Finnish Economic Papers – Volume 8 – Number 2 – Spring 1995

SIMULATING SMALL-SAMPLE PROPERTIES OF THE MAXIMUM LIKELIHOOD COINTEGRATION METHOD: ESTIMATION AND TESTING*

TOR JACOBSON

Department of Statistics, Uppsala University, Box 513, S-751 20 Uppsala, Sweden

This paper analyzes – using Monte Carlo simulation – small-sample properties of the maximum likelihood cointegration method for estimation and inference in cointegrated systems. The simulations of a bivariate system concentrate on the following; the estimator of the cointegrating vector, the trace test for determining cointegrating rank, and the likelihood ratio and Wald tests for linear restrictions on the cointegrating vector. Furthermore, we introduce autoregressive conditional heteroscedasticity, as well as multivariate non-normality in the form of excess skewness and kurtosis, in the error process. All in all, the results suggest that the maximum likelihood method displays desirable features as long as the samples are of reasonable sizes. (JEL C15, C32)

1. Introduction

In the growing literature on cointegration analysis the method initiated by Søren Johansen – maximum likelihood cointegration (MLCI) – is recognized as one of the more important contributions. Its popularity in applied work has several reasons.[1] One being the straightforward treatment of multivariate aspects of the estimation problem, i.e. simultaneous estimation of two or more long run relationships. Another reason is the possibility of inference for the cointegrating vectors that generate these long run relationships.[2]

The aim of the present study is to evaluate small sample properties of various estimators and test statistics involved in the maximum likelihood approach to estimation of cointegrating relationships. With the use of Monte Carlo experimentation we shall investigate the following three statistics:

- the trace test statistics for cointegrating rank
- the estimator of the cointegrating vectors
- the likelihood ratio and Wald test statistics for testing linear restrictions imposed on the cointegrating vectors.

* *Comments on a previous draft from Anders Ågren, Reinhold Bergström, Henry Ohlsson, Anders Warne and an anonymous referee have substantially improved the paper. Remaining errors are of course the author's responsibility. Financial support from Tore Browalds Forskningsstiftelse is gratefully acknowledged.*

[1] *In a recent and very extensive survey Campbell and Perron (1991) have scrutinized in detail the area sometimes referred to as »unit root econometrics».*

[2] *For an excellent introduction to the maximum likelihood cointegration method see Johansen and Juselius (1990). This reference also contains an instructive application on Danish money demand. The theoretical results are found in Johansen (1988, 1989, 1991).*

Since the theoretical results for these statistics rely on asymptotic considerations, one important concern is to establish the behavior for small to moderate samples of sizes empirical research is likely to encounter (50–200 observations). Moreover, the MLCI-method rests on assumptions concerning the order of the vector autoregression as well as a Gaussian error process. Evaluation of robustness for the estimation and testing procedures against mis-specification in these two respects has been the objective of several recent studies, see i.a. Cheung and Lai (1993) and Gonzalo (1994). We propose to examine the robustness of estimation and testing against one form of deviation from the normality assumption. This is performed by introducing non-normality in the form of excess skewness and kurtosis in the error process. Another potential source of serious mis-specification that will be looked into is ARCH-contaminated errors.

The mis-specification analysis presented here differs somewhat from the standard one. Apart from studying the performance for an estimator or a test when the model is mis-specified, we evaluate the possibilities to detect a deficiency giving rise to a badly fitted model and hence to reject an invalid estimation or testing procedure. For example; does excess skewness in the errors lead to an increased bias in the estimate of the cointegrating vector in comparison with the Gaussian error case? Will this show up in a residual analysis with respect to skewness?

The next section contains an introduction to the statistical model underlying the MLCI-method. Section 3 is devoted to the design of the simulation experiments and results are presented in section 4. Some concluding remarks will end the paper.

2. *Estimation and testing procedures*

The brief account below considers estimation of the cointegrating vectors, the trace and maximum eigenvalue tests for cointegrating rank and the likelihood ratio and Wald tests of linear restrictions on the cointegrating vectors. Readers familiar with the method may safely skip this section.

Consider a VAR-representation of a p-dimensional integrated time series x_t according to

$$(1)\quad \Pi(L)\, x_t = \varepsilon_t, \qquad (t = 1, 2, \ldots, T),$$

where $\Pi(L)$ is a $p \times p$ matrix polynomial of order k given by $\Pi(\lambda) = I_p - \Sigma_{j=1}^{k} \Pi_j \lambda^j$, while L is the lag operator, λ a complex number, and the error term ε_t is assumed to be *iid* $N_p(0, \Sigma)$.

A slight reparameterization of (1) yields a vector error correction, VECM, representation for x_t suitable for estimation of the cointegrating relationships. Letting $\Gamma(\lambda) = I_p - \Sigma_{i=1}^{k-1} \Gamma_i \lambda^i$, where $\Gamma_i = -\Sigma_{j=i+1}^{k} \Pi_j$ and $\alpha\beta' = \Pi = -\Pi(1)$, we get

$$(2)\quad \Gamma(L)\, \Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t, \qquad (t = 1, 2, \ldots, T),$$

where Δ is the first difference operator. Writing $\alpha\beta' = \Pi$ reflects an assumption of reduced rank $r < p$ for Π, implying that α and β are $p \times r$ matrices. When $r > 0$, x_t is cointegrated of order (1,1). The cointegrating vectors are found in the r columns of β, whereas the rows of α have an interpretation as »adjustment coefficients» that determine how $\beta' x_{t-1}$ enters in the p equations.

Maximum likelihood estimation of (2) is non-linear in the parameters α and β, whereas estimation conditional on $\beta' x_{t-1}$ is linear and ordinary least squares is adequate. Estimation of β implies reduced rank regression, and in particular, solving an eigenvalue problem with solutions in the form of eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_p$ and eigenvectors $\hat{V} = [\hat{v}_1 \ldots \hat{v}_p]$.

