QED

Queen's Economics Department Working Paper No. 1227

Critical Values for Cointegration Tests

James G. MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-2010

Critical Values for Cointegration Tests

James G. MacKinnon

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

jgm@econ.queensu.ca
http://www.econ.queensu.ca/faculty/mackinnon/

Abstract

This paper provides tables of critical values for some popular tests of cointegration and unit roots. Although these tables are necessarily based on computer simulations, they are much more accurate than those previously available. The results of the simulation experiments are summarized by means of response surface regressions in which critical values depend on the sample size. From these regressions, asymptotic critical values can be read off directly, and critical values for any finite sample size can easily be computed with a hand calculator. Added in 2010 version: A new appendix contains additional results that are more accurate and cover more cases than the ones in the original paper.

January 1990
Reissued January 2010 with additional results

This paper originally appeared as University of California San Diego Discussion Paper 90-4, which is apparently no longer available on the web. It was written while I was visiting UCSD in the fall of 1989 and early 1990 and supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada. I am grateful to all of the econometricians at UCSD for providing a hospitable research environment and to Rob Engle and the late Clive Granger for comments on an earlier version.

Forward

I spent the academic year of 1989-1990 on sabbatical at the University of California San Diego. Since Robert Engle and Clive Granger were there permanently, and David Hendry, Søren Johansen, and Timo Teräsvirta were also visiting, there was inevitably much discussion about cointegration. I was surprised to find that, for most of the tests, accurate critical values did not then exist. I therefore set out to calculate them for the Dickey-Fuller and Engle-Granger tests, and the result was this paper.

The paper originally appeared as University of California San Diego Discussion Paper No. 90-4. For many years, a bitmapped PDF of it that was missing the cover page could be found on the UCSD Economics website, but it seems to have vanished some time during 2009. The paper was later published in a book edited by Rob Engle and Clive Granger; see the references.

I have made this version available as a Queen's Economics Department Working Paper so that researchers who search for the UCSD working paper on the web will be able to find it. I have also added an appendix in which I provide new results that are much more accurate and cover more cases than the ones in the original paper. The enormous increases in computing power over the past twenty years made this quite easy to do. I have also added some additional references to works that did not exist when the paper was originally written.

After I wrote this paper, I developed more advanced methods for calculating approximate distribution functions of test statistics such as the ones dealt with in this paper; see, in particular, MacKinnon (1994, 1996, 2000), MacKinnon, Haug, and Michelis (1999), and Ericsson and MacKinnon (2002). MacKinnon (1996) provides reasonably accurate results for Dickey-Fuller and Engle-Granger tests which cover the same cases as those in the new appendix, although they are not quite as accurate. The "numerical distribution functions" obtained in that paper can be used to compute P values as well as critical values. Nevertheless, the results of this paper continue to be used far more often than the ones from the 1996 paper. Perhaps that is because they do not require the use of a specialized computer program to calculate critical values. I hope that, in future, researchers who prefer the approach of this paper will use the more accurate and more widely applicable results that are now in Tables 2, 3, and 4.

This version of the paper is dedicated to the late Sir Clive Granger, 1934–2009, without whose fundamental contributions it could never have been conceived.

January 2010

1. Introduction

Engle and Granger (1987) suggested several techniques for testing the null hypothesis that two or more series, each of which is I(1), are not cointegrated. This paper is concerned with the most popular of these techniques, which I shall refer to as Engle-Granger (or EG) tests even though they were not the only tests proposed by those authors. EG tests are closely related to some of the tests suggested by Fuller (1976) and Dickey and Fuller (1979) to test the unit root hypothesis; I shall refer to these as Dickey-Fuller or DF tests. EG and DF tests are very easy to calculate, but they suffer from one serious disadvantage: The test statistics do not follow any standard tabulated distribution, either in finite samples or asymptotically.

Engle and Granger (1987), Engle and Yoo (1987), Yoo (1987), and Phillips and Ouliaris (1990) all provide tables for one or more versions of the EG test. But these tables are based on at most 10,000 replications, which means that they are quite inaccurate. Moreover, they contain critical values for only a few finite sample sizes; asymptotic critical values, which are in many cases the most interesting ones, are not provided.

This paper provides tables of critical values for two versions of the EG test and three versions of the DF test. Although they are based on simulation, they should be accurate enough for all practical purposes. The results of the simulation experiments are summarized by means of response surface regressions, in which critical values are related to sample size. The coefficients of the response surface regressions are tabulated in such a way that asymptotic critical values can be read off directly, and critical values for any finite sample size can easily be computed with a hand calculator.

2. Engle-Granger and Dickey-Fuller Tests

Engle-Granger tests are conceptually and computationally quite simple. Let the vector $\boldsymbol{y}_t \equiv [y_{t1}, \ldots, y_{tN}]^\top$ denote the t^{th} observation on N time series, each of which is known to be I(1). If these time series are cointegrated, there exists a vector $\boldsymbol{\alpha}$ such that the stochastic process with typical observation $z_t \equiv [1 \;\; \boldsymbol{y}_t]^\top \boldsymbol{\alpha}$ is I(0). If they are not cointegrated, there will exist no vector $\boldsymbol{\alpha}$ with this property, and any linear combination of y_1 through y_N and a constant will still be I(1).

