{\epsilon}_t, \tag{2}$$

where $\boldsymbol{\Psi}_i$, $i = 1,\ldots,k-1$ are functions of $\boldsymbol{\Phi}_i$, $i = 1,\ldots,k$.

The general form of the loss function minimized by information criteria is given by

$$IC(s) = -2l_T(\boldsymbol{\theta}) + 2c_T(s), \tag{3}$$

where $l_T(\boldsymbol{\theta})$ is the log likelihood of the model, s is the number of free parameters, and $c_T(s)$ is a penalty term promoting model parsimony depending on s and the sample size. For three common information criteria the penalty terms are as follows: s (Akaike's information criterion) (Akaike, 1973), $(s/2)\ln(T)$ (Bayesian information criterion) (Schwarz, 1978), and $s\ln(\ln(T))$ (Hannan–Quinn information criterion [HQ]) (Hannan and Quinn, 1979). The model specification chosen is that for which $IC(k)$ is minimized. In lag order selection it is well known that the estimated lag order for stationary and unit root nonstationary vector autoregressive (VAR) models will be weakly consistent iff $c_T(k) \xrightarrow{p} \infty$ and $c_T(k)/T \xrightarrow{p} 0$ as $T \to \infty$ and $c_T(k)$ is bounded in k where k is the lag order. For a proof of the latter case, for deterministic penalty terms, see Paulsen (1984). Clearly, whereas BIC and HQ are weakly consistent for lag order selection, AIC is not. This is a well-known result for AIC (see, e.g., Shibata, 1976). Note that we choose to have a general expression for the penalty term to accommodate other less widely used criteria such as, e.g., the generalized information criterion (GIC) (see Konishi and Kitagawa, 1996) and the PIC (see Phillips, 1996; Phillips and Ploberger, 1994; Phillips and Ploberger, 1996). Note that whereas AIC, BIC, and HQ have deterministic penalty terms, GIC and PIC have stochastic penalty terms (hence the notation concerning the conditions on the asymptotic behavior of the generic penalty term). These results have been shown by various authors to extend to more general model selection frameworks (see, e.g., Sin and White, 1996; Kapetanios, 2001).

Aznar and Salvador (2000) show that the standard conditions on the penalty terms for weak consistency in lag selection extend to the cointegration frame-

https://doi.org/10.1017/S0266466604204066 Published online by Cambridge University Press

work. In particular, they show that the cointegration rank and the lag order may be jointly weakly consistently estimated iff $c_T(k) \to \infty$ and $c_T(k)/T \to 0$ as $T \to \infty$. This result holds only for information criteria whose penalty terms are deterministic, and therefore criteria such as the GIC are not covered. We also note that the asymptotic properties of PIC have been discussed in Chao and Phillips (1999) where weak consistency of PIC in jointly estimating cointegration rank and VAR lag order is established. Note that this is the first paper to give a consistency result for estimating cointegration rank via an information criterion.

3. THE ASYMPTOTIC DISTRIBUTION OF THE RANK ESTIMATE USING AKAIKE'S INFORMATION CRITERION

Following Pesaran and Pesaran (1997) we specify the number of free parameters for a multivariate cointegrated model with no intercept or time trend to be equal to $s = m^2(k^0 - 1) + 2mr - r^2$ where k^0 is the true lag order of the system. The penalty terms for AIC, BIC, and HQ are then given, respectively, by s, $(s/2)\ln(T)$, and $s\ln(\ln(T))$ or equivalently by $\tilde{s}$, $\tilde{s}/2\ln(T)$, and $\tilde{s}\ln(\ln(T))$ where $\tilde{s} = -(m-r)^2$.

It is clear that AIC is not consistent in rank determination. Nevertheless it is also clear that the probability of picking a rank that is lower than the true rank goes to zero asymptotically, as we also show in the Appendix. The following theorem provides the means for determining the asymptotic probabilities that AIC will pick a higher rank than the true one.

THEOREM. *Consider the* VAR *model of (1) with $\boldsymbol{\mu} = 0$ and known k^0. The asymptotic distribution of the rank estimate obtained through* AIC *is given by*

$$\lim_{T\to\infty} P(\hat{r}_{AIC} = r) = \begin{cases} 0 & \text{if } r < r^0 \\ p_r & \text{if } r \geq r^0, \end{cases}$$

where p_r are given by expression (A.7) in the Appendix and r^0 is the true rank of the model.

Note that we assume a known true lag order for the derivation of the asymptotic distribution of AIC. Unlike the result of Aznar and Salvador (2000) on the joint determination of lag order and cointegration rank, allowing for an unknown lag order in this context would obviously change the asymptotic distribution of the cointegration rank estimated using AIC.