Johansen (1988) derived two likelihood ratio tests for cointegrating rank r. The trace statistic tests a null of r cointegrating vectors against p and is calculated as

$$(3)\quad -2\ln(Q : r \mid p) = -T \sum_{i=r+1}^{p} \ln(1 - \hat{\lambda}_i).$$

The lambda-max statistic tests r against $r + 1$ vectors and reads

$$(4)\quad -2\ln(Q : r \mid r+1) = -T \ln(1 - \hat{\lambda}_{r+1}).$$

The ML-estimates of the cointegrating vectors are taken to be

$$(5) \quad \hat{\beta} = [\hat{v}_1 \dots \hat{v}_r].$$

That is, the eigenvectors corresponding to the r largest eigenvalues.

Reinsel and Ahn (1992) suggest a small-sample adjustment of the trace test. They find, by simulation, that the modified test

$$(6) \quad -2 \ln (Q : r \mid p) = -(T - kp) \sum_{i=r+1}^{p} \ln (1 - \hat{\lambda}_i)$$

has rejection rates that are closer to the nominal significance levels when compared with those of the unadjusted trace test. We will in the simulations evaluate the modified test along with the ordinary trace test.

For given rank r, the likelihood ratio principle leads to standard inference, i.e. test statistics for linear restrictions on α and β have asymptotic χ^2-distributions. In this paper we will consider linear restrictions on the cointegrating vectors β. The hypotheses under consideration can be expressed as: $\Pi = \alpha\varphi' H'$, that is $\beta = H\varphi$, where H $(p \times s)$, $r \leq s \leq p$, is a known matrix that specifies the restriction. The test statistic is given by

$$(7) \quad W_{LR} = -2 \ln (Q : H \neq I_n \mid H = I_n) = T \sum_{i=1}^{r} \ln \left[\frac{(1 - \hat{\lambda}_{Hi})}{(1 - \hat{\lambda}_i)} \right],$$

where $\hat{\lambda}_{Hi}$ and $\hat{\lambda}_i$ are the eigenvalues found as solutions to the eigenvalue problem implied by maximum likelihood estimation of the restricted and unrestricted models. Johansen (1991) derives this test and shows that W_{LR} is asymptotically χ^2 with r $(p - s)$ degrees of freedom.

In Johansen (1989) we find an account for an alternative procedure; the Wald test. The idea is to test the null that β satisfies the restriction $h(\beta) = 0$ by calculating the unrestricted $\hat{\beta}$ and build a decision on the closeness of $h(\hat{\beta})$ to the zero vector using some appropriate acceptance region. Let $\hat{v}$ be the eigenvectors corresponding to the eigenvalues $\hat{\lambda}_{r+1},\dots,\hat{\lambda}_p$ and $D = \text{diag}\,[\hat{\lambda}_1,\dots,\hat{\lambda}_r]$. Under the null of $K'\beta = 0$, where K is a specified matrix of order $(p \times (p - s))$ defining the linear restriction, the asymptotic distribution of

$$(8) \quad w_{\beta}^2 = T \,\text{tr}\, \{[K'\hat{\beta}\,(\hat{D}^{-1} - \text{I})^{-1}\,\hat{\beta}'K]\,[K'\hat{v}\hat{v}'K]\}$$

is again that of a χ^2 with r $(p - s)$ degrees of freedom.

3. *Design of the Monte Carlo study*

As noted in the introduction we will augment the simulation analysis of ARCH and non-normality effects by attempting the question: when poor performance for an estimator or a test is recorded, e.g. a large estimate of the bias $(E(\hat{\beta}) - \beta)$, would then analysis of the residuals indicate mis-specification? Johansen and Juselius (1990) propose univariate pre-analysis testing of the residuals.[3] Here we will consider three multivariate tests; the two multivariate tests for normality based on estimated skewness and kurtosis as suggested by Mardia (1970) and the bivariate ARCH-test proposed by Engle, Granger and Kraft (1984). First some general remarks.

For the sake of comparability we will let the data generating process (DGP) be the well known bivariate process used by Gonzalo (1994) and others. Although being simple in its construction and as Banerjee et al. (1986) puts it: »not very interesting economically», it is, in the slightly extended version presented below, well suited for our purposes. In order to evaluate – with respect to power – the likelihood ratio and the Wald tests in (7) and (8) for linear restrictions on the cointegrating vector, we will set $\beta = (1 \quad -\text{b})'$. The DGP for arbitrary b is

$$(9) \quad \begin{array}{ll} y_t - bx_t = z_t; & z_t = pz_{t-1} + \varepsilon_{zt} \\ ay_t - x_t = w_t; & w_t = w_{t-1} + \varepsilon_{wt}' \end{array}$$

$$\begin{bmatrix} \varepsilon_{zt} \\ \varepsilon_{wt} \end{bmatrix} \equiv iid\, N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta\sigma \\ \theta\sigma & \sigma^2 \end{pmatrix} \right].$$

[3] *Johansen and Juselius propose testing with respect to autocorrelations (Box and Pierce, 1970), normality (Jarque and Bera, 1980) and ARCH-structure (Engle and Bollerslev, 1986).*

Table 3.1: Critical values for the trace test; the case of no constant or linear trend.

T	*Dim*	0.50	0.80	0.90	0.95	0.975	0.99	mean	variance
50	1	0.5859	1.8225	2.8819	4.0419	5.1940	6.8745	1.1173	2.2727
	2	5.2603	8.2088	10.1591	12.0235	13.8444	16.3446	5.9083	11.4511
100	1	0.5990	1.8929	3.0010	4.1569	5.3822	7.1172	1.1606	2.8689
	2	5.4121	8.4474	10.4664	12.3788	14.2776	16.7346	6.0767	11.7691
200	1	0.6013	1.9134	3.0369	4.2798	5.5421	7.2130	1.1736	2.7296
	2	5.4744	8.5486	10.5728	12.4916	14.3560	16.7539	6.1403	11.8740

According to the first equation y_t and x_t are cointegrated with a serially correlated error term provided $\rho < 1$. The second equation dictates how the $I(1)$-ness property of the series is implemented by the random walk w_t. Gonzalos contribution to the DGP in (9) is the parameter a. For $a \neq 0$, Δx_t is not weakly exogenous for the estimation of (α, β) and single equation analysis of the first equation above would leads to inefficient estimation of the cointegrating relationship, see Johansen (1992). The corresponding error correction representation is given by

$$(10)\quad \begin{bmatrix}\Delta y_t\\ \Delta x_t\end{bmatrix} = \frac{1}{ba-\frac{1}{2}}\left[\begin{bmatrix}-\frac{1}{2}(\rho-1)\\ -a(\rho-1)\end{bmatrix}\right.$$

$$\left.[1 \quad -b]\begin{bmatrix}y_{t-1}\\ x_{t-1}\end{bmatrix} + \begin{bmatrix}v_{yt}\\ v_{xt}\end{bmatrix}\right],$$

where $v_{yt} = b\varepsilon_{wt} - \frac{1}{2}\varepsilon_{zt}$ and $v_{xt} = \varepsilon_{wt} - a\varepsilon_{zt}$.

