" Small Sample properties of Residuals Based Tests for Cointegtration: Some New Developments in Econometrics "

Sami Al - Khatrash (*)
Robert Mc Nown (**)

ملخص

مواصفات العينة الصغيرة لاختبارات التكامل المشترك المعتمدة على البواقى :

تطورات حديثة في الاقتصاد القياسي

منذ نشر بحث جرينجز وويز (١٩٨٣) وبحث انجل وجرينجز (١٩٨٧) عن موضوع التكامل المشترك كان هناك حجم كبير من الانتاج المتواصل والأدوات المستحدثة في هذا الحقل.

وقد أفردت نشرة اكسفورد للاقتصاد والاحصاء عددا خاصا عن الموضوع في (١٩٨٦) . انجل وجرينجز اقترحا عدة اختبارات للتكامل المشترك تعتمد على بواقى انحدار المربعات الصغرى واستخدام اختبار ديكى - فيلير (١٩٧٩) لجذر الوحدة في بواقي الانحدار التكاملي المشترك. ويتمثل الهدف من هذا البحث فى عرض نتائج تجربة مونت - كارلو لمواصفات عينة صغيرة لبواقي الاختبارات المبنية علي التكامل المشتراك.

(*) Kuwait University - Department of Economics
(**) University of Colorado, Boulder - Department of Economics

1- INTRODUCTION

Since the publication of Granger and Weiss (1983) paper on cointegration there has been an extraordinary volume of research applying and extending these techinques. Engle and Granger (1987)presented several alternative tests for cointegration based on the residuals from least squares regressions, and their use of the Dicky - Fuller (1979) test for a unit root in the residuals from a cointegrating regression has been widely used in the applied literature. Other unit root tests, such as phillips' (1987) modification of the Dickey - Fuller statistic, have been adapted to the task of cointegration testing. Despite widespread use of these residual based tests, there is scant information on the small sample properties of these procedures.The objective of this paper is to present the results of a Monte Carlo study of the small sample properties of residuals based tests of cointegration.

However the tests investigated include several proposed by Engle and Granger (1987) and Philips and Durlauf's (1987) adaptation of Philips (1987) unit root test. The exact form of the tests studied here are described in the following section. Although there is growing interest in alternative approaches to cointegration testing, such as Johansen's (1988) maximum likelihood procedures and Stock and Watson's (1988) common trends approach, for computational reasons we have limited this study to the residual based tests . Investigation of the small sample properties of these alternative approaches would be a worthwhile extension of this paper.

With the exception of a small simulation study reported in Engle and Granger (1987), we are not aware of any other Monte Carlo evidence on the performance of these cointegration tests. There are, however, several studies of the size and power of the unit root tests, and these studies have relevance to this investigation. These studies are surveyed in section 3. The design of the simulation experiments is descibed in Section 4, and results are summarized in Section 5.

2. Test Descriptions :

Variations on five tests for cointegration are examined in this study. Four of these are tests presented by Engle and Granger (1987), and the fifth is the application of Phillip's. (1987) adaptation of the Dickey - Fuller test to cointegrating regression residuals. Every test begins with the least squares estimation of the cointegrating regression,

$$Y_t = \alpha X_t + c + \mu_t \tag{1}$$

1. DF (Dickey - Fuller). The first test uses the residuals from (1) in the regression,

$$\Delta\mu_t = \emptyset\mu_{t-1} + \varepsilon_t \tag{2}$$

   and tests the significance of $\emptyset$ with the computed t-statistic. The distribution of the test statistic is nonstandard, and critical values have been generated through simulation by Engle and Yoo (1987).

2. ADF (Augmented Dickey - Fuller). The second test extends the DF test to allow for higher order autoregressive processes by using the cointegrating regression residuals in the regression,

$$\Delta\mu_t = \emptyset\mu_{t-1} + b_1\Delta\mu_{t-1} + \cdots + b_p\Delta\mu_{t-p} + \varepsilon_t \tag{3}$$

   and again testing the significance of $\emptyset$. The lag length, p, must be sufficiently long to extract any serial correlation in the residuals. Two alternative lag specifications of four (ADF4) and twelve (ADF12) are examined in this study.

3. ECM (Error correction model test). In this test a triangular form of the error correction model.

$$\Delta Y_t = ß_1 \mu_{t-1} + E_{1t}$$

$$\Delta X_t = ß_2 \mu_{t-1} + \gamma \Delta Y_t + E_{2t}$$

is estimated by least squares. By imposing triangularity, the disturbances are uncorrelated, and assuming normality, the t-statistics for β_1, and β_2 are independent. The test statistic is the sum of the squared t-statistics, $T^2 \beta_1 + T^2 \beta_2$.

4. AECM (Augmented error correction model test). This test is the same as the previous test but with p lags of ΔY_t and ΔX_t in each equation. The test statistic is again the sum of the t-statistics on the error correction terms, $T^2 \beta_1 + T^2 \beta_2$. In this study four lags of ΔY_t and ΔX_t are included in each equation.

5. PD (Phillips and Durlauf test). Phillips (1987) employs a nonparametric adjustment to the Dickey-Fuller statistic, to permit a wide class of dependent and heterogeneous innovation processes. The resulting test statistic has a limiting distribution that is identical to that of the DF statistic under the assumption of iid errors.

