MIXED SIGNALS AMONG TESTS FOR COINTEGRATION

ALLAN W. GREGORY*, ALFRED A. HAUG AND NICOLETTA LOMUTO

Department of Economics, Queen's University, Kingston, Ontario, Canada K7L 3N6
Department of Economics, York University, Toronto, Ontario, Canada M3J 1P3
DATACORP, Providence, Rhode Island, USA

SUMMARY

This paper illustrates that, under the null hypothesis of no cointegration, the correlation of p–values from a single–equation residual–based test (*i.e.* ADF or $\widehat{Z}_\alpha$) with a system–based test (trace or maximum eigenvalue) is very low even as the sample size gets large. With data generating processes under the null or 'near' it, the two types of tests can yield virtually any combination of p–values regardless of sample size. As a practical matter, we also conduct tests for cointegration on 132 data sets from 34 studies appearing in this *Journal* and find substantial differences in p–values for the same data set.

* Correspondence to: Gregory, Department of Economics, Queen's University, Kingston, Ontario, Canada K7L 3N6; Phone: (613) 533-2299; Fax: (613) 533-6668; Email: awg@qed.econ.queensu.ca.

1. INTRODUCTION

The literature has suggested numerous tests for cointegration. Broadly speaking these tests can be classified either as single equation residual–based tests or as system–based tests. Practitioners have applied from the residual–based category mostly the augmented Dickey-Fuller (ADF) test suggested by Engle and Granger (1987) and the $\widehat{Z}_\alpha$ test of Phillips and Ouliaris (1990). From the system–based category, the most commonly used tests are the trace and maximum eigenvalue tests of Johansen (1988, 1991, and 1995) and Reinsel and Ahn (1992), based on the maximum likelihood principle. [Other tests suggested for cointegration can be found, for example, in Stock and Watson (1988), Phillips and Ouliaris (1990), Hansen (1992), Banerjee et al. (1998), Saikonnen and Lütkepohl (2000), and also in Zivot (2000).] The Monte Carlo performance of these tests, along with other tests, has been studied by, among others, Gregory (1994), Toda (1995), and Haug (1996) in terms of test size and test power for finite sample sizes typically used in applied work. No one test was found to dominate. In terms of empirical research, the choice of which test to apply is often dictated by the nature of the investigation. For instance, if interest is concentrated on a particular relation or variable, then single equation residual-based tests are likely used. On the other hand, for truly multivariate settings, then a systems approach is likely to be adopted. In fact, many applied researchers have reported results for two (or more) cointegration tests in order to gain more information regarding the null hypothesis.

For those occasions where more than one test is calculated, one might expect that, as the sample size increases, instances in which one test rejects and another retains would would become less frequent. In this sense the inferences from various tests for cointegration, at least at large samples, would be reinforcing. In this paper we explore testing results under the null hypothesis of no cointegration and under the alternative of cointegration. First, we carry out a Monte Carlo study and calculate (pairwise) for every sample of artificial data the p-values for the ADF, $\widehat{Z}_\alpha$, trace and maximum eigenvalue tests based on the corresponding empirical distribution. For every sample we generate 50,000 replications. Second, we apply the same tests under a standardized methodology to 132 different data

sets from 34 empirical studies that use cointegration tests and are published in the *Journal of Applied Econometrics*. One reason for choosing studies in this journal is that the data sets are readily available on the web from the journal's data archive. We calculate again p-values for every sample of actual data but this time from the approximate asymptotic distributions. For both the artificial and non-artificial data, we record pairwise for every sample when both tests reject the null hypothesis and also when only one test rejects the null. For a given set of simulation replications and for the actual data sets, we aggregate for each possible pair of tests the frequency of both tests rejecting, the frequency of just one test rejecting, and the correlations of p-values for each pair.

In the simulation study, we demonstrate that under the null hypothesis of no cointegration, the correlation of the p–values from the empirical distribution of a residual–based test with those from the empirical distribution of a system–based test is very low. This low correlation is not due to size distortions since we use p-values from the empirical distribution and the sample size we use is very large. Nor is the low correlation an artifact of the experimental design since the simulations are carried out under the null distribution. The implication of this finding for empirical work is that researchers who calculate two tests under the null hypothesis (one from each group) may obtain any combination of p–values, so that substantial differences can persist no matter what the sample size. This situation is not resolved asymptotically.

