Testing for Cointegrating Rank via Model Selection: Evidence from 165 Data Sets

Badi H. Baltagi
Syracuse University. Center for Policy Research, bbaltagi@maxwell.syr.edu

Zijun Wang

ISSN: 1525-3066

Center for Policy Research
Working Paper No. 83

TESTING FOR COINTEGRATING RANK VIA MODEL SELECTION: EVIDENCE FROM 165 DATA SETS

Badi H. Baltagi and Zijun Wang

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu

July 2006

$5.00

CENTER FOR POLICY RESEARCH – Spring 2006

Timothy Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret Austin
Associate Director
Budget and Administration

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi Baltagi .... Economics
Pablo Beramendi .... Political Science
Dan Black .... Economics
Lloyd Blanchard .... Public Administration
William Duncombe .... Public Administration
Gary Engelhardt .... Economics
Deborah Freund .... Public Administration
Madonna Harrington Meyer .... Sociology
Christine Himes .... Sociology
William C. Horrace .... Economics
Bernard Jump .... Public Administration
Duke Kao .... Economics
Eric Kingson .... Social Work
Thomas Kniesner .... Economics
Tom Krebs .... Economics
Jeffrey Kubik .... Economics
Andrew London .... Sociology
Len Lopoo .... Public Administration
Jerry Miner .... Economics
Jan Ondrich .... Economics
John Palmer .... Public Administration
Lori Ploutz-Snyder .... Exercise Science
David Popp .... Public Administration
Grant Reeher .... Political Science
Stuart Rosenthal .... Economics
Ross Rubenstein .... Public Administration
Margaret Usdansky .... Sociology
Michael Wasylenko .... Economics
Janet Wilmoth .... Sociology

GRADUATE ASSOCIATES

Javier Baez .... Economics
Sonali Ballal .... Public Administration
Dana Balter .... Public Administration
Jason Bladen .... Economics
Jesse Bricker .... Economics
Maria Brown .... Social Science
Yong Chen .... Economics
Ginnie Cronin .... Social Science
Ana Dammert .... Economics
Mike Eriksen .... Economics
Katie Fitzpatrick .... Economics
Alexandre Genest .... Public Administration
Julie Anna Golebiewski .... Economics
Nadia Greenhalgh-Stanley .... Economics
Yue Hu .... Economics
Becky Lafrancois .... Economics
Joseph Marchand .... Economics
Neelakshi Medhi .... Social Science
Larry Miller .... Public Administration
Wendy Parker .... Sociology
Emily Pas .... Economics
Shawn Rohlin .... Economics
Cynthia Searcy .... Public Administration
Talisha Searcy .... Public Administration
Jeff Thompson .... Economics
Wen Wang .... Economics
Yoshikun Yamamoto .... Public Administration

STAFF

Kelly Bogart .... Administrative Secretary
Martha Bonney .... Publications/Events Coordinator
Karen Cimilluca .... Administrative Secretary
Kati Foley .... Administrative Assistant
Kitty Nasto .... Administrative Secretary
Candi Patterson .... Computer Consultant
Mary Santy .... Administrative Secretary
Tammy Tanner .... Librarian/Office Coordinator

ABSTRACT

The model selection approach has been proposed as an alternative to the popular tests for cointegration such as the residual-based ADF test and the system-based trace test. Using information criteria, we conduct cointegration tests on 165 data sets used in published studies. The empirical results demonstrate the usefulness of the model selection approach for applied researchers.

JEL Code: C21, C23

1. INTRODUCTION

The growing literature on cointegration analysis has typically applied single-equation residual-based tests such as the augmented Dickey-Fuller test (ADF), the $\hat{Z}_\alpha$ test of Phillips and Ouliaris (1990), and the system-based likelihood ratio test of Johansen (1988, 1991) and Reinsel and Ahn (1992). However, from the viewpoint of practitioners, the determination of cointegrating rank is essentially a model specification problem, which ultimately involves a trade-off between model parsimony (complexity) and fit, given the fact that the true model is rarely, if ever, known. As various information criteria take into account both model fit and parsimony, it is thus quite natural to consider information criteria in the determination of cointegrating rank.