For the model

$$Y_t = \alpha Y_{t-1} + u_t \qquad \alpha = 1$$

Phillips (1987) presents the test statistic:

$$Z\alpha = \frac{T(\hat{\alpha}-1) - 1/2\,(S^2_{Tl} - S^2_u)}{T^{-2}\sum_{1}^{T} Y^2_{t-1}}$$

$$S^2_u = T^{-1}\sum_{1}^{T} (Y_t - Y_{t-1})^2$$

$$\hat{\alpha} = \frac{\sum_{1}^{T} Y_t Y_{t-1}}{\sum_{1}^{T} Y^2_{t-1}}$$

$$S^2_{Tl} = T^{-1}\sum_{1}^{T} u^2_t + 2\,T^{-1}\sum_{k=1}^{l} W_{kl}\sum_{t=k+1}^{T} u_t\,u_{t-k}$$

To test the unit root hypothesis, one compares Za with the critical values from Fuller (1979). Phillips and Durlauf (1987) extend this procedure to tests of a unit root in the cointegrating regression residuals

Two alternative lag specifications of the Phillips-Durlauf test are used in this study to test the null hypothesis of a unit root in the regression residuals. Following Schwert's (1989) recommendation the number of lags, l, in calculating the variance S^2_{Tl} is determined by the formulas:

$$l_4 = \text{int}\,\{4(T/100)^{1/4}\}$$

$$l_{12} = \text{int}\,\{12(T/100)^{1/4}\}.$$

For the sample sizes in this study,

$$l_4 = 4 \text{ and } l_{12} = 12 \text{ when } T = 100$$
$$l_4 = 3 \text{ and } l_{12} = 10 \text{ when } T = 50$$
$$l_4 = 5 \text{ and } l_{12} = 15 \text{ when } T = 250.$$

3. Simulation Studies of Unit Root Tests.

Several studies have examined the performance of unit root test, primarily as tests of nonstationarity of individual series. The interesting questions in this literature are the sensitivity of the distribution of each test statistic to parameters in the null and the power of the tests. Similar issues have been examined in the smaller literature on cointegration testing.

Schwert (1989) investigates the sensitivity of test statistic distributions to the strength of the moving average component in IMA (1,1) processes. Empirical test sizes are computed for Dickey-Fuller, augmented Dickey-Fuller, and phillips tests, and a test based on a nonlinear ARIMA estimator. He finds the Phillips statistics have sizes that differ substantially from thier nominal values, and increasing the lag length does not mitigate this problem. His results tend to favor the augmented Dickey-Fuller tests on the basis of low sensitivity of size to the presence of the moving average component, with the longer lags preferred when the moving average parameter is close to one. This conclusion rests on considerations of size only, and he did not analyze power.

Dickey and Fuller (1981), Dickey (1985), and Taylor (1989) examine the power of the augmented and nonaugmented Dickey-Fuller tests is autoregressive models.

One consistent result is the reduction in test power as the autoregressive root approaches unity. Even with autoregressive root as low as 0.9 Dickey and Fuller (1981) find disturbingly low power for their test, although an intercept is always included in the estimated equations even when excluded from the generating process. Dickey's (1985) results show power is reduced when an irrelevant intercept is included. Although exclusion of the intercept leads to a restrictive alternative (stationary about zero), in applications to regression residuals this is appropriate. Therefore, concerns about low power of the Dickey-Fuller tests based on these articles may be less serious in the context of cointegration testing.

More optimistic results on the power of the ADF test are presented in Taylor (1989). Based on simulations of an ARMA (2,1) model with 150 observations, the null hypothesis of a unit root was rejected in 70% of the cases when the autoregressive root was approximately 0.9 .For a simple AR (1) model without intercept Dickey and Fuller (1981) report rejection frequencies of the mean corrected Dickey-Fuller statistic that are quite sensitive to sample size. With an autoregressive root of 0.9 the null of unit is rejected in 73% of the cases when the sample size is 250 but only 19% of the time with samples of 100 observations. Across these studies the power of tests is sensitive to sample size, the strength of the drift parameter is the data generation process, the inclusion of an irrelevant intercept and the value of the autoregressive root.

Dickey and Fuller (1981) and Dickey (1985) also present some results on test size. Dickey and Fuller find the size of the mean corrected

Dickey-Fuller test differs from its nominal value when the data generation process includes an intercept. However, Dickey concludes the augmented Dickey- Fuller test size does not differ substantially from its nominal value, whether an intercept is included or not. It is possible that these differences stem from the inclusion of a lagged augmentation term in the latter study. In any case sensitivity of test size to an intercept in the estimating equation is not an issue in cointegration testing.

De Jong et al. (1989) find the power of Dickey-Fuller tests to be sensitive to the sample size, the initial value of the process, and the proximity of the root to unity. When the autoregressive root exceeds 0.85, or if the sample size is less than 100, they conclude that the power of the Dickey-Fuller tests are too low to make sharp inferences. For example, with 100 observations and a root greater than 0.9, the maximum frequency of rejections of the unit root hypothesis is 0.35. Arguing that a null of 0.85 is as plausible as null of 1.0, they show that the distributions of test statistics under these alternative nulls overlap to such a large extent that the two hypotheses are empirically indistinguishable.

