Queen's Economics Department Working Paper No. 918

Numerical Distribution Functions for Unit Root and Cointegration Tests

James G. MacKinnon

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

1-1995

Discussion Paper #918

Numerical Distribution Functions for Unit Root and Cointegration Tests

by

James G. MacKinnon
Queen's University

January 1995

Numerical Distribution Functions for Unit Root and Cointegration Tests

by

James G. MacKinnon

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

e-mail: jgm@qed.econ.queensu.ca
Phone: 613 545-2293
FAX: 613 545-6668

This research was supported, in part, by the Social Sciences and Humanities Research Council of Canada. I am grateful to Neil Ericsson for helpful discussions and comments on a preliminary draft.

February, 1995

Abstract

This paper employs response surface regressions based on simulation experiments to calculate distribution functions for some well-known unit root and cointegration test statistics. The principal contributions of the paper are a set of data files that contain estimated response surface coefficients and a computer program for utilizing them. This program, which is freely available via the Internet, can easily be used to calculate both asymptotic and finite-sample critical values and P values for any of the tests. Graphs of some of the tabulated distribution functions are provided. There is also an empirical example.

JEL Classification Number: C22

Keywords: unit root tests, Dickey-Fuller tests, cointegration tests, Engle-Granger tests, response surfaces, critical values, approximate P values, simulation methods

1. Introduction

Tests of the null hypothesis that a time-series process has a unit root have been widely used in recent years, as have tests of the null hypothesis that two or more integrated series are not cointegrated. The most commonly used unit root tests are based on the work of Dickey and Fuller (1979) and Said and Dickey (1984). These are known as Dickey-Fuller (DF) tests and Augmented Dickey-Fuller (ADF) tests, respectively. These tests have nonstandard distributions, even asymptotically. The cointegration tests developed by Engle and Granger (1987) are closely related to DF and ADF tests, but they have different, nonstandard distributions, which depend on the number of possibly cointegrated variables.

Although the asymptotic theory of these unit root and cointegration tests is well developed, it is not at all easy for applied workers to calculate the marginal significance level, or P value, associated with a given test statistic. Until a few years ago (MacKinnon, 1991), accurate critical values for cointegration tests were not available at all. In a recent paper (MacKinnon, 1994), I used simulation methods to estimate the asymptotic distributions of a large number of unit root and cointegration tests. I then obtained reasonably simple approximating equations that may be used to obtain approximate asymptotic P values. In the present paper, I correct two deficiencies of this earlier work. The first deficiency is that the approximating equations are considerably less accurate than the underlying estimated asymptotic distributions. The second deficiency is that, even though the simulation experiments provided information about the finite-sample distributions of the test statistics, the approximating equations were obtained only for the asymptotic case.

The key to overcoming these two deficiencies is to use tables of response surface coefficients, from which estimated quantiles for any sample size may be calculated, instead of equations to describe the distributions in question. In effect, these tables, the construction of which is discussed in Section 4, provide numerical distribution functions. The tables are sufficiently large that it would make no sense to print them, but they are not so large that modern computers should have any trouble dealing with them. A computer program can easily read them and use them to compute critical values or P values, either for the asymptotic case or for any reasonable sample size. Using this program is no harder than using a program to compute the approximations derived in MacKinnon (1994), and it is significantly easier than trying to compute the latter by hand.

Both the tables of estimated response surface coefficients and a computer program called `urcdist` that uses them are available via the Internet. Interested readers should ftp to **qed.econ.queensu.ca** and then go to the directory `pub/uroot`. This directory contains the fortran source code

for `urcdist`, a compiled version of the program for DOS-based personal computers with at least 4 MB of memory and a numeric coprocessor, and 12 zipped files containing the estimated response surface coefficients. The `urcdist` program is run interactively and prompts the user for input. Those who wish to compute large numbers of critical values or P values should write their own main programs to call the routines in `urcdist`.

2. Unit Root and Cointegration Tests

The literature on unit root and cointegration tests is enormous and growing rapidly. Banerjee, Dolado, Galbraith, and Hendry (1993) is a reasonably accessible reference. In this paper, I deal only with the distributions of tests that are well known, widely used, and easy to compute. More recently proposed tests, which have different asymptotic distributions, are not dealt with; see, for example, Pantula, Farias-Gonzales, and Fuller (1994).

A Dickey-Fuller test of the null hypothesis that the series y_t has a unit root may be based on OLS estimates of any of the following regressions:

$$\Delta y_t = (\alpha - 1)y_{t-1} + u_t \tag{1}$$

$$\Delta y_t = \beta_0 + (\alpha - 1)y_{t-1} + u_t \tag{2}$$

$$\Delta y_t = \beta_0 + \beta_1 t + (\alpha - 1)y_{t-1} + u_t \tag{3}$$

$$\Delta y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + (\alpha - 1)y_{t-1} + u_t, \tag{4}$$

where $\Delta y_t \equiv y_t - y_{t-1}$, t is a linear time trend, u_t is an error term, and α is a parameter that equals unity under the null hypothesis. The u_t's must be independent, but neither normality nor homoskedasticity needs to be assumed. The number of observations actually used in the regressions is T.

There are two types of DF tests, one based on t statistics and one based directly on the estimate of α. I shall refer to these as τ tests and z tests, respectively. For the former, the test statistic is the ordinary t statistic for $\alpha - 1$ to equal zero, and for the latter it is $T(\hat{\alpha} - 1)$. Following MacKinnon (1994), I shall refer to the τ statistics based on equations (1) through (4) as τ_{nc}, τ_c, τ_{ct}, and τ_{ctt}, respectively, and to the corresponding z statistics as z_{nc}, z_c, z_{ct}, and z_{ctt}. The subscripts stand for "no constant," "constant," "constant and trend," and "constant, trend, and trend squared."

Equations (1) through (4) impose successively less restrictive assumptions. Equation (1) makes sense for $\alpha < 1$ only if y_t has (population) mean zero. In contrast, equation (2) allows y_t to have a nonzero mean, (3) allows it to have a trend, and (4) allows it to have a trend that changes over time, under both the null and alternative hypotheses. The most commonly

encountered tests are based on equations (2) and (3). Ouliaris, Park, and Phillips (1989) have advocated tests based on (4).

The tests described above require that the u_t's be serially independent. When this assumption is unreasonable, as it often is, there are two asymptotically equivalent approaches. One is to employ "nonparametric" τ or z tests, as proposed by Phillips (1987) and Phillips and Perron (1988). An easier approach is to use "augmented " Dickey-Fuller, or ADF, tests, in which lags of Δy_t are added to equations (1) through (4) so as to whiten the residuals. The τ statistics, computed as ordinary t statistics, remain asymptotically valid in the presence of serial correlation when this is done, provided the number of lags of Δy_t is allowed to increase at an appropriate rate. Asymptotically valid z statistics may be obtained by dividing T times the coefficient on y_{t-1} by one minus the sum of the coefficients on the lags of Δy_t. The results of ADF tests can be quite sensitive to the way the number of lags is chosen; see Ng and Perron (1995).

