Numerical Distribution Functions for Unit Root and Cointegration Tests

James G. MacKinnon

Journal of Applied Econometrics, Vol. 11, No. 6 (Nov. - Dec., 1996), 601-618.

NUMERICAL DISTRIBUTION FUNCTIONS FOR UNIT ROOT AND COINTEGRATION TESTS

JAMES G. MACKINNON

Department of Economics, Queen's University, Kingston, Ontario, Canada K7L 3N6
E-mail: jgm@qed.econ.queensu.ca

SUMMARY

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. An empirical example deals with interest rates and inflation rates in Canada.

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 non-standard distributions, even asymptotically. The cointegration tests developed by Engle and Granger (1987) are closely related to DF and ADF tests, but they have different, non-standard 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 extend the results to allow for up to 12 variables, instead of six, and I correct two deficiencies of the 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; for details, see the note at the end of this paper. The urcdist program is run interactively and prompts the user for input. For those who wish to compute large numbers of critical values or P-values, a set of routines called urcrouts.f is also provided. These users simply need to write their own main programs to call the routine urcval, which in turn reads the appropriate files of response surface coefficients and calls other routines to do the calculations.

2. UNIT ROOT AND COINTEGRATION TESTS

The literature on unit root and cointegration tests is enormous and growing rapidly. Banerjee *et al.* (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 for asymptotic results neither normality nor homoscedasticity needs to be assumed.

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 $T(\hat{\alpha} - 1)$, where T is the number of observations used to estimate the test regression. Following MacKinnon (1994), I shall refer to the τ statistics based on equations (1) to (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) to (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 non-zero 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 equation (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 'non-parametric' 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) to (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 Y denote a $T \times l$ matrix of observations on l time series that are believed to be $I(1)$. Then if y_1 denotes one column of Y, Y_1 denotes the remaining $l - 1$ columns, and X denotes a matrix of non-stochastic regressors such as a constant and possibly one or more trend terms, the equation

$$y_1 = X\beta + Y_1\eta_1 + v \tag{5}$$

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

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

where $\hat{v}_t$ denotes the t_{th} residual from OLS estimation of equation (5), and $\Delta \hat{v}_{t-j} = \hat{v}_{t-j} - \hat{v}_{t-j-1}$. An alternative to including the $\Delta \hat{v}_{t-j}$'s is to use 'non-parametric' 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 η_1 that have to be estimated. Thus $k = l$ when all elements of η_1 are unknown. They also depend on the form of the matrix 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 equations (1) to (4). The tests will be referred to as $\tau_s(k)$ or $z_s(k)$, where $s = $ nc, c, ct, or ctt, as before. Unless $k = 1$, the asymptotic distributions of these tests 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 with various co-authors, which is still in progress, I have in fact applied these techniques to both these types of 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. Second, the observed variation among the 50 or 100 experiments provides an easy way to measure experimental randomness. Third, it was sometimes convenient to be able to divide the experiments among two or more computers. 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: 0·0001, 0·0002, 0·0005, 0·001, ..., 0·010, 0·015, ..., 0·985, 0·990, 0·991, ... 0·999, 0·9995, 0·9998, 0·9999. 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 ith 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, θ^p_3 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 equation (7), it is possible to choose sample sizes somewhat more scientifically. If we write equation (7) as

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

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

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

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

This expression was evaluated for 50,000 sets of randomly chosen T_i's, and several interesting results emerged. First, 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 $\mathbf{Z}^*$ has explaining a constant term, and thus the larger is $\iota'\mathbf{M}^*\iota$. Of course, if some of the values of T_i are too small, equation (7) may not fit satisfactorily. Second, 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. Third, 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 per cent, depending on whether or not θ_3^p in equation (7) was set to zero and on 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 1400 observations for each of the unit root tests and up to 700 observations 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 equation (7) are heteroscedastic. Suppose that $\mathbf{\Omega}$ denotes the covariance matrix of the error vector $\boldsymbol{\varepsilon}$ in equation (8). Since all the experiments are independent, $\mathbf{\Omega}$ is certainly a diagonal matrix, but it is not the identity matrix. The Cragg estimator can be written as

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

Here $\mathbf{W}$ 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 $\hat{\mathbf{\Omega}}$ is a diagonal matrix, the principal diagonal of which consists of the squared residuals from an OLS regression of $\mathbf{q}^p$ on $\mathbf{W}$. 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 $\mathbf{\Omega}$. The variances of the error terms in equation (7) are heteroscedastic only because they vary systematically with T. A more efficient estimate of $\mathbf{\Omega}$ may be obtained by regressing the square roots of the diagonal elements of $\hat{\mathbf{\Omega}}$ on a constant and either $1/T$ or $1/(T - b)$, where b is the number of parameters estimated in the course of computing the test statistic under analysis. In practice, using $1/T$ generally worked well for the z-tests, and using $1/(T - b)$ generally worked well for the τ-tests. The squares of the fitted values from one of these auxiliary regressions are then used as the diagonal elements of $\tilde{\mathbf{\Omega}}$, which replaces $\hat{\mathbf{\Omega}}$ 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