De Jong, et al. (1990) study both power and size of Dickey-Fuller and Phillips tests under autoregressive and moving average error processes. Consistent with Phillips and Perron (1988), They find serious size distortions of the Phillips tests in the presence of negatively autocorrelated moving average errors. Size distortions are small under positive moving average autocorrelations, but the power of this test is much reduced from the white noise case. With autoregressive errors the phillips test has serious size distortions, but the Dickey-Fuller test does not. Although they are not impressed with the power of the either test, they recommend the augmented Dickey-Fuller test over the Phillips test on the bases of lower size sensitivity and greater power.

These Monte Carlo studies of the unit root lead to a number of consistent conclusions. In general, the Dickey-Fuller tests dominate the Phillips tests in terms of power and minimal size distortions. Second, the Dickey-Fuller tests maintain correct size in the present of moving average errors and departures from homoscedasticity and normality (Godfrey and Tremyne, 1988). Third, the power of the Dickey-Fuller tests tends to be rather low with either small sample size or roots close to unity. Although there is little simulation evidence on the application of these tests to cointegrating regression residuals, it is reasonable to expect similar conclusions in this context.

Engle and Granger (1987) present the results from a small simulation study of seven alternative tests for cointegration. Relevant to this study are their results for the DF, ADF, ECM, and AECM tests. They also study the Durbin-Watson test applied to cointegrating regression residuals. They find this test has good power, but the sensitivity of its distribution to the data generating process eliminates it as a viable test. On the basis of power and sensitivity of critical values to parameters within the null, they recommend the ADF test. They find the ADF to have greater power than DF for higher order processes, the former, for example, displaying a 61% rejection frequency of the unit root hypothesis when the root is 0.9, for a sample size of 100, testing at the 5% level. The inclusion of lagged augmentation terms reduces test power when these are irrelevant.

4. The Structure of the Experiments:

The Monte Carlo experiments are designed to address two primary concerns in statistical hypothesis testing - the invariance of test statistic distributions to statistical environment and the power of the tests. The first issue has been addressed in several of the unit root studies through examination of discrepancies between nominal and empirical size. In the experiments of this study, equivalent information is provided through the examination of variation in critical values from those found for the ideal (random walk) environment. Two factors contributed to the decision to provide information in this form: (i) critical values for these tests must be established empirically, even in the ideal environment, and it is important that these baseline critical values are derived by the same generating mechanism as is employed in the other statistical environments; (ii) these reported critical values can be used by researchers for testing cointegrating in a variety of applied situations.

The first set of Monte Carlo experiments are designed to examine the effect of model misspecification on the critical values under the null hypothesis of noncointegration. Eight different nonstationary processes define the eight statistical environments of these experiments within each environment critical values are computed for three different sample sizes (50,100 and 250 observations) and three conventional significance levels (.10, .05 and.01).

In all experiments the innovations are generated as independent, mutually uncorrelated, pseudo-random standard normal deviates, using the International Mathematical and Statistical Library (IMSL Version 1.1) routine RNNOA. Extensive tests of this algorithm have been conducted by Learmonth and Lewis (1973). The data for each series are generated by creating T+20 observations, discarding the first twenty observations to remove the effect of the initial conditions. Each experiment is replicated 10,000 times to create the sampling distribution for the test statistics.

The specific experimental environments are described in Table 1. The first six environments involve two series; the last two contain three. In environments 1 and 7 each series is an independent random walk without drift. Drift was included in a separate set of experiments and was found to have no substantial effect on test critical values (Alkhatrash, 1990). In environment 2 the series are generated as integrated autoregressive processes, and in environments 3 and 8 they follow integrated moving average models. The series of environment 4 are intergrated of order two. Environments 5 and 6 complicate environments 1 and 3 with heteroscedastic disturbances. Heteroscedasticity is created by dividing the sample into intervals of size 20, with the innovation variance in each interval equal to the variance of the previous interval plus 0.2.

Three key issues must be confronted in the design of the power experiments. Frist, there is the mechanism by which cointegration is introduced into the system. Following Granger (1988), the cointegrated series are generated with a common I (1) factor, Wt, and individual stationary components Xt* and Yt*

$$X_t = \gamma X_t^* + (1-\gamma)W_t$$
$$Y_t = \gamma Y_t^* + (1-\gamma)W_t$$

The paramerer γ (set alternately at 0.2 and 0.8) weights the relative contributions of the individual and common factors.

Second, one must choose from the unlimited number of alternative hypotheses a reasonable range of experimental environments for the study of power. In this paper we present results on the power of the tests, in which the nonstationary common factor follows a random walk process. Variation in statisticl environment is introduced through the processes governing the individual stationary components, which include one white noise model, three first order autoregressive models (with parameters 0.5 , 0.8 and 0.9) and two moving average processes (with parameters -.7 and .7). In the power experiments only the sample size of 100 is considered, and tests employ the 5% critical values.Only bivariate models are included, and individual components of the two series always follow the same process. The six environments for the power experiments are summarized in Table 1.