To implement the original form of the EG test, one first has to run the cointegrating regression

$$y_{t1} = \alpha_1 + \sum_{j=2}^{N} \alpha_j y_{tj} + u_t, \tag{1}$$

for a sample of size $T+1$, thus obtaining a vector of coefficients $\hat{\boldsymbol{\alpha}} \equiv [1 \;\; -\hat{\alpha}_1 \ldots -\hat{\alpha}_N]^\top$. One then calculates

$$\hat{z}_t = [1 \;\; \boldsymbol{y}_t]^\top \hat{\boldsymbol{\alpha}} = y_{1t} - \hat{\alpha}_1 - \hat{\alpha}_2 y_{t2} \ldots - \hat{\alpha}_N y_{tN}$$

and tests to see if $\hat{z}_t$ is I(1) using a procedure essentially the same (except for the distribution of the test statistic) as the DF test. The null hypothesis of non-cointegration corresponds to the null hypothesis that $\hat{z}_t$ is I(1). If one rejects the null, one concludes that y_1 through y_N are cointegrated.

To test whether $\hat{z}_t$ is I(1), one may either run the regression

$$\hat{z}_t = \rho \hat{z}_{t-1} + \varepsilon_t \tag{2}$$

and calculate the ordinary t statistic for $\rho = 1$, or run the regression

$$\Delta \hat{z}_t = \gamma \hat{z}_{t-1} + \varepsilon_t, \tag{3}$$

where $\Delta \hat{z}_t \equiv \hat{z}_t - \hat{z}_{t-1}$, and calculate the ordinary t statistic for $\gamma = 0$. In either case, one drops the first observation, reducing the sample size to T. These two procedures evidently yield identical test statistics. Because there is a constant term in (1), there is no need to include one in (2) or (3). The regressand $\hat{z}_t$ and regressor $\hat{z}_{t-1}$ would each have mean zero if both were observed over observations 0 through T. However, because the regression does not make use of the first observation on $\hat{z}_t$ or the last observation on $\hat{z}_{t-1}$, that will not be quite true. But they should both have mean very close to zero except when T is small and either $\hat{z}_0$ or $\hat{z}_T$ is unusually large in absolute value. Hence adding a constant to (2) or (3) would generally have a negligible effect on the test statistic.[1]

The way the EG test is computed is somewhat arbitrary, since any one of the y_j could be given the index 1 and made the regressand of the cointegrating regression (1). As a result, the value (but not the distribution) of the test statistic will differ depending on which series is used as the regressand. One may therefore wish to repeat the procedure with different choices of y_j serving as regressand, thus computing up to N different test statistics, especially if the first one is near the chosen critical value.

If $N = 1$, this procedure is equivalent to one variant of the ordinary DF test (see below), in which one runs the regression

$$\Delta z_t = \alpha_1 + \gamma z_{t-1} + \varepsilon_t$$

and tests for $\gamma = 0$. As several authors have shown (see West (1988) and Hylleberg and Mizon (1989)), the latter has the Dickey-Fuller distribution only when there is no drift term in the data-generating process for z_t, so that $\alpha_1 = 0$. When $\alpha_1 \neq 0$, the test statistic is asymptotically distributed as $\mathrm{N}(0, 1)$, and in finite samples its distribution may or may not be well approximated by the Dickey-Fuller distribution. The original version of the EG test likewise has a distribution that depends on the value of α_1; since all tabulated critical values assume that $\alpha_1 = 0$, they may be quite misleading when that is not the case.

There is a simple way to avoid the dependence on α_1 of the distribution of the test statistic. It is to replace the cointegrating regression (1) by

$$y_{t1} = \alpha_0 t + \alpha_1 + \sum_{j=2}^{N} \alpha_j y_{tj} + u_t, \tag{4}$$

[1] Some changes were made in this paragraph in the 2010 version to correct minor errors in the original paper. The conclusion is unchanged.

that is, to add a linear time trend to the cointegrating regression. The resulting test statistic will now be invariant to the value of α_1, although it will have a different distribution than the one based on regression (1).[2] Adding a trend to the cointegrating regression often makes sense for a number of other reasons, as Engle and Yoo (1990) discuss. There are thus two variants of the Engle-Granger test. The "no-trend" variant uses (1) as the cointegrating regression, and the "with-trend" variant uses (4).

In some cases, the vector $\boldsymbol{\alpha}$ (or at least α_2 through α_N) may be known. We can then just calculate $z_t = y_{t1} - \alpha_2 y_{t2} \ldots - \alpha_N y_{tN}$ and use an ordinary DF test. In this case, it is easiest to dispense with the cointegrating regressions (1) or (4) entirely and simply run one of the following test regressions:

$$\Delta z_t = \gamma z_{t-1} + \varepsilon_t \tag{5}$$

$$\Delta z_t = \alpha_1 + \gamma z_{t-1} + \varepsilon_t \tag{6}$$

$$\Delta z_t = \alpha_0 t + \alpha_1 + \gamma z_{t-1} + \varepsilon_t. \tag{7}$$

The t statistics for $\gamma = 0$ in these three regressions yield the test statistics that Fuller (1976) refers to as $\hat{\tau}$, $\hat{\tau}_\mu$, and $\hat{\tau}_t$, respectively; he provides some estimated critical values on page 373. We will refer to these test statistics as the "no-constant", "no-trend", and "with-trend" statistics, respectively. Note that the tabulated distribution of the no-constant statistic depends on the assumption that $z_0 = 0$, while those of the other two are invariant to z_0. The tabulated distribution of the no-trend statistic depends on the assumption that $\alpha_1 = 0$ (see West (1988) and Hylleberg and Mizon (1989)), while that of the with-trend statistic depends on the assumption that $\alpha_0 = 0$.