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

Standard results about GMM estimation imply that, under the null hypothesis that equation (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 equation (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 equation (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 equation (7), would be expected to reduce the value of equation (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 $\tau_c(3)$ 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 average 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 non-stochastic 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 the response surface estimates appear to be remarkably precise. At one extreme, for example, the 0·05 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 0·05 critical value for the $z_{ctt}(11)$ test is estimated to be $-79\cdot052$ 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. There are some striking regularities which it would be interesting to explain theoretically.

Figures 1 to 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
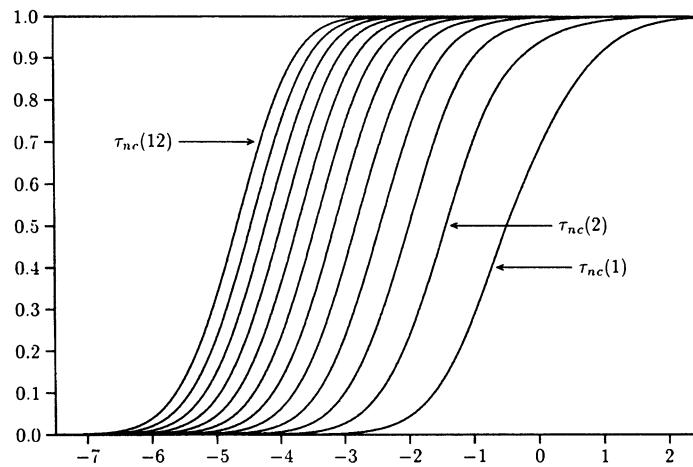
Figure 1. Asymptotic distributions of the τ_{nc} tests
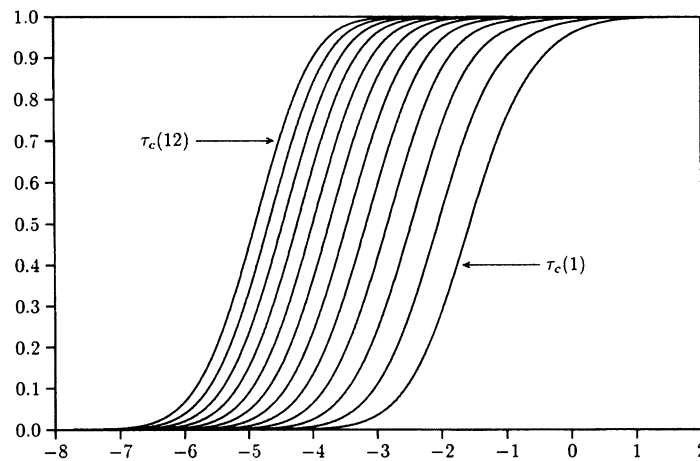
Figure 2. Asymptotic distributions of the τ_c tests
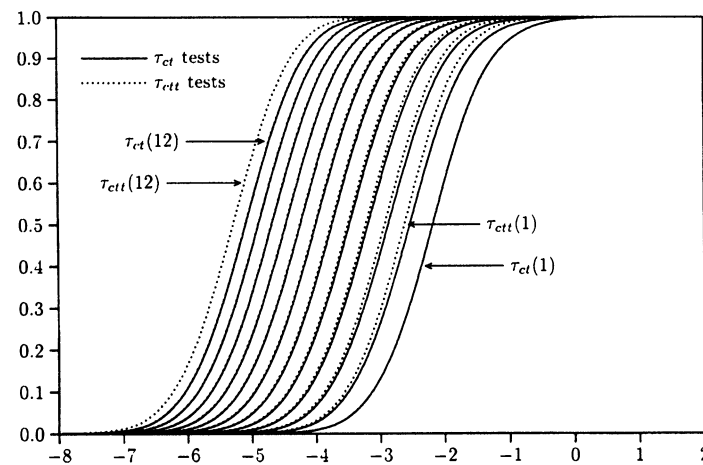
Figure 3. Asymptotic distributions of the τ_{ct} and τ_{ctt} tests
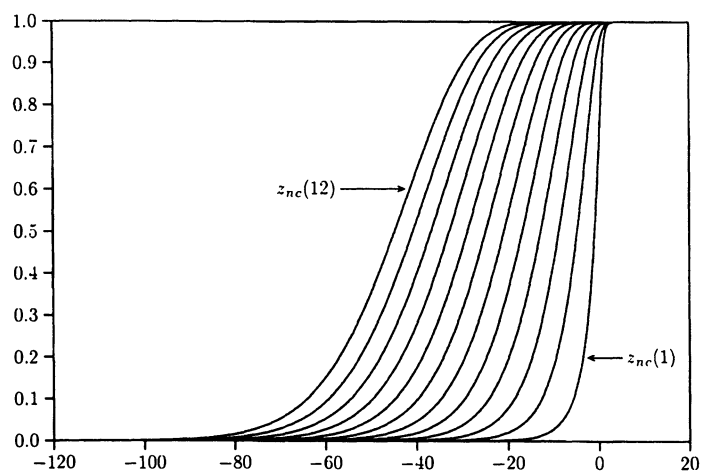
Figure 4. Asymptotic distributions of z_{nc} tests
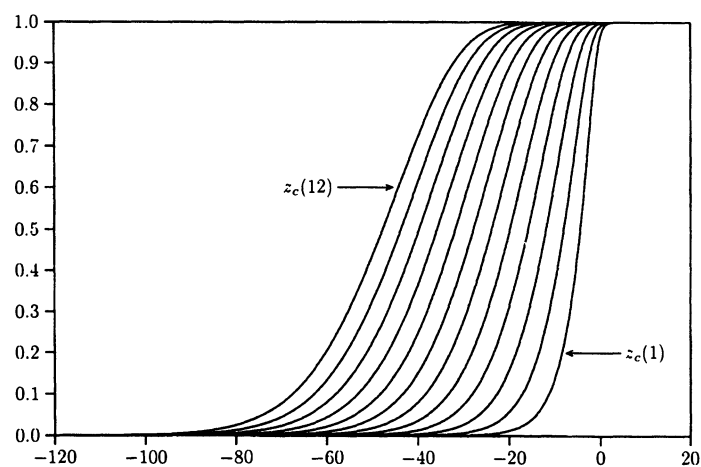
Figure 5. Asymptotic distributions of z_c tests
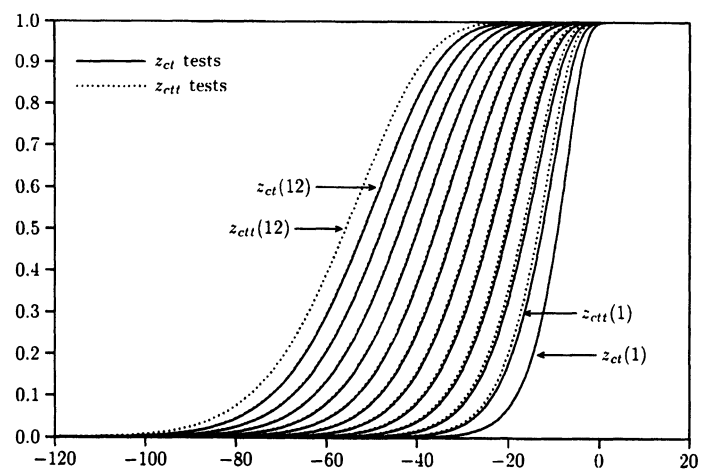
Figure 6. Asymptotic distributions of z_{ct} and z_{ctt} tests
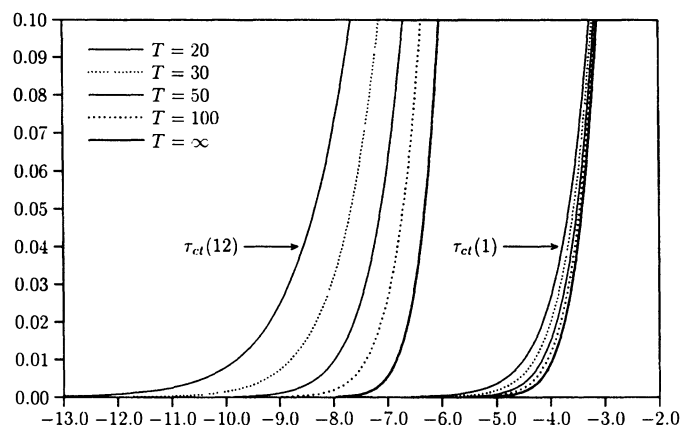
Figure 7. Left-hand tails of distributions of τ_{ct} tests
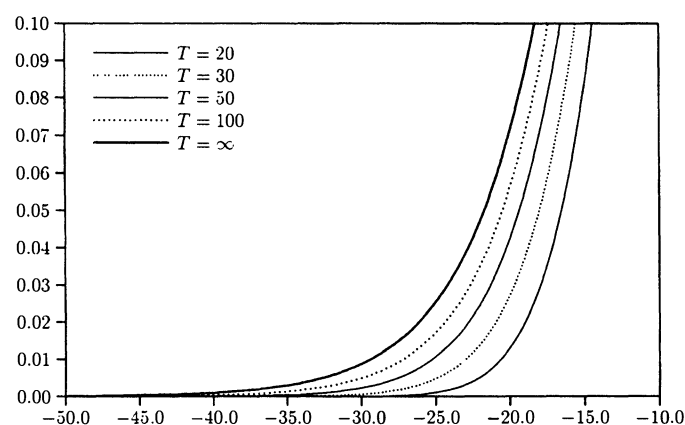
Figure 8. Left-hand tails of distributions of $z_{ct}(1)$ tests