The simulations are performed in double precision FORTRAN 77. The error term ε_{it} is a pseudo standard normal generated by the »normal-by-ratio-of-uniform»-routine (see Ripley, 1987), where the uniforms in turn are calculated using the DRAND-routine (see Schrage, 1979). The number of replications is 5,000 in each subexperiment and nominal test sizes of 5% are used throughout.

Three sample sizes are evaluated; $T = 50$, 100 and 200. We generate 20 pre-sample observations and the larger samples are true extensions of the smaller ones in that e.g. the first 50 observations are common to all three sizes. Moreover, the same random numbers are used across sub-experiments. Critical values for the trace test are presented in Table 3.1. Since the DGP contains neither a constant nor a deterministic trend these critical values have been simulated accordingly and correspond with the original ones given in Johansen (1988).[4]

The simulations are separated in five sections and can be described as follows:

(1) *The trace and the modified trace test:* The trace test statistic in (3) and the modified trace test statistic in (6) are calculated for the hypotheses $r \leq 1$ and $r = 0$. The former hypothesis being true in (9) make the estimated rejection rates interpretable as actual test sizes and for the latter hypothesis rejection rates estimate test powers. The modified test statistics are compared with the critical values for $T = 200$ irrespective of what sample size is considered. Preliminary simulations revealed that the trace test is invariant to the parameters α, β and θ in (9), therefore these have been set to 0, 1 and 0.3 respectively. Remaining choices of parameter values are: $\rho \in \{.9, .7\}$, $\sigma \in \{.25, .5, 1, 2, 4\}$ and sample sizes T $\in \{50, 100, 200\}$.

(2) *The estimate of the cointegrating vector* β*:* According to preliminary simulations the estimated bias is invariant to the choice of parameter b in (9), therefore b is set to 1. Like-

[4] *Table 3.1 has been calculated in the following way; 100,000 stochastic matrices M have been simulated by replacing the m-dimensional Brownian motion B below with an m-dimensional random walk in T steps where the increments are standard normal. Suitable order statistics of the trace of M give the critical values.*
$M = \int_0^1 (dB)\, B' \, [\int_0^1 B\, B'\, du]^{-1} \int_0^1 B\, (dB')$.

wise θ is not important, we let $\theta = 0$. The parameter a determines the vector of loadings α and setting $a \in \{-.2, -.1, 0, .1\}$ yields the following four $\alpha = (\alpha_1\ \alpha_2)'$ vectors: $(\frac{5}{7}[\rho - 1], \frac{2}{7}[\rho - 1])'$, $(\frac{5}{6}[\rho - 1], \frac{1}{6}[\rho - 1])'$, $([\rho - 1], 0)'$, and $(\frac{5}{4}[\rho - 1], \frac{1}{4}[\rho - 1])'$. Remaining parameters are set to: $\rho \in \{.9, .7\}$, $\sigma \in \{1, 2\}$, and $T \in \{50, 100, 200\}$. For each sample we normalize the estimated cointegrating vector so that $\hat{\beta} = (1\ \hat{\beta})'$ and then calculate the bias and the mean squared error. A priori, one would expect bias and dispersion to decrease as the signal to noise ratio σ, and the sample size T, increases.

(3) *The likelihood ratio and the Wald test statistic:* The same values on T and ρ, and σ as in (2) are considered in this round., Morover, $\sigma = 1$ and $\theta = 0.3$. To keep down the number of experiments we let Δx_t be weakly exogenous by setting $a = 0$ and let the cointegrating vector β vary according to $b \in \{.75, .90, .98, 1\}$. The LR test in (7) and the Wald test in (8) are calculated; first with $b = 1$ in order to establish the actual sizes of the tests as well as obtaining size adjusting critical values.[5] The latter are used in the second stage where size adjusted power under $H_A : b \neq 1$ is calculated.

(4) *ARCH:* This section seeks to demonstrate effects on $\hat{\beta}$ and the trace test from an error process governed by autoregressive conditional heteroscedasticity. That is, we will impose a first-order, bivariate ARCH-model – formulated by Engle et al. (1984) – on ε_t in (9). Thus, we let $\varepsilon_t \mid \Psi_{t-1}$ follow a $N(0, H(\varepsilon_{t-1}))$ where Ψ_{t-1} represents past information and $H(\varepsilon_{t-1}) = H_t$ is a (2×2) matrix with elements made up of quadratic forms in ε_{t-1}, i.e.

$$(11)\quad H_t = \begin{bmatrix} c_{01} \\ c_{02} \\ c_{03} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1}^2 \\ \varepsilon_{1t-1}\varepsilon_{2t-1} \\ \varepsilon_{2t-1}^2 \end{bmatrix}$$

[5] *Size adjusted critical values, presented throughout the paper, have been calculated by simply taking the appropriate order statistics for the empirical distributions of test statistics generated under the null hypotheses.*