The asymptotic distribution of the estimate of the cointegration rank depends only on $d = m - r^0$. Tables 1 and 2 show the distribution of the cointegration rank estimate for the case of no deterministic terms and the case of an unrestricted constant obtained through simulation. Brownian motion is simulated using a random walk of 1,000 observations. Five thousand replications have been used.

In the standard case of lag order selection, the asymptotic probability of AIC picking a lag order larger than the true one is quite small and declines rapidly

Table 1. Asymptotic distribution of cointegration rank estimate under AIC when the model contains no deterministic terms

$m - r^0$	r^0	r^0+1	r^0+2	r^0+3	r^0+4	r^0+5	r^0+6	r^0+7	r^0+8	r^0+9
1	0.820	0.180	—	—	—	—	—	—	—	—
2	0.643	0.308	0.049	—	—	—	—	—	—	—
3	0.490	0.362	0.136	0.012	—	—	—	—	—	—
4	0.374	0.414	0.146	0.054	0.012	—	—	—	—	—
5	0.300	0.414	0.198	0.060	0.025	0.003	—	—	—	—
6	0.226	0.391	0.248	0.097	0.030	0.007	0.001	—	—	—
7	0.149	0.403	0.277	0.113	0.041	0.011	0.006	0.000	—	—
8	0.110	0.328	0.318	0.165	0.056	0.012	0.007	0.004	0.000	—
9	0.072	0.311	0.326	0.199	0.061	0.023	0.006	0.002	0.000	0.000

Table 2. Asymptotic distribution of cointegration rank estimate under AIC when the model contains an unrestricted constant

$m - r^0$	r^0	r^0+1	r^0+2	r^0+3	r^0+4	r^0+5	r^0+6	r^0+7	r^0+8	r^0+9
1	0.851	0.149	—	—	—	—	—	—	—	—
2	0.386	0.529	0.085	—	—	—	—	—	—	—
3	0.227	0.394	0.317	0.062	—	—	—	—	—	—
4	0.147	0.349	0.258	0.212	0.034	—	—	—	—	—
5	0.108	0.259	0.305	0.185	0.126	0.017	—	—	—	—
6	0.082	0.259	0.277	0.180	0.107	0.082	0.013	—	—	—
7	0.048	0.224	0.263	0.204	0.122	0.070	0.058	0.011	—	—
8	0.037	0.172	0.263	0.233	0.131	0.077	0.045	0.038	0.004	—
9	0.019	0.138	0.283	0.246	0.156	0.080	0.025	0.029	0.018	0.006

for higher lag orders. The results for the cointegration rank estimate do not follow this pattern. The probabilities of overestimation are quite large and depend crucially on the nature of the deterministic terms included in the model. In most cases the true rank is not even the mode of the asymptotic distribution of the estimate. As $m - r^0$ rises, the problem is further accentuated. In the extreme case considered in the tables, when $m - r^0 = 9$ and a constant is included in the model, the probability of picking the right rank is equal to just 1.9%.

The motivation behind the derivation of AIC is not consistency in the selection of the true model but optimization in terms of goodness of the selected model as measured by the Kullback–Leibler information metric.[2] Therefore, our result does not necessarily imply that the criterion is in general “bad” in selecting cointegration rank because such a judgment would have to be related to a particular modeling purpose. Nevertheless, the optimality properties of AIC hold for stationary models. Currently, there is no compelling theoretical reason motivating the use of AIC for rank determination in cointegration models. Furthermore, we can provide some evidence in favor of methods that are parsimonious in cointegration rank selection such as BIC and PIC. In terms of forecasting and over long horizons, error correction models have been in general shown to have an advantage. However, Christoffersen and Diebold (1998) cast doubt on the notion that error correction models are better forecasting tools even at long horizons, at least with respect to the standard root mean square forecasting error criterion. They also argue that although unit roots are estimated consistently, modeling nonstationary series in (log) levels is likely to produce forecasts that are suboptimal in finite samples relative to a procedure that imposes unit roots, a phenomenon exacerbated by small sample estimation bias. Developing this argument, they suggest that for cointegrated series it is better to overestimate rather than underestimate the number of common trends, or in other words, underestimate the cointegrating rank.

4. CONCLUSION

In this paper the asymptotic distribution of the rank estimator of a cointegrated model using AIC has been derived. The results are rather critical of the Akaike criterion in this context and point toward the use of other criteria such as BIC and PIC. The AIC estimate severely overestimates the rank. The overestimation is accentuated by the presence of deterministic terms in the model and by the magnitude of the difference between the true rank and the dimension of the model.

NOTES

1. We assume that the error correction representation exists by imposing extra conditions such as, e.g., condition (iii) of Chao and Phillips (1999).

2. Note also the efficiency property of AIC in terms of selecting the model order for linear models, discussed by Shibata (1980).

REFERENCES

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki (eds.), *2nd International Symposium on Information Theory*, pp. 267–281. Akademiai Kiado.

