STATIONARITY AND COINTEGRATION TESTS: COMPARISON OF ENGLE -GRANGER AND JOHANSEN METHODOLOGIES

Faik Bilgili*

I. INTRODUCTION

Macroeconomic time series studies are based on the assumption that the underlying time series is stationary. Time series studies, however, show that many time series are not stationary in their levels but stationary in differences. The results of regression a nonstationary time series variable on an another nonstationary time series variable yields often spurious results although there is no meaningful relationship between them.

To avoid spurious results such as biased traditional F and t statistics, one should use stationary variables in their levels or difference stationary variables. Stationarity means that the time series has a constant mean, *m* and finite (bounded) variance, σ^2. A stationary time series has a tendency to frequently to return to the mean value. A nonstationary time series can not be used in estimation of the model to be used in forecasting. In this case one should investigate if these variables have long-run relationship (cointegration). If they are cointegrated, a regression in which nonstationary variables are employed would not suffer from losing any valuable long term information.

Section II explains relationship between stationarity and cointegration. In section III, two methodologies for testing cointegration, Engle-Granger and Johansen methodologies, show the testing procedures for cointegration step by step. Section IV gives an empirical evidence on cointegration by comparing Engle-Granger and Johansen methodologies.

II. STATIONARITY AND COINTEGRATION

Test for unit roots are performed on univariate time series. Considering the univariate time series

$$y^*_t = \varsigma y^*_{t-1} + u_t \tag{1}$$

where y^*_t is measured as deviations of y from its population mean, μ. Expected value of random error is equal to zero, $E(u_t) = 0$ and variance of random error is finite (a scalar),

$\text{var}(u_t) = \sigma^2$ or $E[u_t - E(u_t)]^2 = \sigma^2$.

Consider now the multivariate time series

$$\mathbf{x}_t = \Phi_1\mathbf{x}_{t-1} + \Phi_2\mathbf{x}_{t-1} + \Phi_3\mathbf{x}_{t-1} + \ldots\ldots + \Phi_p\mathbf{x}_{t-p} + \mathbf{e}_t, \tag{2}$$

where $\Phi_1, \Phi_2, \Phi_3, \ldots, \Phi_p\mathbf{x}_{t-p}$ n by n matrices. Eq.(2) can be reparameterized as

$$\Delta\mathbf{x}_t = \Pi_1\Delta\mathbf{x}_{t-1} + \Pi_2\Delta\mathbf{x}_{t-2} + \Pi_3\Delta\mathbf{x}_{t-3} + \ldots + \Pi_{p-1}\Delta\mathbf{x}_{t-p+1} - \phi\mathbf{x}_{t-1} + \mathbf{e}_t. \tag{3}$$

---

* Dr., Erciyes University Faculty of Economics and Administrative Sciences, Dept. of Economics.

In eq.(1), if ς =0, y_t is known as random walk or nonstationary time series.

. In eq.(3), if the matrix $\phi = (\mathbf{I} - \Phi_1 - \Phi_2 - \Phi_3 - \ldots \Phi_p)$ is full rank, then any linear combination of $\mathbf{x}_t$ will be stationary. The test for cointegration is to test for rank of ϕ by testing whether the eigenvalues of estimated ϕ are significantly different from zero.

Studies of macroeconomic theory assumes that there should be a stable long-run relationship among variables. In the existence of long-run relationship, variables can not move too far from each other. If individual time series are not stationary, they can wander too far from each other, and traditional statistics of a projection of one nonstationary variable on another become unreliable[1]. If series are stationary but cointegrated, however, they are expected to move together in the long-run. In short, cointegration means that one or more linear combinations of these variables is stationary even though individually they are not.