From a broader perspective, Granger et al. (1995) recommend the use of model selection procedures rather than formal hypothesis testing when deciding on model specification. They argue that formal testing favors the model chosen to be the null hypothesis, and the choice of significance level is typically arbitrary. These concerns seem applicable in cointegration analysis. For example, the choice between the null of no cointegration and the null of cointegration may result in quite different results in practice. Another attractive feature of the model selection approach is that it makes possible for the joint determination of VAR lag order and cointegrating rank. As is well known, the choice of lag order in a VAR has an important impact on the cointegration test performance (e.g., Boswijk and Franses, 1992). However, since choices of lag order and cointegrating rank are two separate steps in the application of the trace test and other probability-based procedures, it is essentially impossible to comment on the underlying probability distribution of the final results. In contrast, it is possible to determine the lag order and cointegrating rank in one step by minimizing an information criterion over a domain of models with different lag orders and cointegrating ranks.

The application of the model selection approach in cointegration analysis was first suggested and implemented in Phillips and McFarland (1997). Contributions to this line of research also include Aznar and Salvador (2002), and Kapetanios (2004), etc. The simulation evidence provided by Chao and Philips (1999) and Wang and Bessler (2005) suggests that the performance of the model selection approach based on the information criteria SIC and PIC is close to, and sometimes better than that of the trace test. Extending these theoretical results, the primary purpose of this paper is to show how much of the results on cointegration analysis in 165 published data sets can be replicated using the alternative model selection approach. If the results from the model selection approach correspond well with the results from the traditional hypothesis testing, applied researchers would have more confidence in their empirical results. This is important given that researchers, such as Gonzalo and Lee (1998) and Brüggemann and Lütkepohl (2004), have repeatedly reported the missing robustness of system based tests for cointegration. The current paper can also be seen as an extension of Gregory et al. (2004) who examine the correspondence between the performance of single-equation residual-based tests and that of Johansen’s (1988, 1991) likelihood ratio tests.

The rest of the paper is organized as follows. Section 2 gives a brief review of the information criteria and trace test. Section 3 describes the data sources and the related research methodology used in the replications. Section 4 summarizes our findings based on the 165 published data sets. A short summary concludes the paper.

2. INFORMATION CRITERIA AND TRACE TEST

Assume the model to be tested for cointegration has the following error correction form:

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1} \Delta Y_{t-p+1} + \Pi Y_{t-1} + \Psi D_t + \mu + \varepsilon_t, \quad t = 1,\ldots, \mathrm{T}, \qquad (1)$$

where Y_t is a $m \times 1$ vector of stochastic variables, Y_{t-1}, …, Y_{t-p}, are 1 up to p lags of Y_t, ε_t are independent and identically distributed (i.i.d.) random variates following multivariate $N(0, \Sigma)$ with Σ being positive definite, D_t may contain dummy variables, time trend, or weak exogenous stochastic variables, Γ_i, Π, and Ψ are conformable parameter matrices, and μ is a $m \times 1$ vector of parameters (constants).

The three information criteria that have been widely used are Akaike's (1974) information criterion (AIC), Hannan-Quin's (1979) criterion (HQ), Schwarz's (1978) information criterion (SIC), They are computed according to the following equations:

$$AIC = \ln(\det(\tilde{\Omega})) + 2K/T\ , \tag{2}$$

$$HQ = \ln(\det(\tilde{\Omega})) + 2K\ln(\ln(T))/T\ , \tag{3}$$

$$SIC = \ln(\det(\tilde{\Omega})) + K\ln(T)/T\ , \tag{4}$$

where $\tilde{\Omega}$ is the maximum likelihood estimator of Σ with given p and cointegrating rank r. Note that the penalty terms of AIC, HQ and SIC are a simple function of a parameter count (K), which increases with p and r. Wei (1992), Phillips and Ploberger (1996), and Chao and Phillips (1999) argue that the coefficients associated with regressors exhibiting a trend should be penalized more strongly than regressors without a trend. Phillips and Ploberger (1996) and Chao and Phillips (1999) propose the use of the posterior information criterion (PIC):