Up to this point, it has been assumed that the innovations ε_t are serially independent and homoskedastic. These rather strong assumptions can be relaxed without affecting the asymptotic distributions of the test statistics. The test statistics do not even have to be modified to allow for heteroskedasticity, since, as Phillips (1987) has shown, heteroskedasticity does not affect the asymptotic distribution of a wide class of unit root test statistics. They do have to be modified to allow for serial correlation, however. The easiest way to do this is to use **Augmented Dickey-Fuller**, or **ADF**, and **Augmented Engle-Granger**, or **AEG**, tests. In practice, this means that one must add as many lags of $\Delta\hat{z}_t$ to regressions (2) or (3), or of Δz_t to regressions (5), (6), or (7), as are necessary to ensure that the residuals for those regressions appear to be white noise.

A different approach to obtaining unit root tests that are asymptotically valid in the presence of serial correlation and/or heteroskedasticity of unknown form was suggested by Phillips (1987) and extended to the cointegration case by Phillips and Ouliaris (1990). The asymptotic distributions of what Phillips and Ouliaris call the $\hat{Z}_t$ statistic are identical to those of the corresponding DF, ADF, EG, and AEG tests. Phillips and Ouliaris tabulate critical values for two forms of this statistic (corresponding to the no-trend and with-trend versions of the DF and EG statistics) for several values of N. Unfortunately, these critical values are based on only 10,000 replications, so that

[2] It will not be invariant to the value of α_0, however. To achieve that, one would have to add t^2 to the regression; see the appendix.

they suffer from considerable experimental error. Moreover, they are for 500 rather than an infinite number of observations, so that they are biased away from zero as estimates of asymptotic critical values. As can be seen from Table 1 below, this bias is by no means negligible in some cases.

3. The Simulation Experiments

Instead of simply providing tables of estimated critical values for a few specific sample sizes, as previous papers have done, this paper estimates response surface regressions. These relate the 1%, 5% and 10% lower-tail critical values for the test statistics discussed above, for various values of N, to the sample size T.[3] Recall that T refers to the number of observations in the unit root test regression, and that this is one less than the total number of observations available and used in the cointegrating regression. Response surfaces were estimated for thirteen different tests: the no-constant, no-trend and with-trend versions of the DF test, which are equivalent to the corresponding EG tests for $N = 1$, and the no-trend and with-trend versions of the EG test for $N = 2, 3, 4, 5$ and 6. Thus a total of thirty-nine response surface regressions were estimated.

The DF tests were computed using the one-step procedures of regressions (5), (6), and (7), while the EG tests were computed using two-step procedures consisting of regressions (1) or (4) followed by (3). These are the easiest ways to calculate these tests. Note that there is a slight difference between the degrees-of-freedom corrections used to calculate the regression standard errors, and hence t statistics, for the DF tests ($N = 1$) and for the EG tests ($N \geq 2$). If the no-trend and with-trend DF tests were computed in the same way as the corresponding EG tests, they would be larger by factors of $\big((T-1)/(T-2)\big)^{1/2}$ and $\big((T-1)/(T-3)\big)^{1/2}$, respectively.

Conceptually, each simulation experiment consisted of 25,000 replications for a single value of T and a single value of N.[4] The 1%, 5% and 10% empirical quantiles for these data were then calculated, and each of these became a single observation in the response surface regression. The number 25,000 was chosen to make the bias in estimating quantiles negligible, while keeping the memory requirements of the program manageable.

For all values of N except $N = 6$, forty experiments were run for each of the following sample sizes: 18, 20, 22, 25, 28, 30, 32, 40, 50, 75, 100, 150, 200, 250, and 500. For $N = 6$, the sample sizes were 20, 22, 25, 28, 30, 32, 36, 40, 50, 100, 250, and 275. Most of the sample sizes were relatively small because the cost of the experiments was slightly less than proportional to the sample size, and because small sample sizes

[3] The upper tail is not of any interest in this case, and the vast majority of hypothesis tests are at the 1%, 5%, or 10% levels.

[4] In fact, results for $N = 2$, 3, 4, and 5 were computed together to save computer time. Results for $N = 1$ were computed separately because the calculations were slightly different. Results for $N = 6$ were computed separately because it was not decided to extend the analysis to this case until after most of the other calculations had been completed.

provided more information about the shape of the response surfaces.[5] However, a few large values of T were also included so that the response surface estimates of asymptotic critical values would be sufficiently accurate. The total number of replications was 12 million in 480 experiments for $N = 1$ and 15 million in 600 experiments for the other values of N.[6]

Using a correct functional form for the response surface regressions is crucial to obtaining useful estimates. After considerable experimentation, the following form was found to work very well:

$$C_k(p) = \beta_\infty + \beta_1 T_k^{-1} + \beta_2 T_k^{-2} + e_k. \tag{8}$$

Here $C_k(p)$ denotes the estimated p% quantile from the k^{th} experiment, T_k denotes the sample size for that experiment, and there are three parameters to be estimated. The parameter β_∞ is an estimate of the asymptotic critical value for a test at level p, since, as T tends to infinity, T^{-1} and T^{-2} both tend to zero. The other two parameters determine the shape of the response surface for finite values of T.

The ability of (8) to fit the data from the simulation experiments was remarkably good. To test its adequacy, it was compared to the most general specification possible, in which $C_k(p)$ was regressed on 15 dummy variables (12 when $N = 6$), corresponding to the different values of T. It was rejected by the usual F test in only a very few cases where it seemed to have trouble fitting the estimated critical values for the very smallest value(s) of T. The adequacy of (8) was therefore further tested by adding dummy variables corresponding to the smallest values of T, and this test proved slightly more powerful than the first one. When either test provided evidence of model inadequacy, the offending observations ($T = 18$ and in one case $T = 20$ as well) were dropped from the response surface regressions.