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 $k \rightarrow \infty$, 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 to 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 coefficients 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 $\hat{\theta}^p_\infty$ 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

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

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 neighbourhood 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 $-3 \cdot 29$. The estimated quantile closest to this is $\hat{q}(\cdot 07) = -3 \cdot 2773$. Then if 9 points are to be used, (12) would be estimated using the data for $p = 0 \cdot 050$, $0 \cdot 055$, $0 \cdot 060$, $0 \cdot 065$, $0 \cdot 070$, $0 \cdot 075$, $0 \cdot 080$, $0 \cdot 085$, and $0 \cdot 090$.

It may seem rather odd that the regressors in equation (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

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

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 $\delta_2 = \delta_3 = 0$ and no error term. 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 heteroscedasticity 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 $\hat{q}_j \equiv \hat{q}(p_j)$, estimated by maximum likelihood from the same sample of size N, is

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

where $p_j > p_i$, $f(q_i)$ denotes the density of the underlying random variable evaluated at $q(p_i)$, and '$\overset{a}{=}$' 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

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

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

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

these may be combined with correlations estimated using equation (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 simulated estimated quantiles were chosen to be approximately as accurate as the ones from the response surfaces estimated in Section 4.
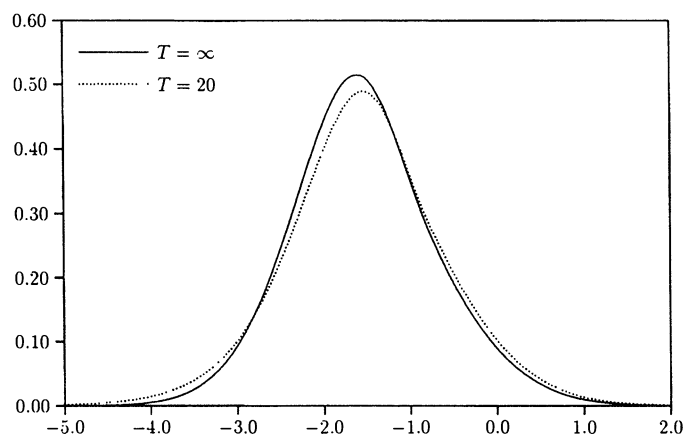
The approximate P-values and critical values which emerge from equations (12) and (13) seem to be just about as accurate as the estimated quantiles on which they are based. This conclusion is based on evaluating 9801 evenly spaced points between 0·01 and 0·99 for the two sets of simulated data and comparing 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

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

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 equation (16):

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

Figure 9. Densities of $\tau_c(1)$ tests
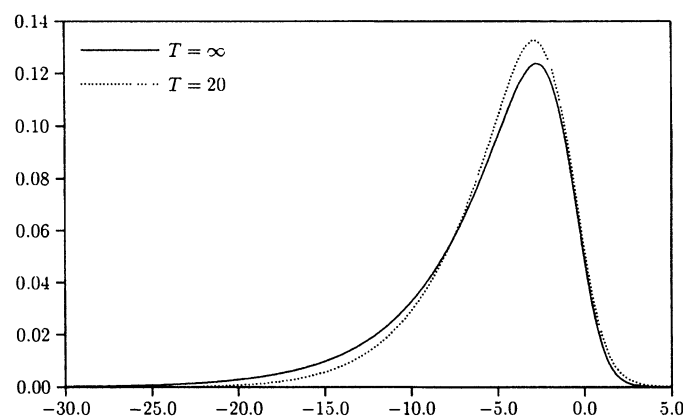
Figure 10. Densities of $z_c(1)$ tests

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

The empirical example of this section is primarily designed to illustrate how the urcdist program may be used. The example concerns interest rates and inflation rates in Canada, using annual data for 1954 to 1994. I examine three different interest rates, for government securities of different maturities, and two different inflation rates. The objective is to answer the following

questions:

(1) Do interest rates have a unit root?
(2) Do inflation rates have a unit root?
(3) Are the interest rate series cointegrated among themselves?
(4) Are the inflation rate series cointegrated among themselves?
(5) Are interest rates and inflation rates cointegrated? If not, then the real interest rate cannot be stationary.