Third, there is the choice of critical values on which the rejection of the null hypothesis is based. Since the experiments on the critical values demonstrate substantial variation in these values across environments for some tests, this decision isn't trivial. In an attempt to imitate applied testing procedures, we employ the baseline critical values from the ideal (random walk environment). Alternatively, one could apply the critical values specific to each alternative environment, but his requires that the statistical environment in the power experiments matches the generating mechanism in the critical value experiments. Such judicious choice of critical values is not likely in practice, since applied researchers will not in general know the statistical environment governing their data. Consequently one critical value (from the baseline environment) is used for each test in the power experiments.

5. Results of Monte Carlo Experiments

The empirical critical values for the alternative environments at .10, .05 and .01 significance levels are reported in Tables 2 through 4. Of primary interest here is the variability of the critical values over the range of environments. Since El is the random walk environment, in which all tests are correctly specified, the critical values for El are taken as the benchmark values against which all others are compared.

Comparing the alternative specifications of the Dickey- Fuller tests, the Dickey-Fuller test with no lags (DF) shows a sharp increase in its critical value for environment 4 (the I(2) case). Surprisingly, the critical values of the 4-lag and 12-lag specifications of the ADF test do not change sharply in this same environment.

The moderate autoregression introduced in E2 has variable effects on the critical values for the DF test, with critical values either larger or smaller than baseline values depending upon sample size and significance level. The augmented versions (ADF4 and ADF12) perform correctly in this environment, with little change in computed critical values compared to E1.

Moving average errors (E3) affect these tests according to expectations. With no augmentation terms the DF critical values are reduced substantially. The 4-lag ADF test has critical values that are somewhat lower than benchmark values, implying that the 4-lag autoregressive approximation is not sufficient to absorb the moving average autocorrelation. Twelve lags, however, are sufficient, since the critical values for ADF12 are virtually identical in E3 and E1. This result confirms Schwert's (1989) recommendation of long lags in the presence of moving average error processes.

Critical values of the ADF tests are larger than the benchmark values for the three-variable environments (E7 and E8), although less for the 12-lag specification than for the 4-lag version. In environment E7 the zero-lag DF test critical values are also too large, but the presence of moving average errors in E8 reverses this result.

The heteroscedasticity introduced in E5 has little effect on any critical values for any of the Dickey-Fuller tests [check Table 3 T=250], a result consistent with Godfrey and Tremayne's (1988) findings for unit root tests on raw data.

For the tests based on the ECMs, the inclusion of augmentation terms again produces greater stability of critical values across environments. The nonaugmented ECM test has unduly large critical values in environments E4 (with I (2) processes) and E2 (containing autoregressive processes). The AECM critical values are also larger than benchmark values, but discrepancies here are not nearly as large as for the ECM test. Critical values for the AECM test are also too large in the three-variable environments (E7 and E8), but this is not consistently so for the ECM test.

Again the introduction of moving average processes (E3) reduces critical values for both tests at the .01 and .05 significance levels, although the effect for the AECM test is relatively small. The inclusion of augmentation terms in the AECM test is designed to cope with autoregressive processes, and the small number of lagged terms employed here (four) is apparently insufficient to approximate the moving average process.

The PD4 and PD12 tests apply Phillips' correction of the Dickey-Fuller statistics to cointegrating regression residuals. Critical values for these tests show considerable variability across environments, even with the 12-lag adjustment factor. Critical values are too small in E4 (the 1 (2) case) particularly for PD12. Although these statistics are designed to deal with heteroscedasticity, their performance in E5 is not encouraging, with critical values substantially below baseline values in every case. With the additional complication of moving average processes (E6), reductions from benchmark critical values are even more severe. Tests on moving average processes without the complication of heteroscedasticity (E3) have critical values that are unpredictably larger or smaller than baseline values for both PD4 and PD12 specifications.

The stability of the critical values for the alternative test statistics across all environments is summarized by the coefficients of variation presented in Table 5. Separate coefficients are computed for each significance level and sample size, since critical values are not expected to be insensitive to these factors. The clear conclusion from the coefficients of variation is that the ADF tests have the most stable critical values, showing smallest sensitivity to the changing statistical environments. The 12-lag specification produced smaller variability than the 4-lag version, except for the .01 level test with 50 observations.

Ranked second behind the ADF tests by this criterion is the AECM test, with coefficients of variation across sample sizes and significance levels sufficiently small to make this test competitive with the ADF tests. However, the omission of augmentation terms in both the DF and ECM tests substantially increases the variability of these tests, implying excessive sensitivity of these tests to the statistical environment.

The coefficients of variation for the Phillips-Durlauf tests are likewise too large, although the 12-lag specification shows more stability across environments than does the 4-lag version.

The sensitivity of the Phillips-Durlauf, DF, and ECM critical values to statistical environment poses serious problems in applying these tests. If one test with conventionally computed critical values, the true test sizes will differ from nominal sizes. Alternatively, one could apply the critical values presented in Tables 2-4, but this requires knowledge that the statistical environment in the application matches the generating mechanism in the Monte Carlo experiments. The results summarized in Tables 2-5 argue for the preference of AECM and ADF tests over the alternatives, by the criterion of stable critical values.