Several alternative functional forms were also tried. Adding additional powers of $1/T$ never seemed to be necessary. In fact, in several cases, fewer powers were necessary, since the restriction that $\beta_2 = 0$ appeared to be consistent with the data. In most cases, one could replace T^{-2} by $T^{-3/2}$ without having any noticeable effect on either the fit of the regression or the estimate of β_∞; the decision to retain T^{-2} rather than $T^{-3/2}$ in (8) was based on very slim evidence in favor of the former. On the other hand, replacing T by either $T - N$ or $T - N - 1$, the numbers of degrees of freedom for

---

[5] The experiments would have required roughly nine hundred hours on a 20 Mh. 386 personal computer. All programs were written in FORTRAN 77. About 70% of the computations were done on the PC, using programs compiled with the Lahey F77L-EM/32 compiler. Some experiments, representing roughly 30% of the total computational burden, were performed on other computers, namely, an IBM 3081G, which was about 7.5 times as fast as the PC, and an HP 9000 Model 840, which was about 15% faster.

[6] The experiments for $N = 6$ were done later than the others and were designed in the light of experience with them. It was decided that the extra accuracy available by doing more experiments for large values of T was not worth the extra cost.

the cointegrating regression in the no-trend and with-trend cases, respectively, often (but not always) resulted in a dramatic deterioration in the fit of the response surface.

The residuals e_k in regression (8) were heteroskedastic, being larger for the smaller sample sizes. This was particularly noticeable when (8) was run for larger values of N. The response surface regressions were therefore estimated by feasible GLS. As a first step, $C_k(p)$ was regressed on 15 (or 12) dummy variables, yielding residuals $\acute{e}_k$. The following regression was then run:

$$\acute{e}_k^2 = \delta_\infty + \delta_1(T_k - d)^{-1} + \delta_2(T_k - d)^{-2} + \text{error}, \tag{9}$$

where d is the number of degrees of freedom used up in the cointegrating regression, and there are three coefficients to be estimated.[7] The inverses of the square roots of the fitted values from (9) were then used as weights for feasible GLS estimation of (8). The feasible GLS estimates were generally much better than the OLS ones in terms of loglikelihood values, but the two sets of estimates were numerically very close.

The final results of this paper are the feasible GLS estimates of regression (8) for 39 sets of experimental data. These estimates are presented in Table 1. The estimates of β_∞ provide asymptotic critical values directly, while values for any finite T can easily be calculated using the estimates of all three parameters. The restriction that $\beta_2 = 0$ has been imposed whenever the t statistic on $\hat{\beta}_2$ was less than one in absolute value.

Estimated standard errors are reported for $\hat{\beta}_\infty$ but not for $\hat{\beta}_1$ or $\hat{\beta}_2$, since the latter are of no interest. What is of interest is the standard error of

$$\hat{C}(p, T) = \hat{\beta}_\infty + \hat{\beta}_1 T^{-1} + \hat{\beta}_2 T^{-2},$$

the estimated critical value for a test at the p% level when the sample size is T. This varies with T and tends to be smallest for sample sizes in the range of 80 to 150. Except for very small values of T (less than about 25), the standard error of $\hat{C}(p, T)$ was always less than the standard error of the corresponding $\hat{\beta}_\infty$, so that, if the standard errors of the $\hat{\beta}_\infty$ were accurate, they could be regarded as upper bounds for the standard errors of $\hat{C}(p, T)$ for most values of T.

However, the standard errors for $\hat{\beta}_\infty$ reported in Table 1 are undoubtedly too small. The problem is that they are conditional on the specification of the response surface regressions. Although the specification (8) performed very well in all cases, other specifications also performed well in many cases, sometimes outperforming (8) insignificantly. Estimates of β_∞ sometimes changed by as much as twice the reported standard error as a result of minor changes in the specification of the response surface that did not significantly affect its fit. Thus it is probably reasonable to think of the

[7] Considerable experimentation preceded the choice of the functional form for regression (9). It was found that omitting d had little effect on the fit of the regression, although on balance it seemed preferable to retain it. In this respect, regression (9) is quite different from regression (8), where using $T_k - d$ rather than T_k sometimes worsened the fit substantially.

actual standard errors on the $\hat{\beta}_\infty$ as being about twice as large as the reported ones. Even so, it seems likely that few if any of the estimated 1% critical values in Table 1 differ from the true value by as much as .01, and extremely unlikely that any of the estimated 5% and 10% critical values differ from their true values by that much.

4. Conclusion

It is hoped that the results in Table 1 will prove useful to investigators testing for unit roots and cointegration. Although the methods used to obtain these results are quite computationally intensive, they are entirely feasible with current personal computer technology. The use of response surface regressions to summarize results is valuable for two reasons. First, this approach allows one to estimate asymptotic critical values without actually using infinitely large samples. Second, it makes it possible to tabulate results for all sample sizes based on experimental results for only a few. Similar methods could be employed in many other cases where test statistics do not follow standard tabulated distributions.