III. COINTEGRATION TESTS BY ENGLE-GRANGER AND JOHANSEN METHODOLOGIES

In this section, I will first introduce theorems and then explain the statistical calculations of the Engle-Granger and Johansen tests. In fact, there are several estimation of cointegration relations, such as, OLS (Engle-Granger, 1987), Augmented Least Squares (Bewley, 1979; Hendry and Richard, 1982), Instrumental Variables (Phillips and Hansen, 1990), Fully Modified Estimator (Park and Phillips, 1988), Non-Parametric Canonical Cointegraton (Park, 1989), Three Step Estimator (Engle and Yoo, 1991), Canonical Cointegration (Bossaerts, 1988), Spectral Regression (Phillips, 1991), Principal Components (Stock and Watson, 1989), Maximum Likelihood (Johansen, 1988), Modified Box-Tiao (Bewley , Orden and Fisher, 1991).[2] In this study, however, I examine the two of them, Engle-Granger and Johansen procedures, since they have been the most commonly used among others for cointegration analysis. OLS estimator is the simplest one to analyze the cointegration analysis while Maximum likelihood estimator is the best if the model is well specified without highly autocorrelated cointegrating errors.

III. 1 Engle-Granger Methodology

Testing for cointegration by Engle-Granger methodology proposes a straightforward test whether variable in $\mathbf{x}_t$ vector are cointegrated. Let y_t and z_t be two variables in $\mathbf{x}_t$ vector and suppose they are integrated of order 1. Engle and Granger methodology tests whether y_t and z_t are cointegrated of order CI (1,1).

Definition: *The components of the vector x_t are said to be cointegrated of order d, b, denoted*

$x_t \sim CI(d,b)$, if

a- all components of $\mathbf{x}_t$ are I(d),

[1] See P. Phillips (1986) and C. Granger and P. Newbold (1974).
[2] See Hargreaves (1994).

b- there exists a vector b($\neq$ 0) so that $m = b'x_t \sim I(d\text{-}b)$, b>0. The vector b is called the co-integrating vector[3].

One can perform Engle-Granger cointegration test as follows[4]:

1- Determine order of integration of variables y_t and z_t. If they are integrated of the same order, one can apply the cointegration test. Eq. (1) can be tested for both y_t and z_t by Dicky-Fuller or Augmented Dicky-Fuller to see if $\varsigma = 0$ for each variable. If it is so, it would mean that variables are not stationary and that their differences might be integrated of order zero.

$$\Delta y_t = a_0 + \varsigma y_{t-1} + e_t. \tag{4}$$

$$\Delta y_t = a_0 + \varsigma y_{t-1} + a_1 t + e_t. \tag{5}$$

$$\Delta^2 y_t = b_0 + \varsigma \Delta y_{t-1} + e_t. \tag{6}$$

$$\Delta^2 y_t = b_0 + \varsigma \Delta y_{t-1} + b_1 t + e_t. \tag{7}$$

$$\Delta^3 y_t = c_0 + \varsigma \Delta^2 y_{t-1} + e_t. \tag{8}$$

$$\Delta^3 y_t = c_0 + \varsigma \Delta^2 y_{t-1} + c_1 t + e_t. \tag{9}$$

$$\Delta^d y_t = m_0 + \varsigma \Delta^{d-1} y_{t-1} + e_t. \tag{10}$$

$$\Delta^d y_t = m_0 + \varsigma \Delta^{d-1} y_{t-1} + m_1 t + e_t. \tag{11}$$

The tests for eqs.(4) and (6) are Dicky-Fuller tests and the tests for eqs. (5) and (7) are Dicky-Fuller with trend variable tests. One can write the same equations for z_t as well. The error term e_t is white noise, if it is serially uncorrelated and its expected mean is equal to zero. If e_t does not seem to be white noise, an Augmented Dicky Fuller (ADF) test will be implemented. The number of lags of ADF are increased until e_t becomes white noise. If $\varsigma = 0$ from eqs.(6) and (7) for each variable, then they are integrated of order one, *I*(1). A series that needs to be differenced *d* times to achieve stationary is said to be integrated of order *d* or *I(d)*.

2- If the variables are found integrated of order one at first step, one can proceed the following regressions and save the residuals.

$$y_t = a_o + a_1 z_t + e_{1,t} \tag{12}$$

$$z_t = b_o + b_1 y_t + e_{2,t}. \tag{13}$$

3- Regress the following regressions and test for unit root for each equation.