Of course, the answers to some of these questions are fairly well known. For example, it would be surprising if interest rates turned out not to have a unit root, if the two inflation rate series turned out not to be cointegrated, or if the three interest rate series turned out not to be cointegrated. However, questions (2) and (5) are by no means settled.

I used annual data partly for purposes of illustration, so that the differences between asymptotic and finite-sample P-values would not be negligible. However, the use of annual data also has several advantages. First, it avoids the problems associated with unit root tests that use seasonally adjusted data; see Ghysels and Perron (1993). Seasonality is not a problem with interest rates, but it is with inflation rates. Second, the use of annual data greatly simplifies the calculation of ADF tests, since the number of lags of the regressand that must be added to the test regression to whiten the residuals is almost always much smaller than for quarterly or monthly data. Third, annual data on inflation rates are much less subject to errors in variables than higher-frequency data, and, as I discuss below, such errors would make standard unit root tests invalid. The disadvantage of using annual data rather than higher frequency data is that the tests may be less powerful. However, as has been shown by Shiller and Perron (1985) and Pierse and Snell (1995), the power loss from using annual data for unit root tests, if the span of the data is held constant, is very modest.

The five series I analyzed are graphed in Figure 11. The three interest rate series are denoted R_s, R_m and R_1. These are, respectively, the rate on 91-day treasury bills, the rate on three to five year federal government bonds, and the rate on federal government bonds with maturities of ten years and over.[1] The two inflation rate series are denoted π_c and π_y. The former is based on the Consumer Price Index, and the latter is based on the GDP deflator.[2] The unit root test regressions were run over the 1955–1994 period (forty years), with available lagged data being used if necessary. The cointegrating regressions (5) were run over the 1954–1994 period, and the subsequent test regressions (6) over the 1955–1994 period. If necessary, unavailable lags of the residuals from the cointegrating regressions were replaced by zeros.

For all five series, I computed all eight varieties of unit root tests with 0, 1, and 2 lags of the regressand in the test regressions. Because it is not possible to present all the results here, Table I shows only the results of the test regressions that seemed most consistent with the data. The value of the test statistic is shown along with the asymptotic P-value, $P(\infty)$, and the finite-sample P-value, $P(40)$. The number following the test used is the number of lags of the regressand included in the ADF test regression. For determining the number of lags to include, I considered t-statistics greater than 2 to provide strong evidence in favour of including the lag, and t-statistics between 1 and 2 to provide weak evidence. When more than one set of results is

---

[1] To be precise, R_s, R_m, and R_1 are the annual averages of CANSIM series B14001, B14010, and B14003, respectively. CANSIM is a registered trademark of Statistics Canada.

[2] Specifically, $\pi_c(t) = 100 \log(P_c(t)/\log P_c(t-1))$, where $P_c(t)$ denotes the annual average of CANSIM series P484000 (CPI, all items, 1986 = 100) for year t, and $\pi_y(t) = 100 \log(P_y(t)/\log P_y(t-1))$, where $P_y(t)$ denotes the annual average of CANSIM series D10000 (nominal GDP) divided by the annual average of series D10373 (real GDP), both for year t.
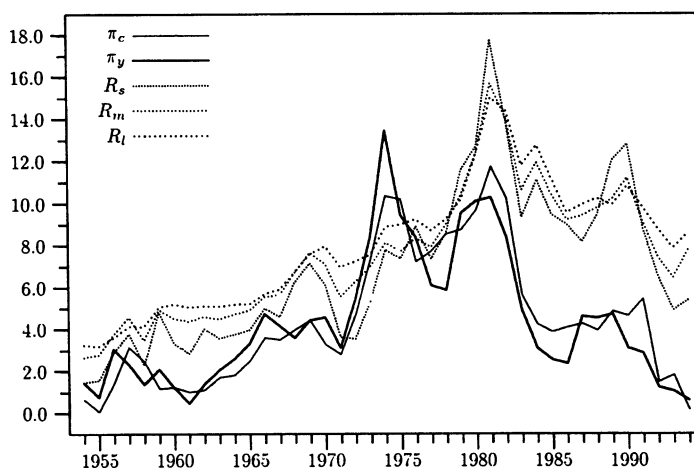
Figure 11. Interest rates and inflation rates in Canada, 1954–1994