Table 1. Response Surface Estimates of Critical Values

N	Variant	Level	Obs.	β_∞	(s.e.)	β_1	β_2
1	no constant	1%	600	−2.5658	(0.0023)	−1.960	−10.04
		5%	600	−1.9393	(0.0008)	−0.398	
		10%	560	−1.6156	(0.0007)	−0.181	
1	no trend	1%	600	−3.4336	(0.0024)	−5.999	−29.25
		5%	600	−2.8621	(0.0011)	−2.738	−8.36
		10%	600	−2.5671	(0.0009)	−1.438	−4.48
1	with trend	1%	600	−3.9638	(0.0019)	−8.353	−47.44
		5%	600	−3.4126	(0.0012)	−4.039	−17.83
		10%	600	−3.1279	(0.0009)	−2.418	−7.58
2	no trend	1%	600	−3.9001	(0.0022)	−10.534	−30.03
		5%	600	−3.3377	(0.0012)	−5.967	−8.98
		10%	600	−3.0462	(0.0009)	−4.069	−5.73
2	with trend	1%	600	−4.3266	(0.0022)	−15.531	−34.03
		5%	560	−3.7809	(0.0013)	−9.421	−15.06
		10%	600	−3.4959	(0.0009)	−7.203	−4.01
3	no trend	1%	560	−4.2981	(0.0023)	−13.790	−46.37
		5%	560	−3.7429	(0.0012)	−8.352	−13.41
		10%	600	−3.4518	(0.0010)	−6.241	−2.79
3	with trend	1%	600	−4.6676	(0.0022)	−18.492	−49.35
		5%	600	−4.1193	(0.0011)	−12.024	−13.13
		10%	600	−3.8344	(0.0009)	−9.188	−4.85
4	no trend	1%	560	−4.6493	(0.0023)	−17.188	−59.20
		5%	560	−4.1000	(0.0012)	−10.745	−21.57
		10%	600	−3.8110	(0.0009)	−8.317	−5.19
4	with trend	1%	600	−4.9695	(0.0021)	−22.504	−50.22
		5%	560	−4.4294	(0.0012)	−14.501	−19.54
		10%	560	−4.1474	(0.0010)	−11.165	−9.88
5	no trend	1%	520	−4.9587	(0.0026)	−22.140	−37.29
		5%	560	−4.4185	(0.0013)	−13.641	−21.16
		10%	600	−4.1327	(0.0009)	−10.638	−5.48
5	with trend	1%	600	−5.2497	(0.0024)	−26.606	−49.56
		5%	600	−4.7154	(0.0013)	−17.432	−16.50
		10%	600	−4.4345	(0.0010)	−13.654	−5.77
6	no trend	1%	480	−5.2400	(0.0029)	−26.278	−41.65
		5%	480	−4.7048	(0.0018)	−17.120	−11.17
		10%	480	−4.4242	(0.0010)	−13.347	
6	with trend	1%	480	−5.5127	(0.0033)	−30.735	−52.50
		5%	480	−4.9767	(0.0017)	−20.883	−9.05
		10%	480	−4.6999	(0.0011)	−16.445	

Explanation of Table 1

N: Number of I(1) series for which null of non-cointegration is being tested.

Level: Level of one-tail test of the unit root null against the alternative of stationarity.

Obs.: Number of observations used in response surface regression. Possible values are 600, 560, 520, and 480. If Obs. = 600, the regression used 40 observations from each of $T = 18$, 20, 22, 25, 28, 30, 32, 40, 50, 75, 100, 150, 200, 250, and 500. If Obs. = 560, observations for $T = 18$ were not used. If Obs. = 520, observations for $T = 18$ and $T = 20$ were not used. If Obs. = 480, the regression used 40 observations from each of $T = 20$, 22, 25, 28, 30, 32, 36, 40, 50, 100, 250, and 275.

β_∞: Estimated asymptotic critical values (with estimated standard errors in parentheses).

β_1: Coefficient on T^{-1} in response surface regression.

β_2: Coefficient on T^{-2} in response surface regression. It was omitted if the t statistic was less than one in absolute value.

For any sample size T, the estimated critical value is

$$\beta_\infty + \beta_1/T + \beta_2/T^2.$$

For example, when $T = 100$, the 5% critical value for the with-trend EG test when $N = 5$ is

$$-4.7154 - 17.432/100 - 16.50/100^2 = -4.8914\ .$$

Appendix (added in 2010 version)

In the twenty years since this paper was written, computer technology has advanced enormously. On the occasion of reissuing the paper, it therefore makes sense to report new results based on a much larger number of replications that cover a broader range of cases. However, for comparability with the original paper, the methodology is largely unchanged. The simulation results could also have been used to provide somewhat more accurate numerical distribution functions than those of MacKinnon (1996), which allow one to compute P values and critical values for tests at any level, but no attempt is made to do so here.

One major difference between the original experiments and the new ones is that the latter involve far more computation. Instead of 25,000 replications, each simulation now involves 200,000. Instead of 40 simulations for each sample size, there are now 500. And instead of the 12 or 15 sample sizes used originally, there are 30. The sample sizes are 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1200, and 1400. Some of these values of T are much larger than the largest ones used originally. This increases the precision of the estimates but made the experiments much more expensive.

The total number of observations for the response surface regressions is usually 15,000 (that is, 30 times 500), although, in a few cases, the simulations for $T = 20$ were dropped because the response surface did not fit well enough. In a very few cases, the simulations for $T = 25$ were dropped as well. Thus a few of the estimates, always for larger values of N, are based on either 14,500 or 14,000 observations.

Because the new response surface regressions are based on 200 times as many simulations as the original ones, any misspecification becomes much more apparent. It was therefore necessary in most cases to add an additional (cubic) term to equation (8), which thus becomes

$$C_k(p) = \beta_\infty + \beta_1 T_k^{-1} + \beta_2 T_k^{-2} + \beta_3 T_k^{-3} + e_k. \tag{A.1}$$

To avoid adding the cubic term, it would usually have been necessary to drop one or more of the smaller sample sizes (often quite a few of them). This was tried in several cases, and the estimate of β_∞ did not change much when enough small values of T_k were dropped so that the estimate of β_3 became insignificant at the 10% level. In general, the standard error of $\hat{\beta}_\infty$ was lower when equation (A.1) was estimated using all the data than when equation (8) was estimated without the observations corresponding to some of the smaller values of T_k.