Engle and Granger (1987) developed tests of the null hypothesis that two or more integrated time series are not cointegrated. Let $\boldsymbol{Y}$ denote a $T \times l$ matrix of observations on l time series that are believed to be $I(1)$. Then if $\boldsymbol{y}_1$ denotes one column of $\boldsymbol{Y}$, $\boldsymbol{Y}_1$ denotes the remaining $l-1$ columns, and $\boldsymbol{X}$ denotes a matrix of nonstochastic regressors such as a constant and possibly one or more trend terms, the equation

$$\boldsymbol{y}_1 = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Y}_1\boldsymbol{\eta}_1 + \boldsymbol{\nu} \tag{5}$$

can be estimated by OLS. If all the variables in $\boldsymbol{Y}$ are cointegrated, equation (5) is a cointegrating equation, and the error vector $\boldsymbol{\nu}$ should be stationary. Otherwise, $\boldsymbol{\nu}$ must have a unit root. Thus the null hypothesis of noncointegration may be tested by using a DF or ADF test on the residuals from OLS estimation of (5). For the ADF case, the test regression is

$$\Delta\hat{\nu}_t = (\alpha - 1)\hat{\nu}_{t-1} + \sum_{j=1}^{J} \gamma_j \Delta\hat{\nu}_{t-j} + \text{residual}, \tag{6}$$

where $\hat{\nu}_t$ denotes the t^{th} residual from OLS estimation of equation (5), and $\Delta\hat{\nu}_{t-j} = \hat{\nu}_{t-j} - \hat{\nu}_{t-j-1}$. An alternative to including the $\Delta\hat{\nu}_{t-j}$'s is to use "nonparametric" τ and z tests, as proposed by Phillips and Ouliaris (1990).

The asymptotic distributions of these tests depend on k, which is equal to one more than the number of elements of $\boldsymbol{\eta}_1$ that have to be estimated, so that $k = l$ when all elements of $\boldsymbol{\eta}_1$ are unknown. They also depend on the form of the matrix $\boldsymbol{X}$, which may be empty or may consist of a constant, a constant and a linear trend, or a constant, a linear trend, and a quadratic trend, by analogy with (1) through (4). Unless $k = 1$, these

asymptotic distributions are not the same as those of the DF and ADF tests. In MacKinnon (1991, 1994), I only considered $k = 1, \ldots, 6$, but in this paper I obtain results for $k = 1, \ldots, 12$.

The techniques of this paper could be applied to other types of cointegration tests, such as the VAR-based ones of Johansen (1991) and the ECM tests discussed by Kremers, Ericsson, and Dolado (1992). In related work, which is joint with Neil Ericsson and is still in progress, I have in fact applied these techniques to ECM tests.

3. The Simulation Experiments

The simulation experiments which are at the heart of this paper are similar to, but considerably more extensive than, those used in my earlier papers. Each experiment involves 200,000 replications. For the unit root tests, there are 100 experiments for each of 14 sample sizes, and for the cointegration tests, there are 50 experiments for each of up to 14 sample sizes. How the sample sizes were chosen will be discussed in the next section. The number of experiments is the same as in MacKinnon (1994). However, in order to obtain estimates that are reasonably accurate even in the tails of the distributions, the number of replications per experiment is four times as large.

There were several reasons for doing 50 or 100 experiments for each set of test statistics instead of a single experiment with 10 million or 20 million replications. First, the finite size of computer memories means that it would have been quite difficult to handle that many replications at once. Secondly, the observed variation among the 50 or 100 experiments provides an easy way to measure experimental randomness. Finally, it was sometimes convenient to be able to interrupt the computer programs without losing results for experiments that had already been completed. The experiments were performed on several different IBM RS/6000 workstations over a period of several months. Because some of the workstations were faster than others, it is difficult to estimate total CPU time. A rough estimate is 1360 hours on the fastest of the machines used (a Model 3AT), or 3130 hours on the slowest (a Model 355).

Because so many random numbers were used, it was vital to use a pseudo-random number generator with a very long period. The generator I used was also used in MacKinnon (1994). It combines two different uniform pseudo-random number generators recommended by L'Ecuyer (1988). The two generators were started with different seeds and allowed to run independently, so that two independent uniform pseudo-random numbers were generated at once. The procedure of Marsaglia and Bray (1964) was then used to transform them into two $N(0, 1)$ variates.

It would have been totally impractical to store all the simulated test statistics. Therefore, for each experiment, 221 quantiles were estimated and stored. These quantiles were: .0001, .0002, .0005, .001, ..., .010, .015, ..., .985, .990, .991, ..., .999, .9995, .9998, .9999. The 221 quantiles provide more than enough information about the shapes of the cumulative distribution functions of the various test statistics. Storing these estimated quantiles for each set of 50 (or 100) experiments required about 11 (or 22) megabytes of disk space.

4. Response Surface Estimation

The estimated finite-sample quantiles from the simulation experiments were used to estimate response surfaces in which the quantiles of the asymptotic distributions of the various test statistics appear as parameters. Consider the estimation of the p quantile for some test statistic. Let $q^p(T_i)$ denote the estimate of that quantile based on the i^{th} experiment, for which the sample size is T_i. Then the response surface regressions have the form

$$q^p(T_i) = \theta^p_\infty + \theta^p_1 T_i^{-1} + \theta^p_2 T_i^{-2} + \theta^p_3 T_i^{-3} + \varepsilon_i. \tag{7}$$

The first parameter here, θ^p_∞, is the p quantile of the asymptotic distribution, which is what we are trying to estimate. The other three parameters allow the finite-sample distributions to differ from the asymptotic ones. In MacKinnon (1991, 1994), equation (7) with $\theta^p_3 = 0$ was employed, and it was generally found to work well. However, it does not always work sufficiently well when the experiments involve very small sample sizes, especially when the number of possibly cointegrated variables is large. That is why, in this paper, $\theta^p_3 = 0$ is not always set to zero.

In MacKinnon (1994), the smallest sample size used in the experiments was $T = 50$, and the choice of sample sizes was rather arbitrary. Since the functional form of the response surface regressions is known to be (7), it is possible to choose sample sizes somewhat more scientifically. If we write (7) as

$$\boldsymbol{q}^p = \boldsymbol{Z\theta} + \boldsymbol{\varepsilon} = \theta^p_\infty + \boldsymbol{Z}^*\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}, \tag{8}$$

it is easy to derive the standard error of the OLS estimate of θ^p_∞. This standard error will be proportional to $(\boldsymbol{\iota}'\boldsymbol{M}^*\boldsymbol{\iota})^{-1/2}$, where $\boldsymbol{\iota}$ is a vector of ones and $\boldsymbol{M}^* = \mathbf{I} - \boldsymbol{Z}^*(\boldsymbol{Z}^{*\prime}\boldsymbol{Z}^*)^{-1}\boldsymbol{Z}^{*\prime}$.

For any possible set of m T_i's, it is easy to evaluate the standard error of θ^p_∞ and the computation cost of performing that set of experiments. As a rough approximation, it appears that the computation cost for a sample of size T is proportional to $T + 16$. Thus what we wish to minimize is

the product of the square of the standard error and the computation cost, which is proportional to

$$(\boldsymbol{\iota}'\boldsymbol{M}^*\boldsymbol{\iota})^{-1}\left(\sum_{i=1}^{m}(T_i+16)\right). \tag{9}$$

This expression was evaluated for 50,000 sets of randomly chosen T_i's, and several interesting results emerged. First of all, it is extremely desirable for there to be some small values of T_i. The smaller the smallest value of T_i, the more trouble $\boldsymbol{Z}^*$ has explaining a constant term, and thus the larger is $\boldsymbol{\iota}'\boldsymbol{M}^*\boldsymbol{\iota}$. Of course, if some of the values of T_i are too small, equation (7) may not fit satisfactorily. Secondly, it is also desirable for there to be some values of T_i that are reasonably large, although it does not appear to be cost-effective to use values as large as 1000. Finally, it does not appear to be cost-effective to use certain intermediate values of T_i, such as ones between 50 and 80 or ones between 100 and 400.

Based on these results, I used the following 14 different values of T_i in the simulations: 20, 25, 30, 35, 40, 45, 50, 80, 90, 100, 400, 500, 600, 700. For cointegration tests with larger values of k, it was generally necessary to omit some of the smaller values of T_i in order to obtain response surfaces that fit acceptably well. Indeed, it was because of this phenomenon that the values 45 and 50 were included. This set of sample sizes worked much better than the set used in MacKinnon (1994). Expression (9) was reduced by up to 54%, depending on whether θ_3^p in (7) was set to zero and how many of the smallest sample sizes had to be dropped.