Table I. Results of unit root tests

Series	Test	Value	$P(\infty)$	$P(40)$	Test	Value	$P(\infty)$	$P(40)$
R_s	$\tau_c, 0$	−2·026	0·276	0·275	$z_c, 0$	−6·598	0·303	0·277
	$\tau_{ctt}, 0$	−2·416	0·616	0·594	$z_{ctt}, 0$	−12·159	0·577	0·507
	$\tau_{ctt}, 1$	−2·734	0·435	0·426	$z_{ctt}, 1$	−19·928	0·201	0·120
R_m	$\tau_c, 0$	−1·813	0·375	0·369	$z_c, 0$	−4·582	0·476	0·455
	$\tau_{ctt}, 0$	−1·745	0·898	0·879	$z_{ctt}, 0$	−9·919	0·719	0·672
	$\tau_{ctt}, 1$	−2·033	0·804	0·780	$z_{ctt}, 1$	−12·856	0·533	0·458
R_l	$\tau_{ctt}, 0$	−2·031	0·805	0·780	$z_{ctt}, 0$	−8·177	0·824	0·795
	$\tau_{ctt}, 1$	−2·494	0·572	0·552	$z_{ctt}, 1$	−14·327	0·446	0·362
π_c	$\tau_{ctt}, 1$	−2·930	0·331	0·331	$z_{ctt}, 1$	−19·094	0·229	0·144
π_y	$\tau_{ctt}, 1$	−3·406	0·139	0·157	$z_{ctt}, 1$	−24·300	0·098	0·041
	$\tau_{ctt}, 2$	−3·602	0·089	0·109	$z_{ctt}, 2$	−37·670	0·008	0·001

presented, the least general test regression cannot be rejected against the more general one(s) at the 0·05 level, but there is weak evidence against the former.

The results for the interest rates are not surprising. There is no evidence at all against the unit root hypothesis for any of the three series. By far the smallest P-value is the finite-sample one for the z_{ctt} test on R_s with one lag included in the test regression, and even it is greater than 0·10. This atypical P-value probably reflects the fact that z-tests tend to be quite unreliable when the test regression includes lags of the regressand, since dividing by one minus the sum of the coefficients on the lags introduces considerable noise when those coefficients are not estimated precisely. Since the finite-sample P-values are valid only for non-augmented Dickey–Fuller tests, it is probably wise to ignore them for ADF tests, especially in the case of z-tests.

The results for the inflation rates are somewhat less clear. Unlike the interest rates, for which there was never any strong evidence that lags of the regressand needed to be included, one lag is needed for π_c, and at least one lag is needed for π_y. There is also strong evidence that both a trend and a squared trend must be included in the test regressions, something that is not surprising in view of Figure 11. For π_c we cannot reject the unit root hypothesis at any conventional level using any of the tests. For π_y, however, we can reject it at the 0·05 level

using finite-sample P-values for both the reported z-tests, and using asymptotic P-values for the z-test with two lags included in the test regression. However, following the argument of the previous paragraph, we probably should not take the z-tests, especially their finite-sample P-values, very seriously. Moreover, since there is no strong evidence that the second lag needs to be included, the only rejection based on asymptotic P-values is doubtful. Thus, on balance, there seems to be little credible evidence that π_y does not have a unit root.

These results differ from those reported by Rose (1988), but they are consistent with the results of Mishkin (1992) and with some of those reported by Gregory and Watt (1995). The fact that I am using annual rather than monthly or quarterly data is important. When I used monthly CPI data for the inflation rate, I was able to reject the unit root hypothesis very convincingly. However, it is not at all clear that the monthly results are more reliable than the annual ones, as the following argument shows.

Suppose that the level of the CPI is measured with error, so that the logarithm of the observed price level, p_t^*, is equal to the logarithm of the true price level, p_t, plus an i.i.d. error term, v_t. This implies that the observed inflation rate is $\pi_t^* = \pi_t + v_t - v_{t-1}$. Now consider, for simplicity, the DF test regression (1). I terms of the observed inflation rate, it will be

$$\Delta\pi_t^* = (\alpha - 1)\pi_{t-1}^* + u_t$$

When this equation is rewritten in terms of the true inflation rate, it becomes

$$\Delta\pi_t + v_t - 2v_{t-1} + v_{t-2} = (\alpha - 1)(\pi_{t-1} + v_{t-1} - v_{t-2}) + u_t \tag{18}$$

It is obvious that the error in the regressand will be negatively correlated with the error in the regressor. In fact, this correlation is $-\sqrt{3}/2 = -0{\cdot}866$. Thus, depending on the variance of v_t relative to that of u_t, there is the potential for the estimate of $\alpha - 1$ in equation (18) to be very seriously biased downwards. Simulations confirm that this bias can indeed be very large. Since i.i.d. measurement errors will tend to average out over time, this type of bias should be very much less severe for annual data than for monthly or even quarterly data. Thus annual data may well yield the most reliable results.