The use of experimental data allows us to keep the data generating process (DGP) the same and simple, as we move from one independent replication to another in the Monte Carlo simulations. With non-experimental data the true DGP is, of course, not known and is likely to change as we move from one data set to another. Also, test statistics will be affected by lag selection methods, bandwidth choices, common factor restrictions, and size correction methods. These issues have been documented, for example, by Reimers (1992), Kremers at al. (1992), Toda and Phillips (1994), Ho and Sorensen (1996), Kim (1998), and Chao and Phillips (1999). We adopt relatively standard techniques used in the literature to account for such sensitivities and find confirming results of our experimental data. Indeed for reasons discussed below, differences in testing results are even more prevalent with the

non-experimental data.

This paper is organized as follows. In section 2 we motivate the issue in a quite simple but artificial example. We present a Monte Carlo experiment to illustrate test differences both under the null of no cointegration and under an alternative. In section 3 we briefly summarize our calculations for actual data from studies appearing in the *Journal of Applied Econometrics* and in section 4 we offer some concluding remarks.

2. SIMULATION RESULTS UNDER THE NULL AND ALTERNATIVE HYPOTHESIS

Before outlining our Monte Carlo design, we provide a motivating example suggested by an anonymous referee. Imagine a situation where two researchers with the same set of data wish to test whether $y_t = \epsilon_t$ with $\epsilon_t \sim NID(0, 1)$. One researcher, wishes to look for departures from the null with the variable (deterministic) x_{1t} and the other researchers uses variable x_{2t}. Under the alternative, we have:

$$y_t = \beta_1 x_{1t} + \epsilon_t \quad \text{for researcher 1}$$

$$y_t = \beta_2 x_{2t} + \epsilon_t \quad \text{for researcher 2.}$$

In this example, while the researchers test the same null hypothesis, the implicit alternative is different. The usual test statistic for testing the common null is

$$Q_1 = \frac{\sum y_t x_{1t}}{\sqrt{\sum x_{1t}^2}} \quad \text{for researcher 1}$$

$$Q_2 = \frac{\sum y_t x_{2t}}{\sqrt{\sum x_{2t}^2}} \quad \text{for researcher 2.}$$

The extent to which the two tests give the same signal regarding the common null hypothesis depends on the correlation of x_1 and x_2:

$$\rho(Q_1, Q_2) = \frac{\sum x_{1t} x_{2t}}{\sqrt{\sum x_{1t}^2 \sum x_{2t}^2}}.$$

In testing for cointegration, there is the same null hypothesis (no cointegration) but the form of the implicit alternative is different. Although we do not have analytical expressions for the correlation of the various tests for cointegration, we propose to investigate this correlation by means of Monte Carlo methods.

We present simulation results from a simple experiment both under the null hypothesis of no cointegration and under the alternative hypothesis of cointegration. The reported simulations are only a small subset of the ones we have carried out and, to the best of our knowledge, the conclusions for the larger sample sizes are representative. Our simulations under the null hypothesis have included cases with no constant, constant, trend, trend–squared, and up to 5 exogenous regressors. The results we report for small samples when the alternative is true are sensitive to the choice of the data generating processes and are given only to help interpret our non-experimental findings.

The data generating process for the reported results is:

$$y_t = \theta x_t + \epsilon_t,$$

$$\epsilon_t = \rho \epsilon_{t-1} + \omega_t$$

$$x_t = x_{t-1} + \eta_t.$$

Under the null hypothesis of no cointegration, $\theta = 0$ and $\rho = 1.0$, and under the alternative hypothesis, $\theta = 1.0$ and $\rho = 0.98$. For both hypotheses, ϵ_t and η_t are uncorrelated NID(0,1) processes. The sample sizes are $T = 250, 500, 1000$, and 2000, and the number of replications, R, is 50000. Two tests from each group are calculated: the ADF and $\widehat{Z}_\alpha$ test from the single–equation residual–based group, and the maximum eigenvalue (MAX) and trace (TR) test from the system–based group. This allows for six possible pairwise comparisons. In the reported cases, constants are included for the two residual–based tests, and a constant is included in the cointegrating vector only, for the two system–based tests (see Johansen, 1991).

