Finnish Economic Papers – Volume 3 – Number 1 – Spring 1990

SOME SMALL SAMPLE PROPERTIES OF COINTEGRATED LABOUR DEMAND MODELS

MIKAEL LINDÉN*

University of Helsinki, SF-00100 Helsinki

The small sample properties of the cointegrated labour demand models are studied with some Monte Carlo experiments. The results indicate that the Engle-Granger two-step estimation procedure gives biased OLS-estimates in small samples for the cointegrated vector and the error correction parameters. Much better OLS-estimates are found with long run restrictions proposed by economic theory.

1. *Introduction*

The theory of cointegrated economic variables has aroused a considerable amount of interest during the past five years in the applied economic time series work. The theoretical results concerning cointegration and its implications for different model specifications are well known nowadays (see, for example, Engle and Granger 1987, Dolado and Jenkinson 1987, Granger 1986, Hylleberg and Mizon 1989).

However, the results concerning the estimation theory and the statistical inference with linear cointegrated models are mixed and not straight-forward. This state of affairs stems from the nonstandard distribution theory of integrated (or nonstationary) series. Although the asymptotic properties of OLS- and ML-estimators and test statistics for cointegrated models are derived, the small sample properties of these models are still unknown. There exist some small sample Monte Carlo studies with two variable models (see, for example, Banerjee et al. 1986, Molina 1986, Stock 1987, Engle and Granger 1987) showing that parameter estimates are very biased and that the cointegration tests used have a low power.

A common problem with all the cointegration statistics in applied work is due to the uncertainty of the statistical properties of the series used in small samples. The following types of questions, then, are important: Are the series trend stationary or not ? Are the series random walk-type around the trend with drift or not with independent white noise errors ? And so on. As long as the distribution theory and the test statistics of the integrated series are non-standard and lack power, the statistical inference may be a hazardous business in small samples (see Phillips and Ouliaris 1987, Schwert 1989, Molinas 1986). Thus, the above type of questions are then difficult to answer.

Another issue that raised a lot of hopes in the cointegration framework was the Engle-Granger two-step estimation method for the error correction models (see Engle and Granger 1987, Stock 1987). This method is a result of the »super consistency» properties of OLS-

* *Thanks to Pekka Ilmakunnas, Ph.D., and Prof. Markku Rahiala for their valuable comments. Special thanks to Prof. David Hendry for permission to use his PC-NAIVE program.*

estimators for the cointegration models. It is, however, totally sensitive to the inference problems mentioned above in small samples. In the multivariate context, the low power of tests and unclear properties of the estimators easily lead to the possibility of arguing erroneously that the error correction variable (i.e., residual from the »cointegrated» or »static long run model») is statistically well founded and the the error correction estimates are true ones.

In this paper the small sample properties of different specifications of one cointegrated labour demand model is studied with some Monte Carlo methods. Two different model specifications for cointegrated and error correction models are used to clarify the statistical problems with the models. The Engle-Granger two-step estimation method is compared to a model setting where a priori long run restriction is given by economic theory.

2. *Cointegrated labour demand model*

The unrestricted labour demand model used is the following

$$(1) \quad \ln H_t = a_o + a_1 \ln VAL_t + a_2 \ln WS_t + \varepsilon_{1t},$$

where $\ln H_t$ is the natural logarithm of workers' hours, $\ln VAL_t$ is the logarithm of value added in 1980 prices and $\ln ws_t$ is the logarithm of hourly real production wage rate (including social security payments) in 1980 prices. The sample is taken from the annual observations of the Finnish manufacturing industry (ISIC 3) in the years 1960–1986.

