9
Dimensionality Effect in Cointegration Analysis

JESÚS GONZALO AND JEAN-YVES PITARAKIS

1 Introduction

During the past decade a considerable amount of research has focused on the issue of stochastic trends in economic variables and subsequently on whether such trends are common to some or all of the variables in question, a phenomenon known as cointegration (Granger, 1981; Engle and Granger, 1987). Despite an abundant literature dealing with the development of cointegration tests and their applications in economics (see Engle and Granger, 1991, for a comprehensive review), an important issue that has often been overlooked is the impact that the system dimension (number of variables) might have on the accuracy of inferences. The issue is not merely a degrees of freedom problem (see Abadir, Hadri and Tzavalis, 1997) and becomes relevant in a wide range of applied fields. Indeed, the analysis of cross country or sectoral comovements of economic variables in the new growth literature or the determination of the number of factors in asset pricing theories in finance, for instance, are some among numerous other examples that involve the handling of very large systems.

Currently a popular approach for conducting inferences about the presence of cointegration is the reduced rank vector autoregressive (VAR) framework proposed by Johansen (1988, 1991) and Ahn and Reinsel (1990) following earlier work by Anderson (1958) and leading to a likelihood ratio statistic (LR thereafter) of the cointegration hypothesis. Although commonly used in applied work, little is known about its properties in large dimensional systems since most of the published simulation studies rarely included higher than bivariate or trivariate systems.

We wish to thank seminar participants at CEMFI, CORE, Queens University, Université de Montréal, University of California San Diego, Universidad Carlos III de Madrid, Universidad Complutense de Madrid, the Tinbergen Institute and two anonymous referees for helpful suggestions and comments that led to a substantial revision of an earlier draft. Financial support from the Spanish Secretary of Education (PB 950298) and the European Union Human Capital and Mobility Program (ERBCHBICT941677) is gratefully acknowledged.

The main objective of our paper is to introduce a set of new tools for inferring the cointegrating rank and examine their relative robustness to the dimensionality problem together with that of the LR based testing strategy. The structure of the paper is as follows. Sections 2 and 3 introduce the new criteria and analyze their asymptotic and finite sample behavior as the system dimension increases. Section 4 focuses on an alternative to testing, namely a model selection approach for estimating the cointegrating rank, and Section 5 concludes.

2 Estimation of the Cointegrating Rank: Test-Based Approaches

2.1 Theoretical Framework and Alternative Test Criteria

From the Granger representation theorem (Engle and Granger, 1987) a p-dimensional vector of I(1) variables X_t with cointegrating rank r $(0 \leq r \leq p)$ admits the following vector error correction model (VECM) representation

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} H_j \Delta X_{t-j} + \epsilon_t, \tag{1}$$

where we assume that ϵ_t is $NID(0, \Omega)$ with $|\Omega| \neq 0$ and k finite. Under the hypothesis of cointegration, the long run impact matrix Π can be written as $\Pi = \alpha\beta'$ where α and β are $p \times r$ matrices with $r = rank(\Pi)$. Thus a test of the cointegration hypothesis is equivalent to a test of the rank of the Π matrix. When $r = 0$, the components of X_t are not cointegrated and the VECM takes the form of a VAR in first differences. Under $0 < r < p$ there exist r linear combinations of the I(1) variables that are stationary and when $r = p$ the vector X_t is in fact a stationary process. In a series of papers Johansen (1988, 1991) and Ahn and Reinsel (1990) have developed a full information maximum likelihood estimation of (1) subject to the constraint that $rank(\Pi) = r$, leading to a likelihood ratio (LR thereafter) test of the cointegration hypothesis. The LR statistic is given by $-T\sum_{i=r+1}^{p}\log(1 - \hat{\lambda}_i)$ where the $\hat{\lambda}_i$s are the ordered eigenvalues of the quantity $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$ with $S_{ij} = (1/T)\sum_t R_{it}R_{jt}'$ and the R_{it}s denoting the respective residuals of the regression of ΔX_t and X_{t-1} on ΔX_{t-1}, ..., ΔX_{t-k+1}. Its asymptotic distribution obtained in Johansen (1988) was shown to be free of nuisance parameters, depending solely on the number of common stochastic trends $(p - r)$ driving the system. It is important to note that Johansen's framework is based on well known techniques in the multivariate analysis literature, namely the reduced rank regression and canonical correlation analysis (see Izenman, 1975, among others). The LR statistic is only one among many alternative test statistics proposed for inferring the rank of possibly rank deficient matrices (Hotelling, 1931; Pillai, 1954). Moreover simulation studies (see Olson, 1974) have shown that these test statistics behave very differently in a finite sample context despite the fact that they all share a common limiting distribution. Our initial objective here is to investigate their properties within the context of cointegration as well. Our interest is in finding a properly sized test statistic for conducting meaningful inferences in large dimensional systems. At this stage it is important to reiterate that the dimensionality problem does not arise solely because of the resulting degrees of freedom limitations. As shown in Abadir *et al.* (1997), when a VAR contains I(1) variables, increasing its dimension will proportionately raise the asymptotic bias of $\hat{\Pi}$ even when all regressors are independent of each other. In addition to the LR, the set of alternative test statistics considered in this paper are given by

(a) $PB = T\sum_{i=r+1}^{p}\hat{\lambda}_i$ (Pillai–Bartlett),
(b) $HL = T\sum_{i=r+1}^{p}(\hat{\lambda}_i/(1-\hat{\lambda}_i))$ (Hotelling–Lawley).

Through a first order expansion of the functional forms it is straightforward to observe that the PB and HL statistics will have the same asymptotic distribution as that of the LR. Indeed, given that under $r=0$ the $\hat{\lambda}_i$s are $O_p(1/T)$ we have $LR = PB + o_p(1)$ and $HL = LR + o_p(1)$, illustrating the fact that the asymptotic distribution of the PB statistic provides also an approximation to that of the LR or HL. Their finite sample distributions however will display important discrepancies both across each test statistic and compared with the asymptotic approximation. This can be noted by observing that the three criteria will satisfy the inequality $PB \leq LR \leq HL$, suggesting that in finite samples and when inferences are based on the asymptotic critical values the HL statistic will reject the null more frequently than the LR or PB and with the PB statistic rejecting it the least frequently. In order to isolate the impact of dimensionality this paper will mainly focus on models with $k=1$. Although in applied work the lag issue raises serious modeling questions incorporating it here would prevent us from isolating the true impact of dimensionality on inferences.