---

[3] Engle and Granger (1991, p.84).
[4] Enders (1995, pp. 373-385).

$$\Delta e_{1t} = a_1 e_{1t-1} + v_{1t}. \tag{14}$$

$$\Delta e_{2t} = a_2 e_{2t-1} + v_{2t}. \tag{15}$$

If it is not possible to reject the null hypotheses that $|a_1| = 0$ and $|a_2| = 0$, one can not reject the hypothesis that the variables are not cointegrated.

III. 2 Johansen Methodology

In maximum likelihood estimation of cointegration vectors, the null hypothesis is, for any $r \leq p$, Ho: rank(Φ) $\leq$ r or $\Phi = a B'$, where a and B are p x r matrices. If there is cointegration among variables, X_t is cointegrated with the cointegration vectors B ($\Phi = a B'$). One can not estimate the parameters of a and B but can estimate the space spanned by B. Now, in estimation of space spanned by B, the theorem is as follows:

Theorem: *The maximum likelihood estimator of the space spanned by B is the space spanned by the r canonical variates corresponding to the r largest squared canonical correlations between the residuals of* $\mathbf{x}_{t-p}$ *and* $\Delta \mathbf{x}_t$ *corrected for the effect of the lagged differences of the* **x** *process.*[5]

One can obtain the largest canonical correlation as follows[6]:

1- After determining the order of p, regress $\Delta \mathbf{x}_t$ on $\Delta \mathbf{x}_{t-1} + \Delta \mathbf{x}_{t-2} + \Delta \mathbf{x}_{t-3} + \ldots + \Delta \mathbf{x}_{t-p+1}$ and save the residuals.

2- Regress $\mathbf{x}_{t-p}$ on $\Delta \mathbf{x}_{t-1} + \Delta \mathbf{x}_{t-2} + \Delta \mathbf{x}_{t-3} + \ldots + \Delta \mathbf{x}_{t-p+1}$ and save the residuals.

3- Let $\mathbf{n}_t$ be residuals from step 1 and $\mathbf{v}_t$ be residuals from step 2.

4- Compute squares of the canonical correlations[7] between $\mathbf{n}_t$ and $\mathbf{v}_t$, as:

$$\Omega_1^2 \rangle \Omega_2^2 \rangle \Omega_3^2 \ldots\ldots \rangle \Omega_p^2.$$

5- Maximal eigenvalue test that uses $(r+1)^{th}$ largest squared canonical correlation is

$$\Omega \max (r, r+1) = -T \ln (1 - \hat{\Omega}_{r+1}^2). \tag{16}$$

6- Or one can obtain the trace test as follows:

$$\Omega \text{ trace } (r) = -T \sum_{i=r+1}^{n} \ln (1 - \hat{\Omega}_i^2). \tag{17}$$

Ω trace and Ω max tests the number of eigenvalues, r, that are statistically different from zero. For instance, in the three variable case, n = 3, Ω trace tests the hypothesis that there is no cointegration, against alternative that

[5] S. Johansen (1988, p.234).
[6] Dickey, Jansen and Thornton (1991, pp. 62-63).
[7] See Hamilton (1995, pp. 630-35) for calculation of canonical correlations.

there are 1, 2, or 3 cointegration vectors. If H_0: $r = 0$ is rejected against H_1: $r > 0$, then H_0 is $r \leq 1$ is tested against hypothesis $r = 2$ or 3. The Ω max is more specific than the Ω trace test, whose null hypothesis is that there are no cointegrating vectors against the hypothesis that there is one cointegrating vector. Or in Ω max tests; Ho: r =1 vs H_1: r =2, Ho: r = 2 vs H_1: r = 3.

IV. EMPIRICAL EVIDENCE

In this section I will conduct all necessary tests indicated in section III for the variables of a basic consumption function below

$$C_t = a_0 + a_1 Y_t + a_2 T_t + e_t \tag{18}$$

where C is real consumption, Y is real GDP and T is real tax revenues. Real consumption is a function of real income and taxes. It has been known from the literature that this function (or slightly a different version) has been tested for many times to understand the consumption behavior. The basic concern is here, however, not to perform this regression indicated by eq.(18) but to see if one can use these variables in the model to estimate the aggregate demand. In other words, obtaining parameter estimates from this model is the second step which is not of interest here in this study. The purpose is to perform the first step which is to conduct the stationary tests and cointegration tests by Engle-Granger and Johansen methodologies.