The OLS-parameter estimates of equation (1) are proposed to be a long run static (equilibrium) relationship. In other words, the variables are cointegrated. Our alternative specification claims that there exist in the long run unit elasticity restrictions between the variables in the following way:

$$(1') \quad \ln H_t = b_o + \ln VAL_t - \ln WS_t + \varepsilon_{2t}$$

$$\Leftrightarrow \ln(WS/q)_t = b_o + \varepsilon_{2t}, \text{ where}$$

$$\ln q_t = \ln VAL_t - \ln H_t \text{ (hourly real productivity).}$$

This restricted relationship implies constant returns to scale production technology, workers' wage share is constant, and there is unit elasticity between productivity and wage rate. Thus, the model (1') is our null hypothesis and it is tested against more general specification 1).[1]

According to the basic theory of cointegration, there has to be a unique linear relationship between the variables $\ln H_t$, $\ln VAL_t$ and $\ln WS_t$ that is stationary. In other words, the residuals, ε_1 or ε_2, in equations 1) and 1') are I(0)-series. Then, if the dependent variable $\ln H_t$ is I(1), the regressors in the OLS-framework must include some I(1) variables or a combination of variables of a higher order of integration which is cointegrated as I(1). If in such a case some of the regressors are I(0), these variables will affect only the short-run behaviour of the dependent variables (see Hylleberg and Mizon 1989, p. 116). The following tables give OLS-results for equation 1) and the time series properties of the used series and proposed restrictions.

The shapes of the autocorrelation functions of the different series show clearly that series $\ln VAL_t$, $\ln WS_t$ and $\ln q_t$ are nonstationary. However, this result is not clear for the dependent variable $\ln H_t$. Irrespective of the above noted skepticism about the properties of the familiar unit root test (DF-, ADF- and CRDW-test and their modifications) the basic series were tested with the following DF-test model

---

[1] *There are many reasons why the model specifications used were preferred over other alternatives.*

i) The model 1') is easily derived from profit maximization with the log-linear production function, that is, $V = aH^\alpha$, $\max\pi(H) = paH^\alpha - wH$. The first order condition is $\alpha aH^{\alpha-1} = \alpha V/H = ws/p$. When this is solved for $\ln H$ we get $\ln H = \ln\alpha + \ln V - \ln WS$, that is $\ln(WS/q) = \ln\alpha$.

ii) Thus, this simple theoretical model clearly gives the a priori restrictions for lnQ and lnWS. The motivation for these restrictions also stems from empirical studies made in Finland with different labour demand models. In her review of these studies, Santamäki (1986) summarized that the production elasticity is between [.75, 1.00] and the wage elasticity is between [−1.00, −0.75] in years 1960−1980.

iii) The properties of the OLS-estimates in small samples for the cointegrated variables are understood to some extent only in the bivariate case. The properties of multivariate models are still very unknown.

$$\Delta \ln x_t = d_o + d_1 t + d_3 \ln x_{t-1} + \mu_t, \quad \mu_t \sim N(0,\sigma^2)$$

$$H_o : d_3 = 0, \; H_1 : d_3 \neq 0.$$

The test t-values for variables were the following: $\ln_t$: −.371, $\ln VAL_t$: −1.58, $\ln WS_t$: −1.90 and $\ln q_t$: −2.33 (critical value with $\alpha = .05$ and $T = 25$ is −3.60, see Fuller 1976, p. 373). These results and some modelling approaches confirmed that the $\ln H_t$ series is nonstationary[3]

These results are supported by the reported Durbin-Watson test values (see above table). When the series are nonstationary, the DW-test value converges to 0 when $t \rightarrow \infty$ (see Phillips 1987, Dolado & Jenkinson 1987). Thus, a small DW-value (<0.25) gives a good description of the properties of the series. The reported normality-test is the Jarque-Bera test. The test was done because the normality of the series is needed for the Monte Carlo experiments in section 3.

The results concerning the cointegration series (the properties of series e_{1t} and $\ln(WS/q)_t$) reveal an interesting issue. Both series are stationary with high certainty, but the cointegration vectors are different (!). The estimated vector (Table 1) gives »less stationary» results, and the estimates are almost twice as small in absolute value compared to the a priori restricted case (see the last rows of Table 1). This disturbing result casts some doubts on the appropriateness of OLS-regressions in small samples as a method of deriving the cointegrated vectors. In section 4., some Monte Carlo studies are done to reveal the small sample properties of the proposed cointegration vectors. It is evident that the estimated cointegration vector is badly biased.