$$PIC = \frac{1}{T}\ln\left|S_{pp} \otimes \tilde{\Omega}^{-1}\right| + \frac{1}{T}\ln\left|H(S_{yy.p} \otimes \tilde{\Omega}^{-1})H^{'}\right| + tr[\hat{\Omega}^{-1}\tilde{\Omega}], \tag{5}$$

where $\hat{\Omega}$ is the maximum likelihood estimator of Σ when model (1) is estimated with the highest lag order and full cointegrating rank. $M_W = I_T - W(W'W)^{-1}W'$, $W = [W_1, \ldots, W_T]'$,

$W_t = (\Delta Y'_{t-1}, ..., \Delta Y'_{t-p})'$, $Y_{-1} = [Y_0, \ldots, Y_{T-1}]'$, $\Delta Y = [\Delta Y_1, \ldots, \Delta Y_T]'$, $S_{yy} = \Delta Y' \Delta Y$, $S_{pp} = W'W$,

$S_{yp} = Y'_{-1} W$, $S_{py} = W' Y_{-1}$, $S_{yy.p} = S_{yy} - S_{yp} S_{pp}^{-1} S_{py}$, and finally,

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

in which $F(r)' = [0_{(m-r)\times r}, I_{m-r}]$. It can be seen from equation (5) that PIC not only takes into account the number of regressors, but also the magnitude of the regressors and the sample information accumulated in the data about the model parameters (Chao and Phillips, 1999, p. 233).

3. DATA SOURCES AND RESEARCH STRATEGY

The 165 data sets used in this paper are from 43 studies published in the *Journal of Applied Econometrics* from 1994 through the fourth issue of 2004. The first 132 data sets were edited and analyzed by Gregory et al. (2004).

We extend the above data sets by including 33 additional data sets used in nine articles published in the *Journal of Applied Econometrics* from the third issue of 2001 to the fourth issue of 2004 (all 165 data sets can be obtained from http://qed.econ.queensu.ca/jae/). The editing and selection criteria used in this paper are the same as in Gregory et al. (2004). Complete details of on these data sets and the model specifications can be found at: http://www.tamu.edu/perc/publication/0519.xls.

While some new topics about cointegration have also appeared in the literature (e.g., seasonal unit roots, I(2) cointegration and fractionally integration), we limit our exercise to the standard cointegration test.

One- and Two-step Procedures

As mentioned earlier, the cointegrating rank and lag order of the VAR can be determined jointly with the model selection approach. We call this the one-step procedure. In a two-step procedure, first the lag order p is either identified from the original studies or estimated by minimizing one of the four information criteria. Conditional on the chosen p, we can apply the trace test or the information criterion to select the cointegrating rank in the second step.

Trend Specification

It is well known that the assumption about the deterministic components in the model may have important effects on cointegration tests. If the original papers explicitly state their assumptions, we simply follow the assumptions. Otherwise, we determine the specification based on the critical values the authors cited to make their inference, or on the reported test statistics and conclusions about the cointegrating ranks. When these test-related signals are not available, we draw our conclusion by estimating the data for alternative trend specifications, selecting the one that gives the same or closest parameter estimates relative to the original papers. If none of these conditions apply, we conduct the two popular specification tests: with and without a trend term in the cointegrating relations.

It is worth noting here that Phillips (1996) has shown how cointegrating rank, lag length, and trend degree in a VAR can be jointly determined using the model selection method. Nevertheless, as the trend types can be identified from the original papers in most studies, we focus on the determination of cointegrating rank in order to make the replication results comparable to the original findings.