The homoscedastic covariance matrix is nested in H_t and occurs when the second term in (11) is equal to zero. Diagonal ARCH is the label Engle et al. use for the simpler model where the off diagonal elements of $[c_{ij}]$ (i, j = 1, 2, 3) are zero. A necessary condition for H_t to be positive definite in the diagonal model – which is the one we propose to use – is the following: c_{01}, c_{11}, c_{03}, c_{33}, $c_{01}c_{03} - c_{02}^2$ and $c_{11}c_{33} - c_{22}^2$ all greater than zero. Moreover, Engle et al. (1984) suggest a test to $H_0 : c_{11} = c_{22} = c_{33} = 0$ against the alternative of conditional heteroscedasticity. The test statistic, which is asymptotically χ_4^2 under the null, is written

$$(12)\quad ARCH = 3T \sum_{i=1}^{2} \sum_{j=1}^{2} R_{ij}^2,$$

where R_{ij}^2 is the coefficient of determination when $\varepsilon_{it}\varepsilon_{jt}$ is regressed on a constant and $\varepsilon_{it-1}\ \varepsilon_{jt-1}$. In the simulations we calculate the following: bias and MSE for $\hat{\beta}$, size and size adjusted power for the trace test and the ARCH test. The parameters θ, σ, T and a in (9) take the same values as in (3) whereas $b = 1$ and $\rho = .7$. Apart from the null, where $c_{ij} = 0$ $(i, j = 1, 2, 3)$, we will consider combinations of c_{11}, $c_{33} \in \{0.1, 0.5, 0.9\}$ and $c_{22} \in \{0, 0.09, 0.4, 0.8\}$.

(5) *Non-normality:* The last section investigates effects of $\hat{\beta}$ and the trace test when the error process ε_t has third and fourth moments exceeding their expected values under normality. In other words, we will deal with excess skewness and kurtosis. To this end we use what is known as the Fleishman–Vale–Maurelli method for generating non-normality. The procedure was originally designed for the univariate case by Fleishman (1978) and was subsequently extended to the multivariate case by Vale and Maurelli (1983). Quiroga (1992) has worked out the details for the bivariate case as well as investigated the properties of the method using simulations. The results in Quiroga and elsewhere indicate less than desired precision when extreme non-normality is sought. Very briefly stated the method transform standard normal variates with some correlation into non-normal, zero mean, unit variance variates with desired correlation, marginal skewnesses γ_{1i} and marginal kurtoses γ_{2i}, i = 1, 2.

Table 4.1: Rejection rates for the trace and the modified trace test; $a = 0$, $b = 1$, $\theta = 0.3$, $\rho = 0.9$.

θ	T	*Size tr*	*s.e.*	*Power tr*	*s.e.*	*Size m.tr*	*s.e.*	*Power m.tr*	*s.e.*
.25	50	.0198	.00197	.1128	.00447	.0122	.00155	.0760	.00383
	100	.0356	.00262	.2228	.00588	.0306	.00244	.1988	.00564
	200	.0434	.00288	.6668	.00667	.0418	.00283	.6516	.00674
.5	50	.0204	.00200	.1162	.00453	.0124	.00156	.0806	.00385
	100	.0366	.00266	.2254	.00591	.0296	.00240	.2030	.00569
	200	.0438	.00289	.6768	.00661	.0426	.00286	.6674	.00666
1	50	.0210	.00203	.1244	.00467	.1420	.00167	.0872	.00399
	100	.0376	.00269	.2548	.00616	.0328	.00252	.2282	.00594
	200	.0460	.00296	.7324	.00626	.0450	.00293	.7218	.00634
2	50	.0252	.00222	.1600	.00518	.0164	.00180	.1186	.00457
	100	.0452	.00294	.3556	.00677	.0390	.00274	.3322	.00666
	200	.0492	.00306	.8802	.00459	.0484	.00304	.8724	.00472
4	50	.0420	.00284	.3462	.00673	.0280	.00233	.2774	.00633
	100	.0518	.00313	.7360	.00623	.0458	.00296	.7098	.00642
	200	.0490	.00305	.9966	.00082	.0480	.00302	.9960	.00089

As in the previous section the simulations concern bias and MSE for $\hat{\beta}$ and size/power for the trace test. Non-normality is evaluated using two multivariate measures defined by Mardia (1970) for a p-dimensional variable x_t with estimated mean $\bar{x}$ and covariance matrix S; they are given by

$$(13)\quad \hat{\gamma}_{1,p} = T^{-2} \sum_{t,s=1}^{T} g_{ts}^3 \text{ and } \hat{\gamma}_{2,p} = T^{-1} \sum_{t=1}^{T} g_{tt}^2,$$

where $g_{ts} = (x_t - \bar{x})\, S^{-1}\, (x_s - \bar{x})$.

Expected values for $\hat{\gamma}_{1,p}$ and $\hat{\gamma}_{2,p}$ are 0 and $p(p + 2)$, respectively, when x_t is multivariate normal. Moreover, Mardia (1970) suggests two tests for departures from multinormality based on these statistics:

$$(14)\quad Q_{\hat{\gamma}1,p} = \frac{T}{6}\hat{\gamma}_{1,p},$$

which is asymptotically χ^2 with $\frac{p}{6}(p+1)(p+2)$ d.f. and

$$(15)\quad Q_{\hat{\gamma}2,p} = [\hat{\gamma}_{2,p} - p\,(p+1)]^2 \Big/ \frac{8p}{T}(p+2)$$

which is asymptotically χ_1^2. Since the Fleishman–Vale–Maurelli method only permits control for univariate skewness and kurtosis and we want to evaluate the multivariate $Q_{\hat{\gamma}1,p}$ and $Q_{\hat{\gamma}2,p}$, the choice of parameters for the simulations become somewhat arbitrary. However, we will choose marginal skewnesses γ_{1y}, γ_{1x} and marginal kurtoses γ_{2y}, γ_{2x} such that multivariate skewness and kurtosis vary in a fruitful way, i.e. from small to large effects. Calculating mean values of $\hat{\gamma}_{1,p}$ and $\hat{\gamma}_{2,p}$ should confirm that this is achieved. Our choice is

D =	γ_{1y}	γ_{2y}	γ_{1x}	γ_{2x}	D =	γ_{1y}	γ_{2y}	γ_{1x}	γ_{2x}
0:	.00	.00	.00	.00	3:	1.50	2.50	1.50	2.50
1:	.00	3.75	.00	3.75	4:	1.75	3.75	.00	.00
2:	1.50	2.50	.00	.00	5:	1.75	3.75	1.75	3.75

$D = 0$ is of course the multinormal case, $D = 1$ gives large excess kurtoses but zero skewnesses, $D = 3$ and $D = 5$ give relatively smaller and larger excess kurtoses and skewnesses. Remaining parameters in (9) are; $a = 0$, $b = 1$, $\rho = 0.8$ and $\theta \in (0, .6)$.

