# Conflicts Among Tests for Cointegration

May 1998

Allan W. Gregory and Alfred A. Haug*

Abstract

This note illustrates that, under the null hypothesis of no cointegration, the correlation of p–values from a single–equation residual–based test (i.e. ADF, $\hat{Z}_t$, or $\hat{Z}_\alpha$) with a system–based test (trace or maximum eigenvalue) is very low. 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.

JEL codes: C12, C15

* Gregory, Department of Economics, Queen's University, Kingston Ontario Canada K7L 3N6; awg@qed.econ.queensu.ca; and Haug, Department of Economics, University of Canterbury, New Zealand and York University, Canada; a.haug@econ.canterbury.ac.nz. We thank Gregor Smith for helpful comments. The first author would like to thank the University of Canterbury and both authors acknowledge financial support from the Social Sciences and Humanities Research Council of Canada.

Tests for cointegration commonly used are of two types: single-equation residual-based tests and system-based tests invoking the likelihood ratio principle. Popular examples of the first type are the augmented Dickey-Fuller (ADF), $\hat{Z}_t$, and $\hat{Z}_\alpha$ tests (see Phillips and Ouliaris, 1990). Researchers following the second approach calculate the trace or the maximum eigenvalue statistic (see Johansen, 1991). Perhaps because no test dominates in finite-sample simulations (see for example, Gregory, 1994 and Haug, 1996), many applied researchers report both types of tests. This seems a reasonable way to gain more information regarding the null hypothesis. One would expect that, as the sample size increases, instances of test conflict would become less frequent. In this sense the inferences from the two types of tests for cointegration, at least for large samples, would be reinforcing.

In this note we demonstrate with simulation methods 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 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 test conflict can persist no matter what the sample size. This test conflict is not resolved asymptotically.

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 endogenous 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 suggest the outcomes that might occur in practice.

The data generating process for the reported results is:

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

$$\varepsilon_t = \rho \varepsilon_{t-1} + \omega_t$$

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

Under the null hypothesis, $\theta = 0$ and $\rho = 1.0$, and under the alternative hypothesis, $\theta = 1.0$ and $\rho = 0.98$. For both hypotheses, $\varepsilon_t$ and $\eta_t$ 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 $\hat{Z}_\alpha$ 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 in addition 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 the null when it is true with probability equal to the p-value. At large sample sizes, these critical values of course 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 conflict with questions of how to choose lag lengths or bandwidths.

A striking illustration of test conflict under the null hypothesis of no cointegration is provided in Figure 1. The six possible pairwise combinations of p-values for the $ADF$, $\hat{Z}_\alpha$, $MAX$, 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 conflicts will be commonplace when the null is true. Although there appears to be some minor con°ict for very large samples for tests within the same group, this is in sharp contrast to the pronounced con°ict found in cross{over combinations.

In Table 1 we present additional evidence at conventional signi¯cance levels. The ¯rst 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% signi¯cance 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 signi¯cance. 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 $ADF=\hat{Z}_{®}$ 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 2% and 3% for tests from the same group, compared to 8% and 7% 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 (½ = :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 $T$ · 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 con°icts 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 $T$ · 1000, the proportion of replications where only one test rejects also rises. Again con°icts of this sort are more pronounced for tests from di®erent groups. Once $T = 2000$, the proportion of test con°ict is small for the 5% and 10% level tests. Di®erent alternatives will, of course, lead to di®erent rejection frequencies and con°ict percentages.

A researcher calculating one test from the residual{based approach and one from the system{based approach may ¯nd dramatic test con°ict even for large sample sizes. One explanation for this is that under the null of no cointegration the p{values can be vastly di®erent even asymptotically. Test con°icts 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 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.

## References

Gregory A. W. , \Testing for Cointegration in Linear Quadratic Models", Journal of Business and Economic Statistics 4 (1994), 347{360.

Haug, A. A. \Tests for Cointegration: A Monte Carlo Comparsion", Journal of Econometrics 71 (1996), 89{115.

Johansen, S. \Estimation and Hypothesis Testing of Cointegrating Vectors in the Presence of a Linear Trend", Econometrica 59 (1991), 1551{1580.

Phillips, P.C.B. and S. Ouliaris, \Asymptotic Properties of Residual Based Tests for Cointegration", Econometrica 58 (1990), 165{193.