Aznar, A. & M. Salvador (2000) Selecting the Rank of the Cointegration Space and the Form of the Intercept Using an Information Criterion. *Econometric Theory* 18, 926–947.

Bierens, H. (1997) Nonparametric cointegration analysis. *Journal of Econometrics* 77, 379–404.

Chao, J. & P.C.B. Phillips (1999) Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* 91, 227–272.

Christoffersen, P.F. & F.X. Diebold (1998) Cointegration and long horizon forecasting. *Journal of Business and Economic Statistics* 16, 450–458.

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

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

Kapetanios, G. (2001) Model selection in threshold models. *Journal of Time Series Analysis* 22, 733–754.

Konishi, S. & G. Kitagawa (1996) Generalised information criteria in model selection. *Biometrika* 83, 875–890.

Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis*. Springer.

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

Pesaran, M.H. & B. Pesaran (1997) *Microfit 4.0*. Oxford University Press.

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

Phillips, P.C.B. & S. Ouliaris (1988) Testing for cointegration using principal components methods. *Journal of Economic Dynamics and Control* 12, 205–230.

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

Phillips, P.C.B. & W. Ploberger (1996) An asymptotic theory of Bayesian inference for time series. *Econometrica* 63, 381–412.

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.

Shibata, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* 8, 147–164.

Sin, C.Y. & H. White (1996) Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics* 71, 207–225.

Snell, A. (1999) Testing for r versus $r - 1$ cointegrating vectors. *Journal of Econometrics* 88, 151–191.

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

APPENDIX

The proof of the theorem requires some of the results derived by Johansen (1988). To simplify matters we will assume that the VAR process has zero mean and that no constant is included in the estimation. Extension to models with deterministic terms is straightforward. We will denote the loss function used by AIC by $AIC(r) = -2l_T(r) + 2\tilde{s}$ where $l_T(r)$ is the log likelihood of the model for cointegration rank, r. Let

https://doi.org/10.1017/S0266466604204066 Published online by Cambridge University Press

$$\Delta \boldsymbol{Y}_t = (\Delta \boldsymbol{y}_{t-1}', \ldots, \Delta \boldsymbol{y}_{t-k^0+1}')', \qquad \boldsymbol{X} = (\Delta \boldsymbol{Y}_1, \ldots, \Delta \boldsymbol{Y}_T)',$$

$$\boldsymbol{Y}_{-p} = (\boldsymbol{y}_{1-p}, \ldots, \boldsymbol{y}_{T-p})', \qquad \Delta \boldsymbol{Y} = (\Delta \boldsymbol{y}_1, \ldots, \Delta \boldsymbol{y}_T)',$$

$$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}', \qquad \boldsymbol{R}_0 = \boldsymbol{M}\Delta \boldsymbol{Y},$$

$$\boldsymbol{R}_1 = \boldsymbol{M}\boldsymbol{Y}_{-p}, \qquad \boldsymbol{S}_{ij} = \boldsymbol{R}_i'\boldsymbol{R}_j/T, \qquad i,j = 0,1.$$

Finally, let $\boldsymbol{G}$ be the lower triangular Cholesky decomposition of $\boldsymbol{S}_{11}$ and $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_m$ be the eigenvalues of $\boldsymbol{G}\boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\boldsymbol{G}'$. Then by, say, Proposition 11.1 of Lütkepohl (1991) the difference in the log likelihood of the cointegrated VAR(k^0) for cointegration ranks r_1, r_0, $r_1 > r_0$ is equal to $-(T/2)\sum_{i=r_0+1}^{r_1} \ln(1 - \hat{\lambda}_i)$. For $r_0 < r_1 < r^0$ it is clear that $(T/2)\sum_{i=r_0+1}^{r_1} \ln(1 - \hat{\lambda}_i) = O_p(T)$, showing that Akaike's criterion will not pick a rank lower than the true one asymptotically in probability.

For $r > r^0$ we first note that $\hat{\lambda}_r \xrightarrow{p} 0$ by Lemma 4 of Johansen (1988). By a simple expansion we then have that $T\ln(1 - \hat{\lambda}_r) = -T\hat{\lambda}_r + o_p(1)$. But by Lemma 6 of Johansen (1988) we have that $T\hat{\lambda}_{r^0+1}, \ldots, T\hat{\lambda}_m$ converge in distribution to the ordered eigenvalues of the equation

$$\left| \lambda \int_0^1 \boldsymbol{W}\boldsymbol{W}'\,du - \int_0^1 \boldsymbol{W}\,d\boldsymbol{W}' \int_0^1 d\boldsymbol{W}\boldsymbol{W}' \right| = 0$$

denoted by $\lambda_1, \ldots, \lambda_{m-r^0}$, where $\boldsymbol{W}$ is a $m - r^0$ standard Brownian motion.