4. Simulation results

The outcome of the simulation experiments are to be found in Tables 4.1 to 4.9. The heading *Bias* always refers to the parameter b in (9). Standard errors refer to the estimates in the neighboring left column. In the case of estimated quantiles $\hat{p}$, i.e. size and power for various tests, the Monte Carlo standard errors are calculated as $\left(\frac{\hat{p}(1-\hat{p})}{5{,}000}\right)^{\frac{1}{2}}$. When $\hat{p} = 1$ or 0 this statistic breaks down and takes the value 0.

Table 4.2: Rejection rates for the trace and the modified trace test; $a = 0$, $b = 1$, $\theta = 0.3$, $\rho = 0.7$.

θ	T	*Size tr*	*s.e.*	*Power tr*	*s.e.*	*Size m.tr*	*s.e.*	*Power m.tr*	*s.e.*
.25	50	.0458	.00296	.5084	.00707	.0342	.00257	.4088	.00695
	100	.0522	.00315	.9684	.00247	.0462	.00297	.9586	.00282
	200	.0466	.00298	1.000	.00000	.0458	.00296	1.000	.00000
.5	50	.0476	.00301	.5212	.00706	.0350	.00260	.4266	.00700
	100	.0530	.00317	.9720	.00233	.0464	.00298	.9634	.00266
	200	.0468	.00299	1.000	.00000	.0456	.00295	1.000	.00000
1	50	.0496	.00307	.5664	.00701	.0378	.0027	.4760	.00706
	100	.0528	.00316	.9810	.00193	.0470	.00299	.9750	.00221
	200	.0478	.00302	1.000	.00000	.0460	.00296	1.000	.00000
2	50	.0548	.00322	.7210	.00634	.0408	.00280	.6454	.00677
	100	.0532	.00317	.9970	.00077	.0464	.00298	.9962	.00087
	200	.0488	.00305	1.000	.00000	.0470	.00299	1.000	.00000
4	50	.0576	.00330	.9658	.00257	.0456	.00295	.9414	.00332
	100	.0530	.00317	1.000	.00000	.0470	.00299	1.000	.00000
	200	.0458	.00296	1.000	.00000	.0446	.00292	1.000	.00000

Three parameters, T, ρ and σ^2 in (9), dominate in importance in the experiments. For the sample size T we find in general that 50 is inadequate and that 200 is sufficient for tests and estimators to have desirable properties. However, the parameter ρ determining the degree of serial correlation in the error process z_t of the cointegrating relationship in the first equation is likewise important. As ρ approaches unity, the cointegrating relation moves from $I(0)$-ness to $I(1)$-ness – from stationarity to non-stationarity –, since z_t approaches a random walk. For $\rho = 1$, (9) is an $I(1)$-system without cointegration. Thirdly, a small error variance σ^2 in the random walk w_t, will give w_t a constant-like behavior and distort tests and estimates. It is therefore a question of having a reasonable mixture of T, ρ and σ in order to get informative results. Setting $\rho = 0.1$ and $\sigma = 4$ would give excellent size and power properties for e.g. the trace test no matter what size of sample is at hand. We try to present a more balanced picture. The results in detail will be presented in same order as the designs were presented above.

(1) *The trace and the modified trace statistic,* Tables 4.1 and 4.2: In spite of having used critical values simulated using small samples of $T = 50$, 100 and 200 and thus corresponding to the sample sizes used here, results in Table 4.1 show that the trace test is undersized for $T = 50$ and to some extent also for $T = 100$ when $\rho = 0.9$. The rejection rates corresponding to power illustrate the fact that the system is close to being $I(1)$ with no cointegration between y_t and x_t. Looking at the relatively stable case in Table 4.2 where $\rho = 0.7$, we find the estimated actual sizes for the trace test all gathered around 5%. Should we construct confidence intervals using ±2 s.e., almost all such intervals would cover the correct size. The small-sample modification, as we have evaluated it, result in even smaller sizes and smaller power.

(2) *The estimate of the cointegrating vector β,* Tables 4.3 and 4.4: A striking feature of these results is the great sampling variability as reflected by large standard errors in comparison with mean biases and the dramatic impact on bias and MSE from smaller sample sizes. As noted in the previous section the parameter a determine the elements of the loadings vector α. Due to the sampling variation, little can be inferred on the impact of a on bias and MSE. One would for instance expect $a = 0.1$ to result in better estimates compared with $a = -0.1$, since the former involves a larger α in absolute value. The contradicting results in Tables 4.4 and 4.5 point at the difficulties in judging simulation results of this kind.

(3) *The likelihood ratio and the Wald test sta-*

Table 4.3: Bias and mean squared error for $\hat{\beta}$; $b = 1$, $\theta = 0$, $\rho = 0.9$.

			$\sigma = 1$			$\sigma = 2$	
a	*T*	*Bias*	*s.e.*	*MSE*	*Bias*	*s.e.*	*MSE*
0	50	.019607	.086351	37.275508	.009804	.043176	9.318891
	100	–.014699	.022110	2.443924	–.007350	.011055	.610981
	200	.001259	.001419	.010068	.000629	.000710	.002517
.1	50	.056195	.039152	7.666091	.024665	.022529	2.537921
	100	.003082	.007662	.293472	.005238	.003726	.069419
	200	.004166	.002775	.038508	.000980	.000692	.002392
–.1	50	–.225554	.183945	69.196449	–.025764	.031053	4.821182
	100	.058493	.038102	7.260829	–.011661	.008406	.353348
	200	–.000718	.001667	.013897	.000193	.000806	.003246
–.2	50	–.078437	.078626	30.909984	–.131573	.107302	57.573848
	100	–.052776	.063449	20.127380	.034121	.022226	2.470698
	200	–.005610	.004226	.089307	–.000419	.000973	.004729

Table 4.4: Bias and mean squared error for $\hat{\beta}$; $b = 1$, $\theta = 0$, $\rho = 0.7$.