Equation (7) was estimated 221 times for each of 96 different test statistics. There were up to 1400 observations for each of the unit root tests and up to 700 for each of the cointegration tests. In MacKinnon (1994), I used the GMM estimator of Cragg (1983) to allow for the fact that the error terms of (7) are heteroskedastic. Suppose that $\boldsymbol{\Omega}$ denotes the covariance matrix of the error vector $\boldsymbol{\epsilon}$ in equation (8). Since all the experiments are independent, $\boldsymbol{\Omega}$ is certainly a diagonal matrix, but it is not the identity matrix. The Cragg estimator can be written as

$$\tilde{\boldsymbol{\theta}} = \left(\boldsymbol{Z}'\boldsymbol{W}(\boldsymbol{W}'\hat{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{W}(\boldsymbol{W}'\hat{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{q}^p. \tag{10}$$

Here $\boldsymbol{W}$ is a matrix of up to 14 zero-one dummy variables, the first equal to 1 when $T_i = 20$, the second equal to 1 when $T_i = 25$, and so on, and $\hat{\boldsymbol{\Omega}}$ is a diagonal matrix, the principal diagonal of which consists of the squared residuals from an OLS regression of $\boldsymbol{q}^p$ on $\boldsymbol{W}$. The estimator (10) can easily be computed by a weighted least squares regression with as many observations as there are distinct values of T_i; see MacKinnon (1994) for details.

The problem with this approach is that it ignores some valuable information about $\boldsymbol{\Omega}$. The variances of the error terms in (7) are not homoskedastic only because they vary systematically with T. A more efficient estimate of $\boldsymbol{\Omega}$ may be obtained by regressing the diagonal elements of $\hat{\boldsymbol{\Omega}}$ on a constant and either powers of $1/T$ or powers of $1/(T-p)$, where p is the number of parameters estimated in the course of computing the test statistic under analysis. In practice, using $1/T$ and its square generally worked well for the z tests, and using $1/(T-p)$ and its square generally worked well for the τ tests. The fitted values from one of these auxiliary regressions are then used as the diagonal elements of $\tilde{\boldsymbol{\Omega}}$, which replaces $\hat{\boldsymbol{\Omega}}$ in the GMM estimator (10) above.

This GMM estimation procedure automatically generates a statistic for testing the specification of the response surface equation (7). This GMM test statistic, which is the minimized value of the objective function, is the quadratic form

$$(\boldsymbol{q}^p - \boldsymbol{Z}\boldsymbol{\theta})'\boldsymbol{W}(\boldsymbol{W}'\tilde{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}\boldsymbol{W}'(\boldsymbol{q}^p - \boldsymbol{Z}\boldsymbol{\theta}). \tag{11}$$

Standard results about GMM estimation imply that, under the null hypothesis that (7) is a correct specification, (11) is asymptotically distributed as $\chi^2(r)$, where r is equal to the number of distinct T_i's (which may be 14 or less) minus the number of parameters in (7).

The GMM test statistic (11) played a key role in the specification of the response surfaces. In order to avoid discontinuities caused by changes in functional form, the same response surface regression was estimated for every one of the 221 quantiles for a given distribution. The average value of the 221 test statistics was used to choose whether to set $\theta_3^p = 0$ in (7) and to determine how many small values of T_i to drop. The objective was to obtain efficient estimates of θ_∞^p. It was therefore desirable to set $\theta_3^p = 0$, if possible, and to throw out as few small T_i's as possible.

On average, for a correctly specified response surface, reducing by one the number of distinct T_i's, or dropping the constraint that $\theta_3^p = 0$ in (7), would be expected to reduce the value of (11) by 1.0, because the mean of a random variable with a $\chi^2(r)$ distribution is r. In most cases, I chose to reject a model when such a change reduced the value of (11) by more than 1.5, an amount that is, admittedly, somewhat arbitrary. For example, for the $\tau_c(2)$ test, estimating model (7) with $\theta_3^p = 0$ using all 14 sample sizes yielded an average GMM test statistic of 8.80. Dropping the data for $T_i = 20$ reduced this to 8.18, and estimating the full model using all the observations reduced it to 7.94. Since both these reductions in the GMM test statistic are well under 1.5, the preferred model has $\theta_3^p = 0$ and is estimated using all the data.

There were some very clear patterns in the response surface estimates. More values of T_i had to be dropped, and/or the restriction that $\theta_3^p = 0$ relaxed, as either k or the number of nonstochastic regressors was increased. Also, it was generally easier to obtain response surfaces that fit well for the τ tests than it was for the z tests. I sometimes used these empirical regularities to help decide which response surface regression and how many T_i's to use, by taking results for nearby tests into account.

All of the response surface estimates appear to be remarkably precise. At one extreme, for example, the .05 asymptotic critical value for the $\tau_c(1)$ test is estimated to be 2.8614 with a standard error of 0.000226. The response surface regression in this case has $\theta_3^p = 0$ and uses all 1400 observations. At the other extreme, the .05 critical value for the $z_{ctt}(11)$ test is estimated to be -79.052 with a standard error of 0.0113. The response surface regression in this case is unrestricted and uses only 450 observations; this was one of just two cases in which observations for $T_i = 40$ had to be dropped. Of course, because the standard errors for $\hat{\theta}_\infty^p$ reported by the GMM estimation procedure are based on the assumption that the response surfaces are specified correctly, they are probably somewhat too small.

5. Numerical Distributions

The principal results of this paper are 21,216 ($= 221 \times 8 \times 12$) sets of response surface regression coefficients. In the next section, I discuss how these may be used to obtain approximate P values or critical values. But first, it seems worthwhile to see what the distributions obtained in this way actually look like.

Figures 1 through 3 graph the asymptotic distribution functions of the $\tau_{nc}(k)$, $\tau_c(k)$, $\tau_{ct}(k)$, and $\tau_{ctt}(k)$ tests for $k = 1, \ldots, 12$. Each plotted curve simply joins the 221 estimated quantiles for a given test statistic, without any smoothing. One striking feature of these figures is the regular and predictable way in which all the curves move to the left as k increases. It seems plausible that we could estimate the asymptotic distributions of τ tests quite accurately for $k = 13$, $k = 14$, and probably even for larger values of k, without doing any more simulation experiments.

Another striking result is evident in Figure 3, where the distributions of the τ_{ct} tests are plotted as solid lines and those of the τ_{ctt} tests are plotted as dotted lines. This was not done simply to save space. From the figure, it is evident that, for large values of k but not for small ones, the distribution of $\tau_{ct}(k)$ is extremely similar to the distribution of $\tau_{ctt}(k-1)$. This also holds for the distributions of $\tau_c(k)$ and $\tau_{ct}(k-1)$ and for those of $\tau_{nc}(k)$ and $\tau_c(k-1)$. It seems plausible to speculate that, as $k \to \infty$, the effect of adding

the next higher level of constant or trend term becomes identical to the effect of adding an additional $I(1)$ variable in the cointegrating regression.

Figures 4 through 6 graph the asymptotic distribution functions of the $z_{nc}(k)$, $z_c(k)$, $z_{ct}(k)$, and $z_{ctt}(k)$ tests for $k = 1, \ldots, 12$. These distributions are much less symmetric than those of the τ tests, but in other respects they are similar. Once again, there is a steady and predictable movement to the left as k increases. Also, as k becomes large, the distribution of $z_{ct}(k)$ becomes extremely similar to the distribution of $z_{ctt}(k-1)$.