2.2 Finite Sample Distributions and Dimensionality

In order to highlight the influence of the system dimension p on the distributions of the test statistics we initially focused on the empirical size of each test under the null of no cointegration. More specifically we generated p dimensional independent random walks and computed the rejection frequencies of the true null of no cointegration ($r_0 = 0$) using the common asymptotic critical values and a 5 percent nominal level. All our experiments have been performed using $N = 10{,}000$ replications and the asymptotic critical values were approximated using a sample of size $T = 10{,}000$. Results for these preliminary experiments are displayed in Table 1 for system dimensions ranging from $p = 2$ to $p = 10$.

Table 1. Empirical size under no cointegration (5% nominal level)
DGP: $\Delta x_{it} = \epsilon_{it}$, $\epsilon_{it} \equiv N(0, 1)$, $N = 10{,}000$ replications

p	LR	PB	HL	LCT	RALR	LR	PB	HL	LCT	RALR
	$T = 30$					$T = 90$				
2	7.00	2.90	12.96	4.86	5.24	5.12	4.00	7.15	4.69	4.60
3	9.42	1.13	24.69	4.63	4.88	6.21	3.18	9.45	4.58	4.77
4	13.04	0.34	44.00	4.22	3.94	6.64	2.58	12.82	4.54	4.62
5	20.75	0.07	72.06	4.44	3.59	8.10	1.71	21.2	4.27	4.27
6	31.66	0.00	91.62	5.04	2.68	10.48	1.12	31.44	4.26	3.90
7	47.44	0.00	99.02	5.78	1.98	12.12	0.72	43.76	4.12	3.78
8	67.42	0.00	100.00	7.90	2.00	14.00	0.20	57.78	3.62	2.82
9	85.00	0.00	100.00	10.46	1.32	19.96	0.12	76.30	4.04	2.72
10	96.69	0.00	100.00	19.41	0.96	25.03	0.05	88.11	3.87	2.23
	$T = 150$					$T = 400$				
2	5.02	4.18	5.78	4.78	4.74	5.10	5.20	6.37	5.10	5.45
3	5.90	4.29	7.91	5.10	4.76	5.13	4.47	5.80	4.70	4.80
4	6.46	3.38	9.98	4.76	4.96	5.17	4.27	5.97	4.67	4.80
5	7.11	3.12	13.13	4.84	4.90	5.93	4.53	8.20	5.37	5.37
6	7.84	2.44	18.50	4.68	4.50	5.87	3.70	8.30	4.40	4.40
7	8.18	1.44	23.52	4.10	3.90	6.70	3.77	9.83	4.97	4.87
8	9.70	1.12	31.40	4.08	3.58	6.50	2.87	11.13	4.63	4.44
9	11.60	0.92	42.26	4.28	3.58	8.27	3.37	15.27	5.47	5.27
10	14.69	0.55	55.75	4.25	3.17	8.70	3.07	19.03	5.30	4.90

A general picture that emerges from Table 1 is the strong negative impact of the system dimension on the size properties of the LR, PB and HL statistics. Although it is natural to expect a reduction in the accuracy of inferences as p increases, the magnitudes of the distortions are striking. The latter may increase substantially even when a system is augmented by a single additional variable. In addition the magnitudes presented in Table 1 also suggest that the inclusion of additional variables necessitates an extremely important incremental increase in the sample size so as to keep the distortions similar in the original and augmented models.

The distortions of the LR and HL statistics reach unacceptable levels as we move from a medium sized ($p = 5$) to larger systems. In the case of the LR statistic under $p = 8$ for instance the frequency of rejection of $r_0 = 0$ is 67.42 percent when $T = 30$, 14 percent when $T = 90$ and close to 10 percent for $T = 150$. In applied work such sample sizes are not uncommon especially when one is also interested in models with structural breaks or thresholds. The PB statistic on the other hand suffers from the opposite problem, being unable to move away from $r = 0$ unless a very large sample size becomes available. Clearly all three (LR, PB, HL) test criteria appear inappropriate for conducting inferences about the presence of cointegration in large dimensional systems, even when moderately large sample sizes are available: the LR and HL statistics will wrongly point towards too many stationary components and the PB towards too few. Although not reported here due to space considerations (available upon request) it is also important to point out that the results presented in this section and throughout the rest of the paper were highly robust to numerous alternative specifications that included deterministic components (i.e. constant and trend terms) in the fitted models.

In finite samples the poor approximation provided by the asymptotic distribution is not a problem novel to this nonstationary multivariate time series framework. Indeed it is also a well documented issue in the multivariate analysis and canonical correlation literature. Since Bartlett (1947) for instance, numerous authors introduced correction factors to the standard LR statistic with the motivation of having the first moment of the finite sample and asymptotic distributions match up to a certain order of magnitude (see Fujikoshi, 1977, for correction factors in the context of standard canonical correlation analysis and Taniguchi, 1991, for similar results in the stationary time series framework). Unfortunately such analytical corrections pose extremely challenging problems in multivariate systems with I(1) components and to our knowledge the issue has been tackled only partially in simple univariate AR(1) models (see Nielsen, 1997). However, the recent derivation of the multivariate joint moment generating function of S_{11} and S_{10} in Abadir and Larsson (1996) will almost surely open the path to further exact results in the nonstationary VAR framework for test statistics that are functions of these two moment matrices. Given the availability of powerful computer resources an alternative and perhaps more accurate strategy has been to use techniques such as response surface regressions for quantile estimation (see MacKinnon, 1994, 1996; MacKinnon, Haug, and Michelis, 1996, for an application to unit root distributions). Within our framework the specific behavior of the LR and PB test statistics may also allow us to design an alternative criterion with good size properties following the simulation path. Indeed a close analysis of the full densities of the LR and PB statistics across numerous sample sizes and system dimensions[1] prompts us to propose an alternative criterion based on the linear combination of LR and PB that minimizes the following distance:

$$\min_{\omega_1,\omega_2} \sum_{p=1}^{P} (q_p(\alpha) - \omega_1 LR(\alpha, T) - \omega_2 PB(\alpha, T))^2$$
$$s.t. \quad \omega_1 + \omega_2 = 1,$$

where $q_p(\alpha)$ denotes the α percent asymptotic critical value in the p-dimensional model and $LR(\alpha,\ T)$ and $PB(\alpha,\ T)$ the finite sample counterparts. In our estimations we used $p = 1, \ldots, 20$ for $T = 90$ and $T = 150$ respectively and an equal weighting $\omega_1 = \omega_2 = 0.5$ between LR and PB gave the best results with excellent diagnostics across all the experiments. This leads us to propose a

linear combination test statistic (LCT thereafter) expressed as $0.5(LR + PB)$. Empirical size estimates corresponding to the LCT statistic under the null of no cointegration are also displayed in Table 1 together with the standard test criteria. Except for the extreme case where $p = 10$ and $T = 30$ the LCT statistic can be seen to track the asymptotic distribution very closely, with empirical size estimates extremely close to the nominal 5 percent level across all system dimensions. To gain further insight on its behavior Figures 1a and 1b also present the difference between the asymptotic and finite sample critical values across various values of the system dimension. The plot corresponding to the LCT criterion remains horizontal across all values of p clearly highlighting the absence of any distortion even within very large system dimensions.

Note that this result is robust to any specification of the covariance matrix of the errors in the VECM since the eigenvalues are invariant to any nonsingular linear transformation of the variables. An important additional advantage of the LCT statistic comes from the fact that it shares the same asymptotic distribution as that of the LR or PB but this time without the distortions plaguing both statistics. As mentioned previously the main distortion characterizing the inferences based on the LR statistic arises from a drastic

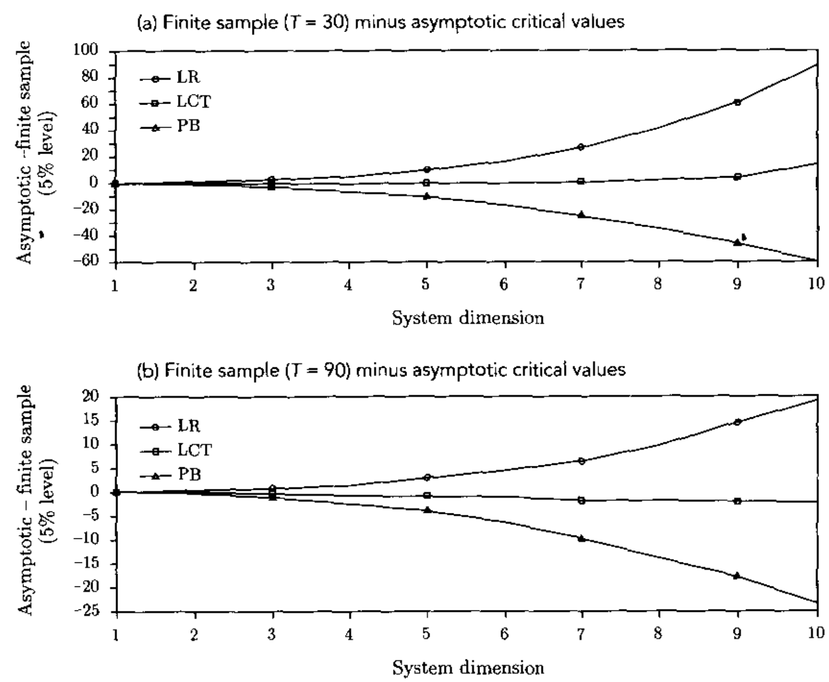

Figure 1

rightwards shift of the distribution as p increases for a given sample size. An intuitive explanation of this phenomenon can be inferred by analysing the $o_p(1)$ terms in the Taylor expansion of $LR = -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i)$ where $\hat{\lambda}_i = O_p(1/T)$ for $i = r+1, \ldots, p$. The latter consist of sums of powers of the estimated eigenvalues which are in turn premultiplied by a factor proportional to the sample size. In finite samples the $\hat{\lambda}_i$s are typically characterized by an important upward bias (the maximum eigenvalue in particular) the cumulated weight of which tends to increase the magnitude of the LR statistic relative to its asymptotic value. Similarly since the PB statistic is the first term in the expansion of LR, its lagging behavior (relative to its asymptotic distribution) can be explained by the fact that in small samples the normalizing factor T is not strong enough to make the statistic's finite sample distribution shift sufficiently rapidly towards its asymptotic counterpart. The LCT statistic offers a compromise between the LR and PB by reducing the weight of the $o_p(1)$ terms by half.

The size distortions characterizing the LR statistic have also been observed in previous studies focusing on possibly misspecified (due to residual autocorrelation for instance) small dimensional systems. A popular ad hoc correction taking the form $((T-p)/T)\ LR$ has been suggested in Reinsel and Ahn (1992) within the context of cointegration testing and was subsequently used in most empirical studies. In order to compare its size behavior with that of the LCT criterion Table 1 also presents corresponding size estimates for this corrected LR statistic (denoted RALR). Although the RALR statistic also appears to track the asymptotic distribution very closely (as judged by the empirical size estimates) even under limited sample sizes, its behavior seriously degenerates as the system dimension increases suggesting that inferences based on RALR might be unable to point to ranks other than zero. An empirical size of 0.96 percent or 2.23 percent under $T = 30$ and $T = 90$ when $p = 10$ for instance is an obvious indication of its limitation.

An extreme scenario that is worth mentioning, because it affects all the tests, is that of a model as in (1) at the border of the estimability region. More specifically the estimability condition in the context of model (1) under $k = 1$ requires $T \geq 2p+1$ (see Brown, 1981) and if the system dimension p is such that $p \to T/2$, it is straightforward to show that $\hat{\lambda}_1 \to 1$ even when the true $\lambda_1 = 0$, a phenomenon that can be qualified as spurious cointegration.