3. *Error correction models*

Although we can cast some doubts on the properties of the estimated cointegrated labour demand model, the second part of the Engle-Granger two-step estimation format, e_{1t-1} as the error correction variable, might work rather well, in addition to the restricted

[2] *Reported »t-values» have a non-standard distribution because of the nonstationary properties of the used series. This leads to a biased parameter inference when the ordinary t-distribution is used.*

[3] *When some modelling strategies were used with these basic series it turned out that random walk with drift best suits the series $\ln VAL_t$, $\ln WS_t$ and $\ln q_t$, but the appropriate model for the $\ln H_t$ series was the IMA(1,1) series without drift with significant MA(1)-coefficient with value .458 (see Linden 1989). This latter result causes some distribution changes in the DF-test (see Schwert (1988), who suggests that in the finite samples critical values for the DF-test used above have to be bigger in absolute value when MA(1)-errors are present).*

Table 1. OLS-parameters for equation (1)[2]

endogenous variable	constant »t»-value	$\ln VAL_t$ »t»-value	$\ln WS_t$ »t»-value	SE	R^2	DW
$\ln H_t$	−.027 (2.13)	.594 (4.17)	−.486 (6.06)	.045	.813	.655

series	Autocorrelations 1	2	3	4	5	10	Normality $\chi^2_N(2)$	DW
$\ln H_t$	.955	.835	.719	.592	.482	−.193	2.931	.225
$\Delta \ln H_t$	.457	.080	.008	−.215	−.234	.431	.521	1.067
$\ln VAL_t$	.993	.983	.976	.967	.956	.931	2.116	.023
$\Delta \ln VAL_t$	.056	−.005	−.168	−.023	−.205	.241	.491	1.459
$\ln WS_t$	.997	.990	.974	.971	.966	.957	.925	.028
$\Delta \ln WS_t$	.017	−.373	−.074	.097	.290	.391	.966	1.653
e_{1t}	.662	.275	.072	.137	.051	−.142	.372	.655
$\ln q_t$	.992	.987	.986	.984	.975	.953	1.265	.024
$\ln(WS/q)_t$	.432	−.156	−.293	−.081	−.033	.271	.391	1.073

$T = 26$ $\qquad \chi^2_N(2)^c_{\alpha=.05} = 5.991$

Table 2. OLS-estimates for error correction models[4]

variable	parameter (t-value)			F-test for the omitted error correction and level variables			
model:	$OLS_{lnWS/q}$	$OLS_{e_{1t}}$	$OLS_{unrest.}$				
ΔlnH_{t-1}	.342	.472	.335				
	(2.201)	(3.111)	(1.671)				
$\Delta lnWS_t$	−.303	−.254	−.295				
	(3.204)	(2.815)	(2.226)				
$\Delta lnVAL_t$	.318	.310	.298				
	(3.981)	(3.712)	(2.616)				
$ln(WS/q)_{t-1}$	−.238			5.71*			
	(2.398)						
e_{1t-1}		−.275		5.93*			
		(2.315)					
lnH_{t-1}			−.288	1.58	5.23*		
			(2.249)				
$lnWS_{t-1}$			−.239		4.85*	2.49+	
			(2.011)				
$lnVAL_{t-1}$			.233		5.74*		2.92+
			(2.171)				
T = 25, 1962–1986, critical values:				F(1,21)	F(1,19)	F(3,19)	F(2,19)
*) significant 5 % level,				4.32	4.38		
+) significant 10 % level,						2.40	2.61