The response surface regressions allow us to graph finite-sample distributions as well as asymptotic ones. Of course, the former depend on the details of how the test statistic is computed and on the strong assumption of i.i.d. normal errors, and they may not be accurate for values of T much smaller than the smallest value used in estimating the response surface. The finite-sample distributions differ most strikingly from the asymptotic ones in the left-hand tails. For the τ tests, these differences are quite modest for $k = 1$, but they increase sharply as k increases. This is illustrated in Figure 7, which shows the left-hand tails of the distributions of $\tau_{ct}(1)$ and $\tau_{ct}(12)$ for various sample sizes. For the z tests, on the other hand, the differences between finite-sample and asymptotic distributions can be substantial even for $k = 1$. This is illustrated in Figure 8, which shows the left-hand tails of the distributions of $z_{ct}(1)$ for various sample sizes.

6. Local Approximations

The response surface coefficient estimates obtained in Section 4 may be used to obtain approximate P values and approximate critical values for 96 sets of asymptotic and finite-sample distributions. In the asymptotic case, the distribution is approximated by the 221 estimated $\hat{\theta}^p_\infty$'s from equation (7). In the finite-sample case, which necessarily requires much stronger assumptions, it is approximated by the fitted values from 221 estimations of one of these equations for a given sample size T.

In order to obtain a P value for any test statistic or a critical value for any desired test size, some procedure for interpolating between the 221 tabulated values is needed. Many such procedures could be devised, but the one I used has some theoretical appeal and seems to work well. First, consider the regression

$$\Phi^{-1}(p) = \gamma_0 + \gamma_1 \hat{q}(p) + \gamma_2 \hat{q}^2(p) + \gamma_3 \hat{q}^3(p) + e_p, \tag{12}$$

where $\Phi^{-1}(p)$ is the inverse of the cumulative standard normal distribution function, evaluated at p. Notice that, if the distribution from which the estimated quantiles were obtained were in fact normal with any mean and

variance, regression (12) would be correctly specified with $\gamma_2 = \gamma_3 = 0$. Since that is not the case here, this regression can only be valid as an approximation.

The observations over which equation (12) is estimated are indexed by p because that is the most natural way to index them. The idea is to estimate it using only a small number of points in the neighborhood of the test statistic that is of interest. For example, suppose the distribution of interest were the asymptotic distribution of $\tau_{ct}(1)$ and the actual value of the test statistic were -3.29. The estimated quantile closest to this is $\hat{q}(.07) = -3.2773$. Then if 9 points are to be used, (12) would be estimated using the data for $p = .050$, .055, .060, .065, .070, .075, .080, .085, and .090.

It may seem a bit odd that the regressors in (12) are stochastic and the regressand is not. However, because the estimated quantiles are very accurate, the errors in variables bias that this induces is trivially small. This point is discussed in MacKinnon (1994), in which regression (12) was used to obtain approximate asymptotic distribution functions for some of the tests dealt with in this paper.

If we are interested in obtaining approximate critical values, equation (12) has to be turned around. Consider the regression

$$\hat{q}_p = \delta_0 + \delta_1 \Phi^{-1}(p) + \delta_2 \big(\Phi^{-1}(p)\big)^2 + \delta_3 \big(\Phi^{-1}(p)\big)^3 + e_p^*. \tag{13}$$

This is not actually the inverse of equation (12). However, if the distribution from which the estimated quantiles were obtained were in fact normal with any mean and variance, equation (13) would be correctly specified with $\delta_2 = \delta_3 = 0$. In that case, equation (12) would have $\gamma_2 = \gamma_3 = 0$, and (13) would be the inverse of (12). It is worth noting that, in both equations (12) and (13), using $\Phi^{-1}(p)$ worked very much better than using p directly.

Regressions (12) and (13) could be estimated by OLS, but this would ignore both heteroskedasticity and serial correlation. In MacKinnon (1994), the former was taken into account, but the latter was ignored. Actually, it is quite easy to take account of both. It is well known (see, for example, Appendix 2 of Cox and Hinkley (1967)) that, asymptotically, the covariance between two quantiles $\hat{q}_i \equiv \hat{q}(p_i)$ and $\hat{q}_i \equiv \hat{q}(p_j)$, estimated by maximum likelihood from the same sample of size N, is

$$\operatorname{Cov}(\hat{q}_i, \hat{q}_j) \stackrel{a}{=} \frac{p_i(1 - p_j)}{N f\big(q(p_i)\big) f\big(q(p_j)\big)}, \tag{14}$$

where $p_j > p_i$, $f(q_i)$ denotes the density of the underlying random variable evaluated at $q(p_i)$, and "$\stackrel{a}{=}$" denotes asymptotic equality. Because the densities of the test statistics we are interested in are not known, equation

(14) is not directly applicable, but it implies that the correlation between $\hat{q}_i$ and $\hat{q}_j$ is

$$\rho(\hat{q}_i, \hat{q}_j) \stackrel{a}{=} \left(\frac{p_i(1-p_j)}{p_j(1-p_i)} \right)^{1/2}. \tag{15}$$

In addition, we have direct estimates of the standard errors of the $\hat{\theta}^p_\infty$'s from the estimation of (7). Using the fact that

$$\mathrm{Cov}(\hat{q}_i, \hat{q}_j) = \rho(\hat{q}_i, \hat{q}_j)\big(\mathrm{Var}(\hat{q}_i)\mathrm{Var}(\hat{q}_j)\big)^{1/2},$$

these may be combined with correlations estimated using (15) to yield an estimated covariance matrix, and regressions (12) and (13) may then be estimated by feasible GLS.

As discussed above, equations (12) and (13) are to be fitted only to a small number of points near the specified test statistic or test size. Experimentation suggests that 9 points is a good number to use. Also, in many cases, it is possible to set γ_3 or δ_3 equal to zero on the basis of a t test. These conclusions were obtained by estimating equations (12) and (13) for 221 estimated quantiles generated from simulations of the $\chi^2(3)$ and $\chi^2(10)$ distributions. These estimated quantiles were approximately as accurate as the ones from the response surfaces estimated in Section 4.

The approximate P values and critical values which emerge from equations (12) and (13) seem to be just about as accurate as the estimated quantiles on which they are based. This conclusion is based on evaluating 9801 evenly spaced points between 0.01 and 0.99 for the two sets of simulated data and comparing the resulting errors with those in the original estimated quantiles. Thus I am very confident that equations (12) and (13) provide a reliable way to obtain approximate P values and critical values. The program `urcdist` uses these equations for this purpose.

Equation (12) can be used to compute approximate densities as well as approximate P values. In order to calculate the P value for some observed test statistic, say τ_*, we simply estimate equation (12) using only values of $\hat{q}(p)$ near τ_* and then compute

$$P^* = \Phi\big(\hat{\gamma}_0 + \hat{\gamma}_1\tau_* + \hat{\gamma}_2\tau_*^2 + \hat{\gamma}_3\tau_*^3\big). \tag{16}$$

Notice that P^* is the value of the cumulative distribution function evaluated at τ_*. Therefore, the approximate density at τ_* is simply the first derivative of (16):

$$f(\tau_*) \cong \phi\big(\hat{\gamma}_0 + \hat{\gamma}_1\tau_* + \hat{\gamma}_2\tau_*^2 + \hat{\gamma}_3\tau_*^3\big)\big(\hat{\gamma}_1 + 2\hat{\gamma}_2\tau_* + 3\hat{\gamma}_3\tau_*^2\big). \tag{17}$$

By using equations (12) and (17) along with the estimated response surface coefficients of Section 4, it is possible to plot the asymptotic or finite-sample

densities of any of the 96 test statistics studied in this paper. For example, Figure 9 plots the asymptotic density of the $\tau_c(1)$ test and its density for $T = 20$, and Figure 10 plots the corresponding densities for the $z_c(1)$ test. Note that, in order to obtain the $\hat{\gamma}_i$'s needed for these plots, 15 points were used when estimating equation (12), and γ_3 was never set to zero. This resulted in somewhat smoother-looking densities than using only 9 points, as the `urcdist` program does.