These tests will be performed for seven countries; Canada, Germany, India, Italy, Japan, Turkey and the USA.

Annual Private Consumption and GDP data were drawn from World Bank Data CD-ROM 1995 for the period 1960 to 1993. The Tax Revenues, data were drawn from World Bank Data CD-ROM 1995 and IFS CD-ROM 1995 for the period 1970 to 1993. Nominal annual values were divided by the GDP deflator to obtain real values of C_t, Y_t, and T_t. The lag number = 4 was determined for these countries by Akaike information criterion (AIC), and Schwartz Bayesian Criterion (SBC).

In Table 1, all Dicky-Fuller test results indicate that all variables for each country are nonstationary except, Y_t (by DF test without trend) of Japan. Table 2 shows that all nonstationary variables are integrated of order one. Tables 3-A and 4-A give the results of cointegration tests by Engle-Granger and Johansen methodologies, respectively. Engle-Granger cointegration test can be run by eq.(19) through eq.(24).

$$C_t = a_o + a_1 Y_t + a_2 T_t + \varepsilon_{1,t}. \tag{19}$$

$$Y_t = b_o + b_1 T_t + b_2 C_t + \varepsilon_{2,t}. \tag{20}$$

$$T_t = c_o + c_1 Y_t + c_2 C_t + \varepsilon_{3,t}. \tag{21}$$

$$\Delta\varepsilon_{1t} = \alpha_1 \varepsilon_{1t-1} + \nu_{1t}. \tag{22}$$

$$\Delta\varepsilon_{2t} = \alpha_2 \varepsilon_{2t-1} + \nu_{2t}. \tag{23}$$

$$\Delta\varepsilon_{3t} = \alpha_3 \varepsilon_{3t-1} + \nu_{3t}. \tag{24}$$

All residuals seem to be white noise, that is, null hypothesis that their means are equal to zero and they are serially uncorrelated are all accepted at 0.05 level. Therefore I did not run ADF tests for residuals. The cointegration results of eqs (22), (23) and (24) are given in Table 3-A. Tables 3-A and 3-B show that there is no cointegration relation among variables of Canada, India and Japan equations and that α_3 of Germany, α_2 of Italy and USA equation are statistically significant, therefore there is cointegration relation among variables in these countries. However α_1 and α_2 of Germany, α_1 and α_3 of Italy and USA equations do not confirm these cointegration results. Results are inconclusive. Under these inconsistent results one may run Johansen cointegration test.

$$\Delta x_t = c + (\Phi - I)x_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + e_t, \tag{25}$$

$$= c + \psi\, x_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + e_t, \tag{26}$$

where $\Gamma_i = (\Pi_1 + \Pi_2 + \ldots + \Pi_{p-1})$, Φ is an $n \times n$ matrix of parameters and I is an $n \times n$ identity matrix. Eq.(26) is same as eq.(3).The number of non-zero eigenvalues of ψ will determine the number of cointegrating vectors. I apply the Johansen methodology that determines the number of non-zero eigenvalues by maximum likelihood method. One or more liner combinations of these variables might be stationary whereas variables in x_t are non-stationary in levels as we see from tables 1 and 2. Eigenvalues of ψ are given in Table 4-A. The results of Johansen methodolgy indicate that there is at least one cointegration relationship among variables for each country except Germany. Furthermore, max and trace tests show that there are two cointegrating vectors in Canada, India and the USA equations and that there are three cointegrating vectors in Italy, Japan and Turkey equations.