3 Performance under the Presence of Cointegration

Our next concern is to evaluate the performance and overall behavior of the alternative test criteria when the systems are cointegrated. Given the highly distorted nature of the PB and HL statistics our analysis will mainly concentrate on the relative performance of the LR, LCT and RALR statistics. Our motivation here is to investigate whether distortions similar to the ones that were observed under $r_0 = 0$ will also be present when $r_0 > 0$ and whether the LCT statistic will maintain its attractive features under such scenarios. Since the sequential testing procedure (see Johansen, 1995, ch. 12) can be used to construct an estimate of the cointegrating rank we will also naturally investigate the ability of the various criteria to correctly detect the true rank and more importantly the influence of the system dimension when conducting such inferences.

So far we have seen that under the null of no cointegration the distorted behavior of the LR statistic manifests itself in the form of too many rejections of the true null $r_0 = 0$. Due to the sequential nature of the test however it is important to note that such distortions will not necessarily remain when testing ranks greater than zero. Specifically, if we let $LR(r)$ denote the LR statistic under r cointegrating relationships (i.e. $LR(r) = -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i)$) and assuming that the sequential test is implemented starting from the hypothesis of no cointegration, we initially compare $LR(0)$ with its corresponding quantile c_p say and conclude $\hat{r} = 0$ if $LR(0) < c_p$. If $LR(0) \geq c_p$ we then proceed with $LR(1)$ and let $\hat{r} = 1$ if $LR(1) < c_{p-1}$ and so on. The fact that a large system dimension induces frequent rejections of the null of no cointegration due to the occurrence of $LR(0) > c_p$ beyond its theoretical level does not mean that the event $LR(1) > c_{p-1}$ will occur as frequently when $r_0 = 1$. An equivalent question to ask is whether in a $p-1$ dimensional system with $r_0 = 0$ the discrepancies characterizing the distribution of $LR(0)$ will be similar to those occurring in the distribution of $LR(1)$ in a p dimensional system with $r_0 = 1$? Intuitively it is the bias in the first eigenvalue that might be causing an inflated value of the LR statistic. Before focusing more closely on this issue we will initially evaluate the influence of the system dimension on the ability of the sequential testing procedure to detect the true rank. In other words, given that a system is characterized by $r_0 > 0$ cointegrating relationship we ask how frequently will the three test statistics point to the true rank across different system dimensions. The DGP we have considered is given by $\Delta X_t = \Pi X_{t-1} + \epsilon_t$ with ϵ_t a $NID(0, I_p)$ random disturbance and $\Pi = diag(\rho-1, 0, \ldots, 0)$. Thus the system is characterized by a single cointegrating relationship, the strength of which is determined by the magnitude of ρ. For this experiment we chose to set $\rho = 0.7$ which corresponds to a magnitude of the unique nonzero population eigenvalue of $\lambda_1 = (1-\rho)/2 = 0.15$. Table 2 displays the relevant correct decision frequencies for values of p ranging from 2 to 10.

Given our previous findings on the behavior of LR under no cointegration we present the LR based frequencies using both asymptotic and finite sample critical values. The discrepancies between the two versions clearly support our analysis of the LR under the null of no cointegration. Although it is natural to expect a reduction in the accuracy of inferences as the system dimension increases, the magnitudes displayed in Table 2 are striking even when only a

Table 2. Correct decision frequencies $r_0 = 1$ (5% nominal level)
DGP: $x_{1t} = 0.7x_{1t-1} + \epsilon_{1t}$, $\Delta x_{it} = \epsilon_{it}$ $i = 2, \ldots, p$, $\epsilon_{it} \equiv N(0, 1)$, $N = 10{,}000$ replications

p	LR(ad)	LR(fsd)	LCT	RALR	LR(ad)	LR(fsd)	LCT	RALR
	$T = 30$				$T = 90$			
2	20.02	15.10	15.06	16.40	87.84	85.88	86.26	86.86
3	13.98	7.94	7.10	7.80	56.48	52.42	49.86	51.68
4	13.84	5.32	4.48	4.60	34.78	29.46	26.36	27.80
5	18.16	4.82	4.48	3.62	25.28	17.02	15.40	16.02
6	24.82	4.30	3.60	2.46	20.46	12.42	10.52	10.84
7	33.40	4.10	3.82	1.60	19.00	9.52	7.08	6.70
8	42.58	3.72	5.30	1.20	19.04	7.56	5.74	4.84
9	40.64	3.44	7.90	0.88	21.32	6.26	5.06	3.84
10	24.42	3.40	14.50	0.92	23.78	5.96	4.42	3.06
	$T = 150$				$T = 400$			
2	94.90	95.08	95.00	95.08	95.52	95.40	95.52	95.54
3	91.84	92.14	91.12	91.52	95.14	95.34	95.32	95.36
4	74.72	71.16	69.18	70.80	94.54	94.56	94.72	94.80
5	59.02	53.12	50.40	52.22	94.86	94.86	95.52	95.58
6	43.66	36.04	32.52	33.54	94.18	94.72	94.88	95.12
7	34.30	25.22	22.20	22.46	94.02	94.68	94.48	94.90
8	30.34	19.38	16.60	16.24	92.16	92.36	91.82	92.56
9	27.24	14.70	12.98	11.92	87.64	86.54	85.90	86.86
10	26.64	13.24	11.34	9.68	84.30	79.88	79.02	80.02

Note: LR(ad) and LR(fsd) refer to the correct decision frequencies based on the asymptotic and finite sample critical values respectively.

single additional variable is added to the system. Under $T = 90$ for instance the correct decision frequencies decrease by an amount close to 50 percent as we move from a bivariate to a trivariate system. A more extreme scenario is a reduction of the correct decision frequency by an amount close to 300 percent as we move from a bivariate to a ten dimensional system. Although these gaps tend to fade as we increase the sample size they remain highly significant. Overall these findings suggest that even the inclusion of a single additional variable might reduce the quality of inferences by an unexpectedly high proportion, an observation valid for all three test statistics. Comparing the magnitudes corresponding to LR(fsd), LCT and RALR it is also clear that for $T \geq 90$ the three statistics display a very similar ability to point to the true rank of $r_0 = 1$ across all system dimensions thus the advantageous size behavior of the LCT criterion observed in Table 1 is not coupled with any sort of distortion in its ability to point to the true rank *relative* to the other test statistics. It is also interesting to note the close agreement between the frequencies of the LCT statistic and those obtained using the "size corrected" LR statistic. This confirms our point about the ability of the LCT statistic to remain close

to the asymptotic distribution across a wide range of sample size/system dimension pairs.