Choice of the Maximum Lag Length

In setting the maximum lag length k to determine the lag order p in the VARs, we use what is stated or implied in the original papers. When such information cannot be found, we set k = 12 and 4 for monthly and quarterly observations, respectively. For all other frequencies, we use (int($\sqrt[3]{T^*}$)), where T^* is the number of observations available. The effective sample size is $T = T^* - k$. For each data set, we conduct all tests starting with the $(k + 1)th$ observation so that the results on cointegrating ranks from the one-step model selection approach are comparable with those from the two-step procedures. This is different than the practice of some of the papers we examined where the tests were conducted starting with the $(p + 1)th$ observation after p is selected out of a maximum of k lags in the first step (thus the effective sample is $(T^* - p)$).

For some data sets, it might be restrictive to assume a constant lag length for all variables over all equations in a VAR model. However, this is a typical assumption made by the studies under replication. For this reason, we follow the practice throughout the exercise. But see Maringer and Winker (2004) for discussions on the impact of allowing for more general lag structures on the VAR analysis.

Significance Levels

While most of the studies chose the 5% significance level in making inferences, 10% and 1% levels were occasionally used (usually in combination with the 5% level). We compare our results with those in the original studies based on the 5% level whenever available.

Finally, the reviewed studies drew conclusions on cointegrating relationships based on a wide range of tests. However, whenever more than one testing procedure is used in the original papers and the trace test is one of them, we compare our results with those based on the trace test, even if the conclusions based on it do not agree with those of the other methods. Of course, this simplification and the other modeling issues considered in this section are likely to complicate the replication efforts.

4. EMPIRICAL RESULTS

Replications

As mentioned before, we consider two specifications on deterministic trends for some data sets when the trend assumptions are not clear in the original papers. Therefore, we test a total of 187 relationships with the 165 data sets (161 were explicitly tested for cointegrating relationships in the original studies), of which one does not allow for any deterministic component, 26 allow for a constant in the cointegration space only, 111 allow for a linear time trend in data and 49 allow for a linear trend both in data and in the cointegration space. The (effective) sample sizes vary considerably across the studies, ranging from 23 to 7673 with a median size of 84.

The first row of Table 1 (Case 1) summarizes the basic results that are based on the one-step model selection procedure. About half of the cointegration test results reported in the published studies hold true if either one of the four information criteria is applied. Specifically, the performance of AIC, HQ and PIC are similar (around 55%), with SIC coming in with 51.6%. These numbers are largely in the same magnitudes as the proportions of agreements between the ADF test and the Johansen likelihood ratio test reported by Gregory et al (2004). For example, at the 5% significance level, the ADF test and the eigenvalue test and the ADF test and the trace test both reject or both fail to reject the null of no cointegration 50% and 45% of the times, respectively. Nevertheless, it should be pointed out that Gregory et al.'s (2004, Table III) results are not totally comparable to ours. They employ unifying but standard procedures in implementing the two competing tests across the data sets while our results are compared against those from the original studies that are featured with enormous heterogeneity in specifying and testing models.

To investigate what can possibly account for the differences between the test results of the model selection approach and the published ones, we also conducted cointegration tests conditioning on the lag orders chosen in the original studies. These were available for 64 relationships. Case 2 of Table 1 summarizes the results. Interestingly, for all four criteria, the new strategy leads to lower percentages of agreements with the published results. Nevertheless, in comparing Cases 1 and 2, there is a complicating factor. Even though we know what lag orders were chosen, we may have not used the same effective number of observations as the original studies, see the section entitled "Choice of Maximum Lag Length". Consequently, it is possible that our replication results were affected by the possible use of different effective sample size, especially for data sets with small samples. However, not all studies report the effective sample size, and we cannot pursue this issue further (if all regressions start from the $(p + 1)^{th}$ observation, we obtain essentially the same results as in Case 2). Having said this, the differences between Cases 1 and 2 are relatively small, from which it is fair to conclude that the choice of lag order does not explain much of the differences between the published results and our replications.