Table II presents results for a number of cointegration tests, which are designed to answer questions (3), (4), and (5). Since these tests are not invariant to the (arbitrary) choice of the regressand in the cointegrating regression (5), I report results for all choices.

The answers to questions (3) and (4) are quite definitive. It is clear that the three interest rate series are cointegrated among themselves, as one would expect. The evidence against the null of

Table II. Results of cointegration tests

Series	Test	Value	$P(\infty)$	$P(40)$	Test	Value	$P(\infty)$	$P(40)$
$R_s: R_m, R_l$	$\tau_c(3),0$	$-3{\cdot}932$	$0{\cdot}030$	$0{\cdot}053$	$z_c(3),0$	$-25{\cdot}454$	$0{\cdot}062$	$0{\cdot}020$
	$\tau_c(3),1$	$-4{\cdot}227$	$0{\cdot}012$	$0{\cdot}028$	$z_c(3),1$	$-44{\cdot}329$	$0{\cdot}002$	$0{\cdot}000$
$R_m: R_s, R_l$	$\tau_c(3),1$	$-4{\cdot}847$	$0{\cdot}001$	$0{\cdot}007$	$z_c(3),1$	$-56{\cdot}030$	$0{\cdot}000$	$0{\cdot}000$
$R_l: R_s, R_m$	$\tau_{ctt}(3),1$	$-5{\cdot}629$	$0{\cdot}001$	$0{\cdot}010$	$z_{ctt}(3),1$	$-69{\cdot}915$	$0{\cdot}000$	$0{\cdot}000$
$\pi_c: \pi_y$	$\tau_{ct}(2),0$	$-5{\cdot}010$	$0{\cdot}001$	$0{\cdot}005$	$z_{ct}(2),0$	$-32{\cdot}844$	$0{\cdot}018$	$0{\cdot}002$
$\pi_y: \pi_c$	$\tau_{ct}(2),0$	$-5{\cdot}210$	$0{\cdot}000$	$0{\cdot}003$	$z_{ct}(2),0$	$-33{\cdot}386$	$0{\cdot}016$	$0{\cdot}002$
$R_s: \pi_c$	$\tau_{ct},0$	$-3{\cdot}203$	$0{\cdot}184$	$0{\cdot}226$	$z_{ct},0$	$-17{\cdot}054$	$0{\cdot}273$	$0{\cdot}187$
$\pi_c: R_s$	$\tau_{ctt},0$	$-3{\cdot}086$	$0{\cdot}413$	$0{\cdot}474$	$z_{ctt},0$	$-15{\cdot}155$	$0{\cdot}357$	$0{\cdot}271$
$R_l: \pi_c$	$t_{ct},0$	$-2{\cdot}475$	$0{\cdot}534$	$0{\cdot}559$	$z_{ct},0$	$-10{\cdot}978$	$0{\cdot}594$	$0{\cdot}535$
	$\tau_{ct},2$	$-2{\cdot}815$	$0{\cdot}352$	$0{\cdot}388$	$z_{ct},2$	$-23{\cdot}235$	$0{\cdot}102$	$0{\cdot}043$
$\pi_c: R_l$	$\tau_{ctt},0$	$-2{\cdot}701$	$0{\cdot}625$	$0{\cdot}670$	$z_{ctt},0$	$-12{\cdot}499$	$0{\cdot}732$	$0{\cdot}676$
	$\tau_{ctt},2$	$-3{\cdot}167$	$0{\cdot}369$	$0{\cdot}434$	$z_{ctt},2$	$-29{\cdot}037$	$0{\cdot}101$	$0{\cdot}028$

noncointegration is very strong when R_m or R_l is the regressand, and it is reasonably strong when R_s is the regressand. It also seems clear that the two inflation rates are cointegrated. The evidence against the null is very strong, and the results are strikingly insensitive to the choice of regressand.