In the experiments examining power of the tests, the series are cointegrated. Each contains a nonstationary common factor, generated as a random walk, and a stationary individual component, generated according to six alternative processes: a white noise process, three autoregressive processes, and two moving average models. The relative importance of the individual component positively related to the parameter , which is set at values of 0.2 and 0.8. Rejection frequencies are computed by use of the empirical critical values established for each test in the previous experiments.

The results of the power experiments are set out in table 6. Since the critical values of the DF, PP and ECM tests have been shown to be sensitive to the underlying statistical environment, the power results for these tests must be interpreted with caution. For example, knowing the distribution of the DF test shifts to the right in the presence of serial correlation, the achievement of power close to one in these subenvironments is misleading. The incorrect size of this test in the presence of serial correlation eliminates this test as a viable candidate for testing cointegration in such environments.

The results for the ADF tests provide a meaningful indication of performance with alternative lag specifications across these statistical environments. When the common factor is strong (γ=.2) the ADF4 test possesses good power in all cases except those involving strong autoregressive errors. However, its power falls substantially when $\gamma = 0.8$ (to .40), and has very low power (.17) when [rho] is .9. The longer lag version of this test (ADF12) has considerably lower power across all environments, which is a caution against elaborate parameterizations of this test. The greatest power is a mere .37, achieved in the white noise environment. Reducing the relative importance of the common factor severely impairs the power of the ADF tests, although the 4-lag version still rejects a false null over fifty percent of the time in the white noise, weak autoregressive, and negative moving average environments. Power for the 12-lag version is inadequate in all cases.

For the tests based on the error correction model, the effect of the parameter γ is reversed. As the individual component is given more weight, the power of the ECM and AECM tests increase. Since augmentation terms are required for this test to have stable critical values, only the power results for the AECM test are meaningful, and here the results are not promising. When γ = .2, the greatest power is only .03, and with γ = .8, the power of the AECM is lower than that of the ADF4 test in every environment.

The power results for Philips-Durlauf tests are encouraging in most environments, but caution is warranted here because of the sensitivity of critical values to the alternative environments. The effect of γ on power is quite unpredictable, and power holds up surprising well with the additional lags in the correction factor. The power is very low only in the environment with strong individual components that are governed by positive moving average processes ($\gamma = .2$, S5).

6. Conclusion:

The Monte Carlo experiments presented here provide evidence on the stability of critical values of several cointegration tests over alternative statistical environments. These results may be equivalently interpreted as evidence of the degree of disparity between nominal and empirical test size. Two tests, the augmented Dickey-Fuller test and the augmented error correction model test, show reasonable stability of critical values over a wide range of statistical environments. This evidence supports this use of these two tests in applied research, affording confidence that there will be little discrepancy between nominal and empirical test size. The remaining tests investigated - the nonaugmented Dickey- Fuller and error correction model tests and the Phillips- Durlauf test-showed considerable variability in their critical values across the environments, and one would be less confident of a close correspondence between their nominal and actual test sizes in empirical research.

Additional evidence on the performance of these tests when the null hypothesis of noncointegration is false is provided by the power experiments. Again the augmented Dickey-Fuller test shows reasonable power, unless the cointegrated series contain strong stationary components, following autoregressive processes with roots close to one. Inclusion of an excessive number of augmentation terms impairs the power of the test. The power of the augmented error correction model

References:

Al-Khatrash, S. (1990) Residuals Based Tests for Cointegration: A Monte Carlo Investigation, Ph.D. Dissertation.

DeJong, D. et al (1992) "The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors" Journal of Econometrics, Vol. 53, pp 323-43.

DeJong, D.N., J.C. Nankervis, N.E. Savin and C.H. Whiteman, (1989) "Integration versus Trend-Stationarity in Macroeconomic Time Series" University of lowa, Department of Economics Working Paper No. 89-310.

DeJong, D.N., J.C. Nankervis, N.E. Savin, and C.H. Whiteman, (1990) 'The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors", University of lowa, Department of Economics Working Paper No. 90-190.

Dickey, D.A. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root" Journal of the American Statistical Association. 74, 427-431.

Dickey, D.A. and W.A. Fuller (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root" Econometrica 94, 1057-1072.

Dickey, D.A. (1985) "Powers of Unit Root Tests" American Statistical Association Proceedings of the Business and Economic Statistics Section. 489.493.

Engle, R.F., and B.S. Yoo (1987) "Forecasting and Testing in Cointegration Systems" Journal of Econometrics 35, 143-159.

Engle, R.F., and C.W.J. Granger (1987) "Co-integration and Error Correction: Represenation, Estimation and Testing" Econometrica 55, 21-276.

Fuller W.A. (1976) Introduction to Statistical Time Series (New York: John Wiley and Sons.

Godfrey, L.G. and A.R. Tremayne (189) "On the Finite Sample Performance of Tests for Unit Roots", University of York, mimeo.

Granger, C.W.J. (1988) "Some Recent Developments in Concept of Causality" Journal of Econometrics, 39, 199-211.