The second major difference between the original experiments and the new ones is that the latter deal with a broader range of cases. The values of N now go from 1 to 12 instead of from 1 to 6. Moreover, an additional variant of the DF and EG tests is included, in which t^2 is added to either the cointegrating regression (4) or the DF test regression (7). This variant of the tests was advocated by Ouliaris, Park, and Phillips (1989). In the notation used by MacKinnon (1996), the tests for which critical values are computed are the τ_c, τ_{ct}, and τ_{ctt} tests. These involve, respectively, a constant term, a constant and a trend, and a constant, trend, and trend squared in either the cointegrating regression or the DF test regression. For $N = 1$ only, as in Table 1, results are also presented for the τ_{nc} test, in which the DF test regression does not contain a constant term. Use of this test is not recommended, however, because it requires highly unrealistic assumptions.

Table 2 contains results for the τ_c test for $N = 1$ to 12 and also for the τ_{nc} test for $N = 1$. Table 3 contains results for the τ_{ct} test for $N = 1$ to 12, and Table 4 contains results for the τ_{ctt} test for $N = 1$ to 12. These tables are to be read in exactly the same way as Table 1, except that there is (usually) one more coefficient to take into account. Note that the β_3 coefficient was set to 0 (and omitted from the table) when the t statistic on $\hat{\beta}_3$ was less than $\sqrt{2}$ in absolute value.

The feasible GLS method of estimating equation (A.1) discussed in the body of this paper yields identical results to the GMM method discussed in MacKinnon (1996) when the variances of the e_k are estimated in the same way. The estimates in Tables 2, 3, and 4 were actually obtained using the latter method. One advantage of this approach is that it automatically yields a GMM overidentification test statistic which would reveal misspecification if it were at all serious.

The standard errors of $\hat{\beta}_\infty$ reported in the tables are undoubtedly too small, because they ignore uncertainty about the specification of the response surface. Nevertheless, it is interesting to compare the standard errors in Table 1 with the ones in the three new tables. For example, consider the τ_{ct} tests with $N = 2$ and $N = 3$. In Table 1, the standard errors for the .05 asymptotic critical values of those two tests are 0.0013 and 0.0011, respectively. In Table 3, they are 0.000054 and 0.000066. The standard error is larger for $N = 3$ than for $N = 2$ because β_3 has to be estimated in the former case but not in the latter. In most cases, the standard errors seem to be smaller by factors of between 15 and 20.

Using the tables, it is easy to calculate a critical value (strictly valid only under the assumption that the errors are IID normal) for any finite sample size T. The estimated critical value is simply

$$\beta_\infty + \beta_1/T + \beta_2/T^2 + \beta_3/T^3.$$

For example, when $T = 100$, the 5% critical value for the τ_{ct} test (that is, the EG test with trend) when $N = 5$ is

$$-4.71537 - 17.3569/100 - 22.660/100^2 + 91.359/100^3 = -4.89111\ .$$

This is very close to the value of -4.8914 that was calculated using the results in Table 1; see the explanation following that table.

Table 2. Critical Values for No Trend Case (τ_{nc} and τ_c)

N	Variant	Level	Obs.	β_∞	(s.e.)	β_1	β_2	β_3
1	τ_{nc}	1%	15,000	−2.56574	(0.000110)	−2.2358	−3.627	
1	τ_{nc}	5%	15,000	−1.94100	(0.000074)	−0.2686	−3.365	31.223
1	τ_{nc}	10%	15,000	−1.61682	(0.000059)	0.2656	−2.714	25.364
1	τ_c	1%	15,000	−3.43035	(0.000127)	−6.5393	−16.786	−79.433
1	τ_c	5%	15,000	−2.86154	(0.000068)	−2.8903	−4.234	−40.040
1	τ_c	10%	15,000	−2.56677	(0.000043)	−1.5384	−2.809	
2	τ_c	1%	15,000	−3.89644	(0.000102)	−10.9519	−22.527	
2	τ_c	5%	15,000	−3.33613	(0.000056)	−6.1101	−6.823	
2	τ_c	10%	15,000	−3.04445	(0.000044)	−4.2412	−2.720	
3	τ_c	1%	15,000	−4.29374	(0.000123)	−14.4354	−33.195	47.433
3	τ_c	5%	15,000	−3.74066	(0.000067)	−8.5631	−10.852	27.982
3	τ_c	10%	15,000	−3.45218	(0.000043)	−6.2143	−3.718	
4	τ_c	1%	15,000	−4.64332	(0.000101)	−18.1031	−37.972	
4	τ_c	5%	15,000	−4.09600	(0.000055)	−11.2349	−11.175	
4	τ_c	10%	15,000	−3.81020	(0.000043)	−8.3931	−4.137	
5	τ_c	1%	15,000	−4.95756	(0.000101)	−21.8883	−45.142	
5	τ_c	5%	15,000	−4.41519	(0.000055)	−14.0406	−12.575	
5	τ_c	10%	15,000	−4.13157	(0.000043)	−10.7417	−3.784	
6	τ_c	1%	15,000	−5.24568	(0.000124)	−25.6688	−57.737	88.639
6	τ_c	5%	15,000	−4.70693	(0.000068)	−16.9178	−17.492	60.007
6	τ_c	10%	15,000	−4.42501	(0.000054)	−13.1875	−5.104	27.877
7	τ_c	1%	15,000	−5.51233	(0.000126)	−29.5760	−69.398	164.295
7	τ_c	5%	15,000	−4.97684	(0.000068)	−19.9021	−22.045	110.761
7	τ_c	10%	15,000	−4.69648	(0.000054)	−15.7315	−6.922	67.721
8	τ_c	1%	15,000	−5.76202	(0.000126)	−33.5258	−82.189	256.289
8	τ_c	5%	15,000	−5.22924	(0.000068)	−23.0023	−24.646	144.479
8	τ_c	10%	15,000	−4.95007	(0.000053)	−18.3959	−7.344	94.872
9	τ_c	1%	15,000	−5.99742	(0.000126)	−37.6572	−87.365	248.316
9	τ_c	5%	15,000	−5.46697	(0.000069)	−26.2057	−26.627	176.382
9	τ_c	10%	14,500	−5.18897	(0.000062)	−21.1377	−9.484	172.704
10	τ_c	1%	15,000	−6.22103	(0.000128)	−41.7154	−102.680	389.330
10	τ_c	5%	15,000	−5.69244	(0.000068)	−29.4521	−30.994	251.016
10	τ_c	10%	15,000	−5.41533	(0.000054)	−24.0006	−7.514	163.049
11	τ_c	1%	14,500	−6.43377	(0.000145)	−46.0084	−106.809	352.752
11	τ_c	5%	15,000	−5.90714	(0.000068)	−32.8336	−30.275	249.994
11	τ_c	10%	15,000	−5.63086	(0.000055)	−26.9693	−4.083	151.427
12	τ_c	1%	15,000	−6.63790	(0.000127)	−50.2095	−124.156	579.622
12	τ_c	5%	15,000	−6.11279	(0.000069)	−36.2681	−32.505	314.802
12	τ_c	10%	15,000	−5.83724	(0.000054)	−29.9864	−2.686	184.116