We now concentrate on deriving the probabilities for $r \geq r^0$. For $r \geq r^0$ we have that $P(\hat{r}_{AIC} = r)$ is asymptotically equivalent to $P(AIC(r) \leq AIC(u),\ r^0 \leq u \leq m)$. The asymptotic equivalence follows by the fact that the criterion will not pick a rank lower than r^0 asymptotically. Clearly, this may be the case in finite samples. Disregarding constant terms with respect to r, the log likelihood is given by $l_T(r) = -T/2\sum_{i=1}^{r} \ln(1 - \hat{\lambda}_i)$. Then,

$$\begin{aligned} &P(AIC(r) \leq AIC(u),\ r^0 \leq u \leq m) \\ &\quad = P\Bigg(T\sum_{i=u+1}^{r} \ln(1 - \hat{\lambda}_i) \leq 2(m-r)^2 - 2(m-u)^2 \quad (r^0 \leq u < r) \ \text{and} \\ &\qquad\quad T\sum_{j=r+1}^{u} \ln(1 - \hat{\lambda}_j) \geq 2(m-u)^2 - 2(m-r)^2 \quad (r < u \leq m)\Bigg). \end{aligned} \tag{A.1}$$

But $T\ln(1 - \hat{\lambda}_r) = -T\hat{\lambda}_r + o_p(1)$. Therefore, for any $\epsilon > 0$ there exists a positive integer M such that for all T larger than M the difference between the probability on the right-hand side of (A.1) and

$$\begin{aligned} &P\Bigg(T\sum_{i=u+1}^{r} \hat{\lambda}_i \geq 2(m-u)^2 - 2(m-r)^2 \quad (r^0 \leq u < r) \ \text{and} \\ &\quad T\sum_{j=r+1}^{u} \hat{\lambda}_j \leq 2(m-r)^2 - 2(m-u)^2 \quad (r < u \leq m)\Bigg) \end{aligned} \tag{A.2}$$

https://doi.org/10.1017/S0266466604204066 Published online by Cambridge University Press

is less than ϵ. But the probability in (A.2) may be written as

$$P\Bigg(\sum_{i=u+1}^{r}\{T\hat{\lambda}_i - 2[m-(i-1)]^2 + 2(m-i)^2\} \geq 0 \quad (r^0 \leq u < r) \text{ and }$$
$$\sum_{j=r+1}^{u}\{T\hat{\lambda}_j - 2[m-(j-1)]^2 + 2(m-j)^2\} \leq 0 \quad (r < u \leq m)\Bigg) \tag{A.3}$$

or

$$P\Bigg(\sum_{i=u+1}^{r}\{T\hat{\lambda}_i + (4i-4m-2)\} \geq 0 \quad (r^0 \leq u < r) \text{ and }$$
$$\sum_{j=r+1}^{u}\{T\hat{\lambda}_j + (4j-4m-2)\} \leq 0 \quad (r < u \leq m)\Bigg). \tag{A.4}$$

By a change of indices and the weak convergence of $T\hat{\lambda}_i$ we get that, asymptotically, the preceding probability is equivalent to

$$P\Bigg(\sum_{i'=1}^{r'-u'}\{\lambda_{i'} + (4(i'+u'+r^0)-4m-2)\} \geq 0 \quad (0 \leq u' < r') \text{ and }$$
$$\sum_{j'=1}^{u'-r'}\{\lambda_{j'} + (4(j'+r'+r^0)-4m-2)\} \leq 0 \quad (r' < u' \leq m-r^0)\Bigg), \tag{A.5}$$

where $i' = i-u$, $j' = j-r$, $r' = r-r^0$, and $u' = u-r^0$. Regrouping terms gives

$$P\Bigg(\sum_{i'=1}^{r'-u'}\{\lambda_{i'} + (4i'+4(u'-d)-2)\} \geq 0 \quad (0 \leq u' < r') \text{ and }$$
$$\sum_{j'=1}^{u'-r'}\{\lambda_{j'} + (4j'+4(r'-d)-2)\} \leq 0 \quad (r' < u' \leq m-r^0)\Bigg), \tag{A.6}$$

where $d = m-r^0$. Define

$$S_l^q = \sum_{i'=1}^{q-l}\{\lambda_{i'} + (4i'+4(l-d)-2)\}, \qquad q > l.$$

Then, the probability in (A.6) may be expressed as

$$P(S_{u'}^{r'} \geq 0 \quad (0 \leq u' < r') \quad \text{and} \quad S_{r'}^{u'} \leq 0 \ \ (r' < u' \leq m-r^0)). \tag{A.7}$$

The joint probability distribution of S_l^q may easily be obtained by simulation using the standard results of Johansen (1988). ■

https://doi.org/10.1017/S0266466604204066 Published online by Cambridge University Press