The replications in Case 3 are limited to those studies that used Johansen's (1988, 1991) likelihood ratio tests, while those in Case 4 are limited to those that use other tests (mainly the residuals-based ADF and $\hat{z}_{\alpha}$ tests). Clearly, AIC only replicate 28.8% of the results if they are based on the Johansen LR test but can replicate a much higher proportion of the results (72.6%) that are based on the residual-based tests. A similar pattern is also observed for the other three information criteria, although the differences between Case 3 and 4 are much smaller for these criteria. However, the comparisons between Cases 3 and 4 do not necessarily suggest that the model selection approach agrees more with the residual-based tests than with the LR test. This is because the results from the published studies are characterized by a rejection of the null of no cointegration. Among the 112 relationships that were tested for the existence of cointegration—not the specific rank, the null hypothesis was rejected in 87 cases. For these data sets, we define the replication of cointegration test as successful as long as an information criterion concludes that the cointegrating rank $r > 0$. As further evidence, Case 5 summarizes the replication results for those studies that use the LR procedure only to test whether or not a set of variables are cointegrated (a subset examined in Case 3). The performance of the model selection approach is now quite close to that in Case 4 replicating the residual-based test results.

Table 2 presents the replication results with the sample divided in other important ways. The first two rows of Table 2 show the impact of sample size on the performance of AIC, HQ, SIC and PIC. The second two rows show the impact of model dimension, while the last two rows show the impact of the inclusion of (weakly) exogenous variables in the system. Focusing on the last two rows, PIC can only replicate 26.9% of the reported results when the testing model contain one or more exogenous variables, in contrast to 57.8% in models without any exogenous variables. This also happens for AIC and HQ but to a lesser extent. SIC seems to be the least affected.

Conflict among the Alternative Tests

As noted earlier, the published studies employed a variety of methods to test for cointegration. Even if a same test was used, empirical implementations might still differ in the choice of lag order or window widths. To eliminate the impact of this heterogeneity on our earlier results, we apply the two-step trace test on all 187 relationships and compare the results with those from the model selection approach. The first row of Table 3 reports the percentage of times that the trace test and AIC, HQ, SIC and PIC select the same cointegrating ranks. The trace test yields the highest frequency of agreement with SIC (70.6%), followed by PIC (57.2%) and HQ (52.4%). The noticeably low percentage of agreements between the trace test and AIC (23.5%) is in line with the finding in the literature that AIC is not consistent and tends to overestimate the cointegrating rank (Kapetanios, 2004; Wang and Bessler, 2005). Among the four information criteria, the percentage of agreements also varies significantly. SIC and PIC are in agreement 62.0% of the time, a result consistent with the simulation evidence that finds the two criteria having similar performance.

6. CONCLUSION

This paper reexamines 165 published data sets and uses the model selection approach in testing for cointegration. The paper finds that AIC, HQ, SIC and PIC provide similar results on cointegrating relationships as the original studies in about 50% of the cases. This relatively low percentage of correspondence between different testing procedures suggests that caution should be taken in testing and interpreting cointegrating relationships. In particular, more effort should be placed on robustness checks in empirical work. Based on the actual data sets, we also find that SIC is in agreement with the widely used trace test more than 70% of the times. Together with the simulation results in the literature, the overall empirical evidence presented here indicates that the model selection approach (especially SIC and PIC) can be a useful complement to the widely used parametric tests in cointegration analysis for applied researchers.

REFERENCES

Akaike H.1973. Information theory and an extension of the maximum likelihood principle. In Petrov BN, Csaki F (eds.) *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest, p. 267-281.

Aznar A, Salvador M. 2002. Selecting the rank of cointegration space and the form of the intercept using an information criterion. *Econometric Theory* 18: 926-947.

Boswijk HP, Franses PH. 1992. Dynamic specification and cointegration. *Oxford Bulletin of Economics and Statistics* 54: 369-381.

Brüggemann R, Lütkepohl H. 2004. Practical problems with reduced rank ML estimators for cointegration parameters and a simple alternative. EUI Working Paper ECO 2004/20.

Chao JC, Phillips PCB. 1999. Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* 91: 227-271.

Gonzalo J, Lee T. 1998. Pitfalls in testing for long run relationships. *Journal of Econometrics* 86: 129-154.