			$\sigma = 1$			$\sigma = 2$	
a	*T*	*Bias*	*s.e.*	*MSE*	*Bias*	*s.e.*	*MSE*
0	50	.025291	.040981	8.396232	.012646	.020491	2.099059
	100	–.000154	.000537	.001441	–.000077	.000268	.000360
	200	.000140	.000261	.000341	.000070	.000131	.000085
.1	50	–.002333	.002367	.028021	–.006376	.004194	.087986
	100	.000107	.000429	.000918	.000004	.000215	.000230
	200	.000167	.000209	.000218	.000070	.000105	.000055
–.1	50	.001109	.002960	.043794	.003167	.002607	.025558
	100	–.000532	.000647	.002090	–.000179	.000323	.000520
	200	.000086	.000314	.000492	.000064	.000157	.000123
–.2	50	–.018308	.015559	1.210491	.000648	.001727	.014902
	100	–.001029	.000759	.002878	–.000310	.000377	.000711
	200	.000005	.000366	.000670	.000051	.000183	.000167

tistic, Tables 4.5 and 4.6: The first table shows the estimated actual sizes under the null for the two tests of the linear restriction $b = 1$. Both tests are severely oversized and the Wald test is worse off. The variance parameter σ^2 seem to matter less ($\sigma = 1$ is therefore reported), whereas T and ρ are more important. Table 4.6 reveals that the two tests have equivalent size adjusted power properties. In the »non-stationary» case, $\rho = 0.9$, there is practically no power at all for the case $b = 0.98$. For the »stable» case, $\rho = .7$, the power is quite impressive for detecting $b = 0.9$, even when a small size such as $T = 100$ is in use. Considering the poor performance with respect to size for the Wald test, it must be concluded that the likelihood ratio

Table 4.5: Rejection rates under H_0 for the LR and the Wald test; $a = 0$, $b = 1$, $\theta = 0.3$, $\sigma = 1$.

ρ	*T*	*Size Lr*	*s.e.*	*Size Wald*	*s.e.*
.9	50	.1650	.00525	.3928	.00691
	100	.1132	.00448	.2398	.00604
	200	.0804	.00385	.1428	.00495
.7	50	.0908	.00406	.1718	.00534
	100	.0718	.00365	.1084	.00440
	200	.0592	.00334	.0760	.00375

Table 4.6: Rejection rates under H_A for the LR and the Wald test using size adjusted critical values; $a = 0$, $\theta = 0.3$, $\sigma = 1$.

ρ	T	b	Power Lr	s.e.	Power Wald	s.e.
.9	50	.98	.0504	.00309	.0520	.00314
		.90	.0822	.00388	.0750	.00373
		.75	.1406	.00492	.1582	.00516
	100	.98	.0570	.00328	.0570	.00328
		.90	.1574	.00515	.1664	.00527
		.75	.3570	.00678	.4276	.00700
	200	.98	.0760	.00375	.0724	.00367
		.90	.4540	.00704	.4796	.00707
		.75	.7930	.00572	.8518	.00503
.7	50	.98	.0730	.00368	.0752	.00373
		.90	.3648	.00681	.4000	.00693
		.75	.6920	.00653	.7804	.00586
	100	.98	.1290	.00474	.1288	.00474
		.90	.7314	.00627	.7530	.00610
		.75	.9600	.00277	.9724	.00232
	200	.98	.3294	.00665	.3296	.00665
		.90	.9574	.00286	.9632	.00266
		.75	.9994	.00035	.9996	.00028

test is to be favoured. Moreover, a minimum of 100 observations seems warranted.

(4) *ARCH*, Table 4.7: The bivariate ARCH test is severely oversized, the extraordinary feature is however that the rejection rate under the null seems quite invariant to the parameters in (9) such as T, σ, ρ and θ, all estimated actual sizes being around 35% – 40% (not reported). The size adjusting critical values calculated in order to evaluate the test under the alternative are all somewhat larger than 30, or approximately $3\chi^2_4$. Some simulations were run calculating $\frac{1}{3}$ *ARCH* rather than *ARCH* in (12) and with resulting estimated sizes of 5% to 10% (not reported).

However, in Table 4.7 we present results for *ARCH* in (12), despite the above noted inability for the test to keep its nominal size under the

Table 4.7: Bias and mean squared error, rejection rates for the trace test and the ARCH test under H_0: no ARCH effect, i.e. $C = 0$ and H_A: ARCH effect with $C \neq 0$; $a = 0$, $b = 1$, $\theta = 0.3$, $\sigma = 1$, $\rho = 0.7$.

T	C	Bias	s.e.	MSE	Size tr	s.e.	Power tr	s.e.	ARCH	s.e.
50	0	.025291	.040981	8.396232	.0464	.00298	.5046	.00707	.0500	–
	1	.005355	.004724	.111598	.0450	.00293	.5106	.00707	.1108	.00444
	2	.002963	.003486	.060770	.0476	.00301	.5454	.00704	.4912	.00707
	3	.006215	.008103	.328250	-0580	.00331	.5996	.00693	.6940	.00652
100	0	–.000154	.000537	.001441	.0518	.00313	.9674	.00251	.0500	–
	1	–.000039	.000541	.001462	.0522	.00315	.9626	.00268	.1862	.00551
	2	.000789	.000890	.003959	.0520	.00314	.9474	.00316	.8012	.00564
	3	.000233	.000870	.003784	.0628	.00343	.9398	.00336	.9386	.00339
200	0	.000070	.000131	.000085	.0460	.00296	1.000	.00000	.0500	–
	1	.000069	.000130	.000085	.0452	.00294	1.000	.00000	.2778	.00633
	2	–.000040	.000271	.000367	.0510	.00311	1.000	.00000	.9726	.00231
	3	–.000555	.000370	.000686	.0616	.00340	.9972	.00075	.9958	.00091

Table 4.8: Bias and mean squared error, rejection rates for the trace test and the multivariate skewness and kurtosis test under H_0: normality, i.e. $D = 0$ and H_A: non-normality with $D \neq 0$; $a = 0$, $b = 1$, $\theta = 0.6$, $\sigma = 1$, $\rho = 0.8$.