Granger, C.W.J. and Weiss, A.A. (1983) "Time Series Analysis of Error Correction Models in s. Karlin T. Amemiys and L.A. Goodman (eds) Studies in Econmetrics, Time Series and

test is dominated by that of the augmented Dickey-Fuller test across all environments. Especially when the common nonstationary component of the cointegrated series is important relative to the individual stationary components, the power of the augmented error correction model test is distressingly low.

From both the critical value and the power experiments the evidence favours the augmented Dickey-Fuller test over the alternatives studied here. In implementing this test one must guard against excessive parameterization, possibly by testing the significance of individual lagged augmentation terms. The performance of the augmented Dickey-Fuller test relative to alternative approaches, such as Johansen's (1988) maximum likelihood procedures and Stock and Watson's (1988) common trends analysis, is an important agenda for future research.

Multivariate Statistics, Academic Press, New York.

Hohansen, S. (1992) "Estimation and Hypothesis Testing of Cointegrating Vectors" Journal of Economic Dynamics and Control. Vol. 59 pp 1551-80.

Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors" Journal of Economic Dynamics and Control, 12.

Learmonth, G.P. and P.A. Lewis 91973) "Statistical Tests of Some Widely Used and Recently Proposed Uniform Random Number Generators" in W.F. Kennedy (ed.) Computer Science and Statistics: 7th Anual Symposium on the Interface. Ames, Liwa: Iowa State University.

Phillips, P. (1991) "Optimal Inference in Cointegrated Systems" Econometrica, Vol 59, pp 283-307.

Phillips, P. 91994) "Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating Coefficients in Error Correction Models" Econometrica, Vol 62, pp 73-93.

Phillips, P.C.B., and P. Perron (1988) "Testing for a Unit Root in Time Series Regressions", Biometrika 75, 335-346.

Phillips, P.C.B. and S.N. Durauf (1987) "Multiple Time Series Regression with Integrated Variables". Review of Economic Studies 53, 473-496.

Phillips, P.C.B. (1987) "Time Series Regression with a Unit Root". Econometrica 55, 599-607.

Said, S.E. and D.A. Dickey (1984) "Testing for Unit Roots in Autoregressive Moving Moving - Average Models with Unknown Order " Biometrica 71, 599-607.

Schwert, G.W. (1989) "Tests for Unit Roots: A Monte Carlo Investigation" Journal of Business and Economic Statistics, 7 147-160.

Stock, J. and Watson, M. (1988) "Testing for Common Trends" Journal of the American Statistical Association, 83, 1097-1107.

Stock, J. and Watson, M. (1993) "A Simple MLE of Cointegrating Vectors in Higher Order Integrated Systems" Econometrica Vol. 61, pp 782-820.

Taylor, M.P. (1989) "On Unit Roots and Real Exchange Rates: Empirical Evidence and Monte Carlo Analysis" Bank of England, Mimeo.

Table 1

Summary Of The Environments And Subenvironments

Of The Experiments

the environments

E1	$X_t = X_{t-1} + e_{x,t}$	arima (0,1,0)
E2	$X_t = X_{t-1} + 0.7(X_{t-1} - X_{t-2}) + e_{xt}$	arima (1,1,0)
E3	$X_t = X_{t-1} + e_{x,t} + 0.7e_{xt}$	arima (0,1,1)
E4	$X_t = 2X_{t-1} - X_{t-2} + e_{x,t}$	arima (0,2,0)
E5	As E1 with heteroscedastic disturbances	
E6	As E3 with heteroscedastic disturbances	
E7	As E1 with 3 variables (X,Y,Z)	
E8	As E3 with 3 variables (X,Y,Z)	

Hall Environments

MA (1)	$\phi = \pm 0.7$
IMA (1,1)	$\phi = \pm 0.7$

The Sub Environments

sub 1	Both X* and Y* are white noise
sub 2	$X_t^* = 0.9 X_{t-1}^* + e_{x^*,t}$
sub 3	$X_t^* = 0.8 X_{t-1}^* + e_{x^*,t}$
sub 4	$X_t^* = 0.5 X_{t-1}^* + e_{x^*,t}$
sub 5	$X_t^* = e_{x^*,t} + 0.7 e_{x^*,t-1}$
sub 6	$X_t^* = e_{x^*,t} - 0.7 e_{x^*,t-1}$