Table 3. Critical Values for Linear Trend Case (τ_{ct})

N	Level	Obs.	β_∞	(s.e.)	β_1	β_2	β_3
1	1%	15,000	−3.95877	(0.000122)	−9.0531	−28.428	−134.155
1	5%	15,000	−3.41049	(0.000066)	−4.3904	−9.036	−45.374
1	10%	15,000	−3.12705	(0.000051)	−2.5856	−3.925	−22.380
2	1%	15,000	−4.32762	(0.000099)	−15.4387	−35.679	
2	5%	15,000	−3.78057	(0.000054)	−9.5106	−12.074	
2	10%	15,000	−3.49631	(0.000053)	−7.0815	−7.538	21.892
3	1%	15,000	−4.66305	(0.000126)	−18.7688	−49.793	104.244
3	5%	15,000	−4.11890	(0.000066)	−11.8922	−19.031	77.332
3	10%	15,000	−3.83511	(0.000053)	−9.0723	−8.504	35.403
4	1%	15,000	−4.96940	(0.000125)	−22.4694	−52.599	51.314
4	5%	15,000	−4.42871	(0.000067)	−14.5876	−18.228	39.647
4	10%	15,000	−4.14633	(0.000054)	−11.2500	−9.873	54.109
5	1%	15,000	−5.25276	(0.000123)	−26.2183	−59.631	50.646
5	5%	15,000	−4.71537	(0.000068)	−17.3569	−22.660	91.359
5	10%	15,000	−4.43422	(0.000054)	−13.6078	−10.238	76.781
6	1%	15,000	−5.51727	(0.000125)	−29.9760	−75.222	202.253
6	5%	15,000	−4.98228	(0.000066)	−20.3050	−25.224	132.030
6	10%	15,000	−4.70233	(0.000053)	−16.1253	−9.836	94.272
7	1%	15,000	−5.76537	(0.000125)	−33.9165	−84.312	245.394
7	5%	15,000	−5.23299	(0.000067)	−23.3328	−28.955	182.342
7	10%	15,000	−4.95405	(0.000054)	−18.7352	−10.168	120.575
8	1%	15,000	−6.00003	(0.000126)	−37.8892	−96.428	335.920
8	5%	15,000	−5.46971	(0.000068)	−26.4771	−31.034	220.165
8	10%	15,000	−5.19183	(0.000054)	−21.4328	−10.726	157.955
9	1%	15,000	−6.22288	(0.000125)	−41.9496	−109.881	466.068
9	5%	15,000	−5.69447	(0.000069)	−29.7152	−33.784	273.002
9	10%	15,000	−5.41738	(0.000054)	−24.2882	−8.584	169.891
10	1%	15,000	−6.43551	(0.000127)	−46.1151	−120.814	566.823
10	5%	15,000	−5.90887	(0.000069)	−33.0251	−37.208	346.189
10	10%	14,500	−5.63255	(0.000063)	−27.2042	−6.792	177.666
11	1%	15,000	−6.63894	(0.000125)	−50.4287	−128.997	642.781
11	5%	15,000	−6.11404	(0.000069)	−36.4610	−36.246	348.554
11	10%	15,000	−5.83850	(0.000055)	−30.1995	−5.163	210.338
12	1%	15,000	−6.83488	(0.000126)	−54.7119	−139.800	736.376
12	5%	15,000	−6.31127	(0.000068)	−39.9676	−37.021	406.051
12	10%	14,000	−6.03650	(0.000074)	−33.2381	−6.606	317.776

Table 4. Critical Values for Quadratic Trend Case (τ_{ctt})