In order to gain further insight on the behavior of the LCT statistic when cointegration is present we next investigated the proximity of the finite sample distributions to their asymptotic counterparts under $r_0 = 1$ by computing the 5 percent critical values of the $LR(1)$ statistic in p-dimensional systems with $r_0 = 1$ and comparing them with their asymptotic counterparts from a $p - 1$ dimensional system with $r_0 = 0$. In a p-dimensional system with $r_0 = 1$ for instance it is clear that the *limiting distribution* of $LR(1) = -T\sum_{i=2}^{p}\log(1 - \hat{\lambda}_i)$ will be equivalent to that of the LR statistic in a $p - 1$ dimensional system with no cointegration. This will obviously be true asymptotically but important distortions might arise when dealing with small samples and/or large system dimensions. Using the same DGP as above Table 3 presents the relevant critical values for different sample sizes. For clarification purposes the first row in the LR column for instance refers to a DGP with $p = 2$ and $r_0 = 1$ while the column labeled $LR(p, r_0 = 0)$ represents the critical value computed using a DGP with $p = 1$ and $r_0 = 0$. The two should obviously coincide asymptotically.

Table 3. 5% critical values of LR, LCT, and RALR under $r_0 = 1$
DGP: $x_{1t} = 0.7x_{1t-1} + \epsilon_{1t}$, $\Delta x_{it} = \epsilon_{it}$ $i = 2, \ldots, p$, $\epsilon_{it} \equiv N(0, 1)$, $N = 10{,}000$ replications

$p - r_0$	LR	LR*	LCT	RALR	LR	LR*	LCT	RALR	Asymptotic
	$T = 30$				$T = 90$				
1	4.56	4.59	4.16	4.02	4.06	4.26	4.02	3.97	4.16
2	13.09	13.22	12.16	11.98	12.59	12.77	12.21	12.17	12.27
3	25.92	26.73	23.34	22.53	24.70	25.04	23.85	23.60	24.29
4	43.21	45.23	38.17	36.18	40.91	41.82	38.64	38.64	40.40
5	66.21	69.30	56.87	52.84	62.23	62.69	58.87	58.08	59.78
6	95.28	99.66	80.41	73.36	87.61	88.24	82.25	80.80	83.68
7	130.85	137.97	107.59	95.66	118.19	118.08	109.92	107.69	111.65
8	174.63	184.05	139.95	121.38	152.55	152.69	140.80	137.29	143.12
9	227.92	238.95	178.31	151.17	191.19	192.82	175.46	169.95	178.29
	$T = 150$				$T = 400$				
1	4.10	4.15	4.07	4.05	4.12	4.10	4.10	4.09	4.16
2	12.42	12.61	12.19	12.17	12.31	12.35	12.30	12.18	12.27
3	24.56	24.50	24.03	23.90	24.31	24.50	24.15	24.00	24.29
4	41.38	41.25	40.15	40.00	40.95	40.42	40.35	40.10	40.40
5	61.31	61.46	59.28	58.86	61.10	61.46	59.09	58.00	59.78
6	86.13	86.30	82.86	82.11	84.95	84.81	83.24	83.00	83.68
7	115.72	115.56	110.74	109.55	113.12	115.56	110.90	110.00	111.65
8	148.81	148.86	141.72	139.88	145.85	148.86	142.14	140.08	143.12
9	186.83	186.79	177.20	174.37	180.25	182.24	178.00	175.24	178.29

* $(p, r_0 = 0)$

It is again clear that the finite sample distribution of the LR statistic lies to the right of its asymptotic counterpart while that of the LCT tracks the latter very closely (although our judgement here is based solely on the 5 percent critical values similar discrepancies occur throughout the whole quantiles). It is however important to also note the discrepancies between the magnitudes appearing in the LR and $LR(p, r_0 = 0)$ columns respectively. Although both suggest that the finite sample distributions of the LR statistic lie to the right of their asymptotic counterpart it is also clear that for the same value of $p-r$ the critical values corresponding to $LR(p, r_0 = 0)$ are much larger. This suggests that the inflated empirical sizes we have observed under the null of no cointegration (i.e. Table 1) will be less pronounced when considering testing sequences involving ranks greater than zero. This is indeed what we observed when we evaluated the probability of pointing to ranks greater than one (i.e. $P[LR(0) > c_p \text{ and } LR(1) > c_{p-1}]$) for a DGP with $r_0 = 1$. It is important to note however that in this case the choice of the "strength" of the cointegrating relationship will play an important role in the sense that the discrepancies between the LR and $LR(p, r_0 = 0)$ columns appearing in Table 3 will not be robust to the chosen magnitude for ρ. Typically when we repeated the same experiment with smaller values of ρ (i.e. stronger cointegration) the discrepancies between LR and $LR(p, r_0 = 0)$ also appeared much stronger. Regarding the LCT statistic under the presence of cointegration it is again clear from Table 3 that its behavior under the presence of cointegration (judged by the closeness of its finite sample distributions to their asymptotic counterpart) does not present any significant difference to that observed under $r_0 = 0$ in the previous section.

4 Estimation of the Cointegrating Rank: Model Selection Procedure

In the previous section we saw that most standard tools for inferring the cointegrating rank were often ineffective when the ratio of the system dimension to the sample size was too large. Although the new criterion we introduced was able to significantly reduce the degree of size distortions plaguing the standard test statistics, by construction and regardless of the quality of the statistic being used the sequential testing strategy cannot lead to consistent estimates of the cointegrating rank because of the constraint imposed by the size of the test. This problem could become particularly intensified in large dimensional systems where the testing sequences are long. When consistency is a desired feature an alternative way of approaching the problem is to view the estimation of the cointegrating rank as a model selection problem where one chooses a model among a portfolio (assumed to contain the true model) of $p + 1$ competing models. The idea of using a model selection procedure for