Table 2 : Empirical Critical Values for Test of Noncointegration, p=0.10

Test Statistics

	DF	ADF_4	ADF_{12}	ECM	AECM	PD_4	PD_{12}
T=50							
E1	3.13	2.88	2.35	11.33	9.35	3.87	3.04
E2	2.74	2.90	2.42	17.59	10.04	3.29	2.68
E3	2.49	2.73	2.35	10.28	9.24	3.72	2.44
E4	3.45	3.01	2.69	19.20	10.57	2.34	1.96
E5	3.11	2.85	2.35	10.97	9.15	3.20	2.89
E6	2.44	2.70	2.35	9.86	9.11	2.49	2.31
E7	3.60	3.17	2.48	11.12	11.01	4.29	3.21
E8	2.79	3.04	2.50	10.07	10.77	4.19	2.44
T=100							
E1	3.10	2.96	2.68	11.07	9.93	3.87	3.50
E2	3.00	2.97	2.72	17.04	10.28	2.91	2.75
E3	2.47	2.80	2.68	10.04	9.57	4.05	3.01
E4	4.79	3.07	2.89	29.95	10.84	2.51	1.85
E5	3.04	2.92	2.66	10.71	9.69	3.09	3.03
E6	2.37	2.77	2.65	9.56	9.37	2.29	2.32
E7	3.51	3.32	2.93	11.02	11.67	4.53	3.74
E8	2.71	3.16	2.94	9.90	11.30	4.89	3.13
T=250							
E1	2.7	3.00	2.89	10.90	10.42	3.80	3.90
E2	3.05	3.01	2.89	16.77	10.55	2.61	2.57
E3	2.44	2.87	2.89	2.95	10.04	4.56	3.95
E4	7.63	3.11	3.01	64.78	11.38	3.37	2.21
E5	2.12	2.94	2.84	10.59	10.10	3.00	2.99
E6	2.12	2.79	2.83	9.21	9.62	2.01	2.07
E7	3.49	3.39	3.23	10.88	12.16	4.35	4.51
E8	2.70	3.22	3.22	9.78	11.511	6.05	4.55

Table 3 : Empirical Critical Values for Test of Noncointegration, p=0.05

	DF	ADF_4	ADF_{12}	ECM	AECM	PD_4	PD_{12}
T=50							
E1	3.45	3.19	2.63	13.72	11.20	4.35	3.49
E2	3.35	3.22	2.71	24.27	12.01	4.03	3.19
E3	2.78	3.03	2.63	13.12	11.03	4.29	2.74
E4	4.20	3.31	2.99	31.40	12.63	2.84	2.42
E5	3.42	3.15	2.62	13.23	11.07	3.55	3.20
E6	2.74	2.98	2.61	12.45	10.88	2.81	2.55
E7	3.95	3.48	2.77	13.53	13.33	4.74	3.74
E8	3.07	3.33	2.81	12.70	13.02	4.81	2.67
T=100							
E1	3.41	3.26	2.97	13.17	11.86	4.55	3.89
E2	3.31	3.28	2.99	23.38	12.25	3.73	3.34
E3	2.79	3.11	2.95	12.77	11.41	4.96	3.33
E4	5.97	3.37	3.19	48.24	12.71	3.07	2.25
E5	3.35	3.20	2.93	12.79	11.60	3.40	3.31
E6	2.65	3.06	2.94	12.16	11.20	2.59	2.59
E7	3.82	3.62	3.21	13.02	13.80	5.17	4.16
E8	2.99	3.44	3.21	12.50	13.54	6.00	3.45
T=250							
E1	2.96	3.32	3.18	13.00	12.41	4.31	4.62
E2	3.37	3.31	3.16	22.63	12.56	3.31	3.28
E3	2.74	3.14	3.16	12.24	11.95	5.89	4.74
E4	9.49	3.39	3.30	104.15	13.26	4.28	2.81
E5	2.24	3.25	3.13	12.75	12.19	3.30	3.30
E6	2.24	3.09	3.10	11.62	11.62	2.26	2.35
E7	3.81	3.69	3.50	12.95	14.33	5.11	5.13
E8	2.96	3.50	3.48	12.41	13.58	8.17	5.38

Table 4 : Empirical Critical Values for Test of Noncointegration, p=0.01

Test Statistics

	DF	ADF_4	ADF_{12}	ECM	AECM	PD_4	PD_{12}
T=50							
E1	4.13	3.77	3.23	18.59	16.43	5.64	5.06
E2	4.69	3.82	3.39	43.23	16.74	5.89	4.49
E3	3.42	3.60	3.21	20.07	15.31	5.73	3.37
E4	5.51	3.95	3.68	85.46	17.23	3.90	3.64
E5	4.12	3.72	3.24	18.43	15.24	4.18	4.00
E6	3.35	3.56	3.22	19.42	15.12	3.39	3.12
E7	4.62	4.07	3.37	19.02	18.59	5.94	5.92
E8	3.76	3.85	3.47	19.89	18.48	6.45	3.26
T=100							
E1	4.01	3.83	3.49	17.75	16.15	6.37	4.81
E2	3.90	3.84	3.53	39.24	16.52	5.71	4.61
E3	3.32	3.63	3.50	18.85	15.94	7.91	4.08
E4	8.51	3.95	3.80	127.83	17.23	4.18	3.02
E5	3.93	3.77	3.45	17.61	15.66	4.03	3.87
E6	3.12	3.58	3.45	18.53	15.71	3.11	3.04
E7	4.40	4.16	3.77	17.59	19.05	6.80	5.36
E8	3.53	3.97	3.74	18.65	18.22	9.73	4.22
T=250							
E1	3.50	3.82	3.68	17.08	16.47	5.91	6.48
E2	3.93	3.84	3.68	37.72	16.93	5.06	4.90
E3	3.31	3.63	3.79	18.36	16.08	10.94	6.90
E4	13.55	3.95	3.90	246.97	17.21	6.10	3.93
E5	2.39	3.78	3.65	17.16	16.53	3.87	3.87
E6	2.40	3.55	3.63	18.61	15.71	2.75	2.87
E7	4.36	4.22	4.06	17.37	18.66	7.29	7.15
E8	3.50	4.01	4.04	17.90	17.99	18.87	8.06