N	Level	Obs.	β_∞	(s.e.)	β_1	β_2	β_3
1	1%	15,000	−4.37113	(0.000123)	−11.5882	−35.819	−334.047
1	5%	15,000	−3.83239	(0.000065)	−5.9057	−12.490	−118.284
1	10%	15,000	−3.55326	(0.000051)	−3.6596	−5.293	−63.559
2	1%	15,000	−4.69276	(0.000124)	−20.2284	−64.919	88.884
2	5%	15,000	−4.15387	(0.000067)	−13.3114	−28.402	72.741
2	10%	15,000	−3.87346	(0.000052)	−10.4637	−17.408	66.313
3	1%	15,000	−4.99071	(0.000125)	−23.5873	−76.924	184.782
3	5%	15,000	−4.45311	(0.000068)	−15.7732	−32.316	122.705
3	10%	15,000	−4.17280	(0.000053)	−12.4909	−17.912	83.285
4	1%	15,000	−5.26780	(0.000125)	−27.2836	−78.971	137.871
4	5%	15,000	−4.73244	(0.000069)	−18.4833	−31.875	111.817
4	10%	15,000	−4.45268	(0.000053)	−14.7199	−17.969	101.920
5	1%	15,000	−5.52826	(0.000125)	−30.9051	−92.490	248.096
5	5%	15,000	−4.99491	(0.000068)	−21.2360	−37.685	194.208
5	10%	15,000	−4.71587	(0.000054)	−17.0820	−18.631	136.672
6	1%	15,000	−5.77379	(0.000126)	−34.7010	−105.937	393.991
6	5%	15,000	−5.24217	(0.000067)	−24.2177	−39.153	232.528
6	10%	15,000	−4.96397	(0.000054)	−19.6064	−18.858	174.919
7	1%	15,000	−6.00609	(0.000125)	−38.7383	−108.605	365.208
7	5%	15,000	−5.47664	(0.000067)	−27.3005	−39.498	246.918
7	10%	14,500	−5.19921	(0.000062)	−22.2617	−17.910	208.494
8	1%	14,500	−6.22758	(0.000143)	−42.7154	−119.622	421.395
8	5%	15,000	−5.69983	(0.000067)	−30.4365	−44.300	345.480
8	10%	15,000	−5.42320	(0.000054)	−24.9686	−19.688	274.462
9	1%	15,000	−6.43933	(0.000125)	−46.7581	−136.691	651.380
9	5%	15,000	−5.91298	(0.000069)	−33.7584	−42.686	346.629
9	10%	15,000	−5.63704	(0.000054)	−27.8965	−13.880	236.975
10	1%	15,000	−6.64235	(0.000125)	−50.9783	−145.462	752.228
10	5%	15,000	−6.11753	(0.000070)	−37.056	−48.719	473.905
10	10%	15,000	−5.84215	(0.000054)	−30.8119	−14.938	316.006
11	1%	14,500	−6.83743	(0.000145)	−55.2861	−152.651	792.577
11	5%	15,000	−6.31396	(0.000069)	−40.5507	−46.771	487.185
11	10%	14,500	−6.03921	(0.000062)	−33.8950	−9.122	285.164
12	1%	15,000	−7.02582	(0.000124)	−59.6037	−166.368	989.879
12	5%	15,000	−6.50353	(0.000070)	−44.0797	−47.242	543.889
12	10%	14,500	−6.22941	(0.000063)	−36.9673	−10.868	418.414

References

Note: References that were not in the original paper are marked with an asterisk.

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*, **84**, 427–431.

Engle, R. F. and C. W. J. Granger (1987), "Co-integration and error correction: Representation, estimation and testing," *Econometrica*, **55**, 251–276.

Engle, R. F. and B. S. Yoo (1987), "Forecasting and testing in co-integrated systems," *Journal of Econometrics*, **35**, 143–159.

Engle, R. F. and B. S. Yoo (1991), "Cointegrated economic time series: An overview with new results," Chapter 12 in *Long-Run Economic Relationships: Readings in Cointegration*, ed. R. F. Engle and C. W. J. Granger. Oxford, Oxford University Press.

Ericsson, N. R. and J. G. MacKinnon (2002), "Distributions of error correction tests for cointegration," *Econometrics Journal*, **5**, 285–318. [∗]

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

Hylleberg, S. and G. E. Mizon (1989), "A note on the distribution of the least squares estimator of a random walk with drift, *Economics Letters*, **29**, 225–230.

MacKinnon, J. G. (1991), "Critical values for cointegration tests," Chapter 13 in *Long-Run Economic Relationships: Readings in Cointegration*, ed. R. F. Engle and C. W. J. Granger. Oxford, Oxford University Press. [∗]

MacKinnon, J. G. (1994), "Approximate asymptotic distribution functions for unit-root and cointegration tests," *Journal of Business and Economic Statistics*, **12**, 167–176. [∗]

MacKinnon, J. G. (1996), "Numerical distribution functions for unit root and cointegration tests," *Journal of Applied Econometrics*, **11**, 601–618. [∗]

MacKinnon, J. G. (2000), "Computing numerical distribution functions in econometrics," in *High Performance Computing Systems and Applications*, ed. A. Pollard, D. Mewhort, and D. Weaver. Amsterdam, Kluwer, 455–470. [∗]

MacKinnon, J. G., A. A. Haug, and L. Michelis (1999), "Numerical distribution functions of likelihood ratio tests for cointegration," *Journal of Applied Econometrics*, **14**, 563–577. [∗]

Ouliaris, S., J. Y. Park, and P. C. B. Phillips (1989), "Testing for a unit root in the presence of a maintained trend," in *Advances in Econometrics*, ed. B. Raj. Boston, Klumer Academic Publishers, 7–28. [∗]

Phillips, P. C. B. (1987), "Time series regression with a unit root," *Econometrica*, **55**, 277–301.

Phillips, P. C. B. and S. Ouliaris (1990), "Asymptotic properties of residual based tests for cointegration," *Econometrica*, **58**, 165–193.

Yoo, B. S. (1987), "Co-integrated time series: Structure, forecasting and testing," unpublished Ph.D. Dissertation, University of California, San Diego.

West, K. D. (1988), "Asymptotic normality, when regressors have a unit root," *Econometrica*, **56**, 1397–1417.