T	*C*	*Bias*	*s.e.*	*MSE*	*Size trace*	*s.e.*	*Power trace*	*s.e.*	*Mean skew*	*Skew test*	*s.e.*	*Mean kurt*	*Kurt test*	*s.e.*
50	0	–.0026	.0091	.4169	.0448	.0029	.4118	.0070	.4335	.0426	.0029	7.7017	.0190	.0019
	1	.0065	.0098	.4831	.0460	.0030	.4036	.0069	1.8335	.4808	.0071	11.5990	.5708	.0070
	2	.0861	.0630	19.8413	.0396	.0028	.3394	.0067	2.5576	.8604	.0049	9.8700	.3352	.0067
	3	–.1128	.0572	16.3809	.1582	.0052	.4824	.0071	3.7181	.9784	.0021	11.3264	.5898	.0070
	4	–.1064	.0613	18.7762	.3466	.0067	.7590	.0061	2.1081	.7592	.0061	9.3497	.2540	.0062
	5	–.1323	.0348	6.0581	.2428	.0061	.5632	.0070	4.7762	.9932	.0012	12.6865	.7466	.0062
100	0	–.0007	.0012	.0071	.0518	.0031	.8920	.0044	.2243	.0464	.0030	7.8491	.0312	.0024
	1	.0002	.0014	.0096	.0496	.0031	.8898	.0044	1.4540	.6212	.0069	13.2095	.8798	.0046
	2	–.0060	.0019	.0182	.0516	.0031	.8122	.0055	2.6320	.9986	.0005	10.5679	.6064	.0069
	3	–.0680	.0083	.3521	.3056	.0065	.8984	.0043	3.9237	1.000	.0000	12.2985	.8778	.0046
	4	–.0433	.0153	1.1732	.6510	.0067	.9882	.0015	2.1564	.9884	.0015	9.9452	.4734	.0071
	5	–.0813	.0287	4.1365	.4884	.0071	.9284	.0036	5.1629	1.000	.0000	14.0912	.9638	.0026
200	0	.0002	.0005	.0012	.0488	.0031	.9998	.0002	.1169	.0486	.0030	7.9195	.0404	.0028
	1	.0003	.0006	.0021	.0494	.0031	1.000	.0000	1.0550	.7326	.0063	14.4815	.9926	.0012
	2	–.0019	.0010	.0051	.0488	.0031	.9994	.0004	2.6967	1.000	.0000	11.0923	.8688	.0048
	3	–.0348	.0010	.0061	.5568	.0070	.9996	.0003	4.0367	1.000	.0000	12.9336	.9940	.0011
	4	–.0307	.0023	.0265	.9340	.0035	1.000	.0000	2.1628	1.000	.0000	10.3095	.7418	.0062
	5	–.0406	.0078	.3089	.7838	.0058	.9998	.0002	5.3740	1.000	.0000	15.0448	.9996	.0003

null. The parameter C = 0, 1, 2, and 3 correspond to the cases of no / small / medium / large ARCH effect. Under the alternative – unlike the null – we find the test to be sensitive to different T in much the same manner as the LR and the Wald tests in the previous experiments. Remaining parameters; σ, ρ and θ, seem to matter less. What is more important, the test is sensitive to the degree of ARCH, even to the point that when $\hat{\beta}$ or the trace test are affected we find satisfactory power for the ARCH test. When $\rho = 0.7$, there are no ARCH effects on $\hat{\beta}$. For $\rho = 0.9$ (not reported) there are some large biases when the ARCH effect is medium or large. As for the trace test one may conjecture a dislocation to the right for the distribution, at least for the two smaller sample sizes.

(5) *Non-normality,* Table 4.8 and 4.9: For this round of experiments we have neglected the procedure of size adjusting the critical values used for the power evaluation. The reason can be seen in rows corresponding to $D = 0$, i.e. multinormality prevailing, where the kurtosis and in particular the skewness test display acceptable sizes. The columns labelled *Mean Skew* and *Mean Kurt* indicate that the chosen parameters D give rise to the desired varying degrees of non-normality. Possibly with an exception for $D = 2$ and $D = 4$; non-normality for the error process in the first equation only.

The same general conclusion as in the case of imposed ARCH structure can be inferred here. When non-normality constitutes a problem for the estimation of β we have power for the multivariate skewness and kurtosis tests such that the statistical model may be correctly rejected. It is interesting to note the large difference in bias between the cases of no correlation $\theta = 0$ and large correlation $\theta = 0.6$ (the intermediate correlation, $\theta = 0.3$, not reported give rise to intermediate results). When $\theta = 0.6$, we see that the biases occurring for substantial non-normality, D = 3, 4 and 5, are large and account for the large MSE's since the standard errors are small in comparison.

Finally, it is remarkable how differently the trace test behave for D = 3, 4 and 5 in the two cases of correlation / no correlation. The test is severely oversized when $\theta = 0.6$ and this feature is emphasized as T increases. With a non-correlated error process there is no impact on

Table 4.9: Bias and mean squared error, rejection rates for the trace test and the multivariate skewness and kurtosis test under H_0: normality, i.e. $D = 0$ and H_A: non-normality with $D \neq 0$; $a = 0$, $b = 1$, $\theta = 0$, $\sigma = 1$, $\rho = 0.8$.