To limit size distortions, we chose relatively large sample sizes, and for all sample sizes, critical and p–values are from the empirical distributions. The p–values for any individual test are calculated by taking the rank order across the R replications and dividing by R. Therefore, regardless of sample size, each test individually rejects a true null hypothesis with probability equal to the p–value. At large sample sizes, these critical values are also extremely close to the nominal values. We then compare the p–values across the six combinations of tests for cointegration for the same replication. Lastly, since the focus is principally on the null distribution, our data generating process is one in which there are

no nuisance parameter problems so that we do not confound the issue of test differences with questions of how to choose lag lengths or bandwidths.

A striking illustration of the potential for mixed testing signals under the null hypothesis of no cointegration is provided in Figure 1. The six possible pairwise combinations of p–values for the ADF, $\widehat{Z}_\alpha$, MAX, and TR tests at $T = 2000$ are graphed. When tests from the same class are graphed, the combinations cluster around the 45° degree line, indicating a close correspondence. However, mixed pairs produce pictures that look like clouds, showing that all combinations are possible. To the extent that one uses the magnitude of the p–value as a measure of the strength of the rejection, these results imply that substantial testing differences will be commonplace when the null is true. Although there appears to be some minor testing discrepancies for very large samples for tests within the same group, this is in sharp contrast to the pronounced differences found in cross–over combinations.

In Table 1 we present additional evidence at conventional significance levels. The first entry in Table 1 shows the proportion of replications where both tests reject (individually) the null hypothesis of no cointegration at the 1%, 5%, and 10% significance levels. The second entry shows the percentage of replications where one and only one of the two tests rejects. Summing these entries and subtracting the total from one gives the proportion of replications in which both tests do not reject the null at the indicated level of significance. At the bottom of the table, the correlations of the p–values are reported.

For the pairs coming from the same group of tests, the high correlation (over 0.9) implies that the joint probability of rejecting the null is slightly lower than what the individual test sizes would predict. For instance, using an individual 5% level ($T = 2000$), the $ADF/\widehat{Z}_\alpha$ and the MAX/TR jointly reject the null 4% and 3% of the time respectively. In contrast, for the same case, the mixed pairs have a low p–value correlation (0.25 and less) and the joint rejection frequencies are just 1% for all possible cross–over combinations. The proportion of rejections for just one of the tests is between 2%-3% for tests from the same group, compared to a 7%-8% range for cross–over combinations. Despite the low correlations under the null hypothesis, test statistics for the mixed pairs are 'far' from

independent. When the number of regressors increases under the null hypothesis, the correlation of the p–values falls for the TR and MAX tests (not reported in the Tables). For instance, in a system with 5 variables this correlation for $T = 2000$ is 0.80. On the other hand, there is no decline in the correlation of p–values for the single–equation residual based tests for the larger systems.

The results under the particular alternative hypothesis in Table 2 ($\rho = .98$) display for smaller sample sizes a similar pattern to those of the simulations under the null. The initial low power (since this alternative is 'close' to the null) implies a low rejection frequency for all tests. However, for sample sizes $T \leq 500$, the proportion of replications in which only one test rejects for the mixed pairs always exceeds the proportion in which both tests reject. The proportion of cases for just one rejection for mixed pairs can be substantial. Once the sample size gets large, there are fewer departures as all tests reject the null more frequently. The paths to consensus for the tests are not monotonic. With one or two exceptions, as the number of observations rises not only does the joint rejection frequency rise but for sample size $T \leq 1000$, the proportion of replications where only one test rejects also rises. Again differences of this sort are more pronounced for tests from different groups. Once $T = 2000$, the proportion of mixed testing signals is small for the 5% and 10% level tests. Of course in finite samples, a different alternative will lead to different rejection frequencies.

3. EVIDENCE FROM APPLIED STUDIES

A natural question arises as to the frequency and severity of testing differences in practice. We compare the results of conducting the ADF, $\widehat{Z}_\alpha$, trace and maximum eigenvalue tests using 132 data sets from 34 studies published in the *Journal of Applied Econometrics*. This provides a pairwise sample size of $N = 132$. The analysis examines the behavior of the individual test statistics, the frequency of test differences at conventional testing levels and the relationship between the probability values of the tests. Our goal is to document the testing relationship with actual data and not to suggest that the authors of these papers have been in any way strategic in their choice of which test for cointegration to apply. As

6

we have argued before, we think most applied researchers tend to view the different tests as rather interchangeable with the choice more dependent on the nature of the investigation.