Table 1: The Test of Stationarity

	C_t		Y_t		T_t	
Countries	**1**	**2**	**1**	**2**	**1**	**2**
Canada	-1.89	-2.52	-1.98	-1.06	-0.62	-1.01
Germany	0.65	-1.45	0.41	-1.64	0.03	-1.31
India	-0.79	-2.27	-0.55	-2.01	0.88	-1.15
Italy	-1.21	-2.50	-1.23	-2.57	-1.62	-2.39
Japan	-2.84	-3.05	-3.54	-3.48	1.07	-0.80
Turkey	-2.21	-3.21	-2.02	-2.76	-1.28	0.69
USA	0.67	-1.57	-0.78	-2.67	-1.03	-2.43

Table 2: The Test of Integration

	C_t		Y_t		T_t	
Countries	**1**	**2**	**1**	**2**	**1**	**2**
Canada	-3.12	-7.51	-3.61	-3.76	-3.71	-3.86
Germany	-3.46	-4.97	-3.66	-3.61	-4.90	-3.32
India	-5.54	-5.44	-4.94	-4.87	-3.22	-4.00
Italy	-6.39	-6.33	-7.01	-6.91	-5.65	-5.60
Japan	-6.76	-6.67	-7.00	-6.89	-4.83	-4.64
Turkey	-10.04	-9.64	-9.08	-8.76	-5.14	-5.02
USA	-3.77	-3.75	-4.50	-4.47	-4.02	-3.89

	significance	**0.01**	**0.05**	**0.10**
1	$\Delta^2 Y_t = \alpha_0 + \delta\,\Delta Y_{t-1} + \varepsilon$	-3.75	-3.00	-2.62
2	$\Delta^2 Y_t = \alpha_0 + \delta\,\Delta Y_{t-1} + b$	-4.38	-3.60	-3.24

Table 3-A: Cointegration Test Results by Engle-Granger Methodology

Countries	τ ,*tau,* value of of α_1	τ ,*tau,* value of of α_2	τ ,*tau,* value of of α_3
Canada	1.24	1.74	2.05
Germany	3.38	1.93	3.73
India	3.09	3.16	3.19
Italy	3.63	9.85	2.36
Japan	3.32	3.31	0.03
Turkey	5.56	6.11	4.86
USA	3.54	3.88	3.36

Table 3-B: Critical Values for Cointegration tests by Engle and Yoo[8]

0.01	0.05	0.10
4.84	4.11	3.73

Table 4-A: Cointegration Test Results by Johansen Methodology

Countries	λ_i	λMax	λTrace
	λ_1= 0.8539	32.70	53.61
Canada	λ_2= 0.6743	19.07	20.90
	λ_3= 0.1023	1.84	1.84
	λ_1= 0.6815	21.74	31.09
Germany	λ_2= 0.3272	7.53	9.35
	λ_3= 0.0914	1.82	1.82
	λ_1= 1.0000	405.57	443.17
India	λ_2= 0.9170	37.33	37.60
	λ_3= 0.0179	0.27	0.27
	λ_1= 0.9998	134.22	182.86
Italy	λ_2= 0.8829	34.31	48.64
	λ_3=0.5915	14.32	14.32

Table 4-A, continued

	λ_1= 0.9220	43.36	78.45
Japan	λ_2= 0.7512	23.65	35.10
	λ_3=0.4901	11.45	11.45
	λ_1= 0.9408	53.70	91.35
Turkey	λ_2= 0.7884	29.51	37.65
	λ_3=0.3483	8.14	8.14
	λ_1= 0.9495	50.76	73.04
USA	λ_2= 0.7029	20.63	22.28
	λ_3=0.0924	1.65	1.65

[8] Engle and Yoo (1991).

Table 4-B: Null and Alternative Hypothesis by Johansen Methodology

λTrace Test		λ Max Test	
Ho:	H_1:	Ho:	H_1:
r = 0	r > 0	r = 0	r = 1
r ≤ 1	r > 1	r = 1	r = 2
r ≤ 2	r > 2	r = 2	r = 3

Table 4-C: Critical Values for λMax and λTrace tests[9]

	λMax			λTrace		
n-r	**0.10**	**0.05**	**0.01**	**0.10**	**0.05**	**0.01**
1	7.56	9.09	12.74	7.56	9.09	12.74
2	13.78	15.75	19.83	17.95	20.16	24.98
3	19.79	21.89	26.40	32.09	35.06	40.19

V. CONCLUSION

Many economic theories imply that a linear combination of variables is stationary although individually they are not. If there is such a stable linear combination among variables, the variables are said to be cointegrated. In existence of cointegration or long-run relationship, the variables have the same stochastic trends and therefore they can not drift too far apart. In time series analysis of macroeconomic studies, hence, one should check for stationarity and cointegration to avoid losing long term information.