estimating the rank of a matrix was also recently investigated in Donald and Cragg (1995) within the context of a stationary multivariate normal framework. A general class of model selection criteria is given by $IC(\ell) = -2\log L_\ell + c_T m_\ell$ where m_ℓ denotes the number of free parameters to be estimated under the hypothesis that there are ℓ cointegrating relationships ($m_\ell = 2p\ell - \ell^2$), L_ℓ is the likelihood function and c_T is a deterministic penalty term. Many well known information theoretic criteria are encompassed in the above specification. Indeed, when $c_T = 2$, $IC(\ell)$ corresponds to the Akaike criterion (Akaike, 1969, 1976), $c_T = \log(T)$ corresponds to Schwarz's BIC criterion (Schwarz, 1978) and when $c_T = 2c\log(\log(T))$ with $c > 1$ we have the Hannan and Quinn (1979) criterion. According to the model selection procedure, r is estimated by $\hat{r}$ where $\hat{r}$ is chosen such that $IC(\hat{r}) = \arg\min\{IC(\ell), \ell = 0, \ldots, p\}$. More commonly, the general expression for the various criteria is given by

$$IC(\ell) = \log\left|\hat{\Omega}(\ell)\right| + \frac{c_T}{T} m_\ell, \tag{2}$$

where c_T and m_ℓ are defined as before and $\hat{\Omega}(\ell)$ corresponds to the estimated error covariance matrix from model (1) under the hypothesis that Rank(Π) $= \ell$. For computational convenience we can focus on a transformed objective function that involves directly the eigenvalues of $S_{00}^{-1}S_{01}S_{11}^{-1}S_{10}$ since those are readily available. Noting that the eigenvalues of $S_{00}^{-1}S_{01}S_{11}^{-1}S_{10}$ are the same as the eigenvalues of $I_p - S_{00}^{-1}\hat{\Omega}$, where $\hat{\Omega}$ is the covariance matrix of the residuals in the unrestricted VECM, it is straightforward to show that one can instead focus on the minimization of

$$\tilde{IC}(\ell) = IC(\ell) - IC(p) = -T\sum_{i=\ell+1}^{p}\log(1-\hat{\lambda}_i) - c_T(p-\ell)^2 \tag{3}$$

where $\tilde{IC}(p) = 0$, thus rendering the approach trivial to implement once estimates of the eigenvalues have been obtained. Typically, as p the system dimension increases the LR portion of the criterion will increase and this latter increase will be balanced by the decrease due to the presence of the penalty term. The following proposition establishes the asymptotic properties of the resulting estimates.

Proposition 4.1. *Letting r_0 denote the true rank of Π in (1) and $\hat{r} = \arg\min_{0\le\ell\le p}\{\tilde{IC}(\ell)\}$, then $\hat{r} \xrightarrow{p} r_0$ iff (i) $\lim_{T\to\infty} c_T = \infty$ and (ii) $\lim_{T\to\infty}(c_T/T) = 0$.*

Proof. See Appendix.

Thus provided that the chosen penalty in (2) or (3) satisfies both of the above requirements weak consistency of the rank estimate will follow. It is true, however, that the above two conditions allow for an extremely wide spectrum of possibilities beyond the conventional AIC, BIC or HQ type of penalties, thus leaving the choice of an appropriate penalty to become an empirical and model specific problem. The inconsistency of rank estimates resulting from constant penalty criteria (violation of condition (i) in proposition 4.1) is due to the fact that the probability of selecting some rank ℓ greater that r_0 does not vanish asymptotically. This is a well known problem in the model selection literature and its impact has often been downweighted for criteria such as the AIC by arguing that although positive, the limiting probability is usually very small. Within our framework, however, this nonzero limiting probability of overranking is much less innocuous and constant penalty criteria may lead to highly distorted results. This is due to the fact that in our setting the nonzero limiting probability of overranking is an increasing function of the system dimension with immediate consequences for the use of constant penalty based criteria in large dimensional systems. To illustrate this feature let us consider a p dimensional system having $r_0 = 0$. From (3) the probability that the model selection criterion will point to a rank greater than 0, say equal to 1 can be written as $P[\tilde{IC}(1) < \tilde{IC}(0)] = P[-T\log(1 - \hat{\lambda}_1) > c(2p - 1)]$ where c represents a constant penalty term. In this setting the quantity $-T\log(1 - \hat{\lambda}_1)$ turns out to coincide with the well known likelihood ratio statistic (also called λ^{max}) for testing $r = 0$ against $r = 1$ proposed in Johansen (1991) and the full numerical distribution of which has been recently calculated in MacKinnon *et al.* (1996). Using the companion program the authors provide it is therefore possible to evaluate very accurately the above limiting probability of overranking across various values of p. Using values of $p = 3, 5, 7$, and 10 respectively the computation of the above probability for the AIC criterion (i.e. $c = 2$) led to values of 48.52 percent, 69.86 percent, 84.10 percent and 95.04 percent clearly suggesting that the AIC or any other constant penalty criterion will persistently overrank, especially as the system dimension increases. This is in sharp contrast with the lag length estimation framework in which the AIC can be shown to point to the true lag as p increases (see Gonzalo and Pitarakis, 1997).

We now turn to the study of the performance of the criteria leading to consistent rank estimates, namely the BIC and HQ. For the sake of comparability we conducted the same experiments as the ones used in the evaluation of the testing procedure. Thus our initial goal is to examine the "size" behavior of the various model selection criteria as the system dimension increases. Results for this experiment are presented in Table 4a using samples of size $T = 30$, 90, 150 and 400.

The figures represent the number of times the criterion selected the true rank $r_0 = 0$. Given its strong penalty the BIC clearly converges to the true rank very rapidly, selecting it close to 100 percent most of the times for $T \geq 90$ and any value of p. Note that this characteristic of the BIC may be solely due to the strength of its penalty and not to its genuine ability to detect the true rank. The HQ criterion requires samples greater than $T = 90$ in order to achieve a correct decision frequency greater than 90 percent regardless of the