The data are from studies published in the *Journal of Applied Econometrics* from 1994 to March/April 2001 dealing directly with testing for cointegration. Other studies analyzed non-stationary data but covered topics other than cointegration and were excluded from the analysis. Since few researchers publish studies on the lack of cointegration, a large number of the tests reject the null hypothesis of no cointegration at conventional levels. As we have seen in the Monte Carlo power experiment, this leads to more mixed signals, since tests that agree on retaining the null hypothesis of no cointegration are not generally present. The frequency of the testing differences in actual applied work (as opposed to published applied work) may therefore be overstated.

We briefly summarize our main findings but direct the interested reader to follow the link at http://qed.econ.queensu.ca/pub/faculty/gregory/ where there is a complete list of the papers, the kinds of data used and a full report of our results. In the original published studies, the methods employed to specify and test the models vary. We impose a unifying but standard methodology so that the results from the different studies are comparable. However, there are important differences across the studies such as sample size which ranges from 27 to 7,693 with a median sample size of 73.5. For the residual-based tests where a dependent variable is required, we follow the choice in the original paper. However, for those investigations where there is no obvious dependent variable, we chose it on the basis of the highest R^2. We did some sensitivity analysis on this approach and it does not seem that our results in any way hinge on this.

While this applied exercise is related and motivated by our simulation experiments in the previous section, the interpretation of our aggregation is quite different. In the simulation exercise, we know the DGP and hold it constant over independent replications. In actual data, we never know what the true DGP is and more importantly it is different across the studies. Finally, since many papers study similar phenomena our aggregation is dependent, precluding simple binomial calculations for standard errors.

Unlike the simulations, tests of cointegration on actual data involve nuisance parame-

7

ters and so choices of lag lengths and window widths arise. There are a number of different methods suggested in the literature for doing this and we have chosen fairly standard ones. Although individual test results may be sensitive to these choices, the overall conclusions would not be much different if alternative selection methods were employed. For the ADF test, lag length is determined from a variation of the BIC criterion suggested by Kim (1998). For the $\widehat{Z}_{\alpha}$ test, we use the Bartlett kernel with an automatic lag length outlined in Newey and West (1994) with no pre-whitening. For the maximum eigenvalue and trace tests, lags for the vector autoregression are selected using a BIC criterion. We also follow Reimers (1992) and Boswijk and Franses (1992) and apply finite sample correction for the number of regressors for the trace and maximum eigenvalue test.

The probability values (from the approximate asymptotic distributions) for the ADF and $\widehat{Z}_{\alpha}$, maximum eigenvalue and trace tests are from MacKinnon (1996) and MacKinnon et al. (1999) respectively. The programs are available at http://qed.econ. queensu.ca/pub/ faculty/mackinnon. Each equation is tested with a constant and a constant and trend but only the constant results are shown here.

Table 3 summarizes the findings for this empirical exercise. First, we note that the results from the empirical studies are, on balance, characterized by a rejection of the null of no cointegration and so are more likely to correspond to our Monte Carlo power findings. Just as in our simulation exercise, there are more mixed testing signals across single equation residual-based tests and system tests than within. For all levels of the test, differences in test signals occurred over 50% of the time for the system and single equation residual-based comparisons. The two system tests were highly correlated (.94)and have high power in terms of the approximate p-values which lead to very similar test conclusions at conventional levels. On the other side, the single equation residual-based test correlations (.69) are much lower than in the Monte Carlo experiments (.93 to .95) with lower power than either of the system tests and so test discrepancies are somewhat more prevalent.

4. CONCLUSION

A researcher calculating one test from the residual–based approach and one from the system–based approach may find dramatic test differences even for large sample sizes. One explanation for this is that under the null of no cointegration the p–values can be vastly different even asymptotically. Such differences are also likely to arise in data generating processes that are 'close' to the null hypothesis since low correlation can occur for modest sized samples. Of course, since both groups of tests are consistent, the rejection frequency will rise as the sample size increases when the null is false. While it seems that the joint distribution of any two tests are free of nuisance parameters (so that it would be possible in principle to approximate appropriate critical values from the joint limiting distribution), our Monte Carlo evidence has indicated that such a joint test can have less power than using either of the tests individually. This is the subject of future research.

From the practical perspective, we think applied researchers who might use these tests interchangeably, will find the differences in inferences both in our simulations and actual data somewhat unsettling. Our impression is that many theoretical econometricians have reached more definitive (although not the same) conclusions regarding the choice of test for cointegration. This study clearly demonstrates how important the choice is for the test conclusions in applied work and points to the need for further theoretical work on this issue.