7. An Empirical Example

To illustrate the use of the `urcdist` program, I calculated a number of unit root tests and used the program to compute the corresponding P values. The series being tested for a unit root is the Canadian 91-day Treasury Bill rate (CANSIM Number B14001). The original data are monthly, but I also aggregated them to quarterly and annual frequencies. The sample period was 1957 to 1993. When lags of Δy_t were added to whiten the residuals, pre-sample observations were used, so that the number of lags did not affect the sample size. For each test statistic, two variants are reported. The first of these uses the smallest number of lagged Δy_t's that appear to be consistent with the data, based on t tests at the 5% level, while the second uses one more lag of Δy_t. When computing finite-sample P values for test regressions which included lags of Δy_t, I used n minus the number of lags rather than n as the sample size.

Table 1. Unit Root Test Results: Annual Data

Statistic Used	Value	Asymptotic P	Finite-sample P
τ_c (no lags)	−1.775	0.3934	0.3866
τ_c (1 lag)	−2.000	0.2871	0.2856
z_c (no lags)	−6.339	0.3214	0.2940
z_c (1 lag)	−7.842	0.2264	0.1970
τ_{ct} (no lags)	−1.770	0.7195	0.6988
τ_{ct} (1 lag)	−2.177	0.5019	0.4874
z_{ct} (no lags)	−9.143	0.4970	0.4453
z_{ct} (1 lag)	−17.135	0.1239	0.0732
τ_{ctt} (no lags)	−2.214	0.7226	0.6954
τ_{ctt} (1 lag)	−2.550	0.5401	0.5204
z_{ctt} (no lags)	−11.410	0.6242	0.5557
z_{ctt} (1 lag)	−19.510	0.2146	0.1226

Table 2. Unit Root Test Results: Monthly Data

Statistic Used	Value	Asymptotic P	Finite-sample P
τ_c (1 lag)	−2.408	0.1395	0.1401
τ_c (2 lags)	−2.213	0.2017	0.2020
z_c (1 lag)	−11.702	0.0896	0.0876
z_c (2 lags)	−9.976	0.1360	0.1338
τ_{ct} (1 lag)	−2.617	0.2726	0.2729
τ_{ct} (2 lags)	−2.310	0.4279	0.4271
z_{ct} (1 lag)	−17.484	0.1158	0.1116
z_{ct} (2 lags)	−14.513	0.2027	0.1977
τ_{ctt} (1 lag)	−3.198	0.2099	0.2111
τ_{ctt} (2 lags)	−2.896	0.3479	0.3479
z_{ctt} (1 lag)	−22.532	0.1321	0.1253
z_{ctt} (2 lags)	−19.230	0.2240	0.2161

Table 1 reports results for the annual data, and Table 2 reports results for the monthly data. Results for the quarterly data were very similar to those for the monthly data, and they are therefore not reported. The annual and monthly results are also quite similar, as might be expected from the analyses of Shiller and Perron (1985) and Pierse and Snell (1995). Because it is not entirely clear how many trends should be included in the regression, three sets of results are reported. These results are all broadly consistent with each other. The P values do vary a certain amount, but the null hypothesis is never rejected at the 5% level. Interestingly, the P values are almost always lower for the z tests than they are for the τ tests based on the same regressions.

The differences between asymptotic and finite-sample P values are always quite small for the τ tests (recall Figure 7), but they are sometimes quite large for the z tests (recall Figure 8). The exceptionally small differences between asymptotic and finite-sample P values for the τ_c tests on annual data are actually a bit misleading. For these tests, the asymptotic and finite-sample distributions happen to cross at values of the test statistic quite close to those observed here. Of course, except for the annual results with no lagged Δy_t's, the finite-sample P values are not really valid, because the actual distributions would depend on the number of lags and the data generating process.

8. Conclusion

In this paper, I have computed response surface coefficients which provide excellent approximations to the asymptotic and finite-sample distributions of 8 varieties of Dickey-Fuller unit root and cointegration tests for up to 12 possibly cointegrated variables. Although the approach of the paper is fundamentally similar to that of MacKinnon (1991, 1994), the paper does contain some innovations. The principal innovation is that the results consist chiefly of tables of estimated coefficients, which are available via the Internet, and a computer program that uses these to calculate critical values and P values. Another innovation is the use of feasible GLS to estimate approximating regressions so as to obtain approximate critical values, approximate P values, and even approximate densities, using a finite number of estimated quantiles from the response surfaces.

References

Banerjee, A., Dolado, J. J., Galbraith, J. W, and Hendry, D. F. (1993). *Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data*, Oxford, Oxford University Press.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London, Chapman and Hall.

Cragg, J. G. (1983), "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 51, 751–763.

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

Engle, R. F., and Granger, C. W. J. (1987), "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

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

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegration in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

Kremers, J. J. M., Ericsson, N. R., and Dolado, J. J. (1992), "The Power of Cointegration Tests," *Oxford Bulletin of Economics and Statistics*, 54, 325–348.

L'Ecuyer, P. (1988), "Efficient and Portable Random Number Generators," *Communications of the ACM*, 31, 742–751.

MacKinnon, J. G. (1991), "Critical Values for Cointegration Tests," in *Long-run Economic Relationships: Readings in Cointegration*, eds. R. F. Engle and C. W. J. Granger, Oxford: Oxford University Press, pp. 267–276.

MacKinnon, J. G. (1994), "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests," *Journal of Business and Economic Statistics*, 12, 167–176.

Marsaglia, G., and Bray, T. A. (1964), "A Convenient Method for Generating Normal Variables," *SIAM Review*, 6, 260–264.

Ng, S., and Perron, P. (1995), "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, forthcoming.

Ouliaris, S., Park, J. Y., and Phillips, P. C. B. (1989), "Testing for a Unit Root in the Presence of a Maintained Trend," in *Advances in Econometrics*, ed. B. Raj, Boston: Klumer Academic Publishers, pp. 7–28.

Pantula, S. G., Farias-Gonzales, G., and Fuller, W. A. (1994), "A Comparison of Unit-Root Test Criteria," *Journal of Business and Economic Statistics*, 12, 449–459.

Pierse, R. G., and Snell, A. J. (1995), "Temporal Aggregation and the Power of Tests for a Unit Root," *Journal of Econometrics*, 65, 333–345.

Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.

Phillips, P. C. B. (1987), "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

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

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

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

Shiller, R. J., and Perron, P. (1985), "Testing the Random Walk Hypothesis: Power versus Frequency of Observation," *Economics Letters*, 18, 381–86.