Table 4a. Model selection criteria—correct decision frequencies $r_0 = 0$
DGP: $\Delta x_{it} = \epsilon_{it}$, $\epsilon_{it} \equiv N(0, 1)$, $i = 1, \ldots, p$, $N = 10{,}000$ replications

p	AIC	BIC	HQ	AIC	BIC	HQ
	$T = 30$			$T = 90$		
2	59.32	89.32	72.38	61.88	97.96	87.90
3	39.96	89.06	61.90	47.48	98.98	85.80
4	24.14	87.08	49.24	34.08	99.56	85.40
5	12.82	84.62	38.14	23.02	99.54	82.40
6	4.48	80.28	24.22	14.88	99.78	81.64
7	1.38	73.86	13.86	9.06	99.88	80.64
8	0.24	63.36	6.08	4.46	99.90	79.66
9	0.00	49.68	1.94	2.14	99.90	76.06
10	0.00	33.40	0.34	0.86	99.96	74.74
	$T = 150$			$T = 400$		
2	63.34	98.82	89.86	63.64	99.64	93.64
3	49.52	99.66	90.14	50.14	99.98	95.08
4	34.84	99.88	90.32	36.98	100.00	95.98
5	24.18	99.98	90.36	27.44	100.00	96.88
6	17.62	99.96	91.66	19.52	100.00	97.50
7	11.22	99.98	91.34	13.32	100.00	98.04
8	6.42	100.00	91.54	8.90	100.00	98.44
9	3.72	100.00	91.42	5.44	100.00	98.50
10	1.92	100.00	92.04	3.30	100.00	98.72

magnitude of the system dimension. Although its penalty term is such that the resulting estimates of the cointegrating rank are consistent the quantity $2\log(\log(T))$ converges to infinity very slowly explaining why the criterion has a rather strong tendency to overrank under smaller sample sizes. For $T \geq 150$ however its ability to point to the true rank improves drastically. At this stage it is also worth pointing out that the frequencies corresponding to the AIC criterion fully support our previous discussion about its behavior as p increases. Although not presented here, under $T = 400$ it pointed to $r = r_0 + 1 = 1$, 50 percent, 70 percent and 95 percent of the times under $p = 3, 5$, and 10 respectively, thus confirming the accuracy of our previous computations.

Bearing in mind their "size" behavior we now turn to the performance of the model selection criteria under cointegrated systems using models identical to the ones considered when evaluating the features of the test statistics (Table 4b).

The first point worth mentioning is the clear unreliability of the BIC criterion which has very little ability to point to the true rank $r_0 = 1$ even for moderately large sample sizes. For $T = 90$ or $T = 150$ for instance and when $p \geq 5$ its frequency of correct decision lies below 10 percent which is inferior to both the HQ and the LCT or RALR statistics. These frequencies confirm

Table 4b. Model selection criteria—correct decision frequencies $r_0 = 1$
DGP: $x_{1t} = 0.7x_{1t-1} + \epsilon_{1t}$, $\Delta x_{it} = \epsilon_{it}$, $i = 2, \ldots, p$, $\epsilon_{it} \equiv N(0, 1)$,
N = 10,000 replications

p	AIC	BIC	HQ	AIC	BIC	HQ
	$T = 30$			$T = 90$		
2	60.90	26.43	50.17	80.27	77.80	88.03
3	44.17	15.40	37.70	62.50	30.97	73.57
4	40.27	13.00	37.13	47.57	9.40	56.93
5	33.60	15.63	41.83	37.20	3.30	43.00
6	24.40	17.70	42.10	28.23	1.20	33.60
7	13.23	23.03	37.60	23.40	0.57	28.73
8	4.40	30.90	27.97	16.77	0.13	26.27
9	1.13	39.17	17.60	12.10	0.23	26.70
10	0.07	46.90	6.80	7.43	0.03	27.00
	$T = 150$			$T = 400$		
2	81.57	96.47	91.23	80.37	97.87	93.03
3	62.93	76.07	90.47	63.33	99.80	94.23
4	48.97	30.40	86.83	49.17	99.93	95.00
5	37.83	9.00	75.90	37.50	99.33	96.40
6	28.70	1.83	58.80	27.53	88.70	96.03
7	20.03	0.33	44.80	19.50	53.47	97.70
8	14.03	0.17	32.00	14.93	21.33	98.17
9	9.33	0.00	27.47	9.13	6.73	98.00
10	6.43	0.03	21.90	5.87	1.60	96.40

that in large dimensional systems the BIC will be unable to move away from $r = 0$ most of the time even if a sufficiently large sample size is available. Turning to the performance of the HQ criterion under values of T for which it displayed good size behavior (i.e. $T \geq 150$) its ability to point to the true rank is impressive when compared with either the standard test based inferences or the other model selection criteria. More importantly the HQ criterion turns out to be the only criterion able to achieve reasonable results even under a large dimensional system. Under $p = 5$ for instance it showed a tendency to outperform the test criteria by magnitudes as high as 25 percent. Similarly under $p = 10$ its performance was at least twice as good as the test based inferences across a wide range of system dimensions.

5 Conclusion

In this paper we studied various approaches for inferring the cointegrating rank by focusing on their robustness to the system dimensionality. We showed that standard (uncorrected) tools such as the LR statistic will lead to highly distorted inferences as the dimension of the system under study increases. We introduced a new test criterion (LCT) with the same limiting distribution as the LR, similar power properties and more importantly with no size distortions across a wide range of system dimensions. This of course is not meant to suggest that large dimensional systems can be dealt with as accurately as bivariate or trivariate systems since the improvements characterizing the LCT statistic do not make it more powerful (in absolute terms) for detecting the presence of cointegration in large systems.

As an alternative to testing we also examined the asymptotic and finite sample properties of a model selection based approach applied to the estimation of the cointegrating rank. Although commonly used information theoretic criteria such as the AIC or BIC were shown to perform poorly within our framework (with their performance deteriorating as the system dimension was allowed to increase), for moderately large sample sizes we found that the HQ criterion displayed excellent properties, particularly in large dimensional systems under which it consistently outperformed the test statistics. It is perhaps true that our results are based on a simple VAR model with no short run dynamics or any form of artificially induced misspecification but in this latter case it is well known for instance that the asymptotic critical values of the LR statistic become invalid and this would naturally have prevented us from isolating and evaluating the relative robustness of the various techniques to the dimensionality problem. More specifically any distortions characterizing the LR statistic under nonstandard conditions (such as MA or AR errors for instance) will also arise for the LCT or the model selection criteria an issue we leave for further research.

Appendix

Proof of Proposition 4.1. The proof follows by showing that under the chosen penalties the probabilities of "over" and "under" ranking vanish asymptotically.