Finally, the lesson from our simple linear regression example shows that pairs of tests with power against different alternatives can, under the null hypothesis display varying degrees of correlation. While we do not have the same formal results for the tests for cointegration, it is quite apparent that "mixed" signals arise from the power differences against different alternatives under which there is no cointegration. Of course, these kinds of issues occur with other tests as well. The major difference is that we understand the interaction between power and test correlation in the regression example but do not for tests for cointegration.

ACKNOWLEDGEMENTS

REFERENCES

Banerjee A, Dolado JJ, Mestre R. 1998. Error–correction mechanism tests for cointegration in a single–equation framework. *Journal of Time Series Analysis* **19**: 267–283.

Boswijk P, Franses PH. 1992. Dynamic specification and cointegration. *Oxford Bulletin of Economics and Statistics* **54**: 369–381.

Chao JC, Phillips PCB. 1999. Model selection in partially nonstationary vector autoregressive processes with reduced rank. *Journal of Econometrics* **91**: 227–271.

Engle RF, Granger CW. 1987. Cointegration and error correction: representation, estimation and testing. *Econometrica* **55**: 251–276.

Gregory AW. 1994. Testing for cointegration in linear quadratic models. *Journal of Business and Economic Statistics* **4**: 347–360.

Hansen BE. 1992. Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* **10**: 321–335.

Haug AA. 1996. Tests for cointegration: a Monte Carlo comparison. *Journal of Econometrics* **71**: 89–115.

Ho MS, Sorensen BE. 1996. Finding cointegration rank in high dimension systems using the Johansen test. An illustration using data based Monte Carlo simulations. *Review of Economics and Statistics* **78**: 7726–732

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

Johansen S. 1991. Estimation and hypothesis testing of cointegrating vectors in the presence of a linear trend. *Econometrica* **59**: 1551–1580.

Johansen S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**: 231–254.

Kim JY. 1998. Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models. *Econometrica* **66**: 359-80.

Kremers JJM, Ericsson NR, Dolado JJ. 1992. The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* **54**: 325–348.

MacKinnon JG. 1996. Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics* **11**: 601–618.

MacKinnon JG, Haug AA, Michelis L. 1999. Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics* **14**: 563–577.

Newey WK, West KD. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* **61**: 631–654.

Phillips PCB, Ouliaris S. 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica* **58**: 165–193.

Reimers HE. 1992. Comparisons of tests for multivariate cointegration. *Statistical Papers* **33**: 335–359.

Reinsel GC, Ahn SK. 1992. Vector AR models with unit roots and reduced rank structure. *Journal of Time Series Analysis* **13**: 353–375.

Saikonnen P, Lütkepohl H. 2000. Testing for cointegration rank of a VAR process with an intercept. *Econometric Theory* **16**: 373–406.

Stock JH, Watson MW. 1988. Testing for common trends. *Journal of the American Statistical Association* **83**: 1097–1107.

Toda HY. 1995. Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions. *Econometric Theory* **11**: 1015–1032.

Toda HY, Phillips PCB. 1994. Vector autoregression and causality: a theoretical overview and simulation study. *Econometric Reviews* **13**: 25–285.

Zivot E. 2000. The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory* **16**: 407–439.

Figure 1

Table 1. Rejection Frequencies: Under the Null Hypothesis

	T	$ADF/\widehat{Z}_\alpha$	ADF/MAX	ADF/TR	$\widehat{Z}_\alpha/MAX$	$\widehat{Z}_\alpha/TR$	MAX/TR
Level							
.01							
	250	.006/.007	.002/.02	.002/.02	.002/.02	.002/.02	.006/.008
	500	.007/.007	.002/.02	.002/.02	.002/.02	.002/.02	.006/.007
	1000	.007/.006	.001/.02	.002/.02	.002/.02	.002/.02	.006/.008
	2000	.007/.006	.002/.02	.001/.02	.002/.02	.002/.02	.006/.008
.05							
	250	.04/.03	.01/.08	.01/.08	.01/.07	.01/.08	.03/.03
	500	.04/.04	.01/.08	.01/.08	.01/.07	.01/.07	.03/.03
	1000	.04/.02	.01/.08	.01/.08	.01/.07	.01/.07	.03/.03
	2000	.04/.02	.01/.08	.01/.08	.01/.07	.01/.07	.03/.03
.10							
	250	.08/.05	.03/.14	.03/.14	.03/.14	.03/.14	.07/.06
	500	.08/.05	.03/.15	.03/.15	.03/.14	.03/.14	.07/.05
	1000	.08/.04	.03/.15	.03/.15	.03/.14	.03/.14	.07/.05
	2000	.08/.04	.03/.14	.03/.15	.03/.13	.03/.14	.07/.05
Corr							
	250	.93	.16	.14	.24	.23	.93
	500	.94	.16	.14	.25	.24	.93
	1000	.95	.16	.14	.25	.23	.93
	2000	.95	.16	.14	.25	.24	.93