Diagnostics:

model	SE	R^2	DW	$F_{AR(2)}$	$\chi^2_N(2)$	$F_{CHOW(2)}$	$F_{ARCH(2)}$
$OLS_{lnws/q}$	.0233	.598	1.92	.146	.241	.853	.753
$OLSe_{1t}$	.0237	.594	2.02	.132	.743	.834	.743
$OLS_{unrest.}$	.0241	.623	1.98	.082	.142	1.081	.622
$C_{\alpha=.05}$, T = 25 lnws/q & e_{1t}:				F(2,19)	$\chi^2_N(2)$	F(2,20)	F(2,16)
				3.52	5.99	3.49	3.63
unrestricted model:				F(2,16)		F(2,17)	F(2,13)
				3.63		3.59	3.81

error correction model in small samples. According to the »super consistency» of the OLS-estimator with cointegrated variables, short run misspecification errors are redundant as the OLS-estimates of the non-stationary variables converge faster to their true values than stationary ones. However, in small finite samples these results give only theoretical comfort and modelling the short run dynamics may help the properties of cointegration parameters, i.e., reducing the biases of the estimates (see Wickens and Breusch 1988, Dolado and Jenkinson 1987).

After some experiments with different models, the following error correction model was found to be reasonable:

$$(2)\quad \Delta lnH_t = \beta_1 \Delta lnH_{t-1} + \beta_2 \Delta lnWS_t + \beta_3 \Delta lnVAL + \alpha' X_{t-1} + \upsilon_t,$$

where X_{t-1} is e_{1t-1} or $ln(WS/q)_{t-1}$ or »unrestricted error correction» variable, i.e., $X_{t-1} = (lnH_{t-1}\ lnWS_{t-1}\ lnVAL_{t-1})'$. Table 2 gives the OLS-estimation results for these different models.

($F_{AR(2)}$ is an LM-test for residual AR(2)-process, $F_{CHOW(2)}$ is a prediction test for the years 1984–1986 and $F_{ARCH(2)}$ is a test for the residual autoregressive heteroscedasticity).

These results support the above formulat-

[4] *The constant term was excluded from all above equations as it was insignificant in the preliminary regressions.*

ed error correction model irrespective of the used error correction variable(s). Generally, the estimation results for the models are very similar for both the short run and the long run coefficients. When the long run coefficients are solved in the unrestricted model we get the following estimates: −.830 for $\ln WS_t$ and .809 for $\ln VAL_t$. These findings give more support to our earlier results concerning the biases of the estimates in the long run static (cointegrated) OLS-equation. The unrestricted error correction model has corrected the long run estimates toward our a priori parameters. F-test values 5.71 and 5.93 give the importance of adding the a priori restricted and residual based error correction variable into models. Both are very significant, indicating that although variable e_{1t-1} is the residual from a very biased long run model it works rather well as an error correction variable. Thus, the second part of the Engle-Granger two-step method seems to have some power to give sensible error correction estimates in small samples. The F-test value 1.58 comes from adding variable $\ln H_{t-1}$ to the difference model. These results support the above interpretation that series $\ln H_t$ is non-stationary. The test values 5.23, 4.85 and 5.74 stem from the position of adding one of the level variables when the other two are already present in the unrestricted error correction model. The values support the results that the proposed cointegration between the series $\ln H_t$, $\ln WS_t$ and $\ln VAL_t$ is well founded: when you add one of the level variables to the model when the other two are present you are always strongly supported by the used test[5].

However, the last two F-test values, 2.49 and 2.92, indicate that the whole level part of the unrestricted model is significant only at the 10 percent level. The former test value is for the whole level part of the unrestricted model and the latter is for adding $\ln WS_{t-1}$ and $\ln VAL_{t-1}$ to the model when $\ln H_{t-1}$ is present.