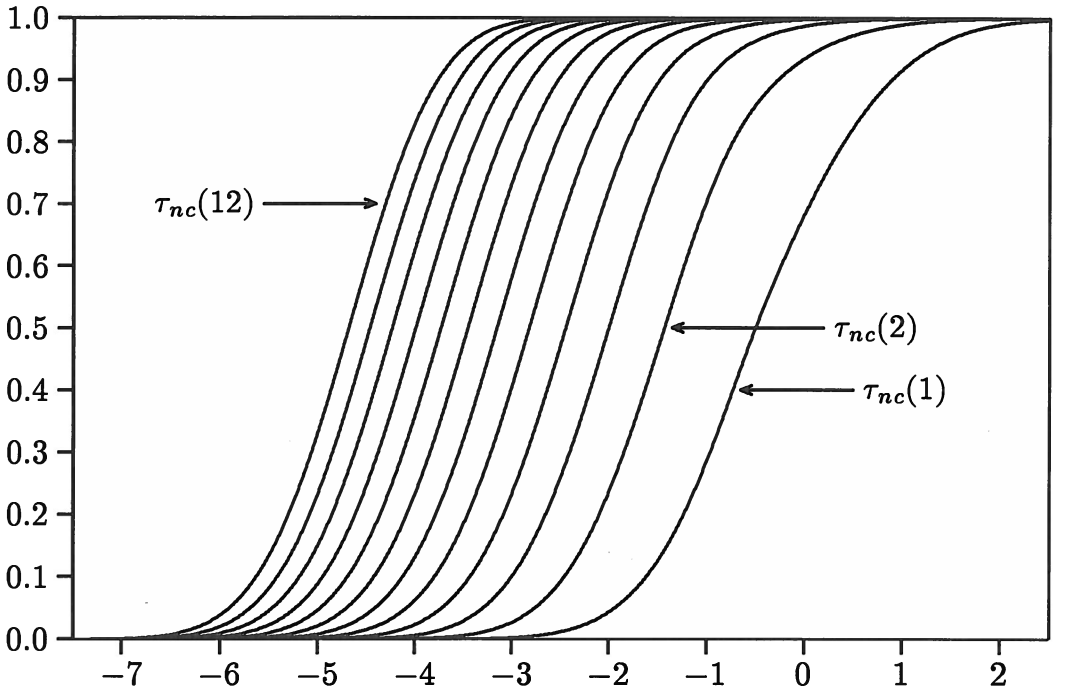

Figure 1. Asymptotic distributions of τ_{nc} tests

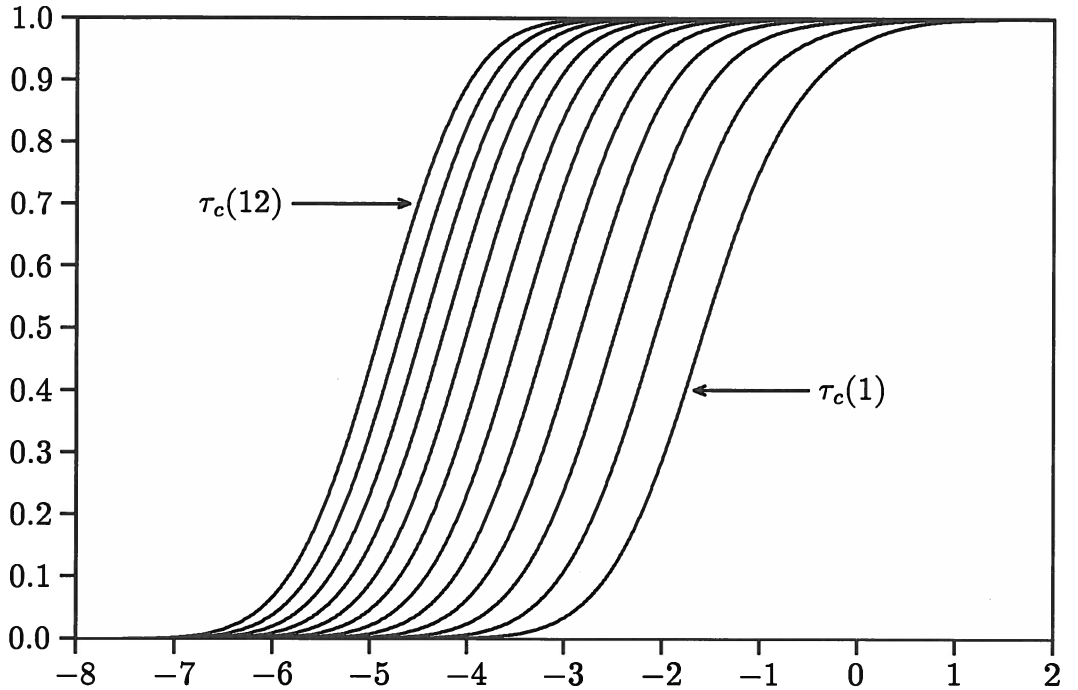

Figure 2. Asymptotic distributions of τ_c tests

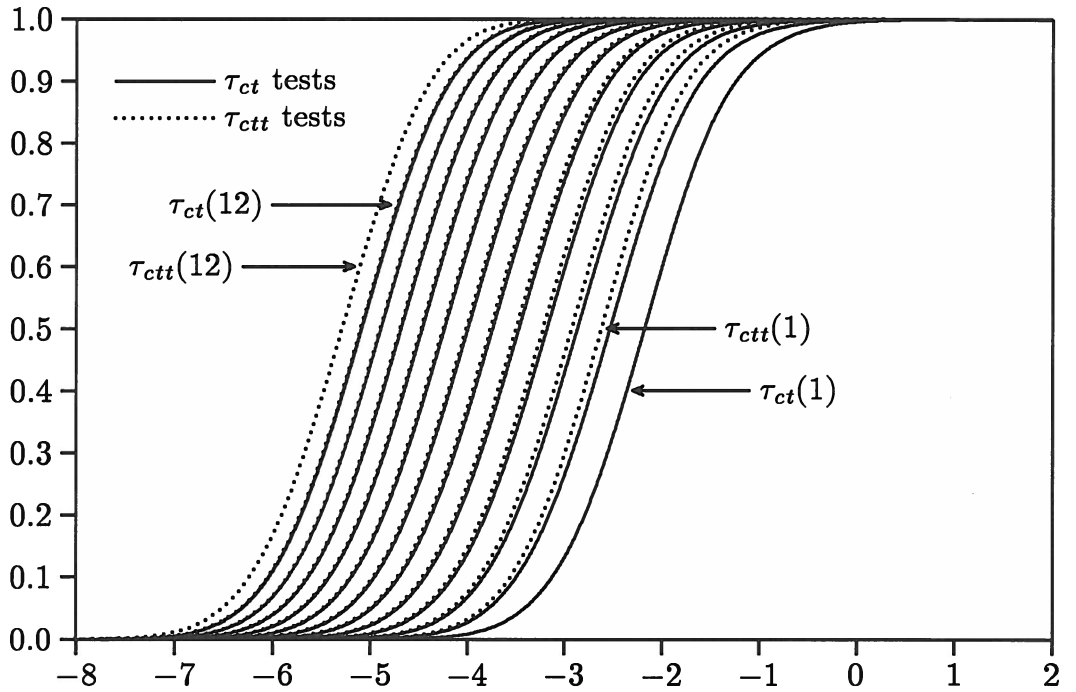

Figure 3. Asymptotic distributions of τ_{ct} and τ_{ctt} tests

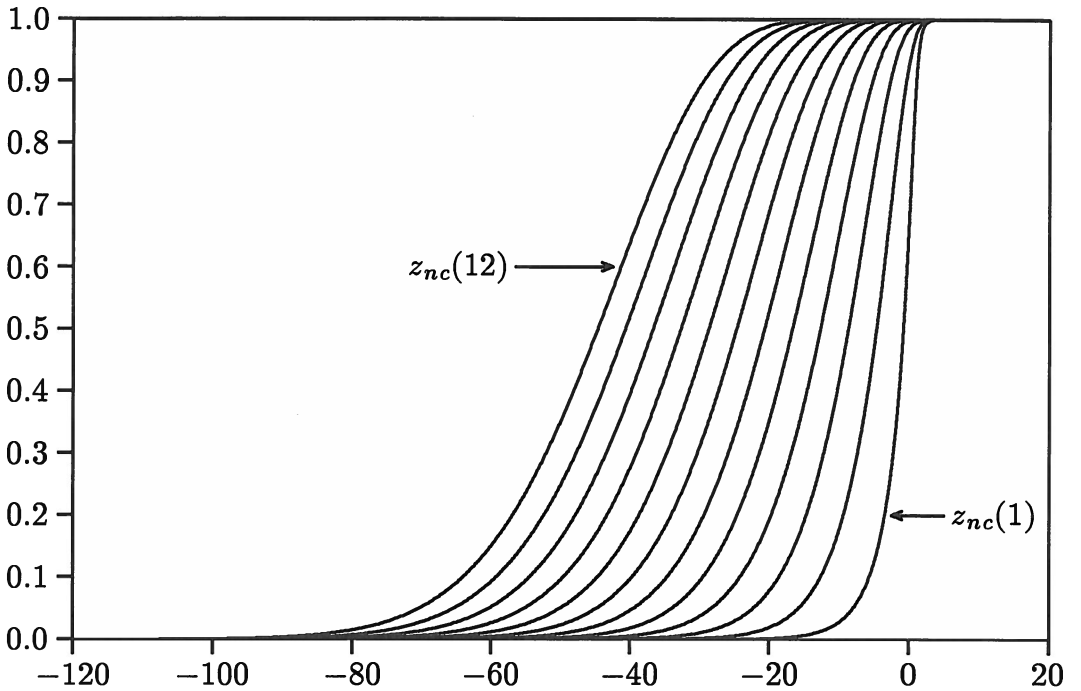