Granger CWJ, King ML, White H. 1995. Comments on testing economic theories and the use of model selection criteria. *Journal of Econometrics* 67: 173-187.

Gregory AW, Haug AA, Lumuto N. 2004. Mixed signals among test for cointegration. *Journal of Applied Econometrics* 19: 89-98.

Hannan EJ, Quinn BG. 1979. The Determinant of the Order of an Autoregression. *Journal of Royal Statistics Society* Series B 41:190-95.

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

Johansen S. 1991. Estimation and hypothesis testing of cointegration vector in Gaussian vector autoregressive models. *Econometrica* 59: 1551-1580.

Kapetanios G. 2004. The asymptotic distribution of the cointegration rank estimator under the Akaike Information Criterion. *Econometric Theory* 20: 735-742.

Maringer D, Winker P. 2004. Optimal lag structure selection in VAR and VEC models. In Welfe A. (Ed.) *New Directions in Macromodeling*. Elsevier, Amsterdam, p. 213-234.

Phillips PCB. 1996. Econometric model determination. *Econometrica* 64: 763-812.

Phillips PCB, Ouliaris S. 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica* 58: 165-193.

Phillips PCB, Ploberger W. 1996. An asymptotic theory of Bayesian inference for time series. *Econometrica* 64: 381-412.

Phillips PCB, McFarland JW. 1997. Forward exchange market unbiasedness: the case of the Australian dollar since 1984. *Journal of International Money and Finance* 16: 885-907.

Reinsel GC, Ahn SK. 1992. Vector AR models with unit roots and reduced rank structures. *Journal of Times Series Analysis* 13: 353-375.

Schwarz G. 1978. Estimating the dimension of a model. *The Annals of Statistics* **8**: 461-464.

Wang Z, Bessler DA. 2005. A Monte Carlo study on the selection of cointegrating rank using information criteria. *Econometric Theory* 21: 593-620.

Wei, CZ. 1992. On predictive least squares principles. *Annals of Statistics* 20, 1-42.

Table 1. Replications of the tests for cointegration

	N	AIC	HQ	SIC	PIC
Case 1	161	54.7	55.3	51.6	55.9
Case 2	161	53.4	53.4	50.3	52.8
Case 3	66	28.8	40.9	45.5	50.0
Case 4	95	72.6	65.3	55.8	60.0
Case 5	17	82.4	64.7	52.9	58.8

Note: Each entry is the percentage of times where the cointegration test results were replicated successfully. In Case 1, we use the one-step model selection procedure to simultaneously determine the lag order and cointegrating rank. Case 2 conditions on the lag orders chosen by the original studies if they are available. Case 3 is limited to those studies that use Johansen's (1988, 1991) likelihood ratio tests, while Case 4 is limited to those that use other tests. Case 5 is the same as Case 3 except that replications are further limited to those studies that only test whether or not a set of variables are cointegrated (not the specific rank).

Table 2. Replications of the tests for cointegration with divided samples

	N	AIC	HQ	SIC	PIC
T <= 83	83	59.0	50.6	47.0	55.4
T > 83	78	47.4	56.4	53.8	50.0
m = 2	59	64.4	57.6	57.6	57.6
m > 2	102	47.1	51.0	46.1	50.0
n = 0	135	57.0	55.6	50.4	57.8
n > 0	26	34.6	42.3	50.0	26.9

Note: T is the sample size, while m and n are the number of endogenous and (weakly) exogenous variables in the system, respectively. Each entry is the percentage of times where the cointegration test results were replicated successfully. The lag orders chosen by the original studies are used if available.

Table 3. The correspondence between alternative methods in testing for cointegration (*N* = 187)

	AIC	HQ	SIC	PIC
Trace	23.5	52.4	70.6	57.2
AIC		58.3	30.5	19.8
HQ			55.1	40.1
SIC				62.0

Note: Each entry is the percentage of cases where the results of cointegration tests are the same from the two competing methods. The two-step trace tests use AIC, HQ, SIC and PIC to select lag order in the first step when they are compared to the corresponding one-step information criteria. The significance level is 5%.