These results are somewhat opposite to the above test results concerning the »importance» of cointegrated variables. Although we have to be very cautious with these test values (see the above footnote 5), the results indicate that there are some problems with the unrestricted model and the interpretation of variable e_{1t-1} as a unique and well-founded error correction variable. The last point arises from the fact that the variable e_{1t-1} and the estimated linear combination of the variables $\ln H_{t-1}$, $\ln WS_{t-1}$ and $\ln VAL_{t-1}$ gives the same information concerning the cointegration. The a priori restricted error correction variable is free of these small sample level variable estimation problems and gives (asymptotically) the same information as the two alternatives above if the a priori restrictions are true. This matter is studied in the next section with some Monte Carlo studies.

4. *Monte Carlo experiments with the cointegrated and error correction models*

The results above concerning the estimated cointegration relationship give support to the interpretation that the parameter estimates are very biased. In the following the magnitude of biases is studied with some small Monte Carlo experiments. As the used series were tested to be normal with the given parameter values, we can always mimic the behaviour of the proposed data generating processes (DGP) when the series are drawn from the normal distribution. In the following experiments we used the following sample sizes and sample replicate numbers 30, 100 and 200.[6]

The bias is defined as $\text{BIAS} = 1/N \sum_{j=1}^{N} (\hat{\beta}_j - \beta_{DGP})$, where N is the number of sample replicates, $\hat{\beta}_j$ is the estimated parameter in different replicates and β_{DGP} is the given parameter value from the OLS-model (Table 1) or the used a priori restriction. The standard deviation (SD) of the bias is defined as

$$\text{SD} = [(1/N-1) \sum_{j=1}^{N} (\hat{\beta}_j - \beta_{DGP} - \text{BIAS})^2]^{1/2}$$

The t_0-rejection-% and t_{DGP}-rejection-% are shares of rejections for following test $\hat{\beta}_j = 0$

[5] *The test for parameter exclusion in the error correction model regression will have the ordinary F-distribution, if the level variables are cointegrated, but will have a non-normal limiting distribution otherwise (see Banerjee et al. 1986).*

[6] *See Henry and Neale (1987) and Linden (1989) for details of the set-up and how to run a Monte Carlo experiment with PC-NAIVE.*

Table 3. Monte Carlo experiment with the estimated cointegration and a priori parameters

variable	DGP-prm	bias (SD) 30	100	200	t_0-reject-% 30	100	200	t_{DGP}-reject-% 30	100	200
model 1):										
$\ln V_t$	.594	.343 (.128)	.385 (.069)	.422 (.049)	100	100	100	93	100	100
lnWS	−.486	−.397 (.151)	−.443 (.084)	−.488 (.058)	100	100	100	90	100	100
model 1'):										
$\ln V_t$	1.0	−.0009 (.0022)	.0038 (.0015)	.0042 (.0009)	100	100	100	13	88	100
$\ln WS_t$	−1.0	−.0062 (.0139)	−.0129 (.0056)	−.0143 (.0034)	100	100	100	36	85	100

Table 4. The biases of the error correction parameters in model 2) when error correction variables are e_{1t-1} or $\ln(WS/q)_{t-1}$

variable	DGP-parm.	bias (SD) 30	100	200	t_0-reject-% 30	100	200	t_{DGP}-reject-% 30	100	200
e_{1t-1}	−.275	−.053 (.127)	−.176 (.079)	−.128 (.036)	70	100	100	7	26	93
$\ln(WS/q)_{t-1}$	−.238	.033 (.055)	.0057 (.034)	.0034 (.022)	73	100	100	3	1	1

and $\hat{\beta}_j = \beta_{DGP}$ over the sample replicates.

The message of these experiments is clear: the estimated cointegration parameters are strongly biased and the a priori restrictions are almost unbiased. The standard deviations for the former biases are big, causing more t_{DGP} -rejections compared to the restricted data generation parameters.