Figure 4. Asymptotic distributions of z_{nc} tests

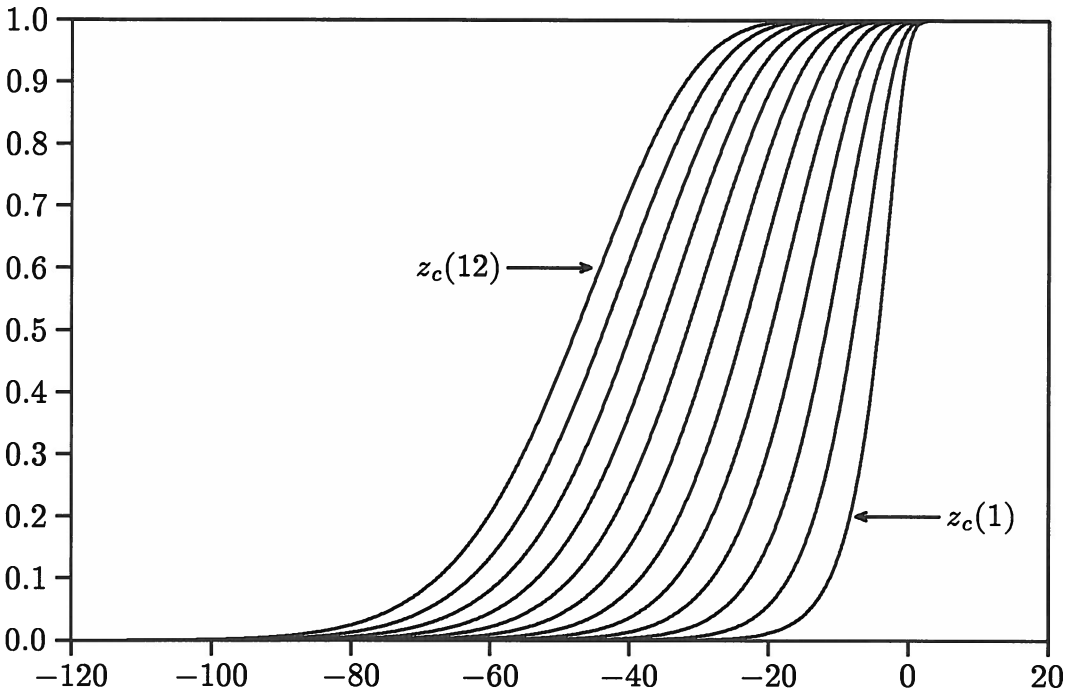

Figure 5. Asymptotic distributions of z_c tests

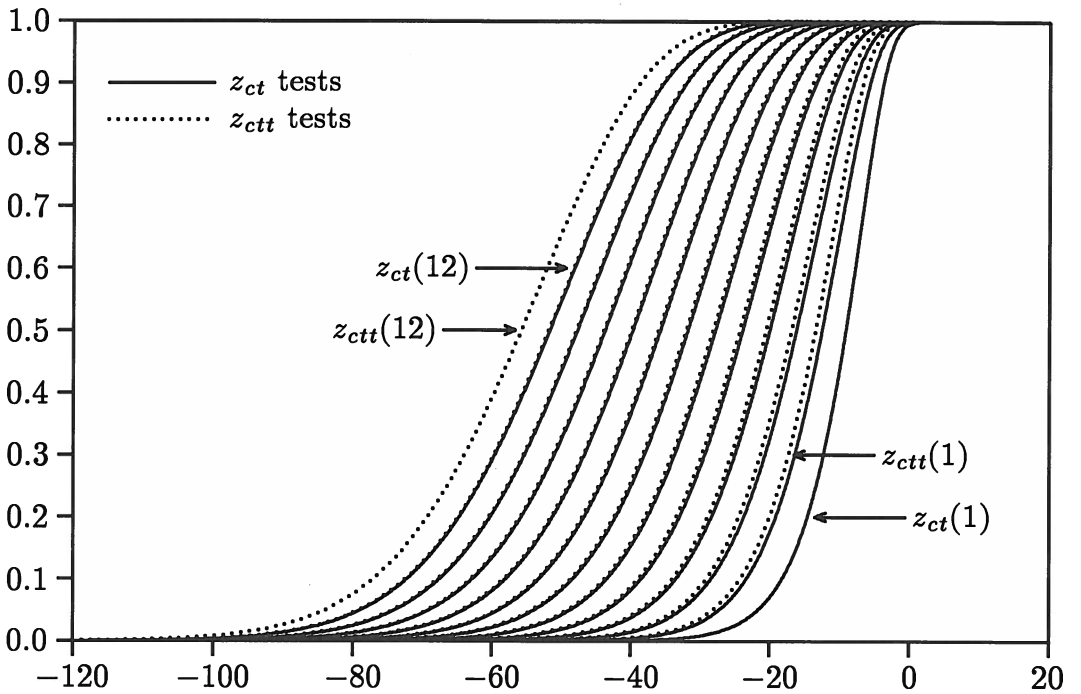

Figure 6. Asymptotic distributions of z_{ct} and z_{ctt} tests

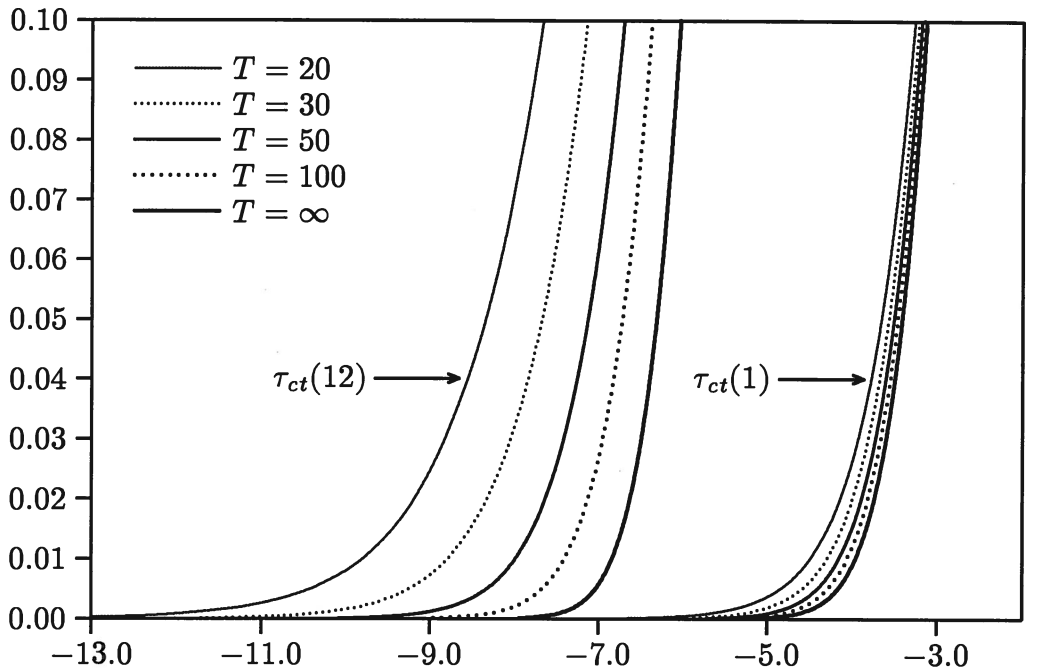