- Case $\ell > r_0$: From (3) we have $P[IC(\ell) < IC(r_0)] = P[-T\sum_{i=r_0+1}^{\ell}\log(1 - \hat{\lambda}_i) > c_T(2p\ell - \ell^2 - 2pr_0 + r_0^2)]$. Since $-T\sum_{i=r_0+1}^{\ell}\log(1 - \hat{\lambda}_i)$ is $O_p(1)$ and the right hand side diverges towards infinity by condition (i) it follows that $\lim_{T\to\infty}P[IC(\ell) < IC(r_0)] = 0$ implying that overranking does not occur asymptotically.
- Case $\ell < r_0$: We have $P[IC(\ell) < IC(r_0)] = P[-\sum_{i=\ell+1}^{r_0}\log(1 - \hat{\lambda}_i) < (c_T/T)(2pr_0 - r_0^2 + \ell^2 - 2p\ell)]$. Since $\text{plim}(-\sum_{i=\ell+1}^{r_0}\log(1 - \hat{\lambda}_i)) > 0$, from condition (ii) the right hand side converges to zero thus leading to $\lim_{T\to\infty}P[IC(\ell) < IC(r_0)] = 0$ and implying that underranking does not occur asymptotically. Taken together these two results imply that $\hat{r} \xrightarrow{p} r_0$ as $T \to \infty$.

In order to show that the requirements (i) and (ii) are necessary let us suppose that c_T is bounded by some constant δ. Condition (ii) still holds and $\lim_{T\to\infty} P[IC(\ell) < IC(r_0)] = 0 \ \forall \ell < r_0$. For $\ell > r_0$ we have $P[IC(\ell) < IC(r_0)] = P[-T\sum_{i=r_0+1}^{\ell} \log(1 - \hat{\lambda}_i) > c_T(2p\ell - \ell^2 - 2pr_0 + r_0^2)]$ which will be nonzero since the right hand side does not tend to infinity when c_T is bounded. There is therefore a positive probability of overranking. In order to show that (ii) is necessary suppose that it fails, $(c_T/T) \to c > 0$. Clearly (i) is satisfied and for $\ell > r_0$ we have $\lim_{T\to\infty} P[IC(\ell) < IC(r_0)] = 0$. When $\ell < r_0$, $\lim_{T\to\infty} P[IC(\ell) < IC(r_0)] = P[-\sum_{i=\ell+1}^{r_0} \log(1 - \hat{\lambda}_i) < c(2pr_0 - r_0^2 + \ell^2 - 2p\ell)]$ and since $c > 0$ the result follows. ■

Note

1. Full density plots for all test criteria and $T \in [30, 5000]$ are available upon request from the authors.

References

Abadir, K. and R. Larsson (1996) "The Joint Moment Generating Function of Quadratic Forms in Multivariate Autoregressive Series," *Econometric Theory*, 12: 682–704.

Abadir, K., K. Hadri, and E. Tzavalis (1997) "The Influence of VAR Dimensions on Estimator Biases," University of Exeter, Department of Economics, Discussion Paper.

Ahn, S. K. and G. C. Reinsel (1990) "Estimation for Partially Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85: 813–23.

Akaike, H. (1969) "Fitting Autoregressive Models for Prediction," *Annals of the Institute of Statistical Mathematics*, 21: 243–7.

Akaike, H. (1976) "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion," in R. K. Mehra and D. G. Lainiotis (eds.), *System Identification: Advances and Case Studies*, New York: Academic Press, pp. 27–96.

Anderson, T. W. (1958) *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley.

Bartlett, M. S. (1947) "Multivariate Analysis," *Journal of the Royal Statistical Society (Supplement)*, 9: 176–97.

Brown, B. W. (1981) "Sample Size Requirements in Full Information Maximum Likelihood Estimation," *International Economic Review*, 22: 443–459.

Donald, S. G. and J. Cragg (1997) "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76: 223–50.

Engle, R. F. and C. W. J. Granger (1987) "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55: 251–76.

Engle, R. F. and C. W. J. Granger (1991) *Long Run Economic Relationships*. Oxford: Oxford University Press.

Fujikoshi, Y. (1977) "Asymptotic Expansions for the Distributions of Some Multivariate Tests," in P. R. Krishnaiah (ed.), *Multivariate* Analysis, vol. 4. Amsterdam: North-Holland Publishing Company, pp. 55–71.

Gonzalo, J. and J. Y. Pitarakis (1997) "Lag Length Estimation in Large Dimensional Systems," University of Reading, Department of Economics, Manuscript.

Granger, C. W. J. (1981) "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16: 121–30.

Hannan, E. J. and Quinn, B. G. (1979) "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society*, 41: 190–5.

Hotelling, H. (1931) "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, 2: 360–78.

Izenman, A. J. (1975) "Reduced Rank Regression for the Multivariate Linear Model," *Journal of Multivariate Analysis*, 5: 248–64.

Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12: 231–54.

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59: 1551–80.

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

MacKinnon, J. G. (1994) "Approximate Asymptotic Distribution Functions for Unit Root and Cointegration Tests," *Journal of Business and Economic Statistics*, 12: 167–76.

MacKinnon, J. G. (1996) "Numerical Distribution Functions for Unit Root and Cointegration Tests," Queen's University, Institute for Economic Research, Discussion Paper no. 918.

MacKinnon, J. G., Haug, A. and L. Michelis (1996) "Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration," Queen's University, Institute for Economic Research, Discussion Paper.

Nielsen, B. (1997) "Bartlett Correction for the Unit Root Test in Autoregressive Models," *Biometrika*, 84: 500–4.

Olson, C. L. (1974) "Comparative Robustness of Six Tests in Multivariate Analysis of Variance," *Journal of the American Statistical Association*, 69: 894–908.

Pillai, K. C. S. (1954) "Some New Test Criteria in Multivariate Analysis," *Biometrika*, 33: 117–21.

Reinsel, G. C. and 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: 353–75.

Schwartz, G. (1978) "Estimating the Dimension of a Model," *Annals of Statistics*, 6: 461–4.

Taniguchi, M. (1991) *Higher Order Asymptotic Theory for Time Series Analysis*, Lecture Notes in Statistics 68. Springer.