Table 5
Coefficient of Variation of the Critical Values

Test 1 : Dickey Fuller

Sig. Lef.	sample sze 50	100	250
0.1	14.34	24.55	55.35
0.05	15.34	29.75	64.09
0.01	17.24	39.95	79.54

Test 2 : Augmented Dickey Fuller

Lag = 4			
0.1	5.44	6.06	6.38
0.05	5.09	5.53	5.80
0.01	4.49	4.90	5.53
Lag = 8			
0.1	4.26	4.99	5.69
0.05	4.17	4.71	5.29
0.01	3.89	4.06	4.89
Lag = 12			
0.1	4.92	4.61	5.49
0.05	4.85	4.25	4.88
0.01	4.89	4.21	4.61

Test 3 : Error Correction :

0.1	29.22	51.27	106.99
0.05	42.15	68	127.26
0.01	77.81	111.36	164.31

Test 4 : Augmented Error Correction Model

0.1	7.99	8.24	8.12
0.05	8.29	7.92	7.16
0.01	8.84	7.45	5.8

Test 5 : Phillips

Lag = 4

0.1	21.35	27.16	34.45
0.05	2.03	28	40.17
0.01	22.11	36.83	67.79

Lag = 8

0.1	15.78	20.93	30.26
0.05	15.83	18.94	29.08
0.01	24	19.87	33.96

Table 6
Emperical Power For Tests
Of Noncointegration At 59% Level 100 Observations :
Environment 1, Random Walk

Test Statistics

	DF	ADF_4	ADF_{12}	ECM	AECM	PD_4	PD_{12}
γ=0.2							
S1	1.00	0.98	0.37	0.71	0.01	0.49	0.25
S2	0.21	0.17	0.12	0.05	0.03	.042	0.41
S3	0.70	0.40	0.19	0.14	0.03	0.54	0.31
S4	1.00	0.86	0.30	0.41	0.01	0.54	0.20
S5	1.00	1.00	0.33	0.95	0.03	0.00	0.01
S6	1.00	0.91	0.33	0.45	0.01	0.42	0.05
γ=0.8							
S1	1.00	0.52	0.04	1.00	0.29	0.99	0.79
S2	0.17	0.12	0.08	0.13	0.09	0.18	0.14
S3	0.56	0.26	0.10	0.44	0.18	0.28	0.29
S4	0.99	0.54	0.09	0.99	0.37	0.76	0.77
S5	1.00	0.18	0.01	1.00	0.11	0.99	0.73
S6	1.00	0.53	0.07	0.99	0.33	0.25	0.61

Call For Paper

United Nations Economic And Social Commission For Western Asia
Centre For Arab Gulf Studies, University Of Exeter

Human Development In The Arab Gulf International Conference ,
September 23-25, 1997
Exeter, Uk.

The Arab Gulf Countries present some of the most interisting insights into the complexity of the relationship between economic growth on the one hand , and human development on the other .

Papers are invited to address the following themes;

1- Sustainability or lack of it;
This theme would cover issues such as institutional capability, education and skills for the labour market and for effective service provision, the economic environment and HD,ecology and sustsinbility,etc.

2-Disparities in Human Development ; this will deal with two sets of issues,
(a) Inequality between countries and regions within a country inequalities between social groups in terms of HD.this will deal with the relationship between social and economic inequalities ; defferent access to services ; and disparities in conditions conducive to human development
(b) Incongruities in social and economic aspects of HD such as good health and social services, but a poor public health and safety health and safety (includingwork safety) record and awareness; expansion in women's education, but not in employment and social roles; high incomes and low economic security; strong local and community institutions and lack of social cohesion at national levels; problems of political participation and lack of individual choice; high material standards, low cultural and technological ones; high output, low self-sufficiency in essentials.

3- Social and cultural aspects of hd :

are HD critical universal. or ought different societies emphasise different sets of criteria? are achievements sequential or simulataneous ? how do paths differ in different societies (reference to arab experiences)? what institutional differences can be identified as being especially conducive to HD?

4- HD and development policy :

Should policy be designed to achieve quantitatively measured targets? or should HD be promoted through built-in policy criteria, monitoring, awareness, some other mechanism , or a mixture of these ? if quantitative targets are appropriate, then what targets, what insturments, and how should development policy be integrated? what institutions if any , can best advance HD? what status should any such institutions have, and should public paricipation be built into them?

Suggestions for other themes, including studies of specific aspects of HD, and case studies are welcome. abstracts (preferably by end of June) to

Mrs Jennifer Davies (HD conference)
Centre For Arab Gulf Studies
Old Library
University Of Exeter
Exeter ex4 4jz, UK

tel.: (01392)264024
fax.:(01392)264023
e-mail:
<j.m.davies@exeter.ac.uk>

You may also contact : a.akil, chief of social development issues and policies division, ESCWA,P.O.927115, amman, jordan. fax : (962-6) 694981
or kamil mahdi , director of centre for arab gulf studies, university of exeter

The Centre for Arab Gulf Studies will provide facilities and accommodation, while respective institutions are invited to cover travel costs wherever possible. a limited travel fund is also available (inquiries to the Centre for Arab Gulf studies).