Note: These results are based on 50,000 replications under the null hypothesis of no cointegration. *Level* refers to the significance level for the individual tests. The first entry is the proportion of these replications in which both tests reject the null hypothesis at the indicated level. The second entry gives the proportion of replications where only one of the tests rejects at the indicated level. The critical values for the individual tests are size-adjusted using the corresponding sample size. *Corr* is the correlation coefficient of the p–values obtained under the null distribution.

Table 2. Rejection Frequencies: Under the Alternative Hypothesis

	T	$ADF/\widehat{Z}_\alpha$	ADF/MAX	ADF/TR	$\widehat{Z}_\alpha/MAX$	$\widehat{Z}_\alpha/TR$	MAX/TR
Level							
.01							
	250	.02/.02	.005/.02	.004/.03	.006/.03	.005/.03	.01/.01
	500	.05/.03	.02/.05	.01/.06	.02/.07	.02/.08	.02/.02
	1000	.28/.09	.12/.19	.11/.21	.13/.25	.12/.28	.10/.08
	2000	.96/.02	.80/.16	.72/.24	.81/.17	.73/.25	.70/.13
.05							
	250	.08/.05	.03/.10	.03/.11	.04/.12	.03/.13	.05/.04
	500	.20/.09	.08/.16	.08/.18	.09/.22	.09/.23	.09/.09
	1000	.66/.11	.39/.30	.35/.35	.41/.37	.37/.41	.33/.15
	2000	.99/.001	.96/.02	.95/.05	.97/.02	.95/.05	.94/.04
.10							
	250	.17/.07	.07/.17	.07/.19	.08/.20	.08/.21	.10/.07
	500	.35/.10	.17/.25	.16/.27	.18/.30	.18/.32	.18/.10
	1000	.83/.07	.58/.29	.54/.32	.59/.31	.56/.35	.51/.15
	2000	1.0/0.0	1.0/0.0	.99/.01	.99/.001	.99/.01	.98/.01

Note: These results are based on 50,000 replications under the alternative hypothesis of cointegration with $\rho = .98$. *Level* refers to the significance level for the individual tests. The first entry is the proportion of these replications in which both tests reject the null hypothesis of no cointegration at the indicated level. The second entry gives the proportion of replications in which only one of the tests is rejecting at the indicated level. The critical values for the individual tests are obtained from the null hypothesis using the corresponding sample size.

Table 3. Test Results in Applied Studies

$N = 132$	$ADF/\widehat{Z}_\alpha$	ADF/MAX	ADF/TR	$\widehat{Z}_\alpha/MAX$	$\widehat{Z}_\alpha/TR$	MAX/TR
Level						
.01						
	.18/.11	.20/.52	.20/.55	.18/.51	.18/.55	.64/.07
.05						
	.29/.12	.32/.50	.31/.55	.29/.53	.27/.60	.74/.10
.10						
	.30/.20	.39/.52	.39/.51	.33/.55	.32/.58	.81/.07
Corr						
	.69	.30	.29	.13	.13	.94

Note: These results (including a constant only) are based on 132 data sets from 34 applied studies appearing in the *Journal of Applied Econometrics* from 1994-2001 (Complete details on this data set and the testing methodology may be found at $http : //qed.econ.queensu.ca/pub/faculty/\ gregory/$). *Level* refers to the significance level for the individual tests. The first entry is the proportion of these replications in which both tests reject the null hypothesis of no cointegration at the indicated level. The second entry gives the proportion of replications where only one of the tests rejects at the indicated level. The critical values for the individual tests are from the asymptotic distributions. *Corr* is the correlation coefficient of the p–values.