Figure 7. Left-hand tails of distributions of τ_{ct} tests

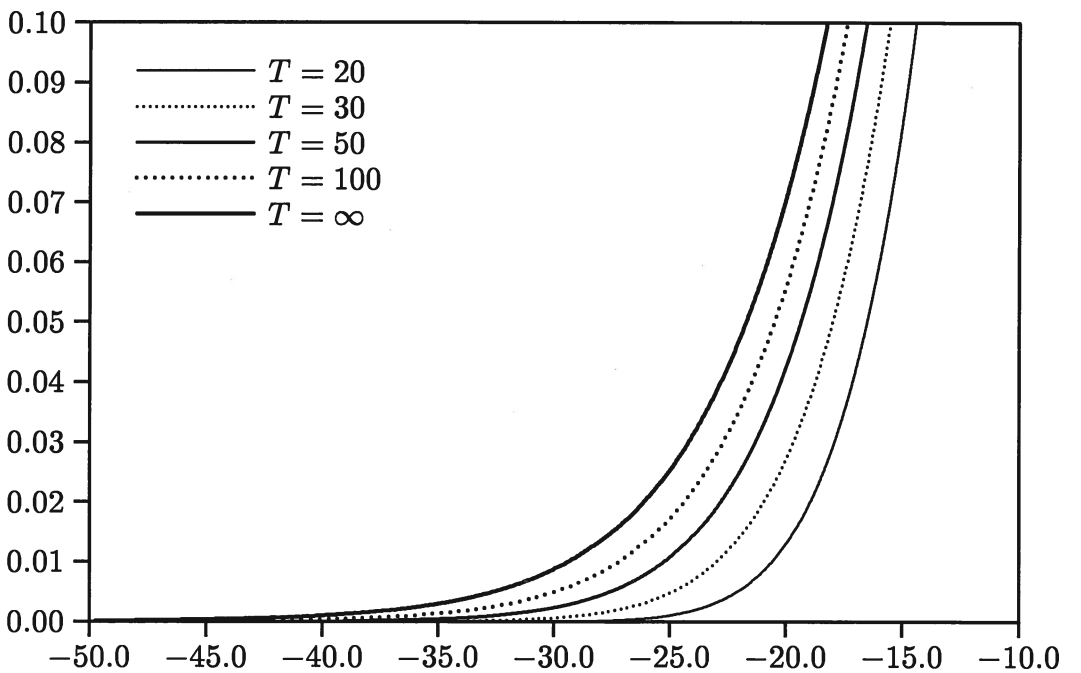

Figure 8. Left-hand tails of distributions of $z_{ct}(1)$ tests

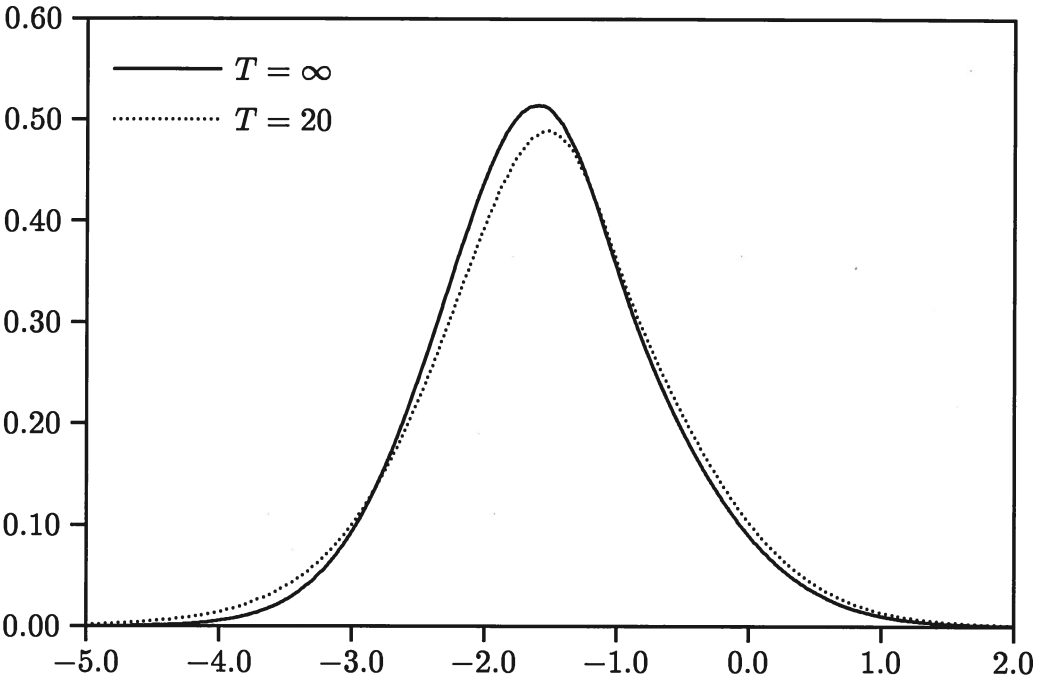

Figure 9. Densities of $\tau_{\mathbf{c}}(1)$ tests

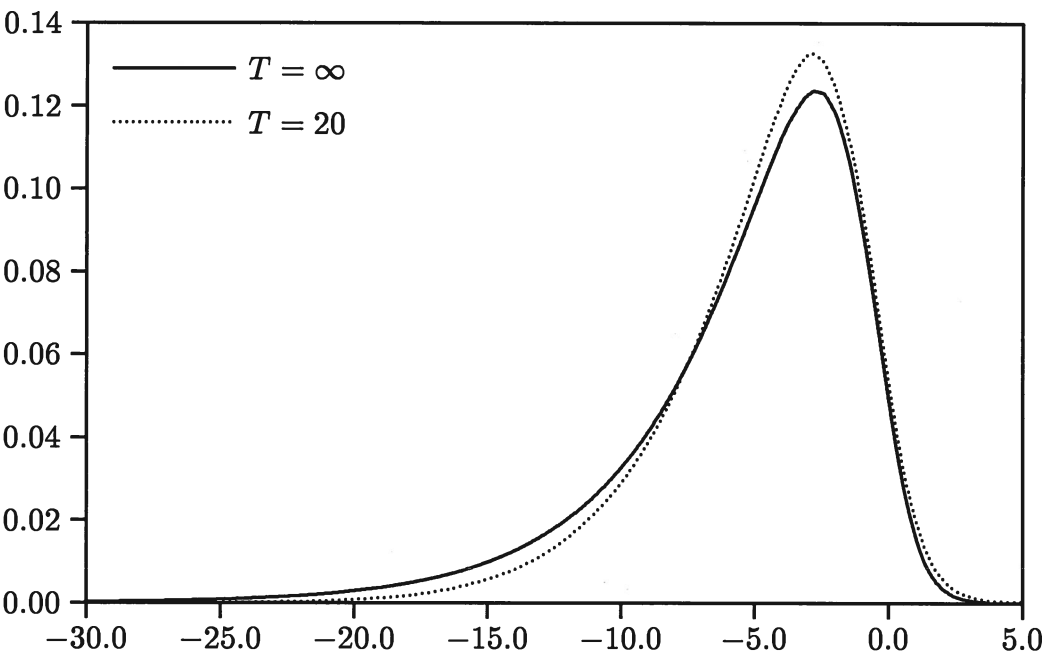

Figure 10. Densities of $z_{\mathbf{c}}(1)$ tests