T	*C*	*Bias*	*s.e.*	*MSE*	*Size trace*	*s.e.*	*Power trace*	*s.e.*	*Mean skew*	*Skew test*	*s.e.*	*Mean kurt*	*Kurt test*	*s.e.*
50	0	.0173	.0128	.8166	.0370	.0027	.2518	.0061	.4319	.0428	.0029	7.6939	.0176	.0019
	1	−.0243	.0256	3.2732	.0356	.0026	.2500	.0061	1.6273	.4710	.0071	11.3239	.5598	.0070
	2	−.1871	.3690	680.6550	.0356	.0026	.2530	.0062	2.1571	.7926	.0057	9.3001	.2510	.0061
	3	−.0597	.0520	13.5427	.0402	.0028	.2584	.0062	3.8505	.9836	.0018	10.8910	.5230	.0071
	4	−.0742	.0598	17.8866	.0424	.0029	.2580	.0062	2.6866	.8794	.0046	9.9359	.3512	.0068
	5	.0019	.0141	.9928	.0422	.0028	.2632	.0062	4.9740	.9956	.0009	12.2204	.6934	.0065
100	0	−.0001	.0010	.0045	.0540	.0031	.6832	.0066	.2244	.0464	.0030	7.8475	.0292	.0024
	1	.0013	.0014	.0111	.0456	.0030	.6834	.0066	1.2311	.5940	.0070	12.7890	.8788	.0046
	2	−.0001	.0011	.0059	.0502	.0031	.6902	.0065	2.2099	.9962	.0009	9.8490	.4788	.0071
	3	.0017	.0018	.0168	.0532	.0032	.6998	.0065	4.1885	1.000	.0000	11.8487	.8328	.0053
	4	.0025	.0066	.2143	.0520	.0031	.6884	.0066	2.8929	.9990	.0004	10.7347	.6404	.0068
	5	−.0129	.0136	.9211	.0516	.0031	.6960	.0065	5.5599	1.000	.0000	13.6359	.9456	.0032
200	0	.0002	.0004	.0008	.0462	.0030	.9984	.0006	.1169	.0506	.0031	7.9195	.0392	.0027
	1	.0002	.0004	.0009	.0470	.0030	.9982	.0006	.8505	.7016	.0064	13.8469	.9932	.0012
	2	.0003	.0004	.0008	.0456	.0030	.9978	.0007	2.2322	1.000	.0000	10.1575	.7602	.0060
	3	−.0059	.0004	.0010	.0478	.0030	.9984	.0006	4.3444	1.000	.0000	12.3897	.9842	.0018
	4	−.0007	.0004	.0009	.0522	.0032	.9986	.0005	2.9779	1.000	.0000	11.1997	.8974	.0043
	5	−.0004	.0004	.0009	.0514	.0031	.9976	.0007	5.8446	1.000	.0000	14.4872	.9994	.0004

the trace test from excess skewness or excess kurtosis.

5. Conclusions

In this paper we have analyzed certain aspects of the maximum likelihood cointegration method for estimation and inference in cointegrated systems. Specifically, using Monte Carlo simulation, the objective has been to investigate the following: the estimator of the cointegrating vector β, the trace test for determining cointegrating rank, the likelihood ratio and Wald tests for linear restrictions on β. Furthermore, we have introduced conditional variances and non-normality in the form of excess skewness and kurtosis in the error process with the following questions in mind: is there an impact on $\hat{\beta}$ and the trace test from these caveats and how do the tests for ARCH and non-normality function in this context?

The main results are:

1. For the bivariate DGP in use we find the trace test undersized for smaller sample sizes when the system is close to being non-stationary without cointegration.
2. Bias and MSE for $\hat{\beta}$ is substantial for smaller sample sizes. However, it is difficult to evaluate the results in this respect since the sampling variability is large as reflected by the estimated standard error.
3. Both the LR and the Wald test of linear restrictions on β are found to be oversized under the null, the latter to a greater extent. Since they display equivalent power – and acceptable for larger T – we favour the LR test.
4. Presence of ARCH affects $\hat{\beta}$ when the system is close to non-stationarity without cointegration, moreover the distribution of the trace statistic is skewed to the right. The ARCH test is oversized under the null but has good power. We conclude that, as far this study goes, it is possible to successfully test for ARCH when it constitutes a potential problem.
5. The previous conclusion is also applicable in the case of non-normality. But with a difference in that the multivariate tests for excess skewness and kurtosis keep their nominal sizes much better.

References

Banerjee, A., J.J. Dolado, D.F. Hendry, and G.W. Smith (1986). »Exploring equilibrium relationships in econometrics through static models: some Monte Carlo evidence.» *Oxford Bulletin of Economics and Statistics,* 48, 253–277.

Box, G.E.P., and D.A. Pierce (1970). »A distribution of residual autocorrelations in autoregressive integrated moving average time series models.» *Journal of the American Statistical Association,* 65, 1509–1526.

Campbell, J.Y., and P. Perron (1991). »Pitfalls and opportunities: what macroeconomists should know about unit roots.» In *NBER Macroeconomics Annual 1991.* Eds. O.J. Blanchard, and S. Fisher. MIT Press.

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

Engle, R.F., and T. Bollerslev (1986). »Modelling the persistence of conditional variances.» *Econometric Reviews,* 5, 1–50.

Engle, R.F., C.W.J. Granger, and D. Kraft (1984). »Combining competing forecasts of inflation using a bivariate ARCH model.» *Journal of Economic Dynamics and Control,* 8, 151–165.

Fleishman, A.I. (1978). »A method for simulating non-normal distributions.» *Psychometrica,* 43, 521–532.

Gonzalo, J. (1994). »Five alternative methods of estimating long-run equilibrium relationships.» *Journal of Econometrics,* 60, 203–230.

Jarque, C.M., and A.K. Bera (1980). »Efficient tests for normality, homoscedasticity and serial independence of regression residuals.» *Economics Letters,* 6, 255–259.

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

— **(1989).** Likelihood based inference on cointegration: theory and applications. Lecture Notes, Institute of Mathematical Statistics, University of Copenhagen.

— **(1991).** »Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models.» *Econometrica,* 59, 1551–1580.

— **(1992).** »Cointegration in partial systems and the efficiency of single-equation analysis.» *Journal of Econometrics,* 52, 389–402.

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

Mardia, K.V. (1970). »Measures of multivariate skewness and kurtosis with applications.» *Biometrika,* 57, 519–530.

Quiroga, A.M. (1992). Studies of the polychoric correlation and other correlation measures for ordinal variables. PhD-thesis, Department of Statistics, Uppsala University.

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

Ripley, B.D. (1987). *Stochastic Simulation.* Wiley.

Shrage, L. (1979). »A more portable Fortran Random Number Generator.» *ACM Transactions on Mathematical Software,* 5, 132–138.

Vale, C.D., and V.A. Maurelli (1983). »Simulating multivariate non-normal distributions.» *Psychometrica,* 48, 465–471.