When similar experiments were done with the error correction models (especially with models that used e_{1t-1} or $\ln(WS/q)_{t-1}$ as the error correction variable), the short run estimates were generally unbiased but the biases for error correction parameters were the following

These findings cast more doubts on the Engle-Granger two-step error correction model estimation method. The estimated DGP-parameter value for e_{1t-1} (−.275) is clearly biased with sample and replicate sizes 100 and 200. On the contrary, the error correction estimate for the a priori restricted error correction variable has a very small bias with all sample and replicate sizes. These results are sensible when compared with the results of the experiment above done with cointegration models, i.e., the estimated cointegration parameters are very biased and sensitive to sample size.

5. *Conclusions*

A very simple labour demand model is proposed and estimated after testing the stationary of used series. Series were found to be nonstationary. These results open the way to one effective model specification, i.e. the error correction model, which can handle nonstationary series as long as a linear stationary relationship exist between the level series. The cointegration between the series was established, but the underlying properties of the series causes the testing and the estimation of the cointegrated vector to be non-standard and fragile in the small samples. With some small-scale Monte Carlo experiments, it was shown that the Engle-Granger two-step cointegration estimation method gives very biased and doubtful results concerning the cointegration and error correction parameters in small samples.

It is emphasized that more Monte Carlo experiments with different parameter values and bigger sample sizes are needed before the final merits of the Engle-Granger method can be determined. However, even these preliminary results support the modelling strategy in which the economic theory has a role to play, i.e., error correction models with the long run restrictions.

References

Banerjee, A., Dolado, J., Hendry, D. and Smith, G. (1986): »Exploring Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence,» *Oxford Bulletin of Economics and Statistics* 48,253 – 277.

Dolado,J.J. and Jenkinson,T. (1987): »Cointegration: A Survey of Recent Developments,» Applied Economics Discussion Papers No. 39, Oxford Univ.

Engle,R.F. and Granger,C.W.J. (1987): »Co-Integration and Error Correction: Representation, Estimation and Testing,» *Econometrica* 55, 251 – 276.

Fuller,W.A. (1976): *Introduction to Statistical Time Series,* Wiley.

Granger, C.W.J. (1986): »Developments in the Study of Co-Integrated Economic Variables,» *Oxford Bulletin of Economics and Statistics* 48, 213 – 228.

Hendry,D.F. and Neale, A.J. (1987): »Monte Carlo Experimentation Using PC-NAIVE.» in Fomby, T. and Rhodes, G. (ed): *Advances in Econometrics* 6,91 – 125, JAI-Press.

Hylleberg, S. and Mizon, G.M. (1989): »Cointegration and Error Correction Mechanisms,» *Economic Journal,* 99 (Conference Papers),113 – 125.

Linden,M. (1989): *Granger Representation Theorem and the Dynamic Labour Demand Models,* (in Finnish), Licentiate thesis, Univ. of Helsinki, mimeo.

Molinas,C. (1986): »A Note of Spurious Regressions with IMA-Errors,» *Oxford Bulletin of Economics and Statistics* 48,279 – 282.

Phillips, P.C.P. (1987): »Time Series Regression with a Unit Root,» *Econometrica* 55,277 – 301.

Phillips, P.C.P. and Ouliaris, S. (1987): »Asymptotic Properties of Residual Based Tests for Cointegration.» Cowles Foundation Working Paper 847.

Santamäki, T. (1986): »Empirical Labour Demand Research in Finland.» (in Finnish), 1 – 22 in *Työllisyystutkimustoimikunnan mietinnön liite, Komiteanmietintö* 1986:40.

Schwert, G.M. (1988): »Test for Unit Roots: A Monte Carlo Investigation.» NBER Technical Working Paper 73.

Schwert, G.M. (1987): Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data,» *Journal of Monetary Economics* 20,73 – 103.

Stock, J.M. (1987): »Asymptotic Properties of Least Square Estimators of Cointegrated Vectors,» *Econometrica* 55,1035 – 1056.

Wickens, M.R. and Breusch, T.S. (1988): »Dynamic Specification, the Long Run and Estimation of Transformed Regression Models,» *Economic Journal* 98 (Conference Papers), 189 – 200.