There are several methods in examining the cointegration analysis. Engle-Granger and Johansen procedures are the most commonly used among others in the literature. In Engle-Granger procedure, one examines the residuals from long-run equilibrium relationship by ordinary least squares method. The variables are cointegrated if these residuals do not have a unit root. Johansen procedure, in estimation cointegration relationship, estimates a vector autoregression in first differences and include the lagged level of the variables in some period t-p.

There are several problems of Engle-Granger methodology. First, in examining residuals from the long-run equation relationship, there is no presumption that any of the three residual series, for instance in three variable case, is preferable to any of the others. One can find a cointegration relationship from residuals of the first regression whereas residuals of second and the third regressions may not yield a cointegration result. In other words, in finite sample case, the test for unit root in the error term sequence from the first regression may not be equivalent to the test for unit root in the error term sequence from another regression. Indeed, results of Section IV confirm this problem.

[9] Source: Enders (1995, p. 420) and see also Johansen and Juselius (1990, p.371).

Engle-Granger methodology relies on a two-step estimator. The first step is to generate the residuals and the second step uses these generated residuals to estimate a regression of first-differenced residuals on lagged residuals. Therefore any error occurred in the first step is carried into second step.

The Johansen maximum likelihood estimators circumvent the use of two-step estimators and can estimate and test for the presence of multiple cointegrating vectors. Some Monte Carlo evidence suggest that Johansen procedure performs better than both single equation methods and alternative multivariate methods. Section IV concludes that Engle-Granger yields some inconclusive results whereas Johansen test finds at least one cointegration relationship among variables for all countries except Germany. When the defectives of Engle-Granger methodology are taken into account, the conclusion of this study may suggest that Johansen methodology dominates the Engle-Granger methodology in cointegration analyses.

BIBLIOGRAPHY

Dickey, David A. and Jansen W. Dennis, Thornton L. Daniel, "A Primer On Cointegration with an Application to Money and Income," **Federal Reserve Bank of St. Louis**, March/April 1991.

Enders, Walter, **Applied Econometric Time Series**, John Wiley & Sons, Inc, New York, 1995.

Engle, R. and C. Granger, "Cointegration and Error Correction: Representation, Estimation, and Testing," In Engle and Granger (eds.), **Long Run Economic Readings in Cointegration**, Oxford University Press, New York, 1991, 81-113.

Engle, R. and Yoo Sam, "Forecasting and Testing in Co-integrated Systems," In Engle and Granger (eds.), **Long Run Economic Relationships. Readings in Cointegration**, Oxford University Press, New York, 1991, 237-67.

Granger, Clive, "Developments in The Study of Cointegrated Economic Variables," In Engle and Granger (eds.), **Long Run Economic Relationships. Readings in Cointegration**, Oxford University Press, New York, 1991, 65-81.

Granger, Clive and P. Newbold, "Spurious Regressions in Econometrics," **Journal of Econometrics**, 2 (1974), 111-20.

Hamilton, James D, **Time Series Analysis**, Princeton University Press, Princeton, New Jersey, 1994.

Hargreaves, Colin, "A Review of Methods of Estimating Cointegrating Relationships," In Colin P. Hargreaves (ed.), **Nonstationary Time Series Analysis and Cointegration**, Oxford University Press, New York, 1994, 87-133.

Johansen, Søren, "Statistical Analysis of Cointegration Vectors," **Journal of Economic Dynamics and Control**, 12 (1988), 231-254.

Johansen, S.and K. Juselius, "Maximum Likelihood Estimation and Inference on Cointegration with Application to the Demand for Money," **Oxford Bulletin of Economics and Statistics**, 52 (1990), 169-209.

Phillips, Peter, "Understanding Spurious Regressions in Econometrics," **Journal of Econometrics**, 33 (1986), 311-40.