The answer to question (5) is also clear. The bottom part of Table II reports results for four of the twelve possible pairs of interest and inflation rates. The results that are not reported were qualitatively very similar to the ones in the table. In all cases, the cointegrating regressions required one trend term when the regressand was an interest rate and two trend terms when the regressand was an inflation rate. In no case did τ-tests yield any evidence of cointegration. In a few cases, such as the last and third-last lines of the table, z-tests did yield weak evidence of it, but this probably just reflects the poor finite-sample performance of these tests in the ADF case. If anything, results using π_y instead of π_c, which are not reported, provided even less evidence of cointegration than the results in the table. Thus, in contrast to the results of Mishkin (1992) for US data, there is no evidence that Canadian interest rates and inflation rates are cointegrated. If they are not, then the real interest rate, however we choose to define it, cannot be stationary; see Rose (1988) and Gregory and Watt (1995).

The purpose of this empirical example was primarily to illustrate the use of the urcdist program. The program makes it easy to calculate P-values, which are much easier to interpret than raw test statistics. This is especially true when the latter come from a number of different distributions, as is the case for this example. If space were limited, the tables could be shortened by omitting the actual test statistics, since all the information is conveyed by the P-values. The tables give a sense of how much asymptotic and finite-sample P-values can differ for samples of modest size. In the case of τ tests, using finite-sample P-values generally leads to modestly smaller or modestly larger P-values, but in the case of z-tests it can lead to substantially smaller ones. It is strongly recommended that finite-sample P-values not be used for z-tests calculated from ADF test regressions, that is, ones which include lags of the regressand. In such cases, z-tests are often not reliable, and using invalid finite-sample P-values is likely to make matters worse.

8. CONCLUSION

In this paper I have computed response surface coefficients which provide excellent approximations to the asymptotic and finite-sample distributions of eight varieties of Dickey–Fuller unit root and cointegration tests for up to 12 possibly cointegrated variables. The paper extends the work of MacKinnon (1991, 1994) by allowing for twice as many possibly cointegrated variables, and it also contains several innovations. The principal innovation is that the results consist chiefly of tables of estimated coefficients, along with a computer program that uses these to calculate critical values and P-values. Both of these are available via the Internet. 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.

It is important to remember that, although the asymptotic distributions obtained in this paper are valid under quite weak assumptions, the finite-sample ones depend on the strong assumption that the error terms are i.i.d. normal and are valid only for DF test regressions which have not been augmented. Finite-sample distributions of ADF tests and of other asymptotically equivalent tests will often differ substantially from the ones obtained here, and they may not converge to the asymptotic distributions as rapidly as the response surface coefficients of this paper suggest.

NOTE

ACKNOWLEDGEMENTS

REFERENCES

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

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

Cragg, J. G. (1983), 'More efficient estimation in the presence of heteroskedasticity of unknown form', *Econometrica*, **51**, 751–763.

Dickey, D. A. and W. A. Fuller (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, **74**, 427–431.

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

Ghysels, E. and P. Perron (1993), 'The effect of seasonal adjustment filters on tests for a unit root', *Journal of Econometrics*, **55**, 57–98.

Gregory, A. W. and D. G. Watt (1995), 'Sources of variation in international real interest rates', *Canadian Journal of Enconomics*, **28**, S120–S140.

Johansen, S. (1991), 'Estimation and hypothesis testing of cointegration in Gaussian vector autoregressive models', *Econometrica*, **59**, 1551–1580.

Kremers, J. J. M., N. R. Ericsson and J. J. Dolado (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 R. F. Engle and C. W. J. Granger (eds), *Long-run Economic Relationships: Readings in Cointegration*, Oxford University Press, Oxford, 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 T. A. Bray (1964), 'A convenient method for generating normal variables', *SIAM Review*, **6**, 260–264.

Mishkin, F. S. (1992), 'Is the Fisher effect for real? A reexamination of the relationship between inflation and interest rates', *Journal of Monetary Economics*, **30**, 195–215.

Ng, S. and P. Perron (1995), 'Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag', *Journal of the American Statistical Association*, **90**, 268–281.

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

Pantula, S. G., G. Farias-Gonzales and W. A. Fuller (1994), 'A comparison of unit-root test criteria', *Journal of Business and Economic Statistics*, **12**, 449–459.

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.

Phillips, P. C. B. and P. Perron (1988), 'Testing for a unit root in time series regression', *Biometrika*, **75**, 335–346.

Pierse, R. G. and A. J. Snell (1995), 'Temporal aggregation and the power of tests for a unit root', *Journal of Econometrics*, **65**, 333–345.

Rose, A. K. (1988), 'Is the real interest rate stable?' *Journal of Finance*, **43**, 1095–1112.

Said, E. S. and D. A. Dickey (1984), 'Testing for unit roots in autoregressive-moving average models of unknown order', *Biometrika*, **71**, 599–607.

Shiller, R. J. and P. Perron (1985), 'Testing the random walk hypothesis: power versus frequency of observation', *Economics Letters